AN INTRODUCTION TO

Dynamic Meteorology

Fifth Edition

James R. Holton
Gregory J. Hakim

An Introduction to Dynamic Meteorology

Fifth Edition

James R. Holton
Gregory J. Hakim

ELSEVIER
AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Table of Contents

Cover image

Title page

Copyright

Dedication

Preface

Chapter 1. Introduction

- 1.1 Dynamic Meteorology
- 1.2 Conservation of Momentum
- 1.3 Noninertial Reference Frames and “Apparent” Forces
- 1.4 Structure of the Static Atmosphere
- 1.5 Kinematics
- 1.6 Scale Analysis
- Suggested References

Chapter 2. Basic Conservation Laws

- 2.1 Total Differentiation
- 2.2 The Vectorial Form of the Momentum Equation in Rotating Coordinates
- 2.3 Component Equations in Spherical Coordinates
- 2.4 Scale Analysis of the Equations of Motion
- 2.5 The Continuity Equation
- 2.6 The Thermodynamic Energy Equation
- 2.7 Thermodynamics of the Dry Atmosphere
- 2.8 The Boussinesq Approximation
- 2.9 Thermodynamics of the Moist Atmosphere
- Suggested References

Chapter 3. Elementary Applications of the Basic Equations

- 3.1 Basic Equations in Isobaric Coordinates
- 3.2 Balanced Flow
- 3.3 Trajectories and Streamlines
- 3.4 The Thermal Wind
- 3.5 Vertical Motion
- 3.6 Surface Pressure Tendency

Chapter 4. Circulation, Vorticity, and Potential Vorticity

4.1 The Circulation Theorem

4.2 Vorticity

4.3 The Vorticity Equation

4.4 Potential Vorticity

4.5 Shallow Water Equations

4.6 Ertel Potential Vorticity in Isentropic Coordinates

Suggested References

Chapter 5. Atmospheric Oscillations: Linear Perturbation Theory

5.1 The Perturbation Method

5.2 Properties of Waves

5.3 Simple Wave Types

5.4 Internal Gravity (Buoyancy) Waves

5.5 Linear Waves of A Rotating Stratified Atmosphere

5.6 Adjustment to Geostrophic Balance

5.7 Rossby Waves

Suggested References

Chapter 6. Quasi-geostrophic Analysis

6.1 The Observed Structure of Extratropical Circulations

6.2 Derivation of the Quasi-Geostrophic Equations

6.3 Potential vorticity derivation of the QG equations

6.4 Potential Vorticity Thinking

6.5 Vertical Motion (w) Thinking

6.6 Idealized Model of a Baroclinic Disturbance

6.7 Isobaric Form of the QG Equations

Suggested References

Chapter 7. Baroclinic Development

7.1 Hydrodynamic Instability

7.2 Normal Mode Baroclinic Instability: A Two-Layer Model

7.3 The Energetics of Baroclinic Waves

7.4 Baroclinic Instability of a Continuously Stratified Atmosphere

7.5 Growth and Propagation of Neutral Modes

Suggested References

Chapter 8. The Planetary Boundary Layer

8.1 Atmospheric Turbulence

8.2 Turbulent Kinetic Energy
8.3 Planetary Boundary Layer Momentum Equations
8.4 Secondary Circulations and Spin Down
Suggested References
Chapter 9. Mesoscale Circulations
9.1 Energy Sources for Mesoscale Circulations
9.2 Fronts and Frontogenesis
9.3 Symmetric Baroclinic Instability
9.4 Mountain Waves
9.5 Cumulus Convection
9.6 Convective Storms
9.7 Hurricanes
Suggested References
Chapter 10. The General Circulation
10.1 The Nature of the Problem
10.2 The Zonally Averaged Circulation
10.3 The Angular Momentum Budget
10.4 The Lorenz Energy Cycle
10.5 Longitudinally Dependent Time-Averaged Flow
10.6 Low-Frequency Variability
10.7 Numerical Simulation of the General Circulation
10.8 Climate Sensitivity, Feedbacks, and Uncertainty
Suggested References
Chapter 11. Tropical Dynamics
11.1 The Observed Structure of Large-Scale Tropical Circulations
11.2 Scale Analysis of Large-Scale Tropical Motions
11.3 Condensation Heating
11.4 Equatorial Wave Theory
11.5 Steady Forced Equatorial Motions
Suggested References
Chapter 12. Middle Atmosphere Dynamics
12.1 Structure and Circulation of the Middle Atmosphere
12.2 The Zonal-Mean Circulation of the Middle Atmosphere
12.3 Vertically Propagating Planetary Waves

12.4 Sudden Stratospheric Warmings
12.5 Waves in the Equatorial Stratosphere
12.6 The Quasi-Biennial Oscillation
12.7 Trace Constituent Transport
Suggested References
Chapter 13. Numerical Modeling and Prediction
13.1 Historical Background
13.2 Numerical Approximation of the Equations of Motion
13.3 The Barotropic Vorticity Equation in Finite Differences
13.4 The Spectral Method
13.5 Primitive Equation Models
13.6 Data Assimilation
13.7 Predictability and Ensemble Forecasting
Suggested References
Appendix A: Useful Constants and Parameters
Appendix B: List of Symbols
Appendix C: Vector Analysis
C.1 Vector Identities
C.2 Integral Theorems
C.3 Vector Operations in Various Coordinate Systems
Appendix D: Moisture Variables
D.1 Equivalent Potential Temperature
D.2 Pseudoadiabatic Lapse Rate
Appendix E: Standard Atmosphere Data
Appendix F: Symmetric Baroclinic Oscillations
Appendix G: Conditional Probability and Likelihood
Index

Copyright

Academic Press is an imprint of Elsevier
225 Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK

© 2013 Elsevier Inc. All rights reserved.
No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.
Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: *www.elsevier.com/permissions*.
This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.
Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.
To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

MATLAB® is a trademark of The MathWorks, Inc., and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.
MATLAB® and Handle Graphics® are registered trademarks of The MathWorks, Inc.

Library of Congress Cataloging-in-Publication Data
Holton, James R., and Hakim, Gregory J., authors.
An introduction to dynamic meteorology.—Fifth edition / Gregory J. Hakim.
pages cm
Includes bibliographical references and index. Revision of Fourth edition by Holton.
ISBN 978-0-12-384866-6 (hardback)
1. Dynamic meteorology. I. Title.
QC880.H65 2012
551.51'5—dc23
2012022179

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

For information on all Academic Press publications visit our website at *http://store.elsevier.com*

Printed in the United States
12 13 14 15 16 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries
www.elsevier.com | www.bookaid.org | www.sabre.org
ELSEVIER BOOK AID International Sabre Foundation

Dedication

To the memory of James R. Holton (1938–2004)

Preface

When Jim Holton was making revisions for the Fourth Edition of this book, I served as a consultant on aspects of the material with which I was most familiar. Toward the end of the revision process, Jim asked whether I would like to join him as a coauthor for the next edition. Although I agreed to this wonderful opportunity and looked forward to the collaboration, it vanished when Jim Holton died on March 3, 2004. The shock of his sudden passing lingers in the dynamic meteorology community, where his influence is difficult to exaggerate. That sphere of influence includes this book on dynamic meteorology, which has served as the standard text on the subject for generations of students and practitioners in the atmospheric and related sciences. It was very difficult to take up revisions without Jim's guidance, but the situation also presented an opportunity to bring a fresh perspective to aspects of the material that had become dated, tracing their origin to lecture notes from classes taught at MIT during the 1960s.

This book serves three main communities: undergraduate and graduate students in the atmospheric sciences, practitioners in the field, and those in related physical sciences who want a definitive, and accessible, introduction to the subject matter. In making revisions, I have tried to draw on my experiences with the book as both a student and an instructor, to streamline and modernize aspects of the text to serve these communities. Major structural changes include Chapter 5 (on waves and the perturbation method, formerly Chapter 7), Chapter 7 (on baroclinic development, formerly Chapter 8), and Chapter 8 (on turbulence and the planetary boundary layer, formerly Chapter 5). The material now flows from the introduction of core concepts (Chapters 1–4) to the development of methods needed to understand the application of core concepts (Chapter 5) to understanding extratropical weather systems (Chapters 6 and 7), followed by advanced topics in later chapters.

In addition to numerous minor changes, specific substantial revisions include the following. In Chapter 1, the treatment of viscous effects and noninertial reference frames is streamlined, and a new section on kinematics is introduced. New sections in Chapter 2 are devoted to the Boussinesq approximation and thermodynamics of the moist atmosphere, which are used in several chapters later in the book. Potential vorticity (PV) is given an expanded treatment in Chapter 4, including a derivation of the Ertel PV as a special case of the Kelvin circulation theorem, a discussion of nonconservative effects, and an illustration of the use of PV to construct maps of the dynamical tropopause. Moreover, PV is used to motivate a derivation of the shallow-water and barotropic equations, which are used extensively in later chapters. In Chapter 5, the discussion of basic wave properties is extended to three dimensions, and a general strategy for solving wave problems is outlined. Stationary Rossby wave solutions are now discussed for both shallow-water and stratified atmospheres at rest, providing the reader with a deeper understanding of the basis for the quasi-geostrophic approximation that follows in the next chapter.

Chapter 6 is largely rewritten, with a novel, and simple, derivation of the quasi-geostrophic equations starting from the Ertel PV conservation equation. In contrast to previous editions, here standard height coordinates are used, which liberates the reader from the mental gymnastics associated with pressure coordinates. A MATLAB code for a quasi-geostrophic model is provided, along with a diagnostic package that allows the reader to simulate and diagnose extratro- pical weather systems; these codes are easily adapted for flow patterns different from those provided.

Baroclinic development is covered in Chapter 7; the treatment is largely the same as in previous editions but offers a richer discussion of "generalized stability." Chapter 9 contains updated information on hurricanes, including a full derivation and discussion of potential intensity theory. A new section in Chapter 10 reviews climate sensitivity and feedbacks, which is a subject of increasing interest in the research community. Finally, Chapter 13 has been revised significantly to include a mathematical treatment of data assimilation, building up from simple examples for scalars to many variables. Kalman filters, variational techniques (3DVAR and 4DVAR), and ensemble Kalman filters are discussed. A new section reviews predictability and ensemble forecasting, including the Liouville equation and primary results on the theory of ensemble prediction. For other information about, and teaching aids for, this book, see booksite.academicpress.com/9780123848666.

I would like to thank Margaret Holton for her friendship and support in undertaking this project. Dale Durran and Cecilia Bitz maintained an excellent errata for the Fourth Edition, which was helpful when working on this revision. I am grateful to the following individuals for offering comments and suggestions on early chapter drafts: Hanin Binder, Bonnie Brown, Dale Durran, Luke Madaus, Max Menchaca, Dave Nolan, Ryan Torn, and Mike Wallace. Errors that remain are mine alone, and I would appreciate hearing about them at `holton.hakim@gmail.com`.

Greg Hakim
Seattle, WA

Chapter 1

Introduction

1.1 Dynamic Meteorology

Dynamic meteorology is the study of air motion in the Earth's atmosphere that is associated with weather and climate. These motions organize into coherent circulation features that affect human activity primarily through wind, temperature, clouds, and precipitation patterns. Short-lived features, lasting from a few minutes to a few days, are related to weather, and some familiar weather examples that we will examine in this book include tropical and extratropical cyclones, organized thunderstorms, and local wind patterns such as those that occur near mountains. Figure 1.1 illustrates the mixing effect of larger weather patterns in the atmosphere, from large areas of convective cloud in the tropics to extratropical cyclones in the higher latitudes of the Northern and Southern Hemispheres. These weather elements occur in the *troposphere*, which is the portion of the atmosphere in contact with the surface. The troposphere normally exhibits a drop in temperature with elevation and contains most of the water vapor, clouds, and precipitation found in the atmosphere. On average, the troposphere extends vertically about 10 kilometers, where the *tropopause* is located. Above the tropopause is the stratosphere, where the temperature increases with elevation due to heating of the air by absorption of ultraviolet radiation by ozone. Most of the topics addressed in this book concern the dynamics of the troposphere and stratosphere.

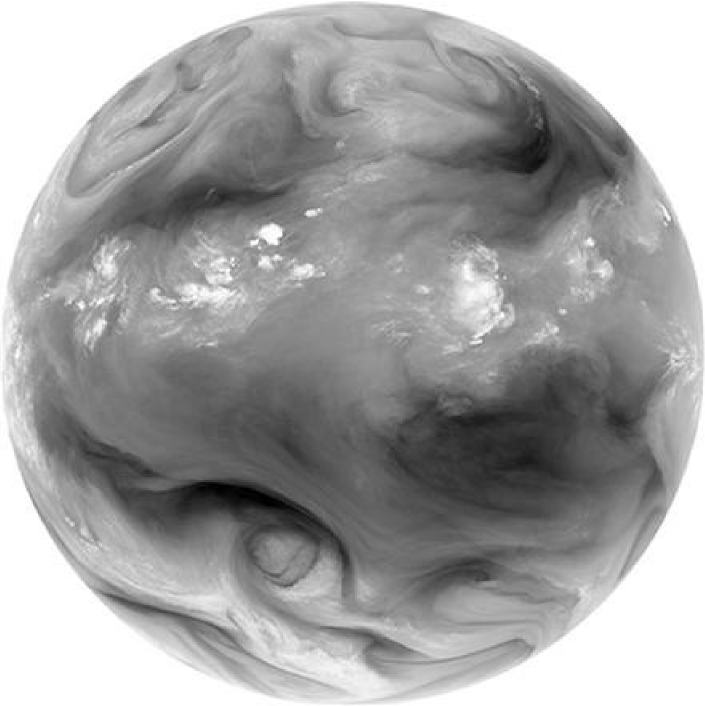


Figure 1.1 Infrared satellite image near a wavelength of 6.7μm, which is known as the "water vapor" channel since it captures the distribution of that field in a layer roughly 5 to 10 km above Earth's surface. Because water vapor is continuously distributed, in contrast to clouds, atmospheric motion is especially well captured. Here we see the convective clouds in the tropics

and the mixing effects of eddies at higher latitude.

Source: NASA.

Over longer time periods, the realm of climate, circulation features may persist from seasons to years over large regions of Earth. Examples of climate variability include shifts in the locations where storms occur, oscillations in large-scale pressure patterns, and planetary patterns of variability associated with the El Niño Southern Oscillation (ENSO) phenomenon of the tropical Pacific Ocean. ENSO reminds us that although dynamic meterology involves the study of air motion in the atmosphere, this motion links to other parts of Earth's system, including the oceans, biosphere, and cryosphere, and plays an active role in the transport of chemical species. Moreover, many of the ideas we present here also apply to the atmospheres of other planets.

Before we set off to explore the landscape of dynamic meteorology, we devote this first chapter to introducing fundamental concepts that will guide the journey. First, note that the laws that govern atmospheric motion satisfy the principle of *dimensional homogeneity*, which means that all terms in the equations expressing these laws must have the same physical dimensions. These dimensions can be expressed in terms of multiples and ratios of four dimensionally independent properties: length, time, mass, and thermodynamic temperature. To measure and compare the scales of terms in the laws of motion, a set of units of measure must be defined for these four fundamental properties. In this text the international system of units (SI) will be used almost exclusively. The four fundamental properties are measured in terms of the SI *base units* shown in Table 1.1. All other properties are measured in terms of SI *derived units*, which are units formed from products or ratios of the base units. For example, velocity has the derived units of meter per second ($\mathrm{m\ s^{-1}}$).

Table 1.1 SI Base Units

Property	Name	Symbol
Length	Meter (meter)	m
Mass	Kilogram	kg
Time	Second	s
Temperature	Kelvin	K

A number of important derived units have special names and symbols. Those that are commonly used in dynamic meteorology are indicated in Table 1.2. In addition, the supplementary unit designating a plane angle, the radian (rad), is required for expressing angular velocity (rad $^{s-1}$) in the SI system.[1] Finally, Table 1.3 lists the symbols frequently used in this book for some of the basic physical quantities. Note that the full three-dimensional velocity vector, $\mathbf{U}$, is related to the horizontal velocity vector, $\mathbf{V}$, by $\mathbf{U} = (\mathbf{V}, w)$ and $\mathbf{U} = (\mathbf{V}, \omega)$ in height and pressure vertical coordinates, respectively. We shall use the term "zonal" to refer to the East–West direction and "meridional" to refer to the North–South direction.

Table 1.2 SI Derived Units with Special Names

Property	Name	Symbol

Frequency	Hertz	Hz(s^{-1})
Force	Newton	N(kg m s^{-2})
Pressure	Pascal	Pa(N m^{-2})
Energy	Joule	J(N m)
Power	Watt	W(J s^{-1})

Table 1.3 Symbols and Units of Basic Physical Quantities

Quantity	Symbol	Units
Three-dimensional velocity vector	$\mathbf{U}$	m s^{-1}
Horizontal velocity vector	$\mathbf{V}$	m s^{-1}
Eastward component of velocity	u	m s^{-1}
Northward component of velocity	v	m s^{-1}
Upward component of velocity	w (ω)	m s^{-1}(Pa s^{-1})
Pressure	P	N m^{-2}
Density	ρ	kg m^{-3}
Temperature	T	K (or °C)

Dynamic meteorology applies the conservation laws of classical physics for momentum (Newton's laws of motion), mass, and energy (First law of thermodynamics) to the atmosphere. An essential aspect of this application involves the *continuum* approximation, whereby the properties of discrete molecules are ignored in favor of a continuous representation involving a local average over a blob of molecules. This approximation is common to all fluids, including liquids and gases, and allows atmospheric properties (e.g., pressure, density, temperature), or "field variables," to be represented as smooth functions taking on unique values in the independent variables of space and time. A "point" in the continuum is regarded as a volume element that is very small compared with the volume of atmosphere under consideration, but that still contains a large number of molecules. The expressions *air parcel* and *air particle* are both commonly used to refer to such a point. In Chapter 2, the fundamental conservation laws are applied to a small volume element of the atmosphere subject to the continuum approximation in order to obtain the governing equations. Our goal here is to provide a survey of the main forces that influence atmospheric motions.

1.2 Conservation of Momentum

Newton's first law states that an object at rest or moving with a constant velocity remains so unless acted upon by an external unbalanced force. Once the forces are identified, Newton's second law states that the temporal change of momentum (i.e., acceleration) is a vector having direction given by the net force (the sum over all forces) and magnitude given by the size of the net force divided by the object's mass. These forces can be classified as either *body forces* or *surface forces*. Body forces act on the center of mass of a fluid parcel and have magnitudes proportional to the mass of the parcel; gravity is an example of a body force. Surface forces act across the boundary surface separating a fluid parcel from its surroundings and have magnitudes independent of the mass of the parcel; the pressure force is an example.

For atmospheric motions of meteorological interest, the forces that are of primary concern are the pressure gradient force, the gravitational force, and the frictional force. These *fundamental* forces determine acceleration as measured relative to coordinates fixed in space. If, as is the usual case, the motion is referred to a coordinate system rotating with Earth, Newton's second law may still be applied provided that certain *apparent* forces, the centrifugal force and the Coriolis force, are included. The fundamental forces are discussed subsequently, and the apparent forces are introduced in Section 1.3.

1.2.1 Pressure Gradient Force

Pressure is defined as the force per unit area acting normal to a surface. In a gas such as the atmosphere, pressure at a point acts equally in all directions due to random molecular motion. Therefore, the magnitude of the net force due to molecules colliding with a surface is independent of the orientation of the surface; note that the *direction* of the net force changes with the orientation of the surface, but the magnitude does not. Placing a wall in a gas, with the pressure on one side different from that on the other, yields a net force that will accelerate the wall toward the side having lower pressure; this net force associated with pressure differences is the essence of the pressure *gradient* force.

Consider now an infinitesimal volume element of air, $\delta V = \delta x \delta y \delta z$, centered at the point x_0, y_0, z_0, as illustrated in Figure 1.2. Due to random molecular motions, momentum is continually imparted to the walls of the volume element by the surrounding air. This momentum transfer per unit time per unit area is just the pressure exerted on the walls of the volume element by the surrounding air. If the pressure at the center of the volume element is designated by p_0, then the pressure on the wall labeled A in Figure 1.2 can be expressed in a Taylor series expansion as

$$p_0 + \frac{\partial p}{\partial x}\frac{\delta x}{2} + \text{higher-order terms}$$

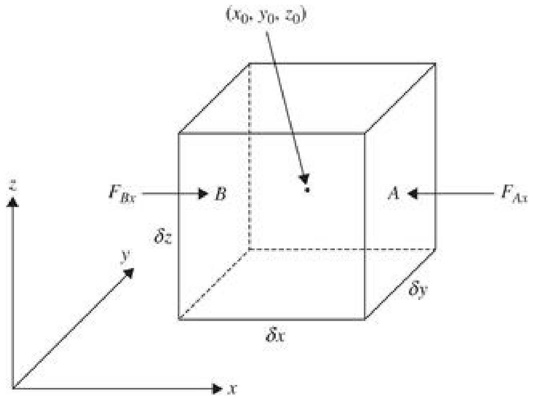


Figure 1.2 The *x* component of the pressure gradient force acting on a fluid element.

Neglecting the higher-order terms in this expansion, the pressure force acting on the volume element at wall *A* is

$$F_{Ax} = -\left(p_0 + \frac{\partial p}{\partial x}\frac{\delta x}{2}\right)\delta y\,\delta z$$

where $\delta y \delta z$ is the area of wall *A*. Similarly, the pressure force acting on the volume element at wall *B* is just

$$F_{Bx} = +\left(p_0 - \frac{\partial p}{\partial x}\frac{\delta x}{2}\right)\delta y\,\delta z$$

Therefore, the net *x* component of this force acting on the volume is

$$F_x = F_{Ax} + F_{Bx} = -\frac{\partial p}{\partial x}\,\delta x\,\delta y\,\delta z$$

Because the net force is proportional to the derivative of pressure, it is referred to as the *pressure gradient force*. The mass *m* of the differential volume element is simply the density ρ times the volume: $m = \rho \delta x \delta y \delta z$. Thus, the *x* component of the pressure gradient force per unit mass is

$$\frac{F_x}{m} = -\frac{1}{\rho}\frac{\partial p}{\partial x}$$

Similarly, it can easily be shown that the *y* and *z* components of the pressure gradient force per unit mass are

$$\frac{F_y}{m} = -\frac{1}{\rho}\frac{\partial p}{\partial y} \qquad \text{and} \qquad \frac{F_z}{m} = -\frac{1}{\rho}\frac{\partial p}{\partial z}$$

so that the total pressure gradient force per unit mass is the vector given by

$$\frac{\mathbf{F}}{m} = -\frac{1}{\rho}\boldsymbol{\nabla} p \tag{1.1}$$

The gradient operator $\boldsymbol{\nabla} = \left(\mathbf{i}\frac{\partial}{\partial x}, \mathbf{j}\frac{\partial}{\partial y}, \mathbf{k}\frac{\partial}{\partial z}\right)$ acts on functions to its right to yield vectors that point toward higher values of the function. It is important to note that (1) the pressure gradient points from low to high pressure, but the pressure gradient *force* points from high to low pressure, and (2) the pressure gradient force is proportional to the *gradient* of the pressure field, not to the pressure itself.

1.2.2 Viscous Force

Any real fluid is subject to internal friction, called viscosity, which causes it to resist the tendency to flow. Viscosity arises when the fluid velocity varies spatially so that random molecular motion accomplishes a net transport of momentum from molecules in faster-moving air parcels to molecules in nearby slower-moving air parcels. This momentum exchange between parcels may be expressed as a viscous force, *F*, acting *along* the face of the air parcel, which produces a *shear* stress, τ, on the parcel per area, *A*,

$$\tau = \frac{F}{A} \tag{1.2}$$

Therefore, the viscous force is given by $F = \tau A$. For a Newtonian fluid, we assume that the shear stress depends linearly on the fluid speed, which is a very good approximation for air. In the vertical direction, for example, variation in the *x* component of the wind, *u*, produces the stress

$$\tau \approx \mu \frac{\partial u}{\partial z} \tag{1.3}$$

where μ is the viscosity coeficient, which depends on the fluid. As in the pressure gradient force discussion, we need the net force acting on the air parcel from viscous effects. Following a similar Taylor-approximation approach as for the pressure gradient derivation, but noting that the force is directed *along* rather than *normal* to the face of the parcel volume, we find that the viscous force per unit mass due to vertical shear of the component of motion in the *x* direction is

$$\frac{1}{\rho}\frac{\partial \tau_{zx}}{\partial z} = \frac{1}{\rho}\frac{\partial}{\partial z}\left(\mu \frac{\partial u}{\partial z}\right) \tag{1.4}$$

For constant μ, the right side may be simplified to $\nu \partial^2 u/\partial z^2$, where $\nu = \mu/\rho$ is the *kinematic viscosity coefficient*. For standard atmosphere[2] conditions at sea level, $\nu = 1.46 \times 10^{-5}\,\mathrm{m^2\,s^{-1}}$. Note that (1.4) represents only the contribution from *x* momentum shear stress in the *z* direction, and the net force in the *x* direction, F_{rx}, also includes contributions from the *x* and *y* directions. The net frictional force components per unit mass in the three Cartesian coordinate directions are

$$\begin{aligned} F_{rx} &= \nu\left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right] \\ F_{ry} &= \nu\left[\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2}\right] \\ F_{rz} &= \nu\left[\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}\right] \end{aligned} \tag{1.5}$$

Each component frictional force represents diffusion of momentum in that coordinate direction, since, for example, $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \nabla \cdot \nabla u = \nabla^2 u$. For any vector **A**, $\nabla \cdot \mathbf{A}$ is a scalar quantity called the *divergence* of A, since it is positive when vectors diverge away from a point; negative divergence is called *convergence*. At a local maximum in *u*, ∇u points toward (converges on) the maximum, and therefore $\nabla^2 u < 0$, resulting in a loss of momentum from the maximum value to the surrounding region. This process is called *downgradient diffusion*, since momentum diffuses down the gradient, from large to small values.

For the atmosphere below 100 km, ν is so small that molecular viscosity is negligible except in a thin layer within a few centimeters of Earth's surface, where the vertical shear is very large. Away from this surface molecular boundary layer, momentum is transferred primarily by turbulent eddy motions, which are discussed in Chapter 8.

1.2.3 Gravitational Force

The sole body force on atmospheric air parcels is due to gravity. Newton's law of universal gravitation states that any two elements of mass in the universe attract each other with a force proportional to their masses and inversely proportional to the square of the distance separating them. Thus, if two mass elements M and m are separated by a distance $r \equiv |\mathbf{r}|$ (with the vector **r** directed toward m as shown in Figure 1.3), then the force exerted by mass M on mass m due to gravitation is

$$\mathbf{F}_g = -\frac{GMm}{r^2}\left(\frac{\mathbf{r}}{r}\right) \tag{1.6}$$

where G is a universal constant called the gravitational constant. The law of gravitation as expressed in (1.6) actually applies only to hypothetical "point" masses, since for objects of finite extent **r** will vary from one part of the object to another. However, for finite bodies, (1.6) may still be applied if $|\mathbf{r}|$ is interpreted as the distance between the centers of mass of the bodies. Thus, if Earth is designated as mass M, and m is a mass element of the atmosphere, then the force per unit mass exerted on the atmosphere by the gravitational attraction of Earth is

$$\frac{\mathbf{F}_g}{m} \equiv \mathbf{g}^* = -\frac{GM}{r^2}\left(\frac{\mathbf{r}}{r}\right) \tag{1.7}$$

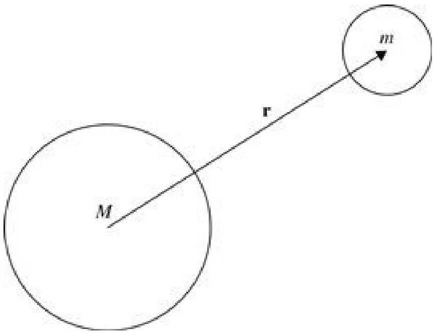


Figure 1.3 Two spherical masses whose centers are separated by a distance r.

In dynamic meteorology it is customary to use the height above mean sea level as a vertical coordinate. If the mean radius of Earth is designated by a and the distance above mean sea level is designated by z, then neglecting the small departure of the shape of Earth from sphericity, $r = a + z$. Therefore, Eq. (1.7) can be rewritten as

$$\mathbf{g}^* = \frac{\mathbf{g}_0^*}{(1 + z/a)^2} \tag{1.8}$$

where $\mathbf{g}_0^* = -(GM/a^2)(\mathbf{r}/r)$ is the gravitational force at mean sea level. For meteorological applications, $z \ll a$, so that with negligible error we can let $\mathbf{g}^* = \mathbf{g}_0^*$ and simply treat the gravitational force as a constant. Note that this treatment of the gravitational force will be modified in Section 1.3.2 to account for centrifugal forces due to Earth's rotation.

1.3 Noninertial Reference Frames and "Apparent" Forces

In formulating the laws of atmospheric dynamics, it is natural to use a *geocentric* reference frame—that is, a frame of reference at rest with respect to rotating Earth. Newton's first law of motion states that a mass in uniform motion relative to a coordinate system fixed in space will remain in uniform motion in the absence of any forces. Such motion is referred to as *inertial motion*, and the fixed reference frame is an inertial, or absolute, frame of reference. It is clear, however, that an object at rest or in uniform motion with respect to rotating Earth is not at rest or in uniform motion relative to a coordinate system fixed in space.

Therefore, motion that appears to be inertial motion to an observer in a geocentric reference frame is really accelerated motion. Hence, a geocentric reference frame is a *noninertial* reference frame. Newton's laws of motion can only be applied in such a frame if the acceleration of the coordinates is taken into account. The most satisfactory way of including the effects of coordinate acceleration is to introduce "apparent" forces in the statement of Newton's second law. These apparent forces are the inertial reaction terms that arise because of the coordinate acceleration. For a coordinate system in uniform rotation, two such apparent forces are required: the centrifugal force and the Coriolis force.

1.3.1 Centripetal Acceleration and Centrifugal Force

To illustrate the essential aspects of noninertial frames, we consider a ball of mass m attached to a string and whirled through a circle of radius r at a constant angular velocity ω. From the point of view of an observer in inertial space the speed of the ball is constant, but its direction of travel is continuously changing so that its velocity is not constant. To compute the acceleration we consider the change in velocity $\delta \mathrm{V}$ that occurs for a time increment δt during which the ball rotates through an angle $\delta\theta$ as shown in Figure 1.4. Because $\delta\theta$ is also the angle between the vectors $\mathbf{V}$ and $\mathrm{V} + \delta \mathrm{V}$, the magnitude of $\delta \mathrm{V}$ is just $|\delta \mathbf{V}| = |\mathbf{V}|$ $\delta\theta$. If we divide by δt and note that in the limit $\delta t \to 0, \delta \mathbf{V}$ is directed toward the axis of rotation, we obtain

$$\frac{D\mathbf{V}}{Dt} = |\mathbf{V}|\frac{D\theta}{Dt}\left(-\frac{\mathbf{r}}{r}\right)$$

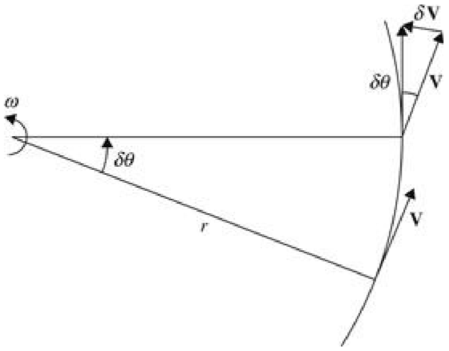


Figure 1.4 Centripetal acceleration is given by the rate of change of the direction of the velocity vector, which is directed toward the axis of rotation, as illustrated here by $\delta \mathrm{V}$.

However, $|\mathbf{V}| = \omega r \text{ and } D\theta/Dt = \omega$, so that

$$\frac{D\mathbf{V}}{Dt} = -\omega^2\mathbf{r} \tag{1.9}$$

Therefore, viewed from fixed coordinates, the motion is one of uniform acceleration directed toward the axis of rotation at a rate equal to the square of the angular velocity times the distance from the axis of rotation. This acceleration is called *centripetal acceleration*. It is caused by the force of the string pulling the ball.

Now suppose that we observe the motion in a coordinate system rotating with the ball. In this rotating system the ball is stationary, but there is still a force acting on the ball—namely, the pull of the string. Therefore, in order to apply Newton's second law to describe the motion relative to this rotating coordinate system, we must include an additional apparent force, the *centrifugal force*, which just balances the force of the string on the ball. Thus, the centrifugal force is equivalent to the inertial reaction of the ball on the string and just equal and opposite to the centripetal acceleration.

To summarize, observed from a fixed system, the rotating ball undergoes a uniform centripetal acceleration in response to the force exerted by the string. Observed from a system rotating along with it, the ball is stationary and the force exerted by the string is balanced by a centrifugal force.

1.3.2 Gravity Revisited

An object at rest on the surface of Earth is not at rest or in uniform motion relative to an inertial reference frame except at the poles. Rather, an object of unit mass at rest on the surface of Earth is subject to a centripetal acceleration directed toward the axis of rotation of Earth given by $-\Omega^2\mathbf{R}$, where **R** is the position vector from the axis of rotation to the object and $\Omega = 7.292 \times 10^{-5}\,\mathrm{rad\,s^{-1}}$ is the angular speed of rotation of Earth.[3] Since, except at the equator and poles, the centripetal acceleration has a component directed poleward along the horizontal surface of Earth (i.e., along a surface of constant *geopotential*), there must be a net horizontal force directed poleward along the horizontal to sustain the horizontal component of the centripetal acceleration.

This force arises because rotating Earth is not a sphere but has assumed the shape of an oblate spheroid in which there is a poleward component of gravitation along a constant geopotential surface just sufficient to account for the poleward component of the centripetal acceleration at each latitude for an object at rest on the surface of Earth. In other words, from the point of view of an observer in an inertial reference frame, geopotential surfaces slope upward toward the equator (Figure 1.5). As a consequence, the equatorial radius of Earth is about 21 km larger than the polar radius.

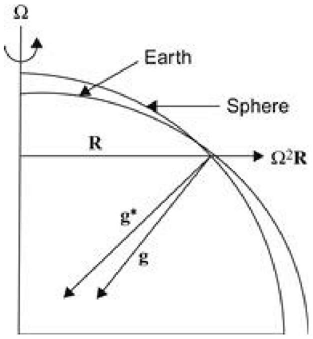


Figure 1.5 Relationship between the true gravitation vector g* and gravity g. For an idealized homogeneous spherical Earth, g* would be directed toward the center of Earth. In reality, g* does not point exactly to the center except at the equator and the poles. Gravity, g, is the vector sum of g* and the centrifugal force and is perpendicular to the level surface of Earth, which approximates an oblate spheroid.

Viewed from a frame of reference rotating with Earth, however, a geopotential surface is everywhere normal to the sum of the true force of gravity, g*, and the centrifugal force $\Omega^2\mathbf{R}$ (which is just the reaction force of the centripetal acceleration). A geopotential surface is thus experienced as a level surface by an object at rest on rotating Earth. Except at the poles, the weight of an object of mass *m* at rest on such a surface, which is just the reaction force of Earth on the object, will be slightly less than the gravitational force mg^* because, as illustrated in Figure 1.5, the centrifugal force partly balances the gravitational force. It is, therefore, convenient to combine the effects of the gravitational force and centrifugal force by defining *gravity* **g** such that

(1.10)

$$\mathbf{g} \equiv -g\mathbf{k} \equiv \mathbf{g}^* + \Omega^2\mathbf{R}$$

where **k** designates a unit vector parallel to the local vertical. Gravity, *g*, sometimes referred to as "apparent gravity," will here be taken as a constant ($g = 9.81 \text{ m s}^{-2}$). Except at the poles and the equator, **g** is not directed toward the center of Earth, but is perpendicular to a geopotential surface as indicated by Figure 1.5. True gravity $\mathbf{g}^*$, however, is not perpendicular to a geopotential surface, but has a horizontal component just large enough to balance the horizontal component of $\Omega^2 \mathbf{R}$.

Gravity can be represented in terms of the gradient of a potential function Φ, which is just the geopotential referred to before:

$$\nabla\Phi = -\mathbf{g}$$

However, because $\mathbf{g}=-g\mathbf{k}$, where $g\equiv|\mathbf{g}|$, it is clear that $\Phi=\Phi(z)$ and $d\Phi/dz=g$. Thus, horizontal surfaces on Earth are surfaces of constant geopotential. If the value of *dz* should be *dz*‘, where *dz*’ is a dummy variable of integration, geopotential is set to zero at mean sea level, the geopotential Φ(*z*) at height *z* is just the work required to raise a unit mass to height *z* from mean sea level:

(1.11)

$$\Phi = \int_0^z g dz$$

Despite the fact that the surface of Earth bulges toward the equator, an object at rest on the surface of rotating Earth does not slide "downhill" toward the pole because, as indicated previously, the poleward component of gravitation is balanced by the equatorward component of the centrifugal force. However, if the object is put into motion relative to Earth, this balance will be disrupted. Consider a frictionless object located initially at the North Pole. Such an object has zero angular momentum about the axis of Earth. If it is displaced away from the pole in the absence of a zonal torque, it will not acquire rotation and thus will feel a restoring force due to the horizontal component of true gravity, which, as indicated before, is equal and opposite to the horizontal component of the centrifugal force for an object at rest on the surface of Earth. Letting *R* be the distance from the pole, the horizontal restoring force for a small displacement is thus $-\Omega^2 R$, and the object's acceleration viewed in the inertial coordinate system satisfies the equation for a simple harmonic oscillator:

(1.12)

$$\frac{d^2R}{dt^2} + \Omega^2 R = 0$$

The object will undergo an oscillation of period $2\pi/\Omega$ along a path that will appear as a straight line passing through the pole to an observer in a fixed coordinate system, but will appear as a closed circle traversed in 1/2 day to an observer rotating with Earth (Figure 1.6). From the point of view of an Earthbound observer, there is an apparent deflection force that causes the object to deviate to the right of its direction of motion at a fixed rate.

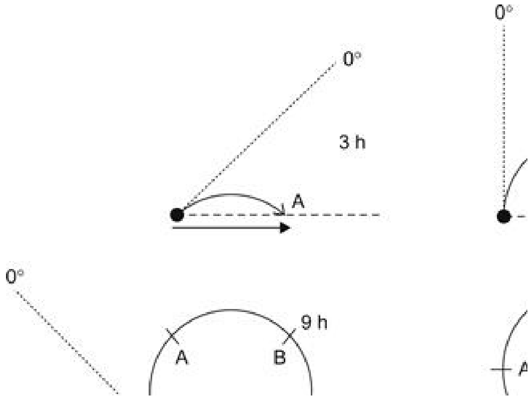


Figure 1.6 Motion of a frictionless object launched from the North Pole along the 0° longitude meridian at $t = 0$, as viewed in fixed and rotating reference frames at 3, 6, 9, and 12 h after launch. The *horizontal dashed line* marks the position that the 0° longitude meridian had at $t = 0$, and the short dashed lines show its position in the fixed reference frame at subsequent 3-h intervals. The *horizontal arrows* show 3-h displacement vectors as seen by an observer in the fixed reference frame. *Heavy curved arrows* show the trajectory of the object as viewed by an observer in the rotating system. Labels A, B, and C show the position of the object relative to the rotating coordinates at 3-h intervals. In the fixed coordinate frame, the object oscillates back and forth along a straight line under the influence of the restoring force provided by the horizontal component of gravitation. The period for a complete oscillation is 24 h (only 1/2 period is shown). To an observer in rotating coordinates, however, the motion appears to be at constant speed and describes a complete circle in a clockwise direction in 12 h.

1.3.3 The Coriolis Force and the Curvature Effect

Newton's second law of motion expressed in coordinates rotating with Earth can be used to describe the force balance for an object at rest on the surface of Earth, provided that an apparent force, the centrifugal force, is included among the forces acting on the object. If, however, the object is in motion along the surface of Earth, additional apparent forces are required in the statement of Newton's second law. The Coriolis force will be given a more formal mathematical treatment in Chapter 2, and our purpose here is to deduce the effect by building upon the centrifugal force discussion of the previous section.

Angular momentum, $\mathbf{m} = \mathbf{r} \times \mathbf{p}$, provides a measures of the rotation traced by the linear momentum vector $\mathbf{p}$ with respect to a set of coordinates, the origin of which defines the position vector $\mathbf{r}$. We note that the dynamically important piece of the angular momentum vector lies parallel to Earth's rotational axis, $m = \mathbf{m}\cos\phi$. For now, assume that the linear momentum vector points eastward, having contributions from eastward air motion, u, and from the planetary rotation, $R\Omega$, where $R = r\cos\phi$ is the distance of the air parcel from the axis of rotation (Figure 1.7). If there is no torque in the east–west direction (i.e., no pressure gradient or viscous forces), then m is conserved following the motion,

$$\frac{Dm}{Dt} = \frac{DR}{Dt}(2R\Omega + u) + R\frac{Du}{Dt} = 0 \tag{1.13}$$

so that

$$\frac{Du}{Dt} = -\frac{(2\Omega R + u)}{R}\frac{DR}{Dt} \tag{1.14}$$

Figure 1.7 shows that moving an air parcel closer to the axis of rotation, $\frac{DR}{Dt} < 0$, while conserving angular momentum, increases the westerly linear momentum, analagous to an ice skater spinning faster as the person's arms are drawn inward.

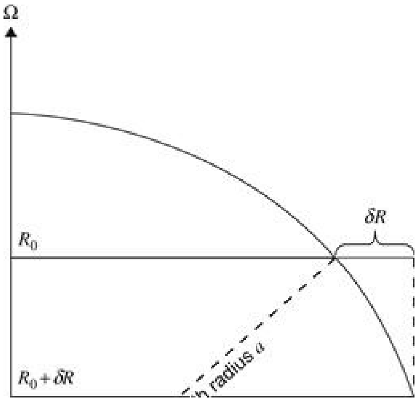


Figure 1.7 For poleward motion, air parcels move closer to the axis of rotation and, through angular momentum conservation, the zonal wind accelerates.

We expand the right side of (1.14) by first noting that

$$\frac{DR}{Dt} = \frac{Dr}{Dt}\cos\phi + r\frac{D}{Dt}\cos\phi = w\cos\phi - v\sin\phi \tag{1.15}$$

where v and w are the northward and upward velocity components, respectively. With this relation, (1.14) becomes

$$\frac{Du}{Dt} = (2\Omega\sin\phi)v - (2\Omega\cos\phi)w - \frac{uw}{r} + \frac{uv}{r}\tan\phi \tag{1.16}$$

The first two terms on the right in (1.16) are the zonal component of the *Coriolis force* due to meridional and vertical motion, respectively. The last two terms on the right are referred to as *metric terms* or *curvature effects* because they arise from the curvature of Earth's surface; since r is large, these terms are negligibly small except near large u.

Suppose now that the object is set in motion in the eastward direction by an impulsive force. Axial angular momentum is not conserved in this case, but considering again the centrifugal force will help expose the meridional component of the Coriolis force. Because the object is now rotating faster than Earth, the centrifugal force on the object will be increased. The excess of the centrifugal force over that for an object at rest is

$$\left(\Omega + \frac{u}{R}\right)^2 \mathbf{R} - \Omega^2\mathbf{R} = \frac{2\Omega u\mathbf{R}}{R} + \frac{u^2\mathbf{R}}{R^2}$$

The terms on the right represent *deflecting* forces, which act outward along the vector **R** (i.e., perpendicular to the axis of rotation). The meridional and vertical components of these forces are obtained by taking meridional and vertical components of **R**, as shown in Figure 1.8, to yield

$$\frac{Dv}{Dt} = -2\Omega u\sin\phi - \frac{u^2}{a}\tan\phi \tag{1.17}$$

$$\frac{Dw}{Dt} = 2\Omega u \cos\phi + \frac{u^2}{a} \tag{1.18}$$

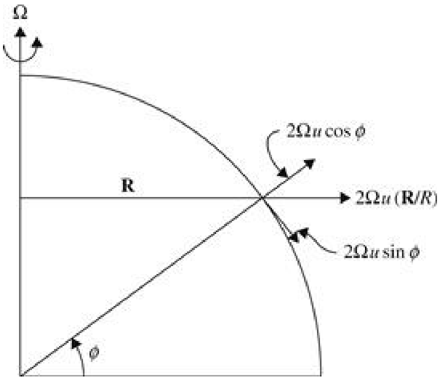


Figure 1.8 Components of the Coriolis force due to relative motion along a latitude circle.

The first terms on the right are the meridional and vertical components, respectively, of the Coriolis forces for zonal motion; the second terms on the right are curvature effects.

For larger-scale motions, the curvature terms can be neglected as an approximation. Therefore, relative horizontal motion produces a horizontal acceleration perpendicular to the direction of motion given by

$$\frac{Du}{Dt} = 2\Omega v \sin\phi = fv \tag{1.19}$$

$$\frac{Dv}{Dt} = -2\Omega u \sin\phi = -fu \tag{1.20}$$

where $f \equiv 2\Omega \sin\phi$ is the *Coriolis parameter*.

Thus, for example, an object moving eastward in the horizontal is deflected equatorward by the Coriolis force, whereas a westward moving object is deflected poleward. In either case the deflection is to the right of the direction of motion in the Northern Hemisphere and to the left in the Southern Hemisphere. The vertical component of the Coriolis force in (1.18) is ordinarily much smaller than the gravitational force so that its only effect is to cause a very minor change in the apparent weight of an object depending on whether the object is moving eastward or westward.

The Coriolis force is negligible for motions with time scales that are very short compared to the period of Earth's rotation (a point that is illustrated by several problems at the end of the chapter). Thus, the Coriolis force is not important for the dynamics of individual cumulus clouds but is essential for an understanding of longer time scale phenomena such as synoptic scale systems. The Coriolis force must also be taken into account when computing long-range missile or artillery trajectories.

As an example, suppose that a ballistic missile is fired due eastward at 43° N latitude $(f = 10^{-4}\,\mathrm{s}^{-1}$ at $43°\mathrm{N})$. If the missile travels 1000 km at a horizontal speed $u_0 = 1000$ m s^{-1}, by how much is the missile deflected from its eastward path by the Coriolis force? Integrating (1.20) with respect to time, we find that

$$v = -fu_0t \tag{1.21}$$

where it is assumed that the deflection is sufficiently small so that we may let f and u_0 be constants. To find the total displacement, we must integrate (1.13) with respect to time:

$$\int_0^t v dt = \int_{y_0}^{y_0+\delta y} dy = -f u_0 \int_0^t t dt$$

Thus, the total displacement is

$$\delta y = -f u_0 t^2/2 = -50\,\mathrm{km}$$

Therefore, the missile is deflected southward by 50 km due to the Coriolis effect. Further examples of the deflection of objects by the Coriolis force are given in some of the problems at the end of the chapter.

The x and y components given in (1.19) and (1.20) can be combined in vector form as

$$\left(\frac{D\mathbf{V}}{Dt}\right)_{Co} = -f\mathbf{k}\times\mathbf{V} \tag{1.22}$$

where $\mathbf{V}\equiv(u,v)$ is the horizontal velocity, **k** is a vertical unit vector, and the subscript Co indicates that the acceleration is due solely to the Coriolis force. Since $-\mathbf{k}\times\mathbf{V}$ is a vector rotated 90° to the right of **V**, (1.22) clearly shows the deflection character of the Coriolis force. The Coriolis force can only change the direction of motion, not the speed.

1.3.4 Constant Angular Momentum Oscillations

Suppose an object initially at rest on Earth at the point (x_0, y_0) is impulsively propelled along the x axis with a speed V at time $t = 0$. Then, from (1.19) and (1.20), the time evolution of the velocity is given by $u = V\cos\, ft$ and $v = -V\sin\, ft$. However, because $u = Dx/Dt$ and $v = Dy/Dt$, integration with respect to time gives the position of the object at time t as

$$x - x_0 = \frac{V}{f}\sin\, ft \quad \text{and} \quad y - y_0 = \frac{V}{f}(\cos\, ft - 1) \tag{1.23}$$

where the variation of f with latitude is neglected. Equations (1.23) show that in the Northern Hemisphere, where f is positive, the object orbits clockwise (anticyclonically) in a circle of radius $R = V/f$ about the point $(x_0, y_0 - V/f)$ with a period given by

$$\tau = 2\pi R/V = 2\pi/f = \pi/(\Omega\sin\phi) \tag{1.24}$$

Thus, an object displaced horizontally from its equilibrium position on the surface of Earth under the influence of the force of gravity will oscillate about its equilibrium position with a period that depends on latitude and is equal to one sidereal day at 30° latitude and 1/2 sidereal day at the pole. Constant angular momentum oscillations (often referred to as “inertial oscillations”) are commonly observed in the oceans, but are apparently not of importance in the atmosphere.

1.4 Structure of the Static Atmosphere

The thermodynamic state of the atmosphere at any point is determined by the values of pressure, temperature, and density (or specific volume) at that point. These field variables are related to one an other by the equation of state for an ideal gas. Letting p, T, ρ, and $\alpha(\equiv \rho^{-1})$ denote pressure, temperature, density, and specific volume, respectively, we can express the equation of state for dry air as

(1.25)

$$p\alpha = RT \quad \text{or} \quad p = \rho RT$$

where R is the gas constant for dry air $(R = 287\,\mathrm{J\,kg^{-1}\,K^{-1}})$.

1.4.1 The Hydrostatic Equation

In the absence of atmospheric motions, the gravity force must be exactly balanced by the vertical component of the pressure gradient force. Thus, as illustrated in Figure 1.9,

(1.26)

$$dp/dz = -\rho g$$

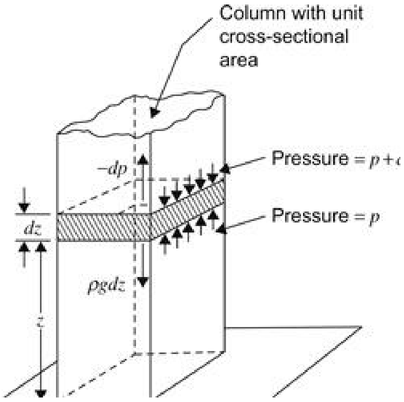


Figure 1.9 Balance of forces for hydrostatic equilibrium. *Small arrows* show the upward and downward forces exerted by air pressure on the air mass in the shaded block. The downward force exerted by gravity on the air in the block is given by ρgdz, whereas the net pressure force given by the difference between the upward force across the lower surface and the downward force across the upper surface is $-dp$. Note that dp is negative, as pressure decreases with height. (*After* *Wallace and Hobbs, 2006*.)

This condition of *hydrostatic balance* provides an excellent approximation for the vertical dependence of the pressure field in the real atmosphere. Only for intense small-scale systems, such as squall lines and tornadoes, is it necessary to consider departures from hydrostatic balance. Integrating (1.26) from a height z to the top of the atmosphere, we find that

(1.27)

$$p(z) = \int_z^\infty \rho g dz$$

so that the pressure at any point is simply equal to the weight of the unit cross-section column of air overlying the point. Thus, mean sea level pressure $p(0) = 1013.25\,\mathrm{hPa}$ is simply the average weight per square meter of the total atmospheric column.[4] It is often useful to express the hydrostatic equation in terms of the geopotential rather than the geometric height. Noting from (1.11) that $d\Phi = gdz$ and from (1.25) that $\alpha = RT/p$, we can express the hydrostatic equation in the form

$$gdz = d\Phi = -(RT/p)dp = -RTd\ln p \tag{1.28}$$

Thus, the variation in geopotential with respect to pressure depends only on temperature. Integration of (1.28) in the vertical yields a form of the *hypsometric equation*:

$$\Phi(z_2) - \Phi(z_1) = g_0(Z_2 - Z_1) = R\int_{p_2}^{p_1} Td\ln p \tag{1.29}$$

Here $Z \equiv \Phi(z)/g_0$ is the *geopotential height*, where $g_0 \equiv 9.80665\,\mathrm{m\,s^{-2}}$ is the global average of gravity at mean sea level. Thus, in the troposphere and lower stratosphere, Z is numerically almost identical to the geometric height z. In terms of Z the hypsometric equation becomes

$$Z_T \equiv Z_2 - Z_1 = \frac{R}{g_0}\int_{p_2}^{p_1} Td\ln p \tag{1.30}$$

where Z_T is the *thickness* of the atmospheric layer between the pressure surfaces p_2 and p_1. Defining a layer mean temperature

$$\langle T\rangle = \left[\int_{p_2}^{p_1} d\ln p\right]^{-1}\int_{p_2}^{p_1} Td\ln p$$

and a layer mean scale height $H \equiv R\langle T\rangle/g_0$, we have from Eq. (1.30)

$$Z_T = H\ln(p_1/p_2) \tag{1.31}$$

Thus, the thickness of a layer bounded by isobaric surfaces is proportional to the mean temperature of the layer. Pressure decreases more rapidly with height in a cold layer than in a warm layer. It also follows immediately from (1.31) that in an isothermal atmosphere of temperature T, the geopotential height is proportional to the natural logarithm of pressure normalized by the surface pressure clearer:

$$Z = H\ln(p_0/p) \tag{1.32}$$

where p_0 is the pressure at $Z = 0$. Thus, in an isothermal atmosphere the pressure decreases exponentially with geopotential height by a factor of e^{-1} per scale height:

$$p(Z) = p(0)e^{-Z/H}$$

1.4.2 Pressure as a Vertical Coordinate

From the hydrostatic equation (1.26), it is clear that a single-valued monotonic relationship exists between pressure and height in each vertical column of the atmosphere. Thus, we may use pressure as the independent vertical coordinate and height (or geopotential) as a dependent variable. The thermodynamic state of the atmosphere is then specified by the fields of $\Phi(x,y,p,t)$ and $T(x,y,p,t)$.

Now the horizontal components of the pressure gradient force given by Eq. (1.1) are evaluated by partial differentiation holding z constant. However, when pressure is used as the vertical coordinate, horizontal partial derivatives must be evaluated holding p constant. Transformation of the horizontal pressure gradient force from height to pressure coordinates may be carried out with the aid of Figure 1.10. Considering only the x,z plane, we see from Figure 1.10 that

$$\left[\frac{(p_0+\delta p)-p_0}{\delta x}\right]_z = \left[\frac{(p_0+\delta p)-p_0}{\delta z}\right]_x \left(\frac{\delta z}{\delta x}\right)_p$$

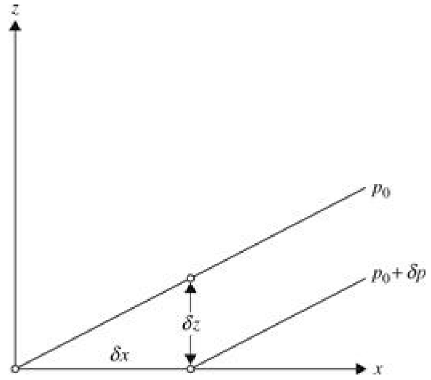


Figure 1.10 Slope of pressure surfaces in the x,z plane.

where subscripts indicate variables that remain constant in evaluating the differentials. For example, in the limit $\delta z \to 0$

$$\left[\frac{(p_0+\delta p)-p_0}{\delta z}\right]_x \to \left(-\frac{\partial p}{\partial z}\right)_x$$

where the minus sign is included because $\delta z < 0 \text{ for } \delta p > 0$.

Taking the limits $\delta x, \delta z \to 0$, we obtain[5]

$$\left(\frac{\partial p}{\partial x}\right)_z = -\left(\frac{\partial p}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_p$$

which after substitution from the hydrostatic equation (1.26) yields

$$-\frac{1}{\rho}\left(\frac{\partial p}{\partial x}\right)_z = -g\left(\frac{\partial z}{\partial x}\right)_p = -\left(\frac{\partial \Phi}{\partial x}\right)_p \quad (1.33)$$

Similarly, it is easy to show that

$$-\frac{1}{\rho}\left(\frac{\partial p}{\partial y}\right)_z = -\left(\frac{\partial \Phi}{\partial y}\right)_p \quad (1.34)$$

Thus, in the *isobaric* coordinate system the horizontal pressure gradient force is measured by the gradient of geopotential at constant pressure. Density no longer appears explicitly in the pressure gradient force; this is a distinct advantage of the isobaric system.

1.4.3 A Generalized Vertical Coordinate

Any single-valued monotonic function of pressure or height may be used as the independent vertical coordinate. For example, in many numerical weather prediction models, pressure normalized by the pressure at the ground, $\sigma \equiv p(x,y,z,t)/p_s(x,y,t)$, is used as a vertical coordinate. This choice guarantees that the ground is a coordinate surface $(\sigma \equiv 1)$ even in the presence of spatial and temporal surface pressure variations. Thus, this so-called σ coordinate system is particularly useful in regions of strong topographic variations.

We now obtain a general expression for the horizontal pressure gradient, which is applicable to any vertical coordinate $s = s(x,y,z,t)$ that is a single-valued monotonic function of height. Referring to Figure 1.11 we see that for a horizontal distance δx, the pressure difference evaluated along a surface of constant s is related to that evaluated at constant z by the relationship

$$\frac{p_C - p_A}{\delta x} = \frac{p_C - p_B}{\delta z}\frac{\delta z}{\delta x} + \frac{p_B - p_A}{\delta x}$$

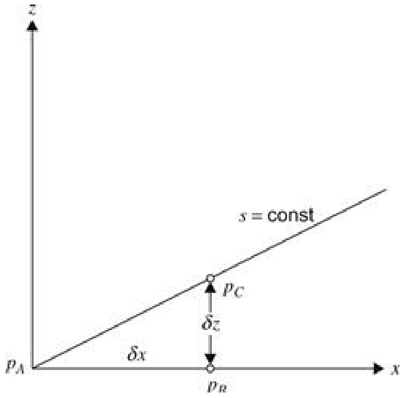


Figure 1.11 Transformation of the pressure gradient force to s coordinates.

Taking the limits as $\delta x, \delta z \to 0$, we obtain

$$\left(\frac{\partial p}{\partial x}\right)_s = \frac{\partial p}{\partial z}\left(\frac{\partial z}{\partial x}\right)_s + \left(\frac{\partial p}{\partial x}\right)_z \tag{1.35}$$

Using the identity $\partial p/\partial z = (\partial s/\partial z)(\partial p/\partial s)$, we can express (1.35) in the alternate form

$$\left(\frac{\partial p}{\partial x}\right)_s = \left(\frac{\partial p}{\partial x}\right)_z + \frac{\partial s}{\partial z}\left(\frac{\partial z}{\partial x}\right)_s\left(\frac{\partial p}{\partial s}\right) \tag{1.36}$$

In later chapters we will apply (1.35) or (1.36) and similar expressions for other fields to transform the dynamical equations to several different vertical coordinate systems.

1.5 Kinematics

Kinematics involves the analysis of motion without reference to forces that change the motion in time. It provides a diagnosis of motion at a particular instant in time, which may in turn prove useful for understanding the dynamics of the flow as it evolves in time. There are many aspects of kinematics, but usually one is interested in the structure of the flow, and here we will limit attention to the horizontal flow. One way to quantify flow structure is to examine linear variations in the flow near an arbitrary point (x_0,y_0). A leading-order Taylor approximation to the wind near the point is

(1.37a)

$$u(x_0+dx,y_0+dy)\approx u(x_0,y_0)+\left.\frac{\partial u}{\partial x}\right|_{(x_0,y_0)}dx+\left.\frac{\partial u}{\partial y}\right|_{(x_0,y_0)}$$

(1.37b)

$$v(x_0+dx,y_0+dy)\approx v(x_0,y_0)+\left.\frac{\partial v}{\partial x}\right|_{(x_0,y_0)}dx+\left.\frac{\partial v}{\partial y}\right|_{(x_0,y_0)}dy$$

Making the following definitions

(1.38a)

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=\delta$$

(1.38b)

$$\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}=\zeta$$

(1.38c)

$$\frac{\partial u}{\partial x}-\frac{\partial v}{\partial y}=d_1$$

(1.38d)

$$\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}=d_2$$

allows the derivatives in (1.37a,b) to be replaced in favor of the elemental quanties $\delta, \zeta, d_1,$ and d_2:

(1.39a)

$$u(x_0+dx,y_0+dy)\approx u(x_0,y_0)+\frac{1}{2}(\delta+d_1)dx+\frac{1}{2}(d_2-\zeta)dy$$

(1.39b)

$$v(x_0+dx,y_0+dy)\approx v(x_0,y_0)+\frac{1}{2}(\zeta+d_2)dx+\frac{1}{2}(\delta-d_1)dy$$

The advantage of this manipulation is that we can now think about the wind near a point as a linear combination of the elemental fluid properties. The vorticity, ζ, represents pure rotation (about the vertical direction); δ represents pure divergence; and $\delta < 0$ is called convergence. Pure deformation is represented by d_1 and d_2, where the wind field contracts, or is *confluent* in one direction, called the axis of contraction, and the wind field stretches in the normal direction, called the axis of dilatation. d_2 represents a 45° rotation of d_1, and therefore they are not independent patterns. The leading, constant, terms in (1.39a,b) represent uniform translation.

By setting all elemental quantities in (1.39a,b) except one to zero, we visualize the spatial distribution of the horizontal wind associated with one elemental quantity (Figure 1.12). Note that the pure deformation pattern represents d_1 only and that, although the vectors appear to converge and diverge, the divergence is exactly zero in this field. This is an important example of the fact that confluence (difluence) in vector fields is not the same as convergence (divergence). In the lower right panel of Figure 1.12, we see a linear combination of vorticity and divergence, illustrating a more complicated pattern than the elemental fields in isolation. By computing the elemental variables at a point, one can visualize the linear variation of the flow near that point through (1.39a,b).

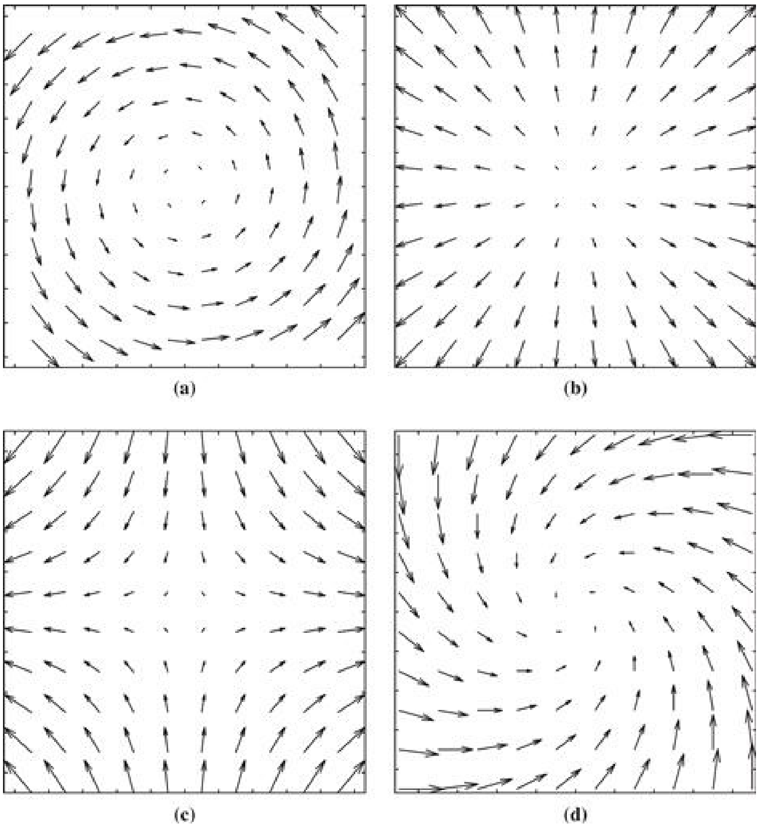


Figure 1.12 Velocity fields associated with pure vorticity (a), pure divergence (b), pure deformation (c), and a mixture of vorticity and convergence (d).

This taste of kinematics highlights the importance of certain properties of the wind field that will be explored in greater depth in future chapters. Divergence is connected to vertical motion by mass conservation, as will be discussed in Chapter 2. Vorticity is fundamental to understanding dynamic meteorology and will be explored in detail in Chapter 3. Deformation is important for creating and destroying boundaries in fluids, such as horizontal temperature contrasts known as *frontal zones*, which will be explored in Chapter 9.

1.6 Scale Analysis

Scale analysis, or scaling, is a convenient technique for estimating the magnitudes of various terms in the governing equations for a particular type of motion. In scaling, typical expected values of the following quantities are specified:

1. Magnitudes of the field variables
2. Amplitudes of fluctuations in the field variables
3. Characteristic length, depth, and time scales on which the fluctuations occur

These typical values are then used to compare the magnitudes of various terms in the governing equations. For example, in a typical midlatitude synoptic[6] cyclone, the surface pressure might fluctuate by 10hPa over a horizontal distance of 1000 km. Designating the amplitude of the horizontal pressure fluctuation by δp, the horizontal coordinates by x and y, and the horizontal scale by L, the magnitude of the horizontal pressure gradient may be estimated by dividing δp by the length L to get

$$\left(\frac{\partial p}{\partial x}, \frac{\partial p}{\partial y}\right) \sim \frac{\delta p}{L} = 10\,\mathrm{hpa}/10^{3}\,\mathrm{km}\ \left(10^{-3}\,\mathrm{Pa\,m}^{-1}\right)$$

Pressure fluctuations of similar magnitudes occur in other motion systems of vastly different scale such as tornadoes, squall lines, and hurricanes. Thus, the horizontal pressure gradient can range over several orders of magnitude for systems of meteorological interest. Similar considerations are also valid for derivative terms involving other field variables. Therefore, the nature of the dominant terms in the governing equations is crucially dependent on the horizontal scale of the motions. In particular, motions with horizontal scales of only a few kilometers tend to have short time scales so that terms involving the rotation of Earth are negligible, while for larger scales they become very important. Because the character of atmospheric motions depends so strongly on the horizontal scale, this scale provides a convenient method for the classification of motion systems. Table 1.4 classifies examples of various types of motions by horizontal scale for the spectral region from 10^{-7} to 10^{7} m. In the following chapters, scaling arguments are used extensively in developing simplifications of the governing equations for use in modeling various types of motion systems.

Table 1.4 Scales of Atmospheric Motions

Type of Motion	**Horizontal Scale (m)**
Molecular mean free path	10^{-7}
Minute turbulent eddies	10^{-2} to 10^{-1}
Small eddies	10^{-1} to 1
Dust devils	1 to 10
Gusts	10 to 10^{2}
Tornadoes	10^{2}
Cumulonimbus clouds	10^{3}
Fronts, squall lines	10^{4} to 10^{5}
Hurricanes	10^{5}
Synoptic cyclones	10^{6}
Planetary waves	10^{7}

Suggested References

Complete reference information is provided in the Bibliography at the end of the book.

Curry and Webster. Curry and Webster, *Thermodynamics of Atmospheres and Oceans*, contains a more advanced discussion of atmospheric statistics.

Durran, 1993. Durran (1993) discusses the constant angular momentum oscillation in detail.

Wallace and Hobbs. Wallace and Hobbs, *Atmospheric Science: An Introductory Survey*, discuss much of the material in this chapter at an elementary level.

Problems

1.1. Neglecting the latitudinal variation in the radius of Earth, calculate the angle between the gravitational force and gravity vectors at the surface of Earth as a function of latitude. What is the maximum value of this angle?

1.2. Calculate the altitude at which an artificial satellite orbiting in the equatorial plane can be a synchronous satellite (i.e., can remain above the same spot on the surface of Earth).

1.3. An artificial satellite is placed in a natural synchronous orbit above the equator and is attached to Earth below by a wire. A second satellite is attached to the first by a wire of the same length and is placed in orbit directly above the first at the same angular velocity. Assuming that the wires have zero mass, calculate the tension in the wires per unit mass of satellite. Could this tension be used to lift objects into orbit with no additional expenditure of energy?

1.4. A train is running smoothly along a curved track at the rate of 50 ms^{-1}. A passenger standing on a set of scales observes that his weight is 10% greater than when the train is at rest. The track is banked so that the force acting on the passenger is normal to the floor of the train. What is the radius of curvature of the track?

1.5. If a baseball player throws a ball a horizontal distance of 100 m at 30° latitude in 4 s, by how much is the ball deflected laterally as a result of the rotation of Earth?

1.6. Two balls 4 cm in diameter are placed 100 m apart on a frictionless horizontal plane at 43° N. If the balls are impulsively propelled directly at each other with equal speeds, at what speed must they travel so that they just miss each other?

1.7. A locomotive of mass 2×10^5kg travels 50 ms^{-1} along a straight horizontal track at 43° N. What lateral force is exerted on the rails? Compare the magnitudes of the upward reaction force exerted by the rails for cases where the locomotive is traveling eastward and westward, respectively.

1.8. Find the horizontal displacement of a body dropped from a fixed platform at a height h at the equator, neglecting the effects of air resistance. What is the numerical value of the displacement for $h = 5$ km?

1.9. A bullet is fired directly upward with initial speed w_0 at latitude ϕ. Neglecting air resistance, by what distance will it be displaced horizontally when it returns to the ground? (Neglect $2\Omega u \cos\phi$ compared to g in the vertical momentum equation.)

1.10. A block of mass $M = 1$kg is suspended from the end of a weightless string. The other end of the string is passed through a small hole in a horizontal platform and a ball of mass $m = 10$ kg is attached. At what angular velocity must the ball rotate on the horizontal platform to balance the weight of the block if the horizontal distance of the ball from the hole is 1 m? While the ball is rotating, the block is pulled down 10 cm. What is the new angular velocity of the ball? How much work is done in pulling down the block?

1.11. A particle is free to slide on a horizontal frictionless plane located at a latitude ϕ on Earth. Find the equation governing the path of the particle if it is given an impulsive northward velocity $v = V_0$ at $t = 0$. Give the solution for the position of the particle as a function of time. (Assume that the latitudinal excursion is sufficiently small that f is constant.)

1.12. Calculate the 1000 to 500hPa thickness for isothermal conditions with temperatures of 273 and 250K, respectively.

1.13. Isolines of 1000 to 500hPa thickness are drawn on a weather map using a contour interval of 60 m. What is the corresponding layer mean temperature interval?

1.14. Show that a homogeneous atmosphere (density independent of height) has a finite height that depends only on the temperature at the lower boundary. Compute the height of a homogeneous atmosphere with surface temperature $T_0 = 273$ K and surface pressure 1000hPa. (Use the ideal gas law and hydrostatic balance.)

1.15. For the conditions of the previous problem, compute the variation of the temperature with respect to height.

1.16. Show that in an atmosphere with uniform *lapse rate* γ (where $\gamma \equiv -dT/dz$) the geopotential height at pressure level p_1 is given by

$$Z = \frac{T_0}{\gamma}\left[1 - \left(\frac{p_0}{p_1}\right)^{-R\gamma/g}\right]$$

where T_0 and p_0 are the sea-level temperature and pressure, respectively.

1.17. Calculate the 1000 to 500hPa thickness for a constant lapse rate atmosphere with $\gamma = 6.5\,\mathrm{K\,km^{-1}}$ and $T\,0 = 273$ K. Compare your results with the results in Problem 1.12.

1.18. Derive an expression for the variation in density with respect to height in a constant lapse rate atmosphere.

1.19. Derive an expression for the altitude variation in the pressure change δp that occurs when an atmosphere with a constant lapse rate is subjected to a height-independent temperature change δT while the surface pressure remains constant. At what height is the magnitude of the pressure change a maximum if the lapse rate is 6.5 K km^{-1}, $T_0 = 300$, and $\delta T = 2\,\mathrm{K}$?

MATLAB Exercises

M1.1. This exercise investigates the role of the curvature terms for high-latitude constant angular momentum trajectories.

(a) Run the **coriolis.m** script with the following initial conditions: initial latitude 60°, initial velocity $u = 0$, $v = 40$ m s^{-1}, run time = 5 days. Compare the appearance of the trajectories for the case with the curvature terms included and the case with the curvature terms neglected. Qualitatively explain the difference that you observe. Why is the trajectory not a closed circle as described in Eq. (1.15) of the text? [*Hint*: Consider the separate effects of the term proportional to tan ϕ and of the spherical geometry.]

(b) Run **coriolis.m** with latitude 60°, $u = 0$, $v = 80$ m/s. What is different from case (a)? By varying the run time, see if you can determine how long it takes for the particle to make a full circuit in each case, and compare this to the time given in Eq. (1.24) for $\phi = 60°$.

M1.2. Using the MATLAB script from Problem M1.1, compare the magnitudes of the lateral deflection for ballistic missiles fired eastward and westward at 43° latitude. Each missile is launched at a velocity of 1000 ms^{-1} and travels 1000km. Explain your results. Can the curvature term be neglected in these cases?

M1.3. This exercise examines the strange behavior of constant angular momentum trajectories near the equator. Run the **coriolis.m** script for the following contrasting cases: (a) latitude 0.5°, $u = 20$ $\mathrm{m\ s}^{-1}$, $v = 0$, run time = 20 days and (b) latitude 0.5°, $u = -20$ $\mathrm{m\ s}^{-1}$,$v = 0$, run time = 20 days. Obviously, eastward and westward motion near the equator leads to very different behavior. Briefly explain why the trajectories are so different in these two cases. By running different time intervals, determine the approximate period of oscillation in each case (i.e., the time to return to the original latitude).

M1.4. More strange behavior near the equator. Run the script **const_ang_ mom_ traj1.m** by specifying initial conditions of latitude = 0, $u = 0$, $v = 50$ $\mathrm{m\ s}^{-1}$, and a time of about 5 or 10 days. Notice that the motion is symmetric about the equator and that there is a net eastward drift. Why does providing a parcel with an initial poleward velocity at the equator lead to an eastward average displacement? By trying different initial meridional velocities in the range of 50 to 250 ms^{-1}, determine the approximate dependence of the maximum latitude reached by the ball on the initial meridional velocity. Also determine how the net eastward displacement depends on the initial meridional velocity. Show your results in a table, or plot them using MATLAB.

---

[1] Note that *Hertz* measures frequency in *cycles* per second, not in radians per second.

[2] The U.S. standard atmosphere is a specified vertical profile of atmospheric structure.

[3] Earth revolves around its axis once every *sidereal* day, which is equal to 23 h 56 min 4 s (86,164 s). Thus, $\Omega = 2\pi/(86{,}164\,\mathrm{s}) = 7.292 \times 10^{-5}\,\mathrm{rad\,s}^{-1}$.

[4] For computational convenience, the mean surface pressure is often assumed to equal 1000hPa.

[5] It is important to note the minus sign on the right in this expression!

[6] The term *synoptic* designates the branch of meteorology that deals with the analysis of observations taken over a wide area at or near the same time. This term is commonly used (as here) to designate the characteristic scale of the disturbances that are depicted on weather maps.

Chapter 2

Basic Conservation Laws

Atmospheric motions are governed by three fundamental physical principles: conservation of mass, conservation of momentum, and conservation of energy. The mathematical relations that express these laws may be derived by considering the budgets of mass, momentum, and energy for an infinitesimal *control volume* in the fluid. Two types of control volume are commonly used in fluid dynamics. In the *Eulerian* frame of reference, the control volume consists of a parallelepiped of sides δx, δy, and δz, whose position is fixed relative to the coordinate axes. Mass, momentum, and energy budgets will depend on fluxes caused by the flow of fluid through the boundaries of the control volume. (This type of control volume was used in Section 1.2.1.) In the *Lagrangian* frame, however, the control volume consists of an infinitesimal mass of "tagged" fluid particles; thus, the control volume moves about following the motion of the fluid, always containing the same fluid particles.

The Lagrangian frame is particularly useful for deriving conservation laws, since such laws may be stated most simply in terms of a particular mass element of the fluid. The Eulerian system is, however, more convenient for solving most problems because in that system the field variables are related by a set of partial differential equations in which the independent variables are the coordinates x, y, z, and t. In the Lagrangian system, however, it is necessary to follow the time evolution of the fields for various individual fluid parcels. Thus, the independent variables are x_0, y_0, z_0, and t, where x_0, y_0, and z_0 designate the position that a particular parcel passed through at a reference time t_0.

2.1 Total Differentiation

The conservation laws to be derived in this chapter contain expressions for the rates of change of density, momentum, and thermodynamic energy following the motion of particular fluid parcels. In order to apply these laws in the Eulerian frame, it is necessary to derive a relationship between the rate of change of a field variable following the motion and its rate of change at a fixed point. The former is called the *substantial*, the *total*, or the *material* derivative (it will be denoted D/Dt). The latter is called the *local* derivative (it is merely the partial derivative with respect to time).

To derive a relationship between the total derivative and the local derivative, it is convenient to refer to a particular field variable (temperature, for example). For a given air parcel the location (x, y, z) is a function of t so that $x = x(t), y = y(t)$, $z = z(t)$. Following the parcel, T may then be considered as truly a function only of time, and its rate of change is just the total derivative DT/Dt. To relate the total derivative to the local rate of change at a fixed point, we consider the temperature measured on a balloon that moves with the wind. Suppose that this temperature is T_0 at the point x_0,y_0, z_0 and time t_0. If the balloon moves to the point $x_0 + \delta x$, $y_0 + \delta y$, $z_0 + \delta z$ in a time increment δt, then the temperature change recorded on the balloon, δT, can be expressed in a Taylor series expansion as

$$\delta T = \left(\frac{\partial T}{\partial t}\right)\delta t + \left(\frac{\partial T}{\partial x}\right)\delta x + \left(\frac{\partial T}{\partial y}\right)\delta y + \left(\frac{\partial T}{\partial z}\right)\delta z + \text{(higher-order terms)}$$

Dividing through by δt and noting that δT is the change in temperature following the motion so that

$$\frac{DT}{Dt} \equiv \lim_{\delta t \to 0} \frac{\delta T}{\delta t}$$

we find that in the limit $\delta t \to 0$

$$\frac{DT}{Dt} = \frac{\partial T}{\partial t} + \left(\frac{\partial T}{\partial x}\right)\frac{Dx}{Dt} + \left(\frac{\partial T}{\partial y}\right)\frac{Dy}{Dt} + \left(\frac{\partial T}{\partial z}\right)\frac{Dz}{Dt}$$

is the rate of change of T following the motion.

If we now let

$$\frac{Dx}{Dt} \equiv u, \frac{Dy}{Dt} \equiv v, \frac{Dz}{Dt} \equiv w$$

then u, v, w are the velocity components in the x, y, z directions, respectively, and

$$\frac{DT}{Dt} = \frac{\partial T}{\partial t} + \left(u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} + w\frac{\partial T}{\partial z}\right) \tag{2.1}$$

Using vector notation, this expression may be rewritten as

$$\frac{\partial T}{\partial t} = \frac{DT}{Dt} - \mathbf{U} \cdot \boldsymbol{\nabla} T$$

where $\mathbf{U} = \mathbf{i}u + \mathbf{j}v + \mathbf{k}w$ is the velocity vector. The term $-\mathbf{U} \cdot \boldsymbol{\nabla} T$ is called the temperature *advection*. It gives the contribution to the local temperature change due to air motion. For example, if the wind is blowing from a cold region toward a warm region, $-\mathbf{U} \cdot \boldsymbol{\nabla} T$ will be negative (cold advection) and the advection term will contribute negatively to the local temperature change. Thus, the local rate of change of temperature equals the rate of change of temperature following the motion (i.e., the heating or cooling of individual air parcels) plus the advective rate of change of temperature.

The relationship between the total derivative and the local derivative given for temperature in (2.1) holds for any of the field variables. Furthermore, the total derivative can be defined following a motion field other than the actual wind field.

Example

We may wish to relate the pressure change measured by a barometer on a moving ship to the local pressure change. The surface pressure decreases by 3 hPa per 180 km in the eastward direction. A ship steaming eastward at 10 km/h measures a pressure fall of 1 hPa per 3 h.

What is the pressure change on an island that the ship is passing? If we take the x axis oriented eastward, then the local rate of change of pressure on the island is

$$\frac{\partial p}{\partial t} = \frac{Dp}{Dt} - u\frac{\partial p}{\partial x}$$

where *Dp/Dt* is the pressure change observed by the ship and *u* is the velocity of the ship. Thus,

$$\frac{\partial p}{\partial t} = \frac{-1\ \mathrm{hPa}}{3\ \mathrm{h}} - \left(10\frac{\mathrm{km}}{\mathrm{h}}\right)\left(\frac{-3\ \mathrm{hPa}}{180\ \mathrm{km}}\right) = -\frac{1\ \mathrm{hPa}}{6\ \mathrm{h}}$$

so that the rate of pressure fall on the island is only half the rate measured on the moving ship.

If the total derivative of a field variable is zero, then that variable is a conservative quantity following the motion. The local change is then entirely due to advection. As shown later, field variables that are approximately conserved following the motion play an important role in dynamic meteorology.

2.1.1 Total Differentiation of a Vector in a Rotating System

The conservation law for momentum (Newton's second law of motion) relates the rate of change of the absolute momentum following the motion in an inertial reference frame to the sum of the forces acting on the fluid. For most applications in meteorology it is desirable to refer the motion to a reference frame rotating with Earth. Transformation of the momentum equation to a rotating coordinate system requires a relationship between the total derivative of a vector in an inertial reference frame and the corresponding total derivative in a rotating system.

To derive this relationship, we let **A** be an arbitrary vector whose Cartesian components in an inertial frame are given by

$$\mathbf{A} = \mathbf{i}'A'_x + \mathbf{j}'A'_y + \mathbf{k}'A'_z$$

and whose components in a frame rotating with an angular velocity Ω are

$$\mathbf{A} = \mathbf{i}A_x + \mathbf{j}A_y + \mathbf{k}A_z$$

Letting $D_a\mathbf{A}/Dt$ be the total derivative of **A** in the inertial frame, we can write

$$\begin{aligned}\frac{D_a\mathbf{A}}{Dt} &= \mathbf{i}'\frac{DA'_x}{Dt} + \mathbf{j}'\frac{DA'_y}{Dt} + \mathbf{k}'\frac{DA'_z}{Dt} \\ &= \mathbf{i}\frac{DA_x}{Dt} + \mathbf{j}\frac{DA_y}{Dt} + \mathbf{k}\frac{DA_z}{Dt} + \frac{D_a\mathbf{i}}{Dt}A_x + \frac{D_a\mathbf{j}}{Dt}A_y + \frac{D_a\mathbf{k}}{Dt}A_z\end{aligned}$$

The first three terms on the preceding line can be combined to give

$$\frac{D\mathbf{A}}{Dt} \equiv \mathbf{i}\frac{DA_x}{Dt} + \mathbf{j}\frac{DA_y}{Dt} + \mathbf{k}\frac{DA_z}{Dt}$$

which is just the total derivative of **A** as viewed in the rotating coordinates (i.e., the rate of change in **A** following the relative motion).

The last three terms arise because the directions of the unit vectors (**i**, **j**, **k**) change their orientation in space as Earth rotates. These terms have a simple form for a rotating coordinate system. For example, consider the eastward directed unit vector:

$$\delta\mathbf{i} = \frac{\partial\mathbf{i}}{\partial\lambda}\delta\lambda + \frac{\partial\mathbf{i}}{\partial\phi}\delta\phi + \frac{\partial\mathbf{i}}{\partial z}\delta z$$

For solid-body rotation $\delta\lambda=\Omega\delta t$, $\delta\phi=0$, $\delta z=0$, so that $\delta\mathbf{i}/\delta t=(\partial\mathbf{i}/\partial\lambda)(\delta\lambda/\delta t)$ and taking the limit $\delta t\rightarrow 0$,

$$\frac{D_a\mathbf{i}}{Dt}=\Omega\frac{\partial\mathbf{i}}{\partial\lambda}$$

But from Figures 2.1 and 2.2, the longitudinal derivative of **i** can be expressed as

$$\frac{\partial\mathbf{i}}{\partial\lambda}=\mathbf{j}\sin\phi-\mathbf{k}\cos\phi$$

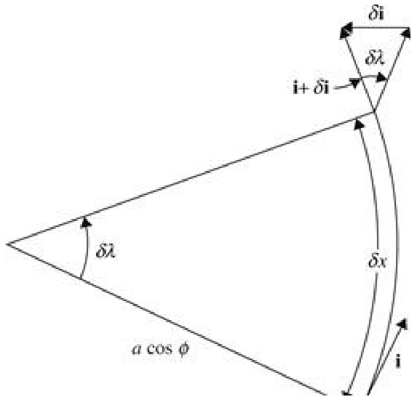


Figure 2.1 Longitudinal dependence of the unit vector i.

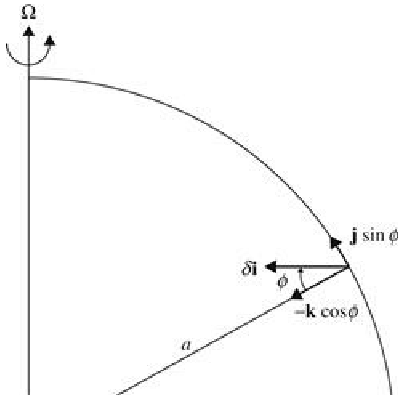


Figure 2.2 Resolution of δi in Figure 2.1 into northward and vertical components.

However, $\boldsymbol{\Omega}=(0,\Omega\cos\phi,\Omega\sin\phi)$ so that

$$\frac{D_a\mathbf{i}}{Dt}=\Omega(\mathbf{j}\sin\phi-\mathbf{k}\cos\phi)=\boldsymbol{\Omega}\times\mathbf{i}$$

In a similar fashion, it can be shown that $D_a\mathbf{j}/Dt=\boldsymbol{\Omega}\times\mathbf{j}$ and $D_a\mathbf{k}/Dt=\boldsymbol{\Omega}\times\mathbf{k}$. Therefore, the total derivative for a vector in an inertial frame is related to that in a rotating frame by the expression

$$\frac{D_a\mathbf{A}}{Dt}=\frac{D\mathbf{A}}{Dt}+\boldsymbol{\Omega}\times\mathbf{A} \tag{2.2}$$

2.2 The Vectorial Form of the Momentum Equation in Rotating Coordinates

In an inertial reference frame, Newton's second law of motion may be written symbolically as

(2.3)

$$\frac{D_a\mathbf{U}_a}{Dt} = \sum\mathbf{F}$$

The left side represents the rate of change of the absolute velocity $\mathbf{U}_a$, following the motion as viewed in an inertial system. The right side represents the sum of the real forces acting per unit mass. In Section 1.3 we found through simple physical reasoning that when the motion is viewed in a rotating coordinate system, certain additional apparent forces must be included if Newton's second law is to be valid. The same result may be obtained by a formal transformation of coordinates in (2.3).

In order to transform this expression into rotating coordinates, we must first find a relationship between $\mathbf{U}_a$ and the velocity relative to the rotating system, which we will designate $\mathbf{U}$. This relationship is obtained by applying (2.2) to the position vector $\mathbf{r}$ for an air parcel on rotating Earth:

(2.4)

$$\frac{D_a\mathbf{r}}{Dt} = \frac{D\mathbf{r}}{Dt} + \boldsymbol{\Omega}\times\mathbf{r}$$

but $D_a\mathbf{r}/Dt \equiv \mathbf{U}_\mathrm{a}$ and $D\,\mathbf{r}/Dt \equiv \mathbf{U}$; therefore, (2.4) may be written as

(2.5)

$$\mathbf{U}_a = \mathbf{U} + \boldsymbol{\Omega}\times\mathbf{r}$$

which states simply that the absolute velocity of an object on rotating Earth is equal to its velocity relative to Earth plus the velocity due to the rotation of Earth.

Now we apply (2.2) to the velocity vector $\mathbf{U}_a$ and obtain

(2.6)

$$\frac{D_a\mathbf{U}_a}{Dt} = \frac{D\mathbf{U}_a}{Dt} + \boldsymbol{\Omega}\times\mathbf{U}_a$$

Substituting from (2.5) into the right side of (2.6) gives

(2.7)

$$\begin{aligned}\frac{D_a\mathbf{U}_a}{Dt} &= \frac{D}{Dt}(\mathbf{U} + \boldsymbol{\Omega}\times\mathbf{r}) + \boldsymbol{\Omega}\times(\mathbf{U} + \boldsymbol{\Omega}\times\mathbf{r})\\ &= \frac{D\mathbf{U}}{Dt} + 2\boldsymbol{\Omega}\times\mathbf{U} - \Omega^2\mathbf{R}\end{aligned}$$

where Ω is assumed to be constant. Here **R** is a vector perpendicular to the axis of rotation, with magnitude equal to the distance to the axis of rotation, so that with the aid of a vector identity,

$$\boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times\mathbf{r}) = \boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times\mathbf{R}) = -\Omega^2\mathbf{R}$$

Equation (2.7) states that the acceleration following the motion in an inertial system equals the rate of change in relative velocity following the relative motion in the rotating frame plus the

Coriolis acceleration due to relative motion in the rotating frame plus the centripetal acceleration caused by the rotation of the coordinates.

If we assume that the only real forces acting on the atmosphere are the pressure gradient force, gravitation, and friction, we can rewrite Newton's second law (2.3) with the aid of (2.7) as

$$\frac{D\mathbf{U}}{Dt} = -2\boldsymbol{\Omega} \times \mathbf{U} - \frac{1}{\rho}\boldsymbol{\nabla} p + \mathbf{g} + \mathbf{F}_r \tag{2.8}$$

where $\mathbf{F}_r$ designates the frictional force (see Section 1.2.2), and the centrifugal force has been combined with gravitation in the gravity term $\mathbf{g}$ (see Section 1.3.2). Equation (2.8) is the statement of Newton's second law for motion relative to a rotating coordinate frame. It states that the acceleration following the relative motion in the rotating frame equals the sum of the Coriolis force, the pressure gradient force, effective gravity, and friction. This form of the momentum equation is basic to most work in dynamic meteorology.

2.3 Component Equations in Spherical Coordinates

For purposes of theoretical analysis and numerical prediction, it is necessary to expand the vectorial momentum equation (2.8) into its scalar components. Since the departure of the shape of Earth from sphericity is entirely negligible for meteorological purposes, it is convenient to expand (2.8) in spherical coordinates so that the (level) surface of Earth corresponds to a coordinate surface. The coordinate axes are then (λ,ϕ,z), where λ is longitude, ϕ is latitude, and z is the vertical distance above the surface of Earth. If the unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are now taken to be directed eastward, northward, and upward, respectively, the relative velocity becomes

$$\mathbf{U} \equiv \mathbf{i}u + \mathbf{j}v + \mathbf{k}w$$

where the components u, v, and w are defined as

$$u \equiv r\cos\phi\frac{D\lambda}{Dt}, \quad v \equiv r\frac{D\phi}{Dt}, \quad w \equiv \frac{Dz}{Dt} \tag{2.9}$$

Here, r is the distance to the center of Earth, which is related to z by $r = a + z$, where a is the radius of Earth. Traditionally, the variable r in (2.9) is replaced by the constant a. This is a very good approximation, since $z \ll a$ for the regions of the atmosphere with which meteorologists are concerned.

For notational simplicity, it is conventional to define x and y as eastward and northward distance, such that $Dx = a\cos\phi\, D\lambda$ and $Dy = a\, D\phi$. Thus, the horizontal velocity components are $u \equiv Dx/Dt$ and $v \equiv Dy/Dt$ in the eastward and northward directions, respectively. The (x, y, z) coordinate system defined in this way is not, however, a Cartesian coordinate system because the directions of the $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ unit vectors are not constant but are functions of position on spherical Earth. This position dependence of the unit vectors must be taken into account when the acceleration vector is expanded into its components on the sphere. Thus, we write

$$\frac{D\mathbf{U}}{Dt} = \mathbf{i}\frac{Du}{Dt} + \mathbf{j}\frac{Dv}{Dt} + \mathbf{k}\frac{Dw}{Dt} + u\frac{D\mathbf{i}}{Dt} + v\frac{D\mathbf{j}}{Dt} + w\frac{D\mathbf{k}}{Dt} \tag{2.10}$$

To obtain the component equations, it is necessary first to evaluate the rates of change of the unit vectors following the motion. We first consider $D\,\mathbf{i}/Dt$. Expanding the total derivative as in (2.1) and noting that $\mathbf{i}$ is a function only of x (i.e., an eastward-directed vector does not change its orientation if the motion is in the north–south or vertical directions), we get

$$\frac{D\mathbf{i}}{Dt} = u\frac{\partial \mathbf{i}}{\partial x}$$

From Figure 2.1 we see by similarity of triangles,

$$\lim_{\delta x \to 0} \frac{|\delta \mathbf{i}|}{\delta x} = \left|\frac{\partial \mathbf{i}}{\partial x}\right| = \frac{1}{a\cos\phi}$$

and that the vector $\partial\mathbf{i}/\partial x$ is directed toward the axis of rotation. Thus, as is illustrated in Figure 2.2,

$$\frac{\partial \mathbf{i}}{\partial x} = \frac{1}{a\cos\phi}(\mathbf{j}\,\sin\phi - \mathbf{k}\cos\phi)$$

Therefore,

$$\frac{D\mathbf{i}}{Dt} = \frac{u}{a\cos\phi}(\mathbf{j}\,\sin\phi - \mathbf{k}\cos\phi) \tag{2.11}$$

Considering now $D\,\mathbf{j}/Dt$, we note that $\mathbf{j}$ is a function only of x and y. Thus, with the aid of Figure 2.3 we see that for eastward motion $|\delta\mathbf{j}| = \delta x/(a/\tan\phi)$. Because the vector $\partial\mathbf{j}/\partial x$ is directed in the negative x direction, we then have

$$\frac{\partial \mathbf{j}}{\partial x} = -\frac{\tan\phi}{a}\mathbf{i}$$

From Figure 2.4 it is clear that for northward motion $|\delta\mathbf{j}| = \delta\phi$, but $\delta y = a\delta\phi$, and $\delta\mathbf{j}$ is directed downward so that

$$\frac{\partial \mathbf{j}}{\partial y} = -\frac{\mathbf{k}}{a}$$

Hence,

$$\frac{D\mathbf{j}}{Dt} = -\frac{u\tan\phi}{a}\mathbf{i} - \frac{v}{a}\mathbf{k} \tag{2.12}$$

Finally, by similar arguments it can be shown that

$$\frac{D\mathbf{k}}{Dt} = \mathbf{i}\frac{u}{a} + \mathbf{j}\frac{v}{a} \tag{2.13}$$

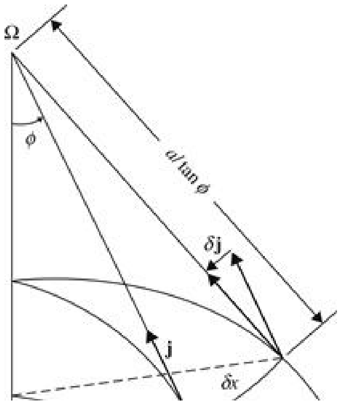


Figure 2.3 Dependence of unit vector j on longitude.

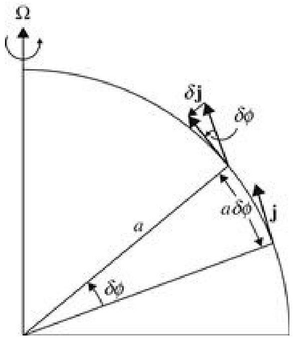


Figure 2.4 Dependence of unit vector j on latitude.

Substituting (2.11), (2.12), and (2.13) into (2.10) and rearranging the terms, we obtain the spherical polar coordinate expansion of the acceleration following the relative motion

$$\begin{aligned}\frac{D\mathbf{U}}{Dt} &= \left(\frac{Du}{Dt} - \frac{uv\tan\phi}{a} + \frac{uw}{a}\right)\mathbf{i} + \left(\frac{Dv}{Dt} + \frac{u^2\tan\phi}{a} + \frac{vw}{a}\right)\mathbf{j} \\ &+ \left(\frac{Dw}{Dt} - \frac{u^2+v^2}{a}\right)\mathbf{k}\end{aligned} \tag{2.14}$$

We next turn to the component expansion of the force terms in (2.8). The Coriolis force is expanded by noting that Ω has no component parallel to **i** and that its components parallel to **j** and **k** are $2\Omega\cos\phi$ and $2\Omega\sin\phi$, respectively. Thus, using the definition of the vector cross-product,

$$\begin{aligned}-2\boldsymbol{\Omega}\times\mathbf{U} &= -2\Omega\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & \cos\phi & \sin\phi \\ u & v & w\end{vmatrix} \\ &= -(2\Omega w\cos\phi - 2\Omega v\sin\phi)\,\mathbf{i} - 2\Omega u\sin\phi\,\mathbf{j} + 2\Omega u\cos\phi\,\mathbf{k}\end{aligned} \tag{2.15}$$

The pressure gradient may be expressed as

$$\nabla p = \mathbf{i}\frac{\partial p}{\partial x} + \mathbf{j}\frac{\partial p}{\partial y} + \mathbf{k}\frac{\partial p}{\partial z} \tag{2.16}$$

and gravity is conveniently represented as

$$\mathbf{g} = -g\,\mathbf{k} \tag{2.17}$$

where g is a positive scalar ($g \cong 9.8 \text{ m s}^{-2}$ at Earth's surface). Finally, recall from Section 1.2.2 that

$$\mathbf{F}_r = \mathbf{i}F_{rx} + \mathbf{j}F_{ry} + \mathbf{k}F_{rz} \tag{2.18}$$

Substituting (2.14) through (2.18) into the equation of motion (2.8) and equating all terms in the **i**, **j**, and **k** directions, respectively, we obtain

$$\frac{Du}{Dt} - \frac{uv\tan\phi}{a} + \frac{uw}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + 2\Omega v\sin\phi - 2\Omega w\cos\phi + F_{rx} \tag{2.19}$$

$$\frac{Dv}{Dt} + \frac{u^2\tan\phi}{a} + \frac{vw}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial y} - 2\Omega u\sin\phi + F_{ry} \tag{2.20}$$

$$\frac{Dw}{Dt} - \frac{u^2 + v^2}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g + 2\Omega u\cos\phi + F_{rz} \tag{2.21}$$

which are the eastward, northward, and vertical component momentum equations, respectively. The terms proportional to $1/a$ on the left sides in (2.19), (2.20), and (2.21) are called *curvature* terms; they are due to the curvature the Earth.[1] Because they are nonlinear (i.e., they are quadratic in the dependent variables), they are difficult to handle in theoretical analyses. Fortunately, as shown in the next section, the curvature terms are unimportant for midlatitude synoptic-scale motions. However, even when the curvature terms are neglected, (2.19), (2.20), and (2.21) are still nonlinear partial differential equations, as can be seen by expanding the total derivatives into their local and advective parts:

$$\frac{Du}{Dt} = \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}$$

with similar expressions for Dv/Dt and Dw/Dt. In general, the advective acceleration terms are comparable in magnitude to the local acceleration term. The presence of nonlinear advection processes is one reason that dynamic meteorology is an interesting and challenging subject.

2.4 Scale Analysis of the Equations of Motion

Section 1.6 discussed the basic notion of scaling the equations of motion in order to determine whether some terms in the equations are negligible for motions of meteorological concern. Elimination of terms on scaling considerations have the advantage of simplifying the mathematics; as shown in later chapters, the elimination of small terms in some cases has the very important property of completely eliminating or *filtering* an unwanted type of motion. The complete equations of motion—(2.19), (2.20), and (2.21)—describe all types and scales of atmospheric motions. Sound waves, for example, are a perfectly valid class of solutions to these equations. However, sound waves are of negligible importance in dynamical meteorology. Therefore, it will be a distinct advantage if, as turns out to be true, we can neglect the terms that lead to the production of sound waves and filter out this unwanted class of motions.

To simplify (2.19), (2.20), and (2.21) for synoptic-scale motions, we define the following characteristic scales of the field variables based on observed values for midlatitude synoptic systems.

$U \sim 10 \text{ m s}^{-1}$	horizontal velocity scale
$W \sim 1 \text{ cm s}^{-1}$	vertical velocity scale
$L \sim 10^6 \text{ m}$	length scale [$\sim 1/(2\pi)$ wavelength]
$H \sim 10^4 \text{ m}$	depth scale
$\delta P/\rho \sim 10^3 \text{ m}^2\, s^{-2}$	horizontal pressure fluctuation scale
$L/U \sim 10^5 \text{ s}$	time scale

Horizontal pressure fluctuation δP is normalized by the density ρ in order to produce a scale estimate that is valid at all heights in the troposphere, despite the approximate exponential decrease with height of both δP and ρ. Note that $\delta P/\rho$ has units of a geopotential. Referring back to (1.31), we see that indeed the magnitude of the fluctuation of $\delta P/\rho$ on a surface of constant height must equal the magnitude of the fluctuation of the geopotential on an isobaric surface. The time scale here is an advective one, which is appropriate for pressure systems that move at approximately the speed of the horizontal wind, as is observed for synoptic-scale motions. Thus, L/U is the time required to travel a distance L at a speed U, and the substantial differential operator scales as $D/Dt \sim U/L$ for such motions.

It should be pointed out here that the synoptic-scale vertical velocity is not a directly measurable quantity. However, as shown in Chapter 3, the magnitude of w can be deduced from knowledge of the horizontal velocity field.

We can now estimate the magnitude of each term in (2.19) and (2.20) for synoptic-scale motions at a given latitude. It is convenient to consider a disturbance centered at latitude $\phi_0 = 45°$ and introduce the notation

$$f_0 = 2\Omega \sin\phi_0 = 2\Omega \cos\phi_0 \cong 10^{-4}\text{s}^{-1}$$

Table 2.1 shows the characteristic magnitude of each term in (2.19) and (2.20) based on the scaling considerations given before. The molecular friction term is so small that it may be neglected for all motions except the smallest- scale turbulent motions near the ground, where vertical wind shears shear can become very large and the molecular friction term must be retained, as discussed in Chapter 8.

Table 2.1 Scale Analysis of the Horizontal Momentum Equations

	A	B	C	D	E	F	G
$x-$Eq.	$\frac{Du}{Dt}$	$-2\Omega v\sin\phi$	$+2\Omega w\cos\phi$	$+\frac{uw}{a}$	$-\frac{uv\tan\phi}{a}$	$=-\frac{1}{\rho}\frac{\partial p}{\partial x}$	$+F_{rx}$
$y-$Eq.	$\frac{Dv}{Dt}$	$+2\Omega u\sin\phi$		$+\frac{vw}{a}$	$+\frac{u^2\tan\phi}{a}$	$=-\frac{1}{\rho}\frac{\partial p}{\partial y}$	$+F_{ry}$
Scales	U^2/L	f_0U	f_0W	$\frac{UW}{a}$	$\frac{U^2}{a}$	$\frac{\delta P}{\rho L}$	$\frac{\nu U}{H^2}$
($\mathrm{m\ s^{-2}}$)	10^{-4}	10^{-3}	10^{-6}	10^{-8}	10^{-5}	10^{-3}	10^{-12}

2.4.1 Geostrophic Approximation and Geostrophic Wind

It is apparent from Table 2.1 that for midlatitude synoptic-scale disturbances the Coriolis force (term B) and the pressure gradient force (term F) are in approximate balance. Retaining only these two terms in (2.19) and (2.20) gives as a first approximation the *geostrophic* relationship

$$-fv \approx -\frac{1}{\rho}\frac{\partial p}{\partial x}; \qquad fu \approx -\frac{1}{\rho}\frac{\partial p}{\partial y} \tag{2.22}$$

where $f \equiv 2\Omega\sin\phi$ is called the *Coriolis parameter*. The geostrophic balance is a *diagnostic* expression that gives the approximate relationship between the pressure field and horizontal velocity in large-scale extratropical systems. The approximation (2.22) contains no reference to time and therefore cannot be used to predict the evolution of the velocity field. It is for this reason that the geostrophic relationship is called a diagnostic relationship.

By analogy to the geostrophic approximation (2.22), it is possible to define a horizontal velocity field, $\mathbf{V}_g \equiv \mathbf{i}u_g + \mathbf{j}v_g$, called the *geostrophic wind*, that satisfies (2.22) identically. In vectorial form,

$$\mathbf{V}_g \equiv \mathbf{k} \times \frac{1}{\rho f}\boldsymbol{\nabla} p \tag{2.23}$$

Thus, knowledge of the pressure distribution at any time determines the geostrophic wind. It should be kept clearly in mind that (2.23) always defines the geostrophic wind; however, only for large-scale motions away from the equator should the geostrophic wind be used as an approximation to the actual horizontal wind field. For the scales used in Table 2.1, the geostrophic wind approximates the true horizontal velocity to within 10 to 15% in midlatitudes.

2.4.2 Approximate Prognostic Equations: The Rossby Number

To obtain prediction equations, it is necessary to retain the acceleration (term A) in (2.19) and (2.20). The resulting approximate horizontal momentum equations are

$$\frac{Du}{Dt} = fv - \frac{1}{\rho}\frac{\partial p}{\partial x} = f\left(v - v_g\right) = fv_a \tag{2.24}$$

$$\frac{Dv}{Dt} = -fu - \frac{1}{\rho}\frac{\partial p}{\partial y} = -f\left(u - u_g\right) = -fu_a \tag{2.25}$$

where (2.23) is used to rewrite the pressure gradient force in terms of the geostrophic wind. Because the acceleration terms in (2.24) and (2.25) are proportional to the difference between the actual wind and the geostrophic wind, they are about an order of magnitude smaller than the Coriolis force and the pressure gradient force, in agreement with our scale analysis. In the final equalities of (2.24) and (2.25) we define the difference between the actual wind and the geostrophic wind as the *ageostrophic wind.*

The fact that the horizontal flow is in approximate geostrophic balance is helpful for diagnostic analysis. However, it makes actual applications of these equations in weather prognosis difficult because acceleration (which must be measured accurately) is given by the small difference between two large terms. Thus, a small error in measurement of either velocity or pressure gradient will lead to very large errors in estimating the acceleration. This problem of numerical weather prediction is discussed in Chapter 13, and the problem of simplifying the equations for physical understanding based on the smallness of the ageostrophic wind will be taken up in Chapter 6.

A convenient measure of the magnitude of the acceleration compared to the Coriolis force may be obtained by forming the ratio of the characteristic scales for the acceleration and Coriolis force terms: $(U^2/L)/(f_0U)$. This ratio is a nondimensional number called the *Rossby number* after the Swedish meteorologist C. G. Rossby (1898–1957) and is designated by

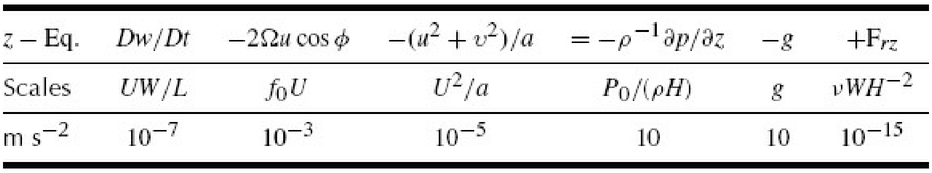
$$\mathbf{Ro} \equiv U/(f_0 L)$$

Thus, the smallness of the Rossby number is a measure of the validity of the geo- strophic approximation.

2.4.3 The Hydrostatic Approximation

A similar scale analysis can be applied to the vertical component of the momentum equation (2.21). Because pressure decreases by about an order of magnitude from the ground to the tropopause, the vertical pressure gradient may be scaled by P_0/H, where P_0 is the surface pressure and H is the depth of the troposphere. The terms in (2.21) may then be estimated for synoptic-scale motions and are shown in Table 2.2. As with the horizontal component equations, we consider motions centered at 45° latitude and neglect friction. The scaling indicates that to a high degree of accuracy the pressure field is in *hydrostatic equilibrium*; that is, the pressure at any point is simply equal to the weight of a unit cross-section column of air above that point.

Table 2.2 Scale Analysis of the Vertical Momentum Equation



z – Eq.	Dw/Dt	$-2\Omega u\cos\phi$	$-(u^2+v^2)/a$	$= -\rho^{-1}\partial p/\partial z$	$-g$	$+F_{rz}$
Scales	UW/L	f_0U	U^2/a	$P_0/(\rho H)$	g	νWH^{-2}
m s^{-2}	10^{-7}	10^{-3}	10^{-5}	10	10	10^{-15}

The preceding analysis of the vertical momentum equation is, however, somewhat misleading. It is not sufficient to show merely that the vertical acceleration is small compared to g. Because only that part of the pressure field that varies horizontally is directly coupled to the horizontal velocity field, it is actually necessary to show that the horizontally varying pressure component is itself in hydrostatic equilibrium with the horizontally varying density field. To show this, it is convenient to first define a standard pressure $p_0(z)$, which is the horizontally averaged pressure at each height, and a corresponding standard density $\rho_0(z)$, defined so that $p_0(z)$ and $\rho_0(z)$ are in *exact* hydrostatic balance:

$$\frac{1}{\rho_0}\frac{dp_0}{dz} \equiv -g \tag{2.26}$$

We may then write the total pressure and density fields as

$$\begin{aligned} p(x,y,z,t) &= p_0(z) + p'(x,y,z,t) \\ \rho(x,y,z,t) &= \rho_0(z) + \rho'(x,y,z,t) \end{aligned} \tag{2.27}$$

where p' and ρ' are deviations from the standard values of pressure and density, respectively. For an atmosphere at rest, p' and ρ' would thus be zero. Using the definitions of (2.26) and (2.27) and assuming that ρ'/ρ_0 is much less than unity in magnitude so that $(\rho_0 + \rho')^{-1} \cong \rho_0^{-1}$ $(1 - \rho'/\rho_0)$, we find that

$$\begin{aligned} -\frac{1}{\rho}\frac{\partial p}{\partial z} - g &= -\frac{1}{(\rho_0 + \rho')}\frac{\partial}{\partial z}(p_0 + p') - g \\ &\approx \frac{1}{\rho_0}\left[\frac{\rho'}{\rho_0}\frac{dp_0}{dz} - \frac{\partial p'}{\partial z}\right] = -\frac{1}{\rho_0}\left[\rho' g + \frac{\partial p'}{\partial z}\right] \end{aligned} \tag{2.28}$$

For synoptic-scale motions, the terms in (2.28) have the magnitudes

$$\frac{1}{\rho_0}\frac{\partial p'}{\partial z} \sim \left[\frac{\delta P}{\rho_0 H}\right] \sim 10^{-1}\,\mathrm{m\,s^{-2}}, \quad \frac{\rho' g}{\rho_0} \sim 10^{-1}\,\mathrm{m\,s^{-2}}$$

Comparing these with the magnitudes of other terms in the vertical momentum equation (refer to Table 2.2), we see that to a very good approximation the perturbation pressure field is in hydrostatic equilibrium with the perturbation density field so that

$$\frac{\partial p'}{\partial z} + \rho' g = 0 \tag{2.29}$$

Therefore, for synoptic-scale motions, vertical accelerations are negligible and the vertical velocity cannot be determined from the vertical momentum equation. However, we show in Chapter 3 that it is, nevertheless, possible to deduce the vertical motion field indirectly. Moreover, in Chapter 6 we show that the vertical motion field may be estimated, and physically understood, from knowledge of the pressure field alone.

2.5 The Continuity Equation

We turn now to the second of the three fundamental conservation principles: conservation of mass. The mathematical relationship that expresses conservation of mass for a fluid is called the *continuity equation*. This section develops the continuity equation using two alternative methods. The first method is based on an Eulerian control volume, whereas the second is based on a Lagrangian control volume.

2.5.1 A Eulerian Derivation

We consider a volume element $\delta x \delta y \delta z$ that is fixed in a Cartesian coordinate frame as shown in Figure 2.5. For such a fixed control volume, the net rate of mass inflow through the sides must equal the rate of accumulation of mass within the volume. The rate of inflow of mass through the left face per unit area is

$$\left[\rho u - \frac{\partial}{\partial x}(\rho u)\frac{\delta x}{2}\right]$$

whereas the rate of outflow per unit area through the right face is

$$\left[\rho u + \frac{\partial}{\partial x}(\rho u)\frac{\delta x}{2}\right]$$

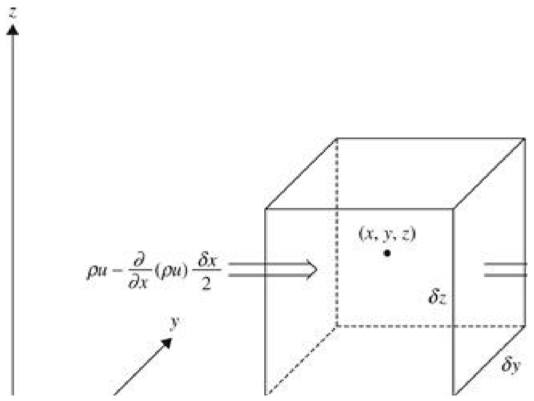

z
(x, y, z)
$\rho u - \frac{\partial}{\partial x}(\rho u)\frac{\delta x}{2}$
δz
y
δy

Figure 2.5 Mass inflow into a fixed (Eulerian) control volume as a result of motion parallel to the *x* axis.

Because the area of each of these faces is $\delta y \delta z$, the net rate of flow into the volume due to the *x* velocity component is

$$\left[\rho u - \frac{\partial}{\partial x}(\rho u)\frac{\delta x}{2}\right]\delta y \delta z - \left[\rho u + \frac{\partial}{\partial x}(\rho u)\frac{\delta x}{2}\right]\delta y \delta z = -\frac{\partial}{\partial x}(\rho u)\delta x \delta y \delta z$$

Similar expressions obviously hold for the *y* and *z* directions. Thus, the net rate of mass inflow is

$$-\left[\frac{\partial}{\partial x}(\rho u) + \frac{\partial}{\partial y}(\rho v) + \frac{\partial}{\partial z}(\rho w)\right]\delta x \delta y \delta z$$

and the mass inflow per unit volume is just $-\nabla \cdot (\rho \mathbf{U})$, which must equal the rate of mass increase per unit volume. Now the increase of mass per unit volume is just the local density change $\partial \rho / \partial t$. Therefore,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{U}) = 0 \tag{2.30}$$

Equation (2.30) is the mass divergence form of the continuity equation.

An alternative form of the continuity equation is obtained by applying the vector identity

$$\nabla \cdot (\rho \mathbf{U}) \equiv \rho \nabla \cdot \mathbf{U} + \mathbf{U} \cdot \nabla \rho$$

and the relationship

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{U} \cdot \boldsymbol{\nabla}$$

to get

$$\frac{1}{\rho}\frac{D\rho}{Dt} + \boldsymbol{\nabla} \cdot \mathbf{U} = 0 \tag{2.31}$$

Equation (2.31) is the velocity divergence form of the continuity equation. It states that the fractional rate of increase of the density *following the motion* of an air parcel is equal to minus the velocity divergence (i.e., convergence). This should be clearly distinguished from (2.30), which states that the *local* rate of change of density is equal to minus the mass divergence.

2.5.2 A Lagrangian Derivation

The physical meaning of divergence can be illustrated by the following alternative derivation of (2.31). Consider a control volume of fixed mass δM that moves with the fluid. Letting $\delta V = \delta x \delta y \delta z$ be the volume, we find that because $\delta M = \rho \delta V = \rho \delta x \delta y \delta z$ is conserved following the motion, we can write

$$\frac{1}{\delta M}\frac{D}{Dt}(\delta M) = \frac{1}{\rho \delta V}\frac{D}{Dt}(\rho \delta V) = \frac{1}{\rho}\frac{D\rho}{Dt} + \frac{1}{\delta V}\frac{D}{Dt}(\delta V) = 0 \tag{2.32}$$

but

$$\frac{1}{\delta V}\frac{D}{Dt}(\delta V) = \frac{1}{\delta x}\frac{D}{Dt}(\delta x) + \frac{1}{\delta y}\frac{D}{Dt}(\delta y) + \frac{1}{\delta z}\frac{D}{Dt}(\delta z)$$

Referring to Figure 2.6, we see that the faces of the control volume in the *y*, *z* plane (designated A and B) are advected with the flow in the *x* direction at speeds $u_{\mathrm{A}} = Dx/Dt$ and $u_{\mathrm{B}} = D(x + \delta x)/Dt$, respectively. Thus, the difference in speeds of the two faces is $\delta u = u_{\mathrm{B}} - u_{\mathrm{A}} = D(x + \delta x)/Dt - Dx/Dt$ or $\delta u = D(\delta x)/Dt$. Similarly, $\delta v = D(\delta y)/Dt$ and $\delta w = D(\delta z)/Dt$. Therefore,

$$\lim_{\delta x, \delta y, \delta z \to 0}\left[\frac{1}{\delta V}\frac{D}{Dt}(\delta V)\right] = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = \boldsymbol{\nabla} \cdot \mathbf{U}$$

so that in the limit $\delta V \to 0$, (2.32) reduces to the continuity equation (2.31); the divergence of the three-dimensional velocity field is equal to the fractional rate of change of the volume of a fluid parcel in the limit $\delta V \to 0$. It is left as a problem for the student to show that the divergence of the *horizontal* velocity field is equal to the fractional rate of change of the horizontal area δA of a fluid parcel in the limit $\delta A \to 0$.

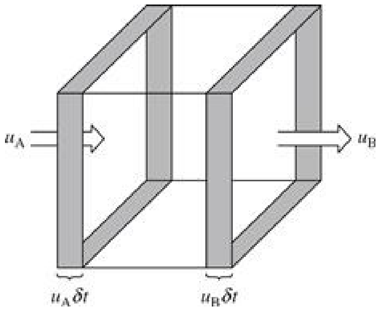


Figure 2.6 Change in Lagrangian control volume (shown by shading) as a result of fluid motion parallel to the *x* axis.

2.5.3 Scale Analysis of the Continuity Equation

Following the technique developed in Section 2.4.3, and again assuming that $|\rho'/\rho_0| \ll 1$, we can approximate the continuity ((2.31)) as

$$\underset{\text{A}}{\frac{1}{\rho_0}\left(\frac{\partial \rho'}{\partial t} + \mathbf{U}\cdot\boldsymbol{\nabla}\rho'\right)} + \underset{\text{B}}{\frac{w}{\rho_0}\frac{d\rho_0}{dz}} + \underset{\text{C}}{\boldsymbol{\nabla}\cdot\mathbf{U}} \approx 0 \tag{2.33}$$

where ρ' designates the local deviation of density from its horizontally averaged value, $\rho_0(z)$. For synoptic-scale motions, $\rho'/\rho_0 \sim 10^{-2}$ so that, using the characteristic scales given in Section 2.4, we find that term A has magnitude

$$\frac{1}{\rho_0}\left(\frac{\partial \rho'}{\partial t} + \mathbf{U}\cdot\boldsymbol{\nabla}\rho'\right) \sim \frac{\rho'}{\rho_0}\frac{U}{L} \approx 10^{-7}\mathrm{s}^{-1}$$

For motions in which the depth scale *H* is comparable to the density scale height, $d\ln\rho_0/dz \sim H^{-1}$, so that term B scales as

$$\frac{w}{\rho_0}\frac{d\rho_0}{dz}\frac{W}{H} \approx 10^{-6}\mathrm{s}^{-1}$$

Expanding term C in Cartesian coordinates, we have

$$\boldsymbol{\nabla}\cdot\mathbf{U} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}$$

For synoptic-scale motions, the terms $\partial u/\partial x$ and $\partial v/\partial y$ tend to be of equal magnitude but opposite sign. Thus, they tend to balance so that

$$\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) 10^{-1}\frac{U}{L} \approx 10^{-6}\,\mathrm{s}^{-1}$$

In addition,

$$\frac{\partial w}{\partial z} \sim \frac{W}{H} \approx 10^{-6}\,\mathrm{s}^{-1}$$

Thus, terms B and C are each an order of magnitude greater than term A, and, to a first approximation, terms B and C balance in the continuity equation. To a good approximation, then

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}+w\frac{d}{dz}\left(\ln \rho_0\right)=0$$

or, alternatively, in vector form

$$\nabla\cdot\left(\rho_0\mathbf{U}\right)=0 \tag{2.34}$$

Thus, for synoptic-scale motions the mass flux computed using the basic state density ρ_0 is nondivergent. This approximation is similar to the idealization of incompressibility, which is often used in fluid mechanics. However, an *incompressible* fluid has a density constant following the motion

$$\frac{D\rho}{Dt}=0$$

Thus, by (2.31) the velocity divergence vanishes ($\nabla\cdot\mathbf{U}=0$) in an incompressible fluid, which is not the same as (2.34). Our approximation (2.34) shows that for purely horizontal flow the atmosphere behaves as though it were an incompressible fluid. However, when there is vertical motion, the compressibility associated with the height dependence of ρ_0 must be taken into account.

2.6 The Thermodynamic Energy Equation

We now turn to the third fundamental conservation principle: the conservation of energy as applied to a moving fluid element. The first law of thermodynamics is usually derived by considering a system in thermodynamic equilibrium—that is, a system that is initially at rest and, after exchanging heat with its surroundings and doing work on the surroundings, is again at rest. For such a system the first law states that *the change in internal energy of the system is equal to the difference between the heat added to the system and the work done by the system.* A Lagrangian control volume consisting of a specified mass of fluid may be regarded as a thermodynamic system. However, unless the fluid is at rest, it will not be in thermodynamic equilibrium. Nevertheless, the first law of thermodynamics still applies.

To show that this is the case, we note that the total thermodynamic energy of the control volume is considered to consist of the sum of the internal energy (due to the kinetic energy of the individual molecules) and the kinetic energy due to the macroscopic motion of the fluid. The rate of change of this total thermodynamic energy is equal to the rate of diabatic heating plus the rate at which work is done on the fluid parcel by external forces.

If we let e designate the internal energy per unit mass, then the total thermodynamic energy contained in a Lagrangian fluid element of density ρ and volume δV is $\rho[e+(1/2)\mathbf{U}\cdot\mathbf{U}]\,\delta V$. The external forces that act on a fluid element may be divided into surface forces, such as pressure and viscosity, and body forces, such as gravity or the Coriolis force. The rate at which work is done on the fluid element by the x component of the pressure force is illustrated in Figure 2.7. Recalling that pressure is a force per unit area and that the rate at which a force does work is given by the dot product of the force and velocity vectors, we see that the rate at which the surrounding fluid does work on the element due to the pressure force on the two boundary surfaces in the y, z plane is given by

$$(pu)_A\delta y\delta z - (pu)_B\delta y\delta z$$

(The negative sign is needed before the second term because the work done on the fluid element is positive if u is negative across face B.) Now, by expanding in a Taylor series, we can write

$$(pu)_B = (pu)_A + \left[\frac{\partial}{\partial x}(pu)\right]_A \delta x + \cdots$$

Thus, the net rate at which the pressure force does work due to the x component of motion is

$$[(pu)_A - (pu)_B]\delta y\delta z = -\left[\frac{\partial}{\partial x}(pu)\right]_A \delta V$$

where $\delta V = \delta x\delta y\delta z$.

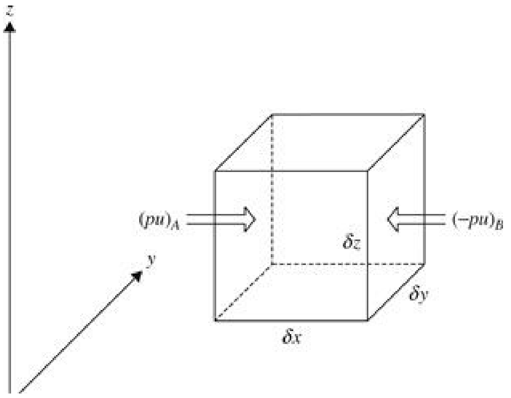


Figure 2.7 Rate of working on a fluid element due to the x component of the pressure force.

Similarly, we can show that the net rates at which the pressure force does work due to the y and z components of motion are

$$-\left[\frac{\partial}{\partial y}(pv)\right]\delta V \text{ and } -\left[\frac{\partial}{\partial z}(pw)\right]\delta V$$

respectively. Hence, the total rate at which work is done by the pressure force is simply

$$-\boldsymbol{\nabla}\cdot(p\mathbf{U})\delta V$$

The only body forces of meteorological significance that act on an element of mass in the atmosphere are the Coriolis force and gravity. However, because the Coriolis force, $-2\boldsymbol{\Omega}\times\mathbf{U}$, is perpendicular to the velocity vector, it can do no work. Thus, the rate at which body forces do work on the mass element is just $\rho\mathbf{g}\cdot\mathbf{U}\,\delta V$.

Applying the principle of energy conservation to our Lagrangian control volume (neglecting effects of molecular viscosity), we thus obtain

$$\frac{D}{Dt}\left[\rho\left(e+\frac{1}{2}\mathbf{U}\cdot\mathbf{U}\right)\delta V\right]=-\boldsymbol{\nabla}\cdot(p\mathbf{U})\delta V+\rho\mathbf{g}\cdot\mathbf{U}\delta V+\rho J\delta V \tag{2.35}$$

Here J is the rate of heating per unit mass due to radiation, conduction, and latent heat release. With the aid of the chain rule of differentiation, we can rewrite (2.35) as

$$\begin{gathered}\rho\delta V\frac{D}{Dt}\left(e+\frac{1}{2}\mathbf{U}\cdot\mathbf{U}\right)+\left(e+\frac{1}{2}\mathbf{U}\cdot\mathbf{U}\right)\frac{D\left(\rho\delta V\right)}{Dt}\\=-\mathbf{U}\cdot\boldsymbol{\nabla}p\,\delta V-p\boldsymbol{\nabla}\cdot\mathbf{U}\,\delta V-\rho gw\,\delta V+\rho J\,\delta V\end{gathered} \tag{2.36}$$

where we have used $\mathbf{g}=-g\,\mathbf{k}$. Now, from (2.32), the second term on the left in (2.36) vanishes so that

$$\rho\frac{De}{Dt}+\rho\frac{D}{Dt}\left(\frac{1}{2}\mathbf{U}\cdot\mathbf{U}\right)=-\mathbf{U}\cdot\boldsymbol{\nabla}p-p\boldsymbol{\nabla}\cdot\mathbf{U}-\rho gw+\rho J \tag{2.37}$$

This equation can be simplified by noting that if we take the dot product of $\mathbf{U}$ with the momentum equation (2.8), we obtain (neglecting friction)

$$\rho\frac{D}{Dt}\left(\frac{1}{2}\mathbf{U}\cdot\mathbf{U}\right)=-\mathbf{U}\cdot\boldsymbol{\nabla}p-\rho gw \tag{2.38}$$

Subtracting (2.38) from (2.37), we obtain

$$\rho\frac{De}{Dt}=-p\boldsymbol{\nabla}\cdot\mathbf{U}+\rho J \tag{2.39}$$

The terms in (2.37) that were eliminated by subtracting (2.38) represent the balance of mechanical energy due to the motion of the fluid element; the remaining terms represent the thermal energy balance.

Using the definition of geopotential (1.11), we have

$$gw=g\frac{Dz}{Dt}=\frac{D\Phi}{Dt}$$

so that (2.38) can be rewritten as

$$\rho\frac{D}{Dt}\left(\frac{1}{2}\mathbf{U}\cdot\mathbf{U}+\Phi\right)=-\mathbf{U}\cdot\boldsymbol{\nabla}p \tag{2.40}$$

which is referred to as the *mechanical energy equation*. The sum of the kinetic energy plus the gravitational potential energy is called the *mechanical energy*. Thus, (2.40) states that following the motion, the rate of change of mechanical energy per unit volume equals the rate at which work is done by the pressure gradient force.

The thermal energy equation (2.39) can be written in more familiar form by noting from (2.31) that

$$\frac{1}{\rho}\nabla\cdot\mathbf{U} = -\frac{1}{\rho^2}\frac{D\rho}{Dt} = \frac{D\alpha}{Dt}$$

and that for dry air the internal energy per unit mass is given by $e = c_v T$, where $c_v (= 717 \text{ J kg}^{-1} \text{ K}^{-1})$ is the specific heat at constant volume. We then obtain

$$c_v \frac{DT}{Dt} + p\frac{D\alpha}{Dt} = J \tag{2.41}$$

which is the usual form of the thermodynamic energy equation. Thus, the first law of thermodynamics indeed is applicable to a fluid in motion. The second term on the left, representing the rate of working by the fluid system (per unit mass), represents a conversion between thermal and mechanical energy. This conversion process enables the solar heat energy to drive the motions of the atmosphere.

2.7 Thermodynamics of the Dry Atmosphere

Taking the total derivative of the equation of state (1.25), we obtain

$$p\frac{D\alpha}{Dt} + \alpha\frac{Dp}{Dt} = R\frac{DT}{Dt}$$

Substituting for $pD\alpha/Dt$ in (2.41) and using $c_p = c_v + R$, where $c_p (= 1004 \text{ J kg}^{-1} \text{ K}^{-1})$ is the specific heat at constant pressure, we can rewrite the first law of thermodynamics as

$$c_p\frac{DT}{Dt} - \alpha\frac{Dp}{Dt} = J \tag{2.42}$$

Dividing through by T and again using the equation of state, we obtain the entropy form of the first law of thermodynamics:

$$c_p\frac{D\ln T}{Dt} - R\frac{D\ln p}{Dt} = \frac{J}{T} \equiv \frac{Ds}{Dt} \tag{2.43}$$

Equation (2.43) gives the rate of change of entropy per unit mass following the motion for a thermodynamically *reversible* process. A reversible process is one in which a system changes its thermodynamic state and then returns to the original state without changing its surroundings. For such a process the entropy s defined by (2.43) is a field variable that depends only on the state of the fluid. Thus, Ds is a perfect differential, and Ds/Dt is to be regarded as a total derivative. However, “heat” is not a field variable, so the heating rate J is not a total derivative.[2]

2.7.1 Potential Temperature

For an ideal gas that is undergoing an *adiabatic* process (i.e., a reversible process in which no heat is exchanged with the surroundings), the first law of thermodynamics can be written in differential form as

$$c_p D\ln T - RD\ln p = D\left(c_p\ln T - R\ln p\right) = 0$$

Integrating this expression from a state at pressure p and temperature T to a state in which the pressure is p_s and the temperature is θ, we obtain, after taking the antilogarithm,

$$\theta = T\left(p_s/p\right)^{R/c_p} \tag{2.44}$$

This relationship in (2.44) is referred to as Poisson’s equation, and the temperature θ defined by (2.44) is called the *potential temperature*. θ is simply the temperature that a parcel of dry air at pressure p and temperature T would have if it were expanded or compressed adiabatically to a standard pressure p_s(usually taken to be 1000 hPa). Thus, every air parcel has a unique value of potential temperature, and this value is conserved for dry adiabatic motion. Because synoptic-scale motions are approximately adiabatic outside regions of active precipitation, θ is a quasi-conserved quantity for such motions.

Taking the logarithm of (2.44) and differentiating, we find that

$$c_p\frac{D\ln\theta}{Dt} = c_p\frac{D\ln T}{Dt} - R\frac{D\ln p}{Dt} \tag{2.45}$$

Comparing (2.43) and (2.45), we obtain

$$c_p\frac{D\ln\theta}{Dt} = \frac{J}{T} = \frac{Ds}{Dt} \tag{2.46}$$

Thus, for reversible processes, changes in fractional potential temperature are indeed proportional to entropy changes. A parcel that conserves entropy following the motion must move along an isentropic (constant θ) surface.

2.7.2 The Adiabatic Lapse Rate

A relationship between the *lapse rate* of temperature (i.e., the rate of *decrease* of temperature with respect to height) and the rate of change of potential temperature with respect to height can be obtained by taking the logarithm of (2.44) and differentiating with respect to height. Using the hydrostatic equation and the ideal gas law to simplify the result gives

$$\frac{T}{\theta}\frac{\partial\theta}{\partial z} = \frac{\partial T}{\partial z} + \frac{g}{c_p} \tag{2.47}$$

For an atmosphere in which the potential temperature is constant with respect to height, the lapse rate is thus

$$-\frac{dT}{dz} = \frac{g}{c_p} \equiv \Gamma_d \tag{2.48}$$

Thus, the dry adiabatic lapse rate is approximately constant throughout the lower atmosphere.

2.7.3 Static Stability

If potential temperature is a function of height, the atmospheric lapse rate, $\frac{T}{\theta}\frac{\partial\theta}{\partial z} = \Gamma_d - \Gamma$, will differ from the adiabatic lapse rate and

$$\frac{T}{\theta}\frac{\partial\theta}{\partial z} = \Gamma_d - \Gamma \tag{2.49}$$

If $\Gamma < \Gamma_d$ so that θ increases with height, an air parcel that undergoes an adiabatic displacement from its equilibrium level will be positively buoyant when displaced downward and negatively buoyant when displaced upward, so that it will tend to return to its equilibrium level and the atmosphere is said to be statically stable or *stably stratified*.

Adiabatic oscillations of a fluid parcel about its equilibrium level in a stably stratified atmosphere are referred to as *buoyancy oscillations*. The characteristic frequency of such oscillations can be derived by considering a parcel that is displaced vertically a small distance δz without disturbing its environment. If the environment is in hydrostatic balance, $\rho_0 g = -dp_0/dz$, where p_0 and ρ_0 are the pressure and density of the environment, respectively. The vertical acceleration of the parcel is

$$\frac{Dw}{Dt} = \frac{D^2}{Dt^2}(\delta z) = -g - \frac{1}{\rho}\frac{\partial p}{\partial z} \tag{2.50}$$

where p and ρ are the pressure and density of the parcel, respectively. In the parcel method it is assumed that the pressure of the parcel adjusts instantaneously to the environmental pressure during the displacement: $p = p_0$. This condition must be true if the parcel is to leave the environment undisturbed. Thus, with the aid of the hydrostatic relationship, pressure can be eliminated in (2.50) to give

$$\frac{D^2}{Dt^2}(\delta z) = g\left(\frac{\rho_0 - \rho}{\rho}\right) = g\frac{\theta}{\theta_0} \tag{2.51}$$

where (2.44) and the ideal gas law have been used to express the buoyancy force in terms of potential temperature. Here θ designates the deviation of the potential temperature of the parcel from its basic state (environmental) value $\theta_0(z)$. If the parcel is initially at level $z = 0$ where the potential temperature is $\theta_0(0)$, then for a small displacement δz, we can represent the environmental potential temperature as

$$\theta_0(\delta z) \approx \theta_0(0) + (d\theta_0/dz)\,\delta z$$

If the parcel displacement is adiabatic, the potential temperature of the parcel is conserved. Thus, $\theta(\delta z) = \theta_0(0) - \theta_0(\delta z) = -(d\theta_0/dz)\delta z$, and (2.51) becomes

$$\frac{D^2}{Dt^2}(\delta z) = -N^2\delta z \tag{2.52}$$

where

$$N^2 = g\frac{d\ln\theta_0}{dz}$$

is a measure of the static stability of the environment. Equation (2.52) has a general solution of the form $\delta z = A$ exp(iNt). Therefore, if $N^2 > 0$, the parcel will oscillate about its initial level with a period $\tau = 2\pi/N$. The corresponding frequency N is the *buoyancy frequency*.[3] For average tropospheric conditions, $N \approx 1.2 \times 10^{-2}\mathrm{s}^{-1}$ so that the period of a buoyancy oscillation is about 8 min.

In the case of $N = 0$, examination of (2.52) indicates that no accelerating force will exist and the parcel will be in neutral equilibrium at its new level. However, if $N^2 < 0$ (potential temperature decreasing with height), the displacement will increase exponentially with time. We thus arrive at the familiar gravitational or static stability criteria for dry air:

$d\theta_0/dz > 0$	statically stable
$d\theta_0/dz = 0$	statically neutral
$d\theta_0/dz < 0$	statically unstable

On the synoptic-scale the atmosphere is always stably stratified because any unstable regions that develop are stabilized quickly by convective overturning.

2.7.4 Scale Analysis of the Thermodynamic Energy Equation

If potential temperature is divided into a basic state $\theta_0(z)$ and a deviation $\theta(x,y,z,t)$ so that the total potential temperature at any point is given by $\theta_{\text{tot}} = \theta_0(z) + \theta(x,y,z,t)$, the first law of thermodynamics (2.46) can be written approximately for synoptic scaling as

$$\frac{1}{\theta_0}\left(\frac{\partial\theta}{\partial t} + u\frac{\partial\theta}{\partial x} + v\frac{\partial\theta}{\partial y}\right) + w\frac{d\ln\theta_0}{dz} = \frac{J}{c_pT} \tag{2.53}$$

where we have used the facts that for $|\theta/\theta_0| \ll 1, |d\theta/dz| \ll d\theta_0/dz$, and

$$\ln\theta_{\text{tot}} = \ln\left[\theta_0\left(1 + \theta/\theta_0\right)\right] \approx \ln\theta_0 + \theta/\theta_0$$

Outside regions of active precipitation, diabatic heating is due primarily to net radiative heating. In the troposphere, radiative heating is quite weak so that typically $J/c_p \leq 1^\circ$ C d^{-1} (except near cloud tops, where substantially larger cooling can occur due to thermal emission by the cloud particles). The typical amplitude of horizontal potential temperature fluctuations in a midlatitude synoptic system (above the boundary layer) is $\theta \sim 4^\circ\mathrm{C}$. Thus,

$$\frac{T}{\theta_0}\left(\frac{\partial\theta}{\partial t} + u\frac{\partial\theta}{\partial x} + v\frac{\partial\theta}{\partial y}\right) \sim \frac{\theta U}{L} \sim 4^\circ\mathrm{C}\,\mathrm{d}^{-1}$$

The cooling due to vertical advection of the basic state potential temperature (usually called adiabatic cooling) has a typical magnitude of

$$w\left(\frac{T}{\theta_0}\frac{d\theta_0}{dz}\right) = w\left(\Gamma_d - \Gamma\right) \sim 4^\circ\mathrm{C}\,\mathrm{d}^{-1}$$

where $w \sim$ 1 cm s^{-1}, and $\Gamma_d - \Gamma$, the difference between dry adiabatic and actual lapse rates, is $\sim 4^\circ$

C km^{-1}.

Thus, in the absence of strong diabatic heating, the rate of change in the perturbation potential temperature is equal to the adiabatic heating or cooling due to vertical motion in the statically stable basic state, and (2.54) can be approximated as

$$\left(\frac{\partial\theta}{\partial t}+u\frac{\partial\theta}{\partial x}+v\frac{\partial\theta}{\partial y}\right)+w\frac{d\theta_0}{dz}\approx 0 \tag{2.54}$$

Alternatively, if the temperature field is divided into a basic state $T_0(z)$ and a deviation $T(x,y,z,t)$, then, since $\theta/\theta_0 \approx T/T_0$, (2.54) can be expressed to the same order of approximation in terms of temperature as

$$\left(\frac{\partial T}{\partial t}+u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}\right)+w\left(\Gamma_d-\Gamma\right)\approx 0 \tag{2.55}$$

2.8 The Boussinesq Approximation

In some situations, air parcel vertical displacements are relatively small, so the density change following the motion is also small. The *Boussinesq approximation* yields a simplified form of the dynamical equations that are appropriate to this situation. In this approximation, density is replaced by a constant mean value, ρ_0, everywhere except in the buoyancy term in the vertical momentum equation. The horizontal momentum equations (2.24) and (2.25) can then be expressed in Cartesian coordinates as

$$\frac{Du}{Dt}=-\frac{1}{\rho_0}\frac{\partial p}{\partial x}+fv+F_{rx} \tag{2.56}$$

and

$$\frac{Dv}{Dt}=-\frac{1}{\rho_0}\frac{\partial p}{\partial y}-fu+F_{ry} \tag{2.57}$$

while, with the aid of equations (2.28) and (2.51), the vertical momentum equation becomes

$$\frac{Dw}{Dt}=-\frac{1}{\rho_0}\frac{\partial p}{\partial z}+g\frac{\theta}{\theta_0}+F_{rz} \tag{2.58}$$

Here, as in Section 2.7.3 θ designates the perturbation departure of potential temperature from its basic state value $\theta_0(z)$. Thus, the total potential temperature field is given by $\theta_{\text{tot}}=\theta(x,y,z,t)+\theta_0(z)$, and the adiabatic thermodynamic energy equation has a form similar to (2.54),

$$\frac{D\theta}{Dt}=-w\frac{d\theta_0}{dz} \tag{2.59}$$

except, from the full material derivative, we see that the vertical advection of perturbation potential temperature is formally included. Finally, the continuity equation (2.34) under the

Boussinesq approximation is

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0 \tag{2.60}$$

2.9 Thermodynamics of the Moist Atmosphere

Although it is convenient as a first approximation to neglect the role of water vapor in atmospheric dynamics, it sometimes plays an important, if not essential, role. Most often it is the change in phase of water, and the accompanying latent heat release, that affects the dynamics, but even just the presence of water vapor can be critical to the buoyancy acceleration. Water vapor is a highly variable constituent in the atmosphere and as such presents a challenge to the ideal gas law, since the gas constant depends on the molecular composition of air. For dry air the ideal gas law is

$$p_d=\rho_d R_d T \tag{2.61}$$

where p_d and ρ_d are the dry air pressure and density, respectively, and R_d is the dry air gas constant. For pure water vapor, the ideal gas law is

$$e=\rho_v R_v T \tag{2.62}$$

where e and ρ_v are the vapor pressure and density, respectively, and R_v is the water vapor gas constant. For moist air (a mixture of dry air and water vapor) we choose

$$p=\rho R_d T_v \tag{2.63}$$

where p and ρ are the pressure and density, respectively, and T_v is the *virtual temperature*, which represents the temperature dry air needs to have the same pressure and density as moist air. This formulation allows use of the dry air gas constant everywhere. By adding (2.61) and (2.62), it can be shown that the virtual temperature is given by

$$T_v=\frac{T}{1-\frac{e}{p}\left(1-\frac{R_d}{R_v}\right)} \tag{2.64}$$

We see that $T_v \geq T$, and the difference is typically on the order of a few degrees Celsius or less; nevertheless, even these small changes can make the difference between an air parcel being stable or unstable.

There are many measures of water vapor in the atmosphere, and we summarize here some of the most common. A measure used earlier in the ideal gas law is the vapor pressure that by Dalton's Law is simply the contribution to the total pressure from water vapor. Saturation vapor pressure is determined solely by temperature as given by the Clausius–Clapeyron equation[4]

$$\frac{1}{e_s}\frac{de_s}{dT}=\frac{L}{R_v T^2} \tag{2.65}$$

For water vapor over a plane surface of liquid water (ice), L represents the latent heat of condensation (sublimation). Saturation vapor pressure depends approximately exponentially on temperature and, for temperatures below freezing, is larger above water surfaces than ice. Relative humidity may be defined as the percentage

$$RH = 100 \times \frac{e}{e_s} \tag{2.66}$$

and e_s may apply to either water or ice saturation; if unspecified, the former may be assumed with confidence.

The water vapor mixing ratio is given by the ratio of the mass of water vapor to the mass of dry air in a volume of air, which takes units of either kg/kg or, more commonly, g/kg. From (2.61) and (2.62), the mixing ratio is given by

$$q = 0.622\frac{e}{p - e} \tag{2.67}$$

where the leading coefficient 0.622 reflects the ratio of the molecular weight of water vapor to dry air.

There are several ways to bring moist air to saturation, one of which is by cooling the air at constant pressure without addition or removal of water vapor. Saturation achieved in this manner gives the dew point temperature.

2.9.1 Equivalent Potential Temperature

We previously applied the parcel method to discuss the vertical stability of a dry atmosphere. We found that the stability of a dry parcel with respect to a vertical displacement depends on the lapse rate of potential temperature in the environment such that a parcel displacement is stable, provided that $\partial\theta/\partial z > 0$ (i.e., the actual lapse rate is less than the adiabatic lapse rate). The same condition also applies to parcels in a moist atmosphere when the relative humidity is less than 100%. If, however, a parcel of moist air is forced to rise, it will eventually become saturated at a level called the lifting condensation level (LCL). A further forced rise will then cause condensation and latent heat release, and the parcel will then cool at the saturated adiabatic lapse rate. If the environmental lapse rate is greater than the saturated adiabatic lapse rate, and the parcel is forced to continue to rise, it will reach a level at which it becomes buoyant relative to its surroundings. It can then freely accelerate upward. The level at which this occurs is called the level of free convection (LFC).

Discussion of parcel dynamics in a moist atmosphere is facilitated by defining a thermodynamic field called the *equivalent potential temperature*. Equivalent potential temperature, designated θ_e, is the potential temperature that a parcel of air would have if all its moisture were condensed and the resultant latent heat used to warm the parcel. The temperature of an air parcel can be brought to its equivalent potential value by raising the parcel from its original level until all the water vapor in the parcel has condensed and fallen out, and then compressing the parcel adiabatically to a pressure of 1000 hPa. Because the condensed water is assumed to fall out, the temperature increase during the compression will be at the dry adiabatic rate and the parcel will arrive back at its original level with a temperature that is higher than its original temperature. Thus, the process is irreversible. Ascent of this type, in which all condensation products are assumed to fall out, is called *pseudoadiabatic* ascent. (It is not a truly adiabatic process because the liquid water that falls out carries a small amount of heat with it.)

A complete derivation of the mathematical expression relating θ_e to the other variables of state is rather involved and will be relegated to Appendix D. For most purposes, however, it is sufficient to use an approximate expression for θ_e that can be immediately derived from the entropy form of the first law of thermodynamics (2.46). If we let q_s denote the mass of water vapor per unit mass of dry air in a saturated parcel (the saturation mixing ratio), then the rate of diabatic heating per unit mass is

$$J = -L_c \frac{Dq_s}{Dt}$$

where L_c is the latent heat of condensation. Therefore, from the first law of thermodynamics

$$c_p \frac{D \ln \theta}{Dt} = -\frac{L_c}{T} \frac{Dq_s}{Dt} \tag{2.68}$$

For a saturated parcel undergoing pseudoadiabatic ascent, the rate of change in q_s following the motion is much larger than the rate of change in T or L_c. Therefore,

$$d \ln \theta \approx -d \left(L_c q_s / c_p T \right) \tag{2.69}$$

Integrating (2.69) from the initial state (θ,q_s,T) to a state where $q_s \approx 0$, we obtain

$$\ln \left(\theta / \theta_e \right) \approx -L_c q_s / c_p T$$

where θ_e, the potential temperature in the final state, is approximately the equivalent potential temperature defined earlier. Thus, θ_e, for a saturated parcel is given by

$$\theta_e \approx \theta \exp \left(L_c q_s / c_p T \right) \tag{2.70}$$

The expression in (2.70) may also be used to compute θ_e for an unsaturated parcel provided that the temperature used in the formula is the temperature that the parcel would have if expanded adiabatically to saturation (i.e., T_{LCL}) and the saturation mixing ratio were replaced by the *actual* mixing ratio of the initial state. Thus, equivalent potential temperature is conserved for a parcel during both dry adiabatic and pseudoadiabatic displacements.

An alternative to θ_e, which is sometimes used in studies of convection, is the *moist static energy*, defined as $h \equiv s + L_c q$, where $s \equiv c_p T + gz$ is the *dry static energy*. It can be shown (Problem 2.10) that

$$c_p T d \ln \theta_e \approx dh \tag{2.71}$$

Thus, moist static energy is approximately conserved when θ_e is conserved.

2.9.2 The Pseudoadiabatic Lapse Rate

The first law of thermodynamics (2.68) can be used to derive a formula for the rate of change in temperature with respect to height for a saturated parcel undergoing pseudoadiabatic ascent. Using the definition of θ (2.44), we can rewrite (2.68) for vertical ascent as

$$\frac{d\ln T}{dz} - \frac{R}{c_p}\frac{d\ln p}{dz} = -\frac{L_c}{c_p T}\frac{dq_s}{dz}$$

which, upon noting that $q_s \equiv q_s$ (T,p) and applying the hydrostatic equation and equation of state, can be expressed as

$$\frac{dT}{dz} + \frac{g}{c_p} = -\frac{L_c}{c_p}\left[\left(\frac{\partial q_s}{\partial T}\right)_p \frac{dT}{dz} - \left(\frac{\partial q_s}{\partial p}\right)_T \rho g\right]$$

Thus, as shown in detail in Appendix D, following the ascending saturated parcel

$$\Gamma_s \equiv -\frac{dT}{dz} = \Gamma_d \frac{[1 + L_c q_s / (RT)]}{[1 + \varepsilon L_c^2 q_s / (c_p R T^2)]} \tag{2.72}$$

where $\varepsilon = 0.622$ is the ratio of the molecular weight of water to that of dry air, $\Gamma_d \equiv g/c_p$ is the dry adiabatic lapse rate, and Γ_s is the *pseudoadiabatic lapse rate*, which is always less than Γ_d. Observed values of Γ_s range from ~ 4 K km^{-1} in warm, humid air masses in the lower troposphere to ~6–7 K km^{-1} in the midtroposphere.

2.9.3 Conditional Instability

Section 2.7.3 showed that for dry adiabatic motions the atmosphere is statically stable provided that the lapse rate is less than the dry adiabatic lapse rate (i.e., the potential temperature increases with height). If the lapse rate Γ lies between dry adiabatic and pseudoadiabatic values ($\Gamma_s < \Gamma < \Gamma_d$), the atmosphere is stably stratified with respect to dry adiabatic displacements but unstable with respect to pseudoadiabatic displacements. Such a situation is referred to as *conditional instability* (i.e., the instability is conditional to saturation of the air parcel).

The conditional stability criterion can also be expressed in terms of the gradient of a field variable θ_e^*, defined as the equivalent potential temperature of a hypothetically saturated atmosphere that has the thermal structure of the actual atmosphere.[5] Thus,

$$d\ln\theta_e^* = d\ln\theta + d\left(L_c q_s / c_p T\right) \tag{2.73}$$

where T is the actual temperature, not the temperature after adiabatic expansion to saturation as in (2.70). To derive an expression for conditional instability, we consider the motion of a saturated parcel in an environment in which the potential temperature is θ_0 at the level z_0. At the level $z_0 - \delta z$ the undisturbed environmental air thus has potential temperature

$$\theta_0 - (\partial\theta / \partial z)\,\delta z$$

Suppose that a saturated parcel that has the environmental potential temperature at $z_0 - \delta z$ is raised to the level z_0. When it arrives at z_0, the parcel will have the potential temperature

$$\theta_1 = \left(\theta_0 - \frac{\partial\theta}{\partial z}\delta z\right) + \delta\theta$$

where $\delta\theta$ is the change in parcel potential temperature due to condensation during ascent through vertical distance δz. Assuming a pseudoadiabatic ascent, we see from (2.69) that

$$\frac{\delta\theta}{\theta} \approx -\delta\left(\frac{L_c q_s}{c_p T}\right) \approx -\frac{\partial}{\partial z}\left(\frac{L_c q_s}{c_p T}\right)\delta z$$

so that the buoyancy of the parcel when it arrives at z_0 is proportional to

$$\frac{(\theta_1-\theta_0)}{\theta_0} \approx -\left[\frac{1}{\theta}\frac{\partial\theta}{\partial z}+\frac{\partial}{\partial z}\left(\frac{L_c q_s}{c_p T}\right)\right]\delta z \approx -\frac{\partial \ln\theta_e^*}{\partial z}\delta z$$

where the last expression arises from the application of (2.73).

The saturated parcel will be warmer than its environment at z_0 provided that $\theta_1 > \theta_0$. Thus, the conditional stability criterion for a saturated parcel is

$$\frac{\partial\theta_e^*}{\partial z}\begin{cases} <0 & \text{conditionally unstable} \\ =0 & \text{saturated neutral} \\ >0 & \text{conditionally stable}\end{cases} \tag{2.74}$$

In Figure 2.8 the vertical profiles of θ, θ_e, and θ_e^* for a typical sounding in the vicinity of an extratropical thunderstorm are shown. It is obvious from this figure that the sounding is conditionally unstable in the lower troposphere. However, this observed profile does not imply that convective overturning will occur spontaneously. The release of conditional instability requires not only that $\partial\theta_e^*/\partial z < 0$, but also parcel saturation at the environmental temperature of the level where the convection begins (i.e., the parcel must reach the LFC). The mean relative humidity in the troposphere is well below 100%, even in the boundary layer. Thus, low-level convergence with resultant forced layer ascent or vigorous vertical turbulent mixing in the boundary layer is required to produce saturation.

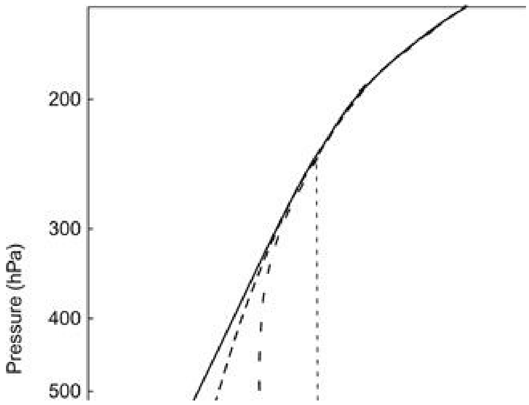


Figure 2.8 Schematic sounding for a conditionally unstable environment, characteristic of midwestern North American thunderstorm conditions, showing the vertical profiles of potential temperature θ, equivalent potential temperature θ_e, and equivalent temperature θ_e^* of a hypothetically saturated atmosphere with the same temperature profile. The *dotted line* shows θ_e for a nonentraining parcel raised from the surface. The *arrow* denotes the LFC for the parcel.

The amount of ascent necessary to raise a parcel to its LFC can be estimated simply from Figure 2.8. A parcel rising pseudoadiabatically from a level $z_0 - \delta z$ will conserve the value of θ_e

characteristic of the environment at $z_0 - \delta z$. However, the buoyancy of a parcel depends only on the difference in density between the parcel and its environment. Thus, in order to compute the buoyancy of the parcel at z_0, it is not correct simply to compare θ_e of the environment at z_0 to $\theta_e(z_0 - \delta z)$ because if the environment is unsaturated, the difference in θ_e between the parcel and the environment may be due primarily to the difference in mixing ratios, not to any temperature (density) difference. To estimate the buoyancy of the parcel, $\theta_e(z_0 - \delta z)$ should instead be compared to $\theta_e^*(z_0)$, which is the equivalent potential temperature that the environment at z_0 would have if it were isothermally brought to saturation. The parcel will thus become buoyant when raised to the level z_0 if $\theta_e(z_0 - \delta z) > \theta_e^*(z_0)$, for then the parcel temperature will exceed the temperature of the environment at z_0.

From Figure 2.8, we see that θ_e for a parcel raised from about 960 hPa will intersect the θ_e^* curve near 850 hPa, whereas a parcel raised from any level much above 850 hPa will not intersect θ_e^* no matter how far it is forced to ascend. It is for this reason that low-level convergence is usually required to initiate convective overturning over the oceans. Only air near the surface has a sufficiently high value of θ_e to become buoyant when it is forcibly raised. Convection over continental regions, however, can be initiated without significant boundary layer convergence, as strong surface heating can produce positive parcel buoyancy all the way to the surface. Sustained deep convection, however, requires mean low-level moisture convergence.

Suggested References

Curry and Webster. Curry and Webster's *Thermodynamics of Atmospheres and Oceans* contains an excellent treatment of atmospheric thermodynamics.

Pedlosky. Pedlosky, *Geophysical Fluid Dynamics*, discusses the equations of motion for a rotating coordinate system and has a thorough discussion of scale analysis at a graduate level.

Salby. Salby's *Fundamentals of Atmospheric Physics* contains a thorough development of the basic conservation laws at the graduate level.

Problems

2.1. A ship is steaming northward at a rate of 10 km h^{-1}. The surface pressure increases toward the northwest at the rate of 5 Pa km^{-1}. What is the pressure tendency recorded at a nearby island station if the pressure aboard the ship decreases at a rate of 100 Pa/3 h?

2.2. The temperature at a point 50 km north of a station is 3° C cooler than at the station. If the wind is blowing from the northeast at 20 m s^{-1} and the air is being heated by radiation at the rate of 1° C h^{-1}, what is the local temperature change at the station?

2.3. Derive the relationship

$$\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) = -\Omega^2 \mathbf{R}$$

which was used in Eq. (2.7).

2.4. Derive the expression given in Eq. (2.13) for the rate of change of **k** following the motion.

2.5. Suppose a 1-kg parcel of dry air is rising at a constant vertical velocity. If the parcel is being heated by radiation at the rate of 10^{-1} W kg^{-1}, what must the speed of rise be to maintain the parcel at a constant temperature?

2.6. Derive an expression for the density ρ that results when an air parcel initially at pressure p_s and density ρ_s expands adiabatically to pressure p.

2.7. An air parcel that has a temperature of 20° C at the 1000-hPa level is lifted dry adiabatically. What is its density when it reaches the 500-hPa level?

2.8. Suppose an air parcel starts from rest at the 800-hPa level and rises vertically to 500 hPa while maintaining a constant 1° C temperature excess over the environment. Assuming that the mean temperature of the 800- to 500-hPa layer is 260 K, compute the energy released due to the work of the buoyancy force. Assuming that all the released energy is realized as kinetic energy of the parcel, what will the vertical velocity of the parcel be at 500 hPa?

2.9. Show that for an atmosphere with an adiabatic lapse rate (i.e., constant potential temperature) the geopotential height is given by

$$Z = H_\theta \left[1 - (p/p_0)^{R/c_p}\right]$$

where p_0 is the pressure at $Z = 0$ and $H_\theta \equiv c_p \theta / g_0$ is the total geopotential height of the atmosphere.

2.10. The azimuthal velocity component in some hurricanes is observed to have a radial dependence given by $v_\lambda = V_0 (r_0/r)^2$ for distances from the center given by $r \geq r_0$. Letting $V_0 = 50\,\mathrm{m\,s^{-1}}$ and $r_0 = 50$ km, find the total geopotential difference between the far field $(r \rightarrow \infty)$ and $r = r_0$, assuming gradient wind balance and $f_0 = 5 \times 10^{-5}\,\mathrm{s^{-1}}$. At what distance from the center does the Coriolis force equal the centrifugal force?

2.11. In the *isentropic* coordinate system (see Section 4.6), potential temperature is used as the vertical coordinate. Because potential temperature in adiabatic flow is conserved following the motion, isentropic coordinates are useful for tracing the actual paths of travel of individual air parcels. Show that the transformation of the horizontal pressure gradient force from z to θ coordinates is given by

$$\frac{1}{\rho} \nabla_z p = \nabla_\theta M$$

where $M \equiv c_p T + \Phi$ is the *Montgomery streamfunction*.

2.12. French scientists have developed a high-altitude balloon that remains at constant potential temperature as it circles Earth. Suppose such a balloon is in the lower equatorial stratosphere where the temperature is isothermal at 200 K. If the balloon were displaced vertically from its equilibrium level by a small distance δz, it would tend to oscillate about the equilibrium level. What would be the period of this oscillation?

2.13. Derive the approximate thermodynamic energy equation (2.55) using the scaling arguments of Sections 2.4 and 2.7.

MATLAB Exercises

M2.1. The MATLAB script **standard_T_p.m** defines and plots the temperature and the lapse rate associated with the U.S. Standard Atmosphere as *functions* of height. Modify this script to compute the pressure and potential temperature, and plot these in the same format used for temperature and lapse rate. [*Hint*: To compute pressure, integrate the hydrostatic equation from the surface upward in increments of δz.] Show that if we define a mean scale height for the layer between z and $z+\delta z$ by letting $H=R[T(z)+T(z+\delta z)]/(2g)$, then $p(z+\delta z)=p(z)\exp[-\delta z/H]$. (Note that as you move upward layer by layer, you must use the local height-dependent value of H in this formula.)

M2.2. The MATLAB script **thermo_profile.m** is a simple script to read in data giving pressure and temperature for a tropical mean sounding. Run this script to plot temperature versus pressure for data in the file **tropical_temp.dat**. Use the hypsometric equation to compute the geopotential height corresponding to each pressure level of the data file. Compute the corresponding potential temperature and plot graphs of the temperature and potential temperature variations with pressure and with geopotential height.

---

[1] It can be shown that when r is replaced by a, as here (the traditional approximation) the Coriolis terms proportional to cos ϕ in (2.19) and (2.21) must be neglected if the equations are to satisfy angular momentum conservation.

[2] For a discussion of entropy and its role in the second law of thermodynamics, see Curry and Webster (1999), for example.

[3] N is often referred to as the Brunt–Väisälä frequency.

[4] For a derivation see, for example, Curry and Webster (1999, p. 108).

[5] Note that θ_e^* is *not* the same as θ_e except in a saturated atmosphere.

Chapter 3

Elementary Applications of the Basic Equations

In addition to the geostrophic wind, which was discussed in Chapter 2, there are other approximate expressions for the relationships among velocity, pressure, and temperature fields that are useful in the analysis of weather systems. We consider estimates of the wind that derive from fundamental force balances, trajectory and streamline analysis, the thermal wind, and estimates for vertical motion and surface pressure tendency.

3.1 Basic Equations in Isobaric Coordinates

The topics are discussed most conveniently using a coordinate system in which pressure is the vertical coordinate. Thus, before introducing the elementary applications of the present chapter, it is useful to present the dynamical equations in isobaric coordinates.

3.1.1 The Horizontal Momentum Equation

The approximate horizontal momentum equations (2.24) and (2.25) may be written in vectorial form as

$$\frac{D\mathbf{V}}{Dt}+f\mathbf{k}\times\mathbf{V}=-\frac{1}{\rho}\boldsymbol{\nabla}p \tag{3.1}$$

where $\mathbf{V} = \mathbf{i}u + \mathbf{j}v$ is the *horizontal* velocity vector. In order to express (3.1) in isobaric coordinate form, we transform the pressure gradient force using (1.33) and (1.34) to obtain

$$\frac{D\mathbf{V}}{Dt}+f\mathbf{k}\times\mathbf{V}=-\boldsymbol{\nabla}_p\Phi \tag{3.2}$$

where ∇_p is the horizontal gradient operator applied with pressure held constant.

Because p is the independent vertical coordinate, we must expand the total derivative as

$$\begin{aligned}\frac{D}{Dt}&\equiv\frac{\partial}{\partial t}+\frac{Dx}{Dt}\frac{\partial}{\partial x}+\frac{Dy}{Dt}\frac{\partial}{\partial y}+\frac{Dp}{Dt}\frac{\partial}{\partial p}\\&=\frac{\partial}{\partial t}+u\frac{\partial}{\partial x}+v\frac{\partial}{\partial y}+\omega\frac{\partial}{\partial p}\end{aligned} \tag{3.3}$$

Here $\omega \equiv Dp/Dt$ (usually called the "omega" vertical motion) is the pressure change following the motion, which plays the same role in the isobaric coordinate system that $w\equiv Dz/Dt$ plays in height coordinates. We note that the partial derivatives with respect to x and y are taken with pressure held constant—that is, on isobaric surfaces.

From (3.2) we see that the isobaric coordinate form of the geostrophic relationship is

$$f\mathbf{V}_g = \mathbf{k}\times\boldsymbol{\nabla}_p\Phi \tag{3.4}$$

One advantage of isobaric coordinates is easily seen by comparing (2.23) and (3.4). In the latter equation, density does not appear. Thus, a given geopotential gradient implies the same geostrophic wind at any height, whereas a given horizontal pressure gradient implies different values of the geostrophic wind depending on the density. Furthermore, if f is regarded as a constant, the horizontal divergence of the geostrophic wind at constant pressure is zero:

$$\boldsymbol{\nabla}_p\cdot\mathbf{V}_g = 0$$

3.1.2 The Continuity Equation

It is possible to transform the continuity equation (2.31) from height coordinates to pressure coordinates. However, it is simpler to directly derive the isobaric form by considering a Lagrangian control volume $\delta V = \delta x\delta y\delta z$ and applying the hydrostatic equation $\delta p = -\rho g\delta z$ (note that $\delta p < 0$) to express the volume element as $\delta V = -\delta x\delta y\delta p/(\rho g)$. The mass of this fluid element, which is conserved following the motion, is then $\delta M = \rho\delta V = -\delta x\delta y\delta p/g$. Thus,

$$\frac{1}{\delta M}\frac{D}{Dt}(\delta M) = \frac{g}{\delta x\delta y\delta p}\frac{D}{Dt}\left(\frac{\delta x\delta y\delta p}{g}\right) = 0$$

After differentiating, using the chain rule, and changing the order of the differential operators, we obtain[1]

$$\frac{1}{\delta x}\delta\left(\frac{Dx}{Dt}\right) + \frac{1}{\delta y}\delta\left(\frac{Dy}{Dt}\right) + \frac{1}{\delta p}\delta\left(\frac{Dp}{Dt}\right) = 0$$

or

$$\frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} + \frac{\delta\omega}{\delta p} = 0$$

Taking the limit $\delta x, \delta y, \delta p \to 0$ and observing that δx and δy are evaluated at constant pressure, we obtain the continuity equation in the isobaric system:

$$\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)_p + \frac{\partial\omega}{\partial p} = 0 \tag{3.5}$$

This form of the continuity equation contains no reference to the density field and does not involve time derivatives. The simplicity of (3.5) is one of the chief advantages of the isobaric coordinate system.

3.1.3 The Thermodynamic Energy Equation

The first law of thermodynamics (2.43) can be expressed in the isobaric system by letting $Dp/Dt = \omega$ and expanding DT/Dt using (3.3):

$$c_p\left(\frac{\partial T}{\partial t}+u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}+\omega\frac{\partial T}{\partial p}\right)-\alpha\omega=J$$

This may be rewritten as

$$\left(\frac{\partial T}{\partial t}+u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}\right)-S_p\omega=\frac{J}{c_p} \tag{3.6}$$

where, with the aid of the equation of state and Poisson’s equation (2.44), we have

$$S_p\equiv\frac{RT}{c_p p}-\frac{\partial T}{\partial p}=-\frac{T}{\theta}\frac{\partial\theta}{\partial p} \tag{3.7}$$

which is the static stability parameter for the isobaric system. Using (2.49) and the hydrostatic equation, (3.7) may be rewritten as

$$S_p=(\Gamma_d-\Gamma)/\rho g$$

Thus, S_p is positive provided that the lapse rate is less than dry adiabatic. However, because density decreases approximately exponentially with height, S_p increases rapidly with height. This strong height dependence of the stability measure S_p is a minor disadvantage of isobaric coordinates.

3.2 Balanced Flow

Despite the apparent complexity of atmospheric motion systems as depicted on synoptic weather charts, the pressure (or geopotential height) and velocity distributions in meteorological disturbances are actually related by rather simple approximate force balances. In order to gain a qualitative understanding of the horizontal balance of forces in atmospheric motions, we idealize by considering flows that are steady state (i.e., time independent) and have no vertical component of velocity. Furthermore, it is useful to describe the flow field by expanding the isobaric form of the horizontal momentum equation (3.2) into its components in a so-called *natural* coordinate system.

3.2.1 Natural Coordinates

The natural coordinate system is defined by the orthogonal set of unit vectors **t**, **n**, and **k**. Unit vector **t** is oriented parallel to the horizontal velocity at each point; unit vector **n** is normal to the horizontal velocity and is oriented so that it is positive to the left of the flow direction; and unit vector **k** is directed vertically upward. In this system the horizontal velocity may be written $\mathbf{V}=V\mathbf{t}$, where V, the horizontal speed, is a nonnegative scalar defined by $V\equiv Ds/Dt$, where $s(x,y,t)$ is the distance along the curve followed by a parcel moving in the horizontal plane. The acceleration following the motion is thus

$$\frac{D\mathbf{V}}{Dt}=\frac{D(V\mathbf{t})}{Dt}=\mathbf{t}\frac{DV}{Dt}+V\frac{D\mathbf{t}}{Dt}$$

The rate of change of **t** following the motion may be derived from geometrical considerations with the aid of Figure 3.1:

$$\delta\psi = \frac{\delta s}{|R|} = \frac{|\delta \mathbf{t}|}{|\mathbf{t}|} = |\delta \mathbf{t}|$$

Here R is the *radius of curvature* following the parcel motion, and we have used the fact that $|\mathbf{t}| = 1$. By convention, R is taken to be positive when the center of curvature is in the positive **n** direction. Thus, for $R > 0$, the air parcels turn toward the left following the motion, and for $R < 0$, the air parcels turn toward the right following the motion.

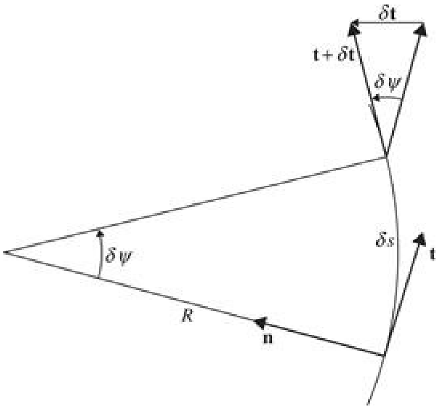


Figure 3.1 Rate of change of the unit tangent vector t following the motion.

Noting that in the limit $\delta s \rightarrow 0$, $\delta \mathbf{t}$ is directed parallel to **n**, the preceding relationship yields $D\mathbf{t}/Ds = \mathbf{n}/R$. Thus,

$$\frac{D\mathbf{t}}{Dt} = \frac{D\mathbf{t}}{Ds}\frac{Ds}{Dt} = \frac{\mathbf{n}}{R}V$$

and

$$\frac{D\mathbf{V}}{Dt} = \mathbf{t}\frac{DV}{Dt} + \mathbf{n}\frac{V^2}{R} \tag{3.8}$$

Therefore, the acceleration following the motion is the sum of the rate of change of speed of the air parcel and its centripetal acceleration due to the curvature of the trajectory. Because the Coriolis force always acts normal to the direction of motion, its natural coordinate form is simply

$$-f\mathbf{k} \times \mathbf{V} = -fV\mathbf{n}$$

whereas the pressure gradient force can be expressed as

$$-\nabla_p \Phi = -\left(\mathbf{t}\frac{\partial \Phi}{\partial s} + \mathbf{n}\frac{\partial \Phi}{\partial n}\right)$$

The horizontal momentum equation may thus be expanded into the following component equations in the natural coordinate system:

$$\frac{DV}{Dt} = -\frac{\partial \Phi}{\partial s} \tag{3.9}$$

$$\frac{V^2}{R}+fV=-\frac{\partial\Phi}{\partial n} \tag{3.10}$$

Equations (3.9) and (3.10) express the force balances parallel to and normal to the direction of flow, respectively. For motion parallel to the geopotential height contours, $\partial\Phi/\partial s=0$, and the speed is constant following the motion. If, in addition, the geopotential gradient normal to the direction of motion is constant along a trajectory, (3.10) implies that the radius of curvature of the trajectory is also constant. In that case the flow can be classified into several simple categories depending on the relative contributions of the three terms in (3.10) to the net force balance.

3.2.2 Geostrophic Flow

Flow in a straight line ($R\rightarrow\pm\infty$) parallel to height contours is referred to as *geostrophic motion*. In geostrophic motion the horizontal components of the Coriolis force and pressure gradient force are in exact balance so that $V=V_g$, where the geostrophic wind V_g is defined by[2]

$$fV_g=-\partial\Phi/\partial n \tag{3.11}$$

This balance is indicated schematically in Figure 3.2. The actual wind can be in exact geostrophic motion only if the height contours are parallel to latitude circles. As discussed in Section 2.4.1, the geostrophic wind is generally a good approximation to the actual wind in extratropical synoptic-scale disturbances. However, in some of the special cases treated later, this is not true.

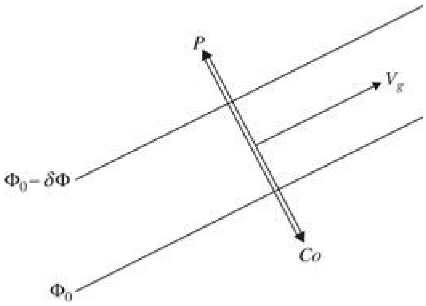


Figure 3.2 Balance of forces for geostrophic equilibrium. The pressure gradient force is designated by P and the Coriolis force by Co.

3.2.3 Inertial Flow

If the geopotential field is uniform on an isobaric surface so that the horizontal pressure gradient vanishes, (3.10) reduces to a balance between Coriolis force and centrifugal force:

$$V^2/R+fV=0 \tag{3.12}$$

Equation (3.12) may be solved for the radius of curvature

$$R=-V/f$$

Since from (3.9), the speed must be constant in this case, the radius of curvature is also constant (neglecting the latitudinal dependence of f). Thus, the air parcels follow circular paths in an anticyclonic sense.[3] The period of this oscillation is

$$P = \left|\frac{2\pi R}{V}\right| = \frac{2\pi}{|f|} = \frac{\frac{1}{2}\text{ day}}{|\sin\phi|} \tag{3.13}$$

P is equivalent to the time that is required for a Foucault pendulum to turn through an angle of 180°. Thus, it is often referred to as one-half *pendulum day*.

Because both the Coriolis force and the centrifugal force due to the relative motion are caused by inertia of the fluid, this type of motion is traditionally referred to as an *inertial oscillation*, and the circle of radius $|R|$ is called the inertia circle. It is important to realize that the “inertial flow” governed by (3.12) is not the same as inertial motion in an absolute reference frame. The flow governed by (3.12) is just the constant angular momentum oscillation referred to in Section 1.3.4. In this flow the force of gravity, acting orthogonal to the plane of motion, keeps the oscillation on a horizontal surface. In true inertial motion, all forces vanish and the motion maintains a uniform absolute velocity.

In the atmosphere, motions are nearly always generated and maintained by pressure gradient forces; the conditions of uniform pressure required for pure inertial flow rarely exist. In the oceans, however, currents are often generated by transient winds blowing across the surface, rather than by internal pressure gradients. As a result, significant amounts of energy occur in currents that oscillate with near inertial periods. An example recorded by a current meter near the island of Barbados is shown in Figure 3.3.

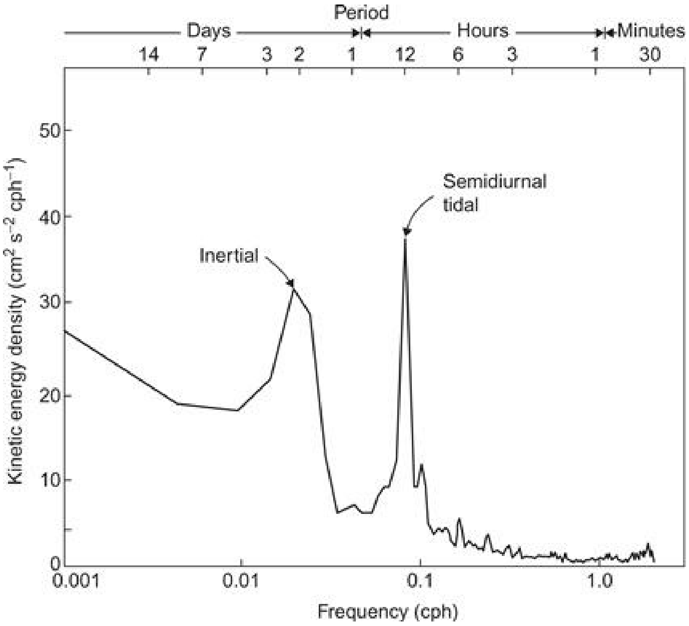


Figure 3.3 Power spectrum of kinetic energy at 30-m depth in the ocean near Barbados (13° N). Ordinate shows kinetic energy density per unit frequency interval (cph − 1 designates cycles per hour). This type of plot indicates the manner in which the total kinetic energy is partitioned among oscillations of different periods. Note the strong peak at 53 h, which is the period of an inertial oscillation at 13° latitude.

After Warsh et al., 1971. Copyright © American Meteorological Society. Reprinted with permission.

3.2.4 Cyclostrophic Flow

If the horizontal scale of a disturbance is small enough, the Coriolis force may be neglected in (3.10) compared to the pressure gradient force and the centrifugal force. The force balance normal to the direction of flow is then

$$\frac{V^2}{R} = -\frac{\partial \Phi}{\partial n}$$

If this equation is solved for V, we obtain the speed of the *cyclostrophic wind*

$$V = \left(-R\frac{\partial \Phi}{\partial n}\right)^{1/2} \tag{3.14}$$

As indicated in Figure 3.4, cyclostrophic flow may be either cyclonic or anticyclonic. In both cases the pressure gradient force is directed toward the center of curvature and the centrifugal force away from the center of curvature.

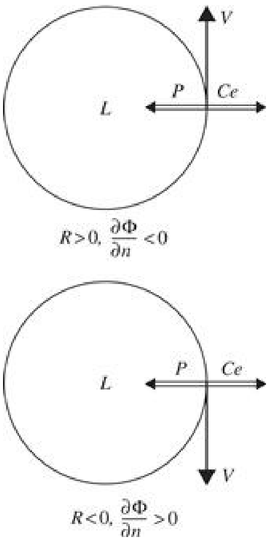

V
L
P
Ce
$R>0,\ \frac{\partial \Phi}{\partial n}<0$
L
P
Ce
V
$R<0,\ \frac{\partial \Phi}{\partial n}>0$

Figure 3.4 Force balance in cyclostrophic flow: *P* designates the pressure gradient; *Ce* designates the centrifugal force.

The cyclostrophic balance approximation is valid provided that the ratio of the centrifugal force to the Coriolis force is large. This ratio $V/(fR)$ is equivalent to the Rossby number discussed in Section 2.4.2. As an example of cyclostrophic scale motion, we consider a typical tornado. Suppose that the tangential velocity is 30 m s^{-1} at a distance of 300 m from the center of the vortex. Assuming that $f = 10^{-4}\ \text{s}^{-1}$, the Rossby number is just $\text{Ro} = V/|fR| \approx 10^3$, which implies that the Coriolis force can be neglected in computing the balance of forces for a tornado. However, the majority of tornadoes in the Northern Hemisphere are observed to rotate in a cyclonic (counterclockwise) sense. This is due to the fact that they are embedded in environments that favor cyclonic rotation (see Section 9.6.1). Smaller-scale vortices, however, such as dust devils and water spouts, do not have a preferred direction of rotation. According to data collected by Sinclair (1965), they are observed to be anticyclonic as often as cyclonic.

3.2.5 The Gradient Wind Approximation

Horizontal frictionless flow that is parallel to the height contours so that the tangential acceleration vanishes ($DV/Dt = 0$) is called *gradient flow*. Gradient flow is a three-way balance among the Coriolis force, the centrifugal force, and the horizontal pressure gradient force. Like geostrophic flow, pure gradient flow can exist only under very special circumstances. It is always possible, however, to define a gradient wind, which at any point is just the wind component parallel to the height contours that satisfies (3.10). For this reason, (3.10) is commonly referred to as the gradient wind equation. Because (3.10) takes into account the centrifugal force due to the curvature of parcel trajectories, the gradient wind is often a better approximation to the actual wind than the geostrophic wind.

The gradient wind speed is obtained by solving (3.10) for V to yield

$$\begin{aligned} V &= -\frac{fR}{2} \pm \left(\frac{f^2R^2}{4} - R\frac{\partial\Phi}{\partial n}\right)^{1/2} \\ &= -\frac{fR}{2} \pm \left(\frac{f^2R^2}{4} + fRV_g\right)^{1/2} \end{aligned} \tag{3.15}$$

where in the lower expression (3.11) is used to express $\partial\Phi/\partial n$ in terms of the geostrophic wind. Not all the mathematically possible roots of (3.15) correspond to physically possible solutions, since it is required that V be real and nonnegative. In Table 3.1 the various roots of (3.15) are classified according to the signs of R and $\partial\Phi/\partial n$ in order to isolate the physically meaningful solutions.

Table 3.1 Classification of Roots of the Gradient Wind Equation in the Northern Hemisphere

Sign $\partial\Phi/\partial n$	$R > 0$	$R < 0$
Positive ($V_g < 0$)	Positive root[a]: unphysical	Positive root: antibaric flow (anomalous low)
	Negative root: unphysical	Negative root: unphysical
Negative ($V_g > 0$)	Positive root: cyclonic flow (regular low)	Positive root: ($V > -fR/2$): anticyclonic flow (anomalous high)
	Negative root: unphysical	Negative root: ($V < -fR/2$): anticyclonic flow (regular high)

[a] The terms “positive root” and “negative root” in columns 2 and 3 refer to the sign taken in the final term in Eq. (3.15).

The force balances for the four permitted solutions are illustrated in Figure 3.5. Equation (3.15) shows that in cases of both regular and anomalous highs the pressure gradient is limited by the requirement that the quantity under the radical be nonnegative—that is,

$$\left|fV_g\right| = \left|\frac{\partial\Phi}{\partial n}\right| < \frac{|R|f^2}{4} \tag{3.16}$$

Thus, the pressure gradient in a high must approach zero as $|R| \rightarrow 0$. It is for this reason that the pressure field near the center of a high is always flat and the wind gentle compared to the region near the center of a low.

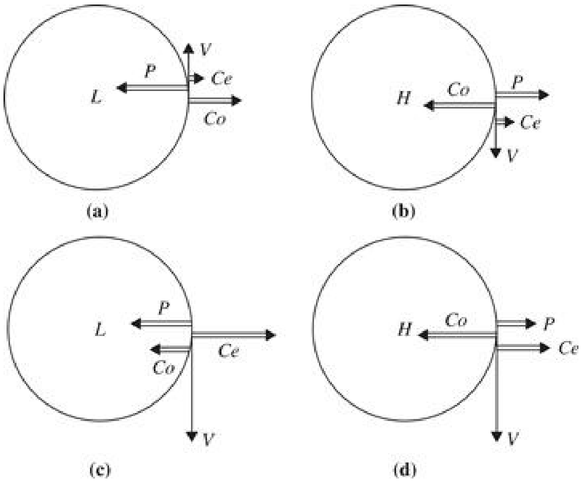


Figure 3.5 Force balances in the Northern Hemisphere for the four types of gradient flow: (a) regular low, (b) regular high, (c) anomalous low, and (d) anomalous high.

The absolute angular momentum about the axis of rotation for the circularly symmetric motions shown in Figure 3.5 is given by $VR+fR^2/2$. From (3.15) it is verified readily that regular gradient wind balances have positive absolute angular momentum in the Northern Hemisphere, whereas anomalous cases have negative absolute angular momentum. Because the only source of negative absolute angular momentum is the Southern Hemisphere, the anomalous cases are unlikely to occur, except perhaps close to the equator.

In all cases except the anomalous low (Figure 3.5c) the horizontal components of the Coriolis and pressure gradient forces are oppositely directed. Such flow is called *baric*. The anomalous low is antibaric; the geostrophic wind V_g defined in (3.11) is negative for an anomalous low and is clearly not a useful approximation to the actual speed.[4] Furthermore, as shown in Figure 3.5, gradient flow is cyclonic only when the centrifugal force and the horizontal component of the Coriolis force have the same sense ($Rf>0$); it is anticyclonic when these forces have the opposite sense ($Rf<0$). Since the direction of anticyclonic and cyclonic flow is reversed in the Southern Hemisphere, the requirement that $Rf>0$ for cyclonic flow holds regardless of the hemisphere considered.

The definition of the geostrophic wind (3.11) can be used to rewrite the force balance normal to the direction of flow (3.10) in the form

$$V^2/R+fV-fV_g=0$$

Dividing through by *fV* shows that the ratio of the geostrophic wind to the gradient wind is

$$\frac{V_g}{V}=1+\frac{V}{fR} \tag{3.17}$$

For normal cyclonic flow ($fR>0$), V_g is larger than V, whereas for anticyclonic flow ($fR<0$), V_g is smaller than V. Therefore, the geostrophic wind is an overestimate of the balanced wind in a region of cyclonic curvature and an underestimate in a region of anticyclonic curvature. For midlatitude synoptic systems, the difference between gradient and geostrophic wind speeds generally does not exceed 10 to 20%. [Note that the magnitude of $V/(fR)$ is just the Rossby number.] For tropical disturbances, the Rossby number is in the range of 1 to 10, and the gradient wind formula must be applied rather than the geostrophic wind. Equation (3.17) also shows that the antibaric anomalous low, which has $V_g < 0$, can exist only when $V/(fR) < -1$. Thus, antibaric flow is associated with small-scale intense vortices such as tornadoes.

3.3 Trajectories and Streamlines

In the natural coordinate system used in the previous section to discuss balanced flow, $s(x, y, t)$ was defined as the distance along the curve in the horizontal plane traced out by the path of an air parcel. The path followed by a particular air parcel over a finite period of time is called the *trajectory* of the parcel. Thus, the radius of curvature R of the path s referred to in the gradient wind equation is the radius of curvature for a parcel trajectory. In practice, R is often estimated by using the radius of curvature of a geopotential height contour, as this can be estimated easily from a synoptic chart. However, the height contours are actually *streamlines* of the gradient wind (i.e., lines that are everywhere parallel to the instantaneous wind velocity).

It is important to distinguish clearly between streamlines, which give a “snapshot” of the velocity field at any instant, and trajectories, which trace the motion of individual fluid parcels over a finite time interval. In Cartesian coordinates, horizontal trajectories are determined by the integration of

$$\frac{Ds}{Dt} = V(x, y, t) \tag{3.18}$$

over a finite time span for each parcel to be followed, whereas streamlines are determined by the integration of

$$\frac{dy}{dx} = \frac{v(x, y, t_0)}{u(x, y, t_0)} \tag{3.19}$$

with respect to x at time t_0. (Note that since a streamline is parallel to the velocity field, its slope in the horizontal plane is just the ratio of the horizontal velocity components.) Only for *steady-state* motion fields (i.e., fields in which the local rate of change of velocity vanishes) do the streamlines and trajectories coincide. However, synoptic disturbances are not steady-state motions. They generally move at speeds of the same order as the winds that circulate about them. To gain an appreciation for the possible errors involved in using the curvature of the streamlines instead of the curvature of the trajectories in the gradient wind equation, it is necessary to investigate the relationship between the curvature of the trajectories and the curvature of the streamlines for a moving pressure system.

We let $\beta(x, y, t)$ designate the angular direction of the wind at each point on an isobaric surface, and R_t and R_s designate the radii of curvature of the trajectories and streamlines, respectively. Then, from Figure 3.6, $\delta s = R\,\delta\beta$ so that in the limit $\delta s \to 0$

$$\frac{D\beta}{Ds} = \frac{1}{R_t} \quad \text{and} \quad \frac{\partial \beta}{\partial s} = \frac{1}{R_s} \tag{3.20}$$

where $D\beta/Ds$ means the rate of change of wind direction along a trajectory (positive for counterclockwise turning) and $\partial\beta/\partial s$ is the rate of change of wind direction along a streamline at any instant. Thus, the rate of change of wind direction following the motion is

$$\frac{D\beta}{Dt} = \frac{D\beta}{Ds}\frac{Ds}{Dt} = \frac{V}{R_t} \tag{3.21}$$

or, after expanding the total derivative,

$$\frac{D\beta}{Dt} = \frac{\partial\beta}{\partial t} + V\frac{\partial\beta}{\partial s} = \frac{\partial\beta}{\partial t} + \frac{V}{R_s} \tag{3.22}$$

Combining (3.21) and (3.22), we obtain a formula for the local turning of the wind:

$$\frac{\partial\beta}{\partial t} = V\left(\frac{1}{R_t} - \frac{1}{R_s}\right) \tag{3.23}$$

Equation (3.23) indicates that the trajectories and streamlines will coincide only when the local rate of change of the wind direction vanishes.

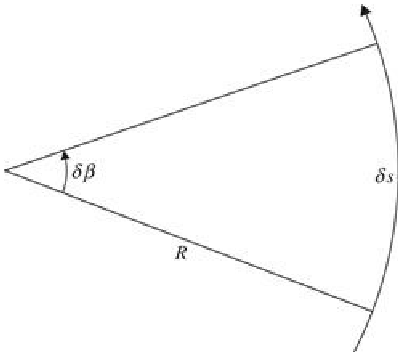


Figure 3.6 Relationship between the change in angular direction of the wind $\delta\beta$ and the radius of curvature R.

In general, midlatitude synoptic systems move eastward as a result of advection by upper-level westerly winds. In such cases there is a local turning of the wind due to the motion of the system even if the shape of the height contour pattern remains constant as the system moves. The relationship between R_t and R_s in such a situation can be determined easily for an idealized circular pattern of height contours moving at a constant velocity **C**. In this case the local turning of the wind is entirely due to the motion of the streamline pattern so that

$$\frac{\partial\beta}{\partial t} = -\mathbf{C}\cdot\nabla\beta = -C\frac{\partial\beta}{\partial s}\cos\gamma = -\frac{C}{R_s}\cos\gamma$$

where γ is the angle between the streamlines (height contours) and the direction of motion of the system. Substituting the preceding equation into (3.23) and solving for R_t with the aid of (3.20), we obtain the desired relationship between the curvature of the streamlines and the curvature of the trajectories:

$$R_t = R_s\left(1 - \frac{C\cos\gamma}{V}\right)^{-1} \tag{3.24}$$

Equation (3.24) can be used to compute the curvature of the trajectory anywhere on a moving pattern of streamlines. In Figure 3.7 the curvatures of the trajectories for parcels initially located due north, east, south, and west of the center of a cyclonic system are shown both for the case of a wind speed greater than the speed of movement of the height contours and for the case of a wind speed less than the speed of movement of the height contours. In these examples the plotted trajectories are based on a geostrophic balance so that the height contours are equivalent to streamlines. It is also assumed for simplicity that the wind speed does not depend on the distance from the center of the system.

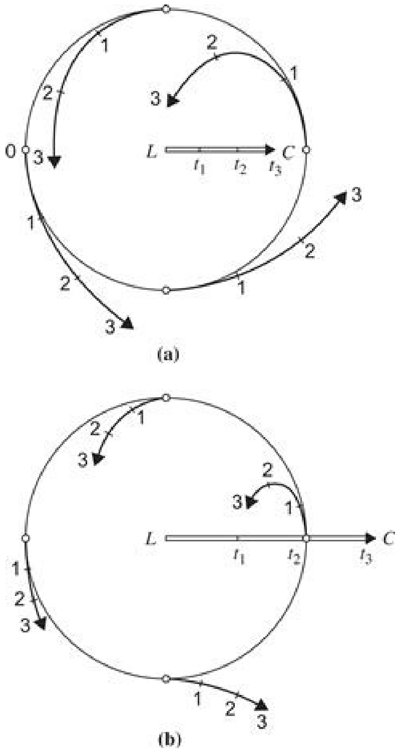


Figure 3.7 Trajectories for moving circular cyclonic circulation systems in the Northern Hemisphere with (a) $V = 2C$ and (b) $2V = C$. Numbers indicate positions at successive times. L designates a pressure minimum.

In the case shown in Figure 3.7b, there is a region south of the low center where the curvature of the trajectories is opposite that of the streamlines. Because synoptic-scale pressure systems usually move at speeds comparable to the wind speed, the gradient wind speed computed on the basis of the curvature of the height contours is often no better an approximation to the actual wind speed than the geostrophic wind. In fact, the actual gradient wind speed will vary along a height contour with the variation of the trajectory curvature.

3.4 The Thermal Wind

The geostrophic wind must have vertical shear in the presence of a horizontal temperature gradient, as can be shown easily from simple physical considerations based on hydrostatic equilibrium. Since the geostrophic wind (3.4) is proportional to the geopotential gradient on an isobaric surface, a geostrophic wind directed along the positive y axis that increases in magnitude with height requires that the slope of the isobaric surfaces with respect to the x axis also must increase with height, as shown in Figure 3.8. According to the hypsometric equation (1.30), the thickness δz corresponding to a (positive) pressure interval δp is

$$\delta z \approx g^{-1} R T \delta \ln p \tag{3.25}$$

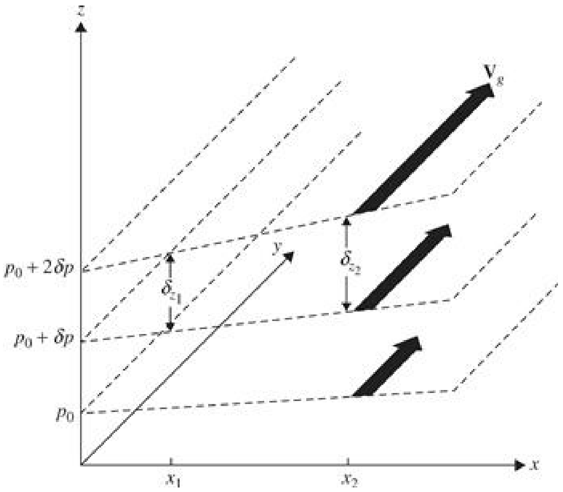


Figure 3.8 Relationship between vertical shear of the geostrophic wind and horizontal thickness gradients. (Note that $\delta p < 0$.)

Thus, the thickness of the layer between two isobaric surfaces is proportional to the mean temperature in the layer, T. In Figure 3.8 the mean temperature T_1 of the column denoted δz_1 must be less than the mean temperature T_2 for the column denoted δz_2. Thus, an increase with height of a positive x directed pressure gradient must be associated with a positive x directed temperature gradient. The air in a vertical column at x_2, because it is warmer (less dense), must occupy a greater depth for a given pressure drop than the air at x_1.

Equations for the rate of change with height of the geostrophic wind components are derived most easily using the isobaric coordinate system. In isobaric coordinates the geostrophic wind (3.4) has components given by

$$v_g = \frac{1}{f}\frac{\partial \Phi}{\partial x} \quad \text{and} \quad u_g = -\frac{1}{f}\frac{\partial \Phi}{\partial y} \tag{3.26}$$

where the derivatives are evaluated with pressure held constant. Also, with the aid of the ideal gas law, we can write the hydrostatic equation as

$$\frac{\partial \Phi}{\partial p} = -\alpha = -\frac{RT}{p} \tag{3.27}$$

Differentiating (3.26) with respect to pressure, and applying (3.27), we obtain

$$p\frac{\partial v_g}{\partial p} \equiv \frac{\partial v_g}{\partial \ln p} = -\frac{R}{f}\left(\frac{\partial T}{\partial x}\right)_p \tag{3.28}$$

$$p\frac{\partial u_g}{\partial p} \equiv \frac{\partial u_g}{\partial \ln p} = \frac{R}{f}\left(\frac{\partial T}{\partial y}\right)_p \tag{3.29}$$

or in vectorial form

$$\frac{\partial \mathbf{V}_g}{\partial \ln p} = -\frac{R}{f}\mathbf{k} \times \nabla_p T \tag{3.30}$$

Equation (3.30) is often referred to as the *thermal wind* equation. However, it is actually a relationship for the vertical wind *shear* (i.e., the rate of change of the geostrophic wind with respect to ln p). Strictly speaking, the term *thermal wind* refers to the vector difference between geostrophic winds at two levels. Designating the thermal wind vector by $\mathbf{V}_T$, we may integrate (3.30) from pressure level p_0 to level p_1 $(p_1 < p_0)$ to get

$$\mathbf{V}_T \equiv \mathbf{V}_g(p_1) - \mathbf{V}_g(p_0) = -\frac{R}{f}\int_{p_0}^{p_1} (\mathbf{k} \times \nabla_p T)\, d \ln p \tag{3.31}$$

Letting $\langle T \rangle$ denote the mean temperature in the layer between pressure p_0 and p_1, the x and y components of the thermal wind are thus given by

$$u_T = -\frac{R}{f}\left(\frac{\partial \langle T \rangle}{\partial y}\right)_p \ln\left(\frac{p_0}{p_1}\right); \quad v_T = \frac{R}{f}\left(\frac{\partial \langle T \rangle}{\partial x}\right)_p \ln\left(\frac{p_0}{p_1}\right) \tag{3.32}$$

Alternatively, we may express the thermal wind for a given layer in terms of the horizontal gradient of the geopotential difference between the top and the bottom of the layer:

$$u_T = -\frac{1}{f}\frac{\partial}{\partial y}(\Phi_1 - \Phi_0); \quad v_T = \frac{1}{f}\frac{\partial}{\partial x}(\Phi_1 - \Phi_0) \tag{3.33}$$

The equivalence of (3.32) and (3.33) can be verified readily by integrating the hydrostatic equation (3.27) vertically from p_0 to p_1 after replacing T by the mean $\langle T \rangle$. The result is the hypsometric equation (1.29):

$$\Phi_1 - \Phi_0 \equiv g Z_T = R\langle T \rangle \ln\left(\frac{p_0}{p_1}\right) \tag{3.34}$$

The quantity Z_T is the *thickness* of the layer between p_0 and p_1 measured in units of geopotential meters. From (3.34) we see that the thickness is proportional to the mean temperature in the layer. Hence, lines of equal Z_T(isolines of thickness) are equivalent to the isotherms of mean temperature in the layer. It can also be used to estimate the mean horizontal temperature advection in a layer, as shown in Figure 3.9. It is clear from the vector form of the thermal wind relation

$$\mathbf{V}_T = \frac{1}{f}\mathbf{k} \times \nabla(\Phi_1 - \Phi_0) = \frac{g}{f}\mathbf{k} \times \nabla Z_T = \frac{R}{f}\mathbf{k} \times \nabla\langle T \rangle \ln\left(\frac{p_0}{p_1}\right) \tag{3.35}$$

that the thermal wind blows parallel to the isotherms (lines of constant thickness) with the warm air to the right facing downstream in the Northern Hemisphere. Thus, as is illustrated in Figure 3.9a, a geostrophic wind that turns counterclockwise with height (backs) is associated with cold-air advection. Conversely, as shown in Figure 3.9b, clockwise turning (veering) of the geostrophic wind with height implies warm advection by the geostrophic wind in the layer.

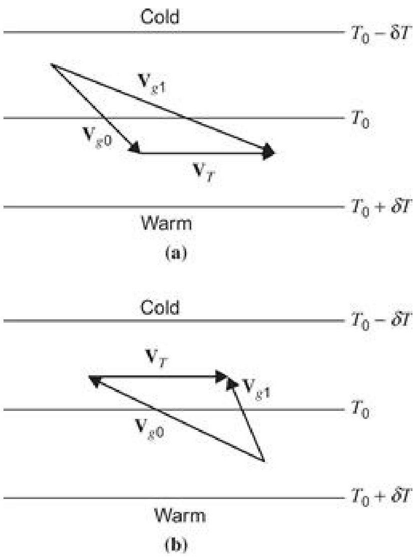


Figure 3.9 Relationship between turning of geostrophic wind and temperature advection: (a) backing of the wind with height and (b) veering of the wind with height.

It is therefore possible to obtain a reasonable estimate of the horizontal temperature advection and its vertical dependence at a given location solely from data on the vertical profile of the wind given by a single sounding. Alternatively, the geostrophic wind at any level can be estimated from the mean temperature field, provided that the geostrophic velocity is known at a single level. Thus, for example, if the geostrophic wind at 850 hPa is known and the mean horizontal temperature gradient in the layer 850 to 500 hPa is also known, the thermal wind equation can be applied to obtain the geostrophic wind at 500 hPa.

3.4.1 Barotropic and Baroclinic Atmospheres

A *barotropic* atmosphere is one in which the density depends only on the pressure, $\rho = \rho(p)$, so that isobaric surfaces are also surfaces of constant density. For an ideal gas, the isobaric surfaces will also be isothermal if the atmosphere is barotropic. Thus, $\nabla_p T = 0$ in a barotropic atmosphere, and the thermal wind equation (3.30) becomes $\partial \mathbf{V}_g / \partial \ln p = 0$, which states that the geostrophic wind is independent of height in a barotropic atmosphere. Thus, barotropy provides a very strong constraint on the motions in a rotating fluid; the large-scale motion can depend only on horizontal position and time, not on height.

An atmosphere in which density depends on both the temperature and the pressure, $\rho = \rho(p, T)$, is referred to as a *baroclinic* atmosphere. In a baroclinic atmosphere the geostrophic wind generally has vertical shear, and this shear is related to the horizontal temperature gradient by the thermal wind equation (3.30). Obviously, the baroclinic atmosphere is of primary importance in dynamic meteorology. However, as shown in later chapters, much can be learned by study of the simpler barotropic atmosphere.

3.5 Vertical Motion

As mentioned previously, for synoptic-scale motions the vertical velocity component is typically of the order of a few centimeters per second. Routine meteorological soundings, however, only give the wind speed to an accuracy of about a meter per second. Thus, in general the vertical velocity is not measured directly but must be inferred from the fields that are measured directly.

Two methods for inferring the vertical motion field are the kinematic method, based on the equation of continuity, and the adiabatic method, based on the thermodynamic energy equation. Both methods are usually applied using the isobaric coordinate system so that $\omega(p)$ is inferred rather than $w(z)$. These two measures of vertical motion can be related to each other with the aid of the hydrostatic approximation.

Expanding Dp/Dt in the (x, y, z) coordinate system yields

(3.36)

$$\omega \equiv \frac{Dp}{Dt} = \frac{\partial p}{\partial t} + \mathbf{V}\cdot\nabla p + w\left(\frac{\partial p}{\partial z}\right)$$

Now, for synoptic-scale motions, the horizontal velocity is geostrophic to a first approximation. Therefore, we can write $\mathbf{V} = \mathbf{V}_g + \mathbf{V}_a$, where $\mathbf{V}_a$ is the *ageostrophic* wind and $|\mathbf{V}_a| \ll |\mathbf{V}_g|$. However, $\mathbf{V}_g = (\rho f)^{-1}\mathbf{k}\times\nabla p$, so that $\mathbf{V}_g\cdot\nabla p = 0$. Using this result plus the hydrostatic approximation, (3.36) may be rewritten as

(3.37)

$$\omega = \frac{\partial p}{\partial t} + \mathbf{V}_a\cdot\nabla p - g\rho w$$

Comparing the magnitudes of the three terms on the right in (3.37), we find that for synoptic-scale motions

$$\begin{gathered} \partial p/\partial t \sim 10\ \mathrm{hPa\ d^{-1}} \\ \mathbf{V}_a\cdot\nabla p \sim \left(1\ \mathrm{m\ s^{-1}}\right)\left(1\ \mathrm{Pa\ km^{-1}}\right) \sim 1\ \mathrm{hPa\ d^{-1}} \\ g\rho w \sim 100\ \mathrm{hPa\ d^{-1}} \end{gathered}$$

Thus, it is quite a good approximation to let

(3.38)

$$\omega = -\rho g w$$

3.5.1 The Kinematic Method

One method of deducing the vertical velocity is based on integrating the continuity equation in the vertical. Integration of (3.5) with respect to pressure from a reference level p_s to any level p yields

(3.39)

$$\begin{aligned} \omega(p) &= \omega(p_s) - \int_{p_s}^{p}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)_p dp \\ &= \omega(p_s) + (p_s - p)\left(\frac{\partial\langle u\rangle}{\partial x} + \frac{\partial\langle v\rangle}{\partial y}\right)_p \end{aligned}$$

Here the angle brackets denote a pressure-weighted vertical average:

$$\langle\ \rangle \equiv (p - p_s)^{-1}\int_{p_s}^{p}(\ )dp$$

With the aid of (3.38), the averaged form of (3.39) can be rewritten as

$$w(z) = \frac{\rho(z_s)\, w(z_s)}{\rho(z)} - \frac{p_s - p}{\rho(z) g}\left(\frac{\partial \langle u \rangle}{\partial x} + \frac{\partial \langle v \rangle}{\partial y}\right) \tag{3.40}$$

where z and z_s are the heights of pressure levels p and p_s, respectively.

Application of (3.40) to infer the vertical velocity field requires knowledge of the horizontal divergence. To determine the horizontal divergence, the partial derivatives $\partial u/\partial x$ and $\partial v/\partial y$ are generally estimated from the fields of u and v by using *finite difference* approximations (see Section 13.2.1). For example, to determine the divergence of the horizontal velocity at the point x_0, y_0 in Figure 3.10, we write

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \approx \frac{u(x_0+d) - u(x_0-d)}{2d} + \frac{v(y_0+d) - v(y_0-d)}{2d} \tag{3.41}$$

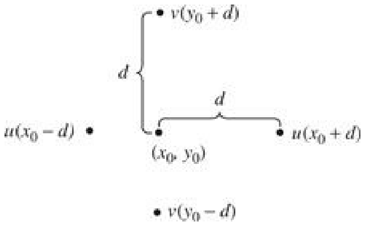


Figure 3.10 Grid for estimation of the horizontal divergence.

However, for synoptic-scale motions in midlatitudes, the horizontal velocity is nearly in geostrophic equilibrium. Except for the small effect due to the variation of the Coriolis parameter (see Problem 3.19), the geostrophic wind is nondivergent; that is, $\partial u/\partial x$ and $\partial v/\partial y$ are nearly equal in magnitude but opposite in sign. Thus, the horizontal divergence is due primarily to the small departures of the wind from geostrophic balance (i.e., the ageostrophic wind). A 10% error in evaluating one of the wind components in (3.41) can easily cause the estimated divergence to be in error by 100%. For this reason, the continuity equation method is not recommended for estimating the vertical motion field from observed horizontal winds.

3.5.2 The Adiabatic Method

A second method for inferring vertical velocities, which is not so sensitive to errors in the measured horizontal velocities, is based on the thermodynamic energy equation. If the diabatic heating J is small compared to the other terms in the heat balance, (3.6) yields

$$\omega = S_p^{-1}\left(\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y}\right) \tag{3.42}$$

Because temperature advection can usually be estimated quite accurately in midlatitudes by using geostrophic winds, the adiabatic method can be applied when only geopotential and temperature data are available. A disadvantage of the adiabatic method is that the local rate of change of temperature is required. Unless observations are taken at close intervals in time, it may be difficult to accurately estimate $\partial T/\partial t$ over a wide area. This method is also rather inaccurate in situations where strong diabatic heating is present, such as storms in which heavy rainfall occurs over a large area. Chapter 6 presents an alternative method for estimating ω, based on the so-called *omega equation*, that does not suffer from these difficulties.

3.6 Surface Pressure Tendency

The development of a negative surface *pressure tendency* is a classic warning of an approaching cyclonic weather disturbance. A simple expression that relates the surface pressure tendency to the wind field, and thus in theory might be used as the basis for short-range forecasts, can be obtained by taking the limit $p \rightarrow$ 0 in (3.39) to get

$$\omega(p_s) = -\int_0^{p_s} (\nabla \cdot \mathbf{V})\, dp \tag{3.43}$$

followed by substituting from (3.37) to yield

$$\frac{\partial p_s}{\partial t} \approx -\int_0^{p_s} (\nabla \cdot \mathbf{V})\, dp \tag{3.44}$$

Here we have assumed that the surface is horizontal so that $w_s = 0$, and have neglected advection by the ageostrophic surface velocity in accord with the scaling arguments in Section 3.5.1.

According to (3.44), the surface pressure tendency at a given point is determined by the total convergence of mass into the vertical column of atmosphere above that point. This result is a direct consequence of the hydrostatic assumption, which implies that the pressure at a point is determined solely by the weight of the column of air above that point. Temperature changes in the air column will affect the heights of upper-level pressure surfaces, but not the surface pressure.

Although, as stated earlier, the tendency equation might appear to have potential as a forecasting aid, its utility is severely limited due to the fact that, as discussed in Section 3.5.1, $\nabla \cdot \mathbf{V}$ is difficult to compute accurately from observations because it depends on the ageostrophic wind field. In addition, there is a strong tendency for vertical compensation. Thus, when there is convergence in the lower troposphere, there is divergence aloft, and vice versa. The net integrated convergence or divergence is then a small residual in the vertical integral of a poorly determined quantity.

Nevertheless, (3.44) does have qualitative value as an aid in understanding the origin of surface pressure changes and the relationship of such changes to the horizontal divergence. This can be illustrated by considering (as one possible example) the development of a thermal cyclone. We suppose that a heat source generates a local warm anomaly in the midtroposphere (Figure 3.11a). Then, according to the hypsometric equation (3.34), the heights of the upper-level pressure surfaces are raised above the warm anomaly, resulting in a horizontal pressure gradient force at the upper levels, which drives a divergent upper-level wind. By (3.44) this upper-level divergence will initially cause the surface pressure to decrease, thus generating a surface low below the warm anomaly (Figure 3.11b). The horizontal pressure gradient associated with the surface low then drives a low-level convergence and vertical circulation, which tends to compensate the upper-level divergence. The degree of compensation between upper divergence and lower convergence will determine whether the surface pressure continues to fall, remains steady, or rises.

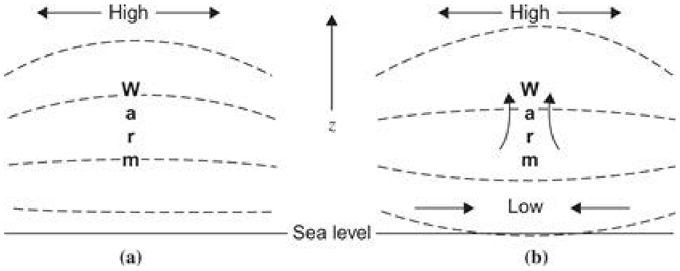


Figure 3.11 Adjustment of surface pressure to a midtropospheric heat source. *Dashed lines* indicate isobars. (a) Initial height increase at upper-level pressure surface. (b) Surface response to upper-level divergence.

The thermally driven circulation of the preceding example is by no means the only type of circulation possible (e.g., cold-core cyclones are important synoptic-scale features). However, it does provide insight into how dynamical processes at upper levels are communicated to the surface and how the surface and upper troposphere are dynamically connected through the divergent circulation. This subject is considered in detail in Chapter 6.

Equation (3.44) is a lower boundary condition that determines the evolution of pressure at constant height. If the isobaric coordinate system of dynamical equations (3.2), (3.5), (3.6), and (3.27) is used as the set of governing equations, the lower boundary condition should be expressed in terms of the evolution of geopotential (or geopotential height) at constant pressure. Such an expression can be obtained simply by expanding $D\Phi/Dt$ in isobaric coordinates

$$\frac{\partial \Phi}{\partial t} = -\mathbf{V}_a \cdot \boldsymbol{\nabla}\Phi - \omega\frac{\partial \Phi}{\partial p}$$

and substituting from (3.27) and (3.43) to get

$$\frac{\partial \Phi_s}{\partial t} \approx -\frac{RT_s}{p_s}\int_0^{p_s}(\boldsymbol{\nabla}\cdot\mathbf{V})\,dp \tag{3.45}$$

where we have again neglected advection by the ageostrophic wind.

In practice, the boundary condition (3.45) is difficult to use because it should be applied at pressure p_s, which is itself changing in time and space. In simple models it is usual to assume that p_s is constant (usually 1000 hPa) and to let $\omega = 0$ at p_s, For modern forecast models, an alternative coordinate system is generally employed in which the lower boundary is always a coordinate surface. This approach is described in Section 10.3.1.

Problems

3.1. An aircraft flying a heading of 60°(i.e., 60° to the east of north) at air speed 200 m s^{-1} moves relative to the ground due east (90°) at 225 m s^{-1}. If the plane is flying at constant pressure, what is its rate of change in altitude (in meters per kilometer horizontal distance) assuming a steady pressure field, geostrophic winds, and $f = 10^{-4}\ \text{s}^{-1}$?

3.2. The actual wind is directed 30° to the right of the geostrophic wind. If the geostrophic wind is 20m s^{-1}, what is the rate of change of wind speed? Let $f = 10^{-4}\ \text{s}^{-1}$.

3.3. A tornado rotates with constant angular velocity ω. Show that the surface pressure at the center of the tornado is given by

$$p = p_0 \exp\left(\frac{-\omega^2 r_0^2}{2RT}\right)$$

where p_0 is the surface pressure at a distance r_0 from the center and T is the temperature (assumed constant). If the temperature is 288 K and pressure and wind speed at 100 m from the center are 1000 hPa and 100 m s^{-1}, respectively, what is the central pressure?

3.4. Calculate the geostrophic wind speed (m s^{-1}) on an isobaric surface for a geopotential height gradient of 100 m per 1000 km, and compare with all possible gradient wind speeds for the same geopotential height gradient and a radius of curvature of ± 500 km. Let $f = 10^{-4}$ s^{-1}.

3.5. Determine the maximum possible ratio of the normal anticyclonic gradient wind speed to the geostrophic wind speed for the same pressure gradient.

3.6. Show that the geostrophic balance in isothermal coordinates may be written

$$f\mathbf{V}_g = \mathbf{k} \times \nabla_T (RT \ln p + \Phi)$$

3.7. Determine the radii of curvature for the trajectories of air parcels located 500 km to the east, north, south, and west of the center of a circular low-pressure system, respectively. The system is moving eastward at 15 m s^{-1}. Assume geostrophic flow with a uniform tangential wind speed of 15 m s^{-1}.

3.8. Determine the normal gradient wind speeds for the four air parcels of Problem 3.7. Using the radii of curvature computed in the previous problem, compare these speeds with the geostrophic speed. (Let $f = 10^{-4}$ s^{-1}.) Use the gradient wind speeds calculated here to recompute the radii of curvature for the four air parcels referred to in Problem 3.7. Use these new estimates of the radii of curvature to recompute the gradient wind speeds for the four air parcels. What fractional error is made in the radii of curvature by using the geostrophic wind approximation in this case? [Note that further iterations could be carried out but would converge rapidly.]

3.9. Show that as the pressure gradient approaches zero, the gradient wind reduces to the geostrophic wind for a normal anticyclone and to inertial flow (Section 3.2.3) for an anomalous anticyclone.

3.10. The mean temperature in the layer between 750 and 500 hPa decreases eastward by 3° C per 100 km. If the 750-hPa geostrophic wind is from the southeast at 20 m s^{-1}, what is the geostrophic wind speed and direction at 500 hPa? Let $f = 10^{-4}$ s^{-1}.

3.11. What is the mean temperature advection in the 750- to 500-hPa layer in Problem 3.10?

3.12. Suppose that a vertical column of the atmosphere at 43° N is initially isothermal from 900 to 500 hPa. The geostrophic wind is 10 m s^{-1} from the south at 900 hPa, 10 m s^{-1} from the west at 700 hPa, and 20 m s^{-1} from the west at 500 hPa. Calculate the mean horizontal temperature gradients in the two layers 900 to 700 hPa and 700 to 500 hPa. Compute the rate of advective temperature change in each layer. How long would this advection pattern have to persist in order to establish a dry adiabatic lapse rate between 600 and 800 hPa? (Assume that the lapse rate is constant between 900 and 500 hPa and that the 800- to 600-hPa layer thickness is 2.25 km.)

3.13. An airplane pilot crossing the ocean at 45° N latitude has both a pressure altimeter and a radar altimeter, the latter measuring his absolute height above the sea. Flying at an air speed of 100 m s^{-1}, he maintains altitude by referring to his pressure altimeter set for a sea-level pressure of 1013 hPa. He holds an indicated 6000-m altitude. At the beginning of a 1-h period he notes that his radar altimeter reads 5700 m, and at the end of the hour he notes that it reads 5950 m. In what direction and approximately how far has he drifted from his heading?

3.14. Work out a gradient wind classification scheme equivalent to Table 3.1 for the Southern Hemisphere (*f*i0) case.

3.15. In the geostrophic momentum approximation (Hoskins, 1975) the gradient wind formula for steady circular flow (3.17) is replaced by the approximation

$$VV_gR^{-1}+fV=fV_g$$

Compare the speeds V computed using this approximation with those obtained in Problem 3.8 using the gradient wind formula.

3.16. How large can the ratio $V_g/(fR)$ be before the geostrophic momentum approximation differs from the gradient wind approximation by 10% for cyclonic flow?

3.17. The planet Venus rotates about its axis so slowly that to a reasonable approximation the Coriolis parameter may be set equal to zero. For steady, frictionless motion parallel to latitude circles, the momentum equation (2.20) then reduces to a type of cyclostrophic balance:

$$\frac{u^2\tan\phi}{a}=-\frac{1}{\rho}\frac{\partial p}{\partial y}$$

By transforming this expression to isobaric coordinates, show that the thermal wind equation in this case can be expressed in the form

$$\omega_r^2(p_1)-\omega_r^2(p_0)=\frac{-R\ln(p_0/p_1)}{(a\sin\phi\cos\phi)}\frac{\partial\langle T\rangle}{\partial y}$$

where R is the gas constant, a is the radius of the planet, and $\omega_r \equiv u/(a\cos\phi)$ is the relative angular velocity. How must $\langle T\rangle$ (the vertically averaged temperature) vary with respect to latitude in order for ω_r to be a function only of pressure? If the zonal velocity at about 60 km above the equator ($p_1 = 2.9\times10^5$ Pa) is 100 m s^{-1} and the zonal velocity vanishes at the surface of the planet ($p_0 = 9.5\times10^6$ Pa), what is the vertically averaged temperature difference between the equator and the pole, assuming that ω_r depends only on pressure? The planetary radius is a = 6100 km, and the gas constant is R = 187 J kg^{-1} K^{-1}.

3.18. Suppose that during the passage of a cyclonic storm the radius of curvature of the isobars is observed to be 800 km at a station where the wind is veering (turning clockwise) at a rate of 10° per hour. What is the radius of curvature of the trajectory for an air parcel that is passing over the station? (The wind speed is 20 m s^{-1}.)

3.19. Show that the divergence of the geostrophic wind in isobaric coordinates on the spherical earth is given by

$$\nabla \cdot \mathbf{V}_g = -\frac{1}{fa}\frac{\partial \Phi}{\partial x}\left(\frac{\cos\phi}{\sin\phi}\right) = -v_g\left(\frac{\cot\phi}{a}\right)$$

(Use the spherical coordinate expression for the divergence operator given in Appendix C.)

3.20. The following wind data were received from 50 km to the east, north, west, and south of a station, respectively: 90°, 10 $\mathrm{m\,s^{-1}}$; 120°, 4 $\mathrm{m\,s^{-1}}$; 90°, 8 $\mathrm{m\,s^{-1}}$; and 60°, 4 $\mathrm{m\,s^{-1}}$. Calculate the approximate horizontal divergence at the station.

3.21. Suppose that the wind speeds given in Problem 3.20 are each in error by ±10%. What would be the percentage error in the calculated horizontal divergence in the worst case?

3.22. The divergence of the horizontal wind at various pressure levels above a given station is shown in the following table. Compute the vertical velocity at each level assuming an isothermal atmosphere with temperature 260 K and letting $w = 0$ at 1000 hPa.

Pressure (hPa)	**$\nabla \cdot \mathbf{V}$ ($\times 10^{-5}\ \mathrm{s^{-1}}$)**
1000	+ 0.9
850	+ 0.6
700	+ 0.3
500	0.0
300	−0.6
100	−1.0

3.23. Suppose that the lapse rate at the 850-hPa level is 4 $\mathrm{K\,km^{-1}}$. If the temperature at a given location is decreasing at a rate of 2 $\mathrm{K\,h^{-1}}$, the wind is westerly at 10 $\mathrm{m\,s^{-1}}$, and the temperature decreases toward the west at a rate of 5 K/100 km, compute the vertical velocity at the 850-hPa level using the adiabatic method.

MATLAB Exercises

M3.1. For the situations considered in Problems M2.1 and M2.2, make further modifications in the MATLAB scripts to compute the vertical profiles of density and of the static stability parameter S_p defined in (3.7). Plot these in the interval from $z = 0$ to $z = 15$ km. You will need to approximate the vertical derivative in S_p using a finite difference approximation (see Section 13.2.1).

M3.2. The objective of this exercise is to gain an appreciation for the difference between trajectories and streamlines in synoptic-scale flows. An idealized representation of a midlatitude synoptic disturbance in an atmosphere with no zonal mean flow is given by the simple sinusoidal pattern of geopotential,

$$\Phi(x,y,t) = \Phi_0 + \Phi' \sin\left[k(x-ct)\right] \cos ly$$

where Φ_0 (p) is the standard atmosphere geopotential dependent only on pressure, Φ' is the magnitude of the geopotential wave disturbance, c is the phase speed of zonal propagation of the wave pattern, and k and l are the wave numbers in the x and y directions, respectively. If it is assumed that the flow is in geostrophic balance, then the geopotential is proportional to the stream function. Let the zonal and meridional wave numbers be equal ($k = l$) and define a perturbation wind amplitude $U' \equiv \Phi' k/f_0$, where f_0 is a constant Coriolis parameter. Trajectories are then given by the paths in (x, y) space obtained by solving the coupled set of ordinary differential equations:

$$\frac{Dx}{Dt} = u = -f_0^{-1}\frac{\partial\Phi}{\partial y} = +U'\sin\left[k\left(x-ct\right)\right]\sin ly$$
$$\frac{Dy}{Dt} = v = +f_0^{-1}\frac{\partial\Phi}{\partial x} = +U'\cos\left[k\left(x-ct\right)\right]\cos ly$$

[Note that U' (taken to be a positive constant) denotes the amplitude of both the x and y components of the disturbance wind.] The MATLAB script **trajectory_1.m** provides an accurate numerical solution to these equations for the special case in which the zonal mean wind vanishes. Three separate trajectories are plotted in the code. Run this script letting $U' = 10\ \mathrm{m\,s^{-1}}$ for cases with $c = 5$, 10, and 15 $\mathrm{m\,s^{-1}}$. Describe the behavior of the three trajectories for each of these cases. Why do the trajectories have their observed dependence on the phase speed c at which the geopotential height pattern propagates?

M3.3. The MATLAB script **trajectory_2.m** generalizes the case of Problem M3.2 by adding a mean zonal flow. In this case the geopotential distribution is specified to be $\Phi\left(x,y,t\right) = \Phi_0 - f_0\bar{U}l^{-1}\sin ly + \Phi'\sin\left[k\left(x-ct\right)\right]\cos ly$.

(a) Solve for the latitudinal dependence of the mean zonal wind for this case.

(b) Run the script with the initial x position specified as $x = -2250$ km and $U' = 15\ \mathrm{m\,s^{-1}}$. Do two runs, letting $\bar{U} = 10\ \mathrm{m\,s^{-1}}$, $c = 5\ \mathrm{m\,s^{-1}}$ and $\bar{U} = 5\ \mathrm{m\,s^{-1}}$, $c = 10\ \mathrm{m\,s^{-1}}$, respectively. Determine the zonal distance that the ridge originally centered at $x = -2250$ km has propagated in each case. Use this information to briefly explain the characteristics (i.e., the shapes and lengths) of the 4-day trajectories for each of these cases.

(c) Which combination of initial position, $\bar{U}$ and c, will produce a straight-line trajectory?

M3.4. The MATLAB script **trajectory_3.m** can be used to examine the dispersion of a cluster of N parcels initially placed in a circle of small radius for a geopotential distribution representing the combination of a zonal mean jet plus a propagating wave with NE to SW tilt of trough and ridge lines. The user must input the mean zonal wind amplitude, $\bar{U}$, the disturbance horizontal wind, U', the propagation speed of the waves, c, and the initial y position of the center of the cluster.

(a) Run the model for three cases specifying $y = 0$, $U' = 15\ \mathrm{m\,s^{-1}}$, and $\bar{U} = 10, 12$, and $15\ \mathrm{m\,s^{-1}}$, respectively. Compute 20-day trajectories for all parcel clusters. Give an explanation for the differences in the dispersion of the parcel cluster for these three cases.

(b) For the situation with $c = 10\ \mathrm{m\,s^{-1}}$, $U' = 15\ \mathrm{m\,s^{-1}}$, and $\bar{U} = 12\ \mathrm{m\,s^{-1}}$, run three additional cases with the initial y specified to be 250, 500, and 750 km.

Describe how the results differ from the case with $y = 0$ km and $\bar{U} = 12\ \mathrm{m\,s^{-1}}$, and give an explanation for the differences in these runs.

---

[1] From now on g will be regarded as a constant.

[2] Note that although the actual speed V must always be positive in the natural coordinates, V_g, which is proportional to the height gradient normal to the direction of flow, may be negative, as in the "anomalous" low shown later in Figure 3.5c.

[3] Anticyclonic flow is a clockwise rotation in the Northern Hemisphere and counterclockwise in the Southern Hemisphere. Cyclonic flow has the opposite sense of rotation in each hemisphere.

[4] Remember that in the natural coordinate system the speed V is positive definite.

Chapter 4

Circulation, Vorticity, and Potential Vorticity

In classical mechanics the principle of conservation of angular momentum is often invoked in the analysis of motions that involve rotation. This principle provides a powerful constraint on the behavior of rotating objects. Analogous conservation laws also apply to the rotational field of a fluid. However, it should be obvious that in a continuous medium, such as the atmosphere, the definition of "rotation" is subtler than that for rotation of a solid object.

Circulation and vorticity are the two primary measures of rotation in a fluid. Circulation, which is a scalar integral quantity, is a *macroscopic* measure of rotation for a finite area of the fluid. Vorticity, however, is a vector field that gives a *microscopic* measure of the rotation at any point in the fluid. Potential vorticity extends the concept of vorticity to include thermodynamic constraints on the motion, yielding a powerful framework for interpreting atmospheric dynamics.

4.1 The Circulation Theorem

The *circulation*, *C*, about a closed contour in a fluid is defined as the line integral evaluated along the contour of the component of the velocity vector that is locally tangent to the contour:

$$C \equiv \oint \mathbf{U} \cdot d\mathbf{l} = \oint |\mathbf{U}| \cos \alpha \, dl$$

where **l** (*s*) is a position vector extending from the origin to the point *s*(*x*,*y*,*z*) on the contour *C*, and *d* **l** represents the limit of $\delta \mathbf{l} = \mathbf{l}(s + \delta s) - \mathbf{l}(s)$ as $\delta s \to 0$. Hence, as indicated in Figure 4.1, *d* **l** is a displacement vector locally tangent to the contour. By convention the circulation is taken to be positive if $C > 0$ for counterclockwise integration around the contour.

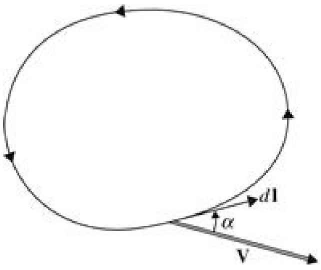


Figure 4.1 Circulation about a closed contour.

That circulation is a measure of rotation is demonstrated readily by considering a circular ring of fluid of radius *R* in solid-body rotation at angular velocity Ω about the *z* axis. In this case, $U = \Omega \times \mathbf{R}$, where **R** is the distance from the axis of rotation to the ring of fluid. Thus, the circulation about the ring is given by

$$C \equiv \oint \mathbf{U} \cdot d\mathbf{l} = \int_0^{2\pi} \Omega R^2 d\lambda = 2\Omega\pi R^2$$

In this case the circulation is just 2π times the angular momentum of the fluid ring about the axis of rotation. Alternatively, note that $C/(\pi R^2) = 2\Omega$, so that the circulation divided by the area enclosed by the loop is just twice the angular speed of rotation of the ring. Unlike angular momentum or angular velocity, circulation can be computed without reference to an axis of rotation; it can thus be used to characterize fluid rotation in situations where "angular velocity" is not easily defined.

The circulation theorem is obtained by taking the line integral of Newton's second law for a closed chain of fluid particles. In the absolute coordinate system, the result (neglecting viscous forces) is

$$\oint \frac{D_a \mathbf{U}_a}{Dt} \cdot d\mathbf{l} = -\oint \frac{\nabla p \cdot d\mathbf{l}}{\rho} - \oint \nabla \Phi \cdot d\mathbf{l} \tag{4.1}$$

where the gravitational force is represented as the gradient of geopotential $\mathbf{\Phi}^*$, defined in terms of true gravity, so that $\nabla\Phi^* = -\mathbf{g}^* = g^*\mathbf{k}$. The integrand on the left side can be rewritten as[1]

$$\frac{D_a \mathbf{U}_a}{Dt} \cdot d\mathbf{l} = \frac{D}{Dt}(\mathbf{U}_a \cdot d\mathbf{l}) - \mathbf{U}_a \cdot \frac{D_a}{Dt}(d\mathbf{l})$$

or after observing that since $\mathbf{l}$ is a position vector, $D_a\, \mathbf{l}/Dt \equiv \mathbf{U}\ a$,

$$\frac{D_a \mathbf{U}_a}{Dt} \cdot d\mathbf{l} = \frac{D}{Dt}(\mathbf{U}_a \cdot d\mathbf{l}) - \mathbf{U}_a \cdot d\mathbf{U}_a \tag{4.2}$$

Substituting (4.2) into (4.1) and using the fact that the line integral about a closed loop of a perfect differential is zero, so that

$$\oint \nabla \Phi \cdot d\mathbf{l} = \oint d\Phi = 0$$

and noting that

$$\oint \mathbf{U}_a \cdot d\mathbf{U}_a = \frac{1}{2} \oint d(\mathbf{U}_a \cdot \mathbf{U}_a) = 0$$

we obtain the circulation theorem:

$$\frac{DC_a}{Dt} = \frac{D}{Dt} \oint \mathbf{U}_a \cdot d\mathbf{l} = -\oint \rho^{-1} dp \tag{4.3}$$

The term that is on the right side in (4.3) is called the solenoidal term, where $dp = \nabla p \cdot d\mathbf{l}$ is the pressure increment along an increment of arc length. For a barotropic fluid, the density is a function only of pressure, and the solenoidal term is zero. Thus, in a barotropic fluid the absolute circulation is conserved following the motion. This result, called *Kelvin's circulation theorem*, is a fluid analog of angular momentum conservation in solid-body mechanics.

For meteorological analysis, it is more convenient to work with the relative circulation C rather than the absolute circulation; a portion of the absolute circulation, C_e, is due to the rotation of Earth about its axis. To compute C_e, we apply Stokes's theorem to the vector $\mathbf{U}_e$, where $\mathbf{U}_e = \boldsymbol{\Omega} \times \mathbf{r}$ is the velocity of Earth at position $\mathbf{r}$:

$$C_e = \oint \mathbf{U}_e \cdot d\mathbf{l} = \int_A\int (\nabla \times \mathbf{U}_e) \cdot \mathbf{n}\, dA$$

where A is the area enclosed by the contour and the unit normal $\mathbf{n}$ is defined by the counterclockwise sense of the line integration using the “right rule." Thus, for the contour of Figure 4.1, $\mathbf{n}$ would be directed out of the page. If the line integral is computed in the horizontal plane, $\mathbf{n}$ is directed along the local vertical (Figure 4.2). Now, by a vector identity (see Appendix C),

$$\nabla \times \mathbf{U}_e = \nabla \times (\boldsymbol{\Omega} \times \mathbf{r}) = \nabla \times (\boldsymbol{\Omega} \times \mathbf{R}) = \boldsymbol{\Omega}\nabla \cdot \mathbf{R} = 2\boldsymbol{\Omega}$$

so that $(\nabla \times \mathbf{U}_e) \cdot \mathbf{n} = 2\Omega \sin\phi \equiv f$ is just the Coriolis parameter. Hence, the circulation in the horizontal plane due to the rotation of Earth is

$$C_e = 2\Omega \langle \sin\phi \rangle\, A = 2\Omega A_e$$

where $\langle \sin\phi \rangle$ denotes an average over the area element A, and A_e is the projection of A in the equatorial plane, as illustrated in Figure 4.2. Thus, the relative circulation may be expressed as

$$C = C_a - C_e = C_a - 2\Omega A_e \tag{4.4}$$

Differentiating (4.4) following the motion and substituting from (4.3), we obtain the *Bjerknes circulation theorem*:

$$\frac{DC}{Dt} = -\oint \frac{dp}{\rho} - 2\Omega\,\frac{DA_e}{Dt} \tag{4.5}$$

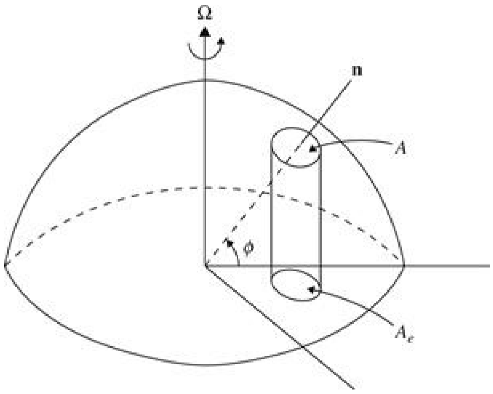


Figure 4.2 Area A_e subtended on the equatorial plane by horizontal area A centered at latitude ϕ.

For a barotropic fluid, (4.5) can be integrated following the motion from an initial state (designated by subscript 1) to a final state (designated by subscript 2), yielding the circulation change

$$C_2 - C_1 = -2\Omega\left(A_2 \sin\phi_2 - A_1 \sin\phi_1\right) \tag{4.6}$$

Equation (4.6) indicates that in a barotropic fluid the relative circulation for a closed chain of fluid particles will be changed if either the horizontal area enclosed by the loop changes or the latitude changes. Furthermore, a negative absolute circulation in the Northern Hemisphere can develop only if a closed chain of fluid particles is advected across the equator from the Southern Hemisphere. The anomalous gradient wind balances discussed in Section 3.2.5 are examples of systems with negative absolute circulations (see Problem 4.6).

Examples

Suppose that the air within a circular region of radius 100 km centered at the equator is initially motionless with respect to Earth. If this circular air mass were moved to the North Pole along an isobaric surface preserving its area, the circulation about the circumference would be

$$C = -2\Omega\pi r^2 \sin(\pi/2) - \sin(0)$$

Thus, the mean tangential velocity at the radius $r = 100$ km would be

$$V = C/(2\pi r) = -\Omega r \approx -7\ \mathrm{m\ s^{-1}}$$

The negative sign here indicates that the air has acquired anticyclonic relative circulation.

In a baroclinic fluid, circulation may be generated by the pressure-density solenoid term in (4.3). This process can be illustrated effectively by considering the development of a sea breeze circulation, as shown in Figure 4.3. For the situation depicted, the mean temperature in the air over the ocean is colder than the mean temperature over the adjoining land. Thus, if the pressure is uniform at ground level, the isobaric surfaces above the ground will slope downward toward the ocean, while the isopycnic surfaces (surfaces of constant density) will slope downward toward the land.

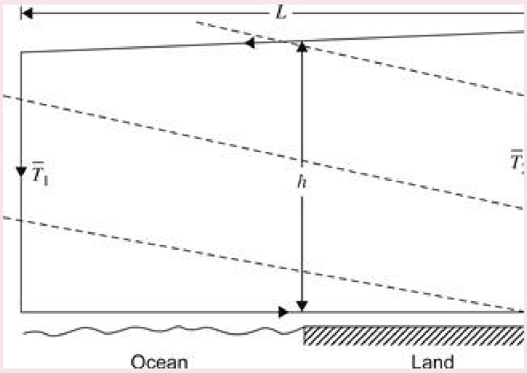


Figure 4.3 Application of the circulation theorem to the sea breeze problem. The closed *heavy solid line* is the loop about which the circulation is to be evaluated. *Dashed lines* indicate surfaces of constant density.

To compute the acceleration as a result of the intersection of the pressure-density surfaces, we apply the circulation theorem by integrating around a circuit in a vertical plane perpendicular to the coastline. Substituting the ideal gas law into (4.3), we obtain

$$\frac{DC_a}{Dt} = -\oint RT\, d\ln p$$

For the circuit shown in Figure 4.3, there is a contribution to the line integral only for the vertical segments of the loop, since the horizontal segments are taken at constant pressure. The resulting rate of increase in the circulation is

$$\frac{DC_a}{Dt} = R\ln\left(\frac{p_0}{p_1}\right)\left(\overline{T}_2 - \overline{T}_1\right) > 0$$

Letting $\langle v\rangle$ be the mean tangential velocity along the circuit, we find that

$$\frac{D\langle v\rangle}{Dt} = \frac{R\ln\left(p_0/p_1\right)}{2(h+L)}\left(\overline{T}_2 - \overline{T}_1\right) \tag{4.7}$$

If we let $p_0 = 1000\,\mathrm{hPa}$, $p_1 = 900\,\mathrm{hPa}$, $\overline{T}_2 - \overline{T}_1 = 10^\circ\,\mathrm{C}$, $L = 20\,\mathrm{km}$, and $h = 1$ km, (4.7) yields an acceleration of about $7\times10^{-3}\,\mathrm{m\,s}^{-2}$. In the absence of frictional retarding forces, this would produce a wind speed of 25 m s^{-1} in about 1 h. In reality, as the wind speed increases, the frictional force reduces the acceleration rate, and temperature advection reduces the land–sea temperature contrast so that a balance is obtained between the generation of kinetic energy by the pressure-density solenoids and frictional dissipation.

4.2 Vorticity

Vorticity, the microscopic measure of rotation in a fluid, is a vector field defined as the curl of velocity. The absolute vorticity $\boldsymbol{\omega}_a$ is the curl of the absolute velocity, whereas the relative vorticity $\boldsymbol{\omega}$ is the curl of the relative velocity:

$$\boldsymbol{\omega}_a \equiv \boldsymbol{\nabla}\times\mathbf{U}_a, \qquad \boldsymbol{\omega} \equiv \boldsymbol{\nabla}\times\mathbf{U}$$

so that in Cartesian coordinates,

$$\boldsymbol{\omega} = \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}, \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}, \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)$$

A scale analysis of the scalar components of the vorticity vector for synoptic-scale motions reveals that the dominant contributions to the horizontal components, $\frac{\partial u}{\partial z}$ and $\frac{\partial v}{\partial z}$,[2] scale like $\frac{U}{H}$, whereas the vertical components, $\frac{\partial u}{\partial y}$ and $\frac{\partial v}{\partial x}$, scale like $\frac{U}{L}$. Therefore, the vertical component of vorticity is of size $\frac{H}{L}$ relative to the horizontal component, or about $10\,\mathrm{km}/1000\,\mathrm{km} = 0.01$. We see that the vorticity vector points mainly in the horizontal direction, and this is true even if we include the planetary contribution, which has size $\frac{fH}{U} = \mathrm{Ro}^{-1}\frac{H}{L} \sim 0.1$, where Ro is the Rossby number. For a westerly jet stream with winds increasing upward in the troposphere, the vorticity vector points mostly toward the north.

As a result of this analysis, it may seem strange, but in fact for large-scale dynamic meteorology we are in general concerned only with the vertical components of absolute and relative vorticity, which are designated by η and ζ, respectively:

$$\eta \equiv \mathbf{k}\cdot\left(\boldsymbol{\nabla}\times\mathbf{U}_a\right), \qquad \zeta \equiv \mathbf{k}\cdot\left(\boldsymbol{\nabla}\times\mathbf{U}\right)$$

In the remainder of this book, η and ζ are referred to as absolute and relative vorticities, respectively, without adding the explicit modifier “vertical component of." We defer a full explanation for why we focus on the relatively small vertical component of vorticity until after we have discussed potential vorticity, but the essential point is that, on synoptic and larger scales, the vertical component of vorticity is tightly coupled to the vertical gradient of potential temperature, which is very large compared to the horizontal gradient.

Regions of positive ζ are associated with cyclonic storms in the Northern Hemisphere; regions of negative ζ are associated with cyclonic storms in the Southern Hemisphere. In both cases ζ has the same sign as *f*, the local value of the planetary rotation, so that we may uniformly define cyclonic by $f\zeta > 0$. Thus, the distribution of relative vorticity is an excellent diagnostic for weather analysis.

The difference between absolute and relative vorticity is *planetary vorticity*, which is just the local vertical component of the vorticity of Earth due to its rotation; $\mathbf{k}\cdot\mathbf{\nabla}\times\mathbf{U}_e = 2\Omega\sin\phi \equiv f$. Thus, $\eta = \zeta + f$ or, in Cartesian coordinates,

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}, \qquad \eta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} + f$$

The relationship between relative vorticity and relative circulation *C* discussed in the previous section can be clearly seen by considering an alternative approach in which the vertical component of vorticity is defined as the circulation about a closed contour in the horizontal plane divided by the area enclosed, in the limit where the area approaches zero:

$$\zeta \equiv \lim_{A\to 0}\left(\oint \mathbf{V}\cdot d\mathbf{l}\right)A^{-1} \tag{4.8}$$

This latter definition makes explicit the relationship between circulation and vorticity discussed in the introduction to this chapter. The equivalence of these two definitions of ζ is demonstrated by considering the circulation about a rectangular element of area $\delta x\delta y$ in the (*x*,*y*) plane, as shown in Figure 4.4. Evaluating **V** · **dl** for each side of the rectangle in the figure yields the circulation

$$\begin{aligned}\delta C &= u\delta x + \left(v + \frac{\partial v}{\partial x}\delta x\right)\delta y - \left(u + \frac{\partial u}{\partial y}\delta y\right)\delta x - v\delta y\\ &= \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)\delta x\delta y\end{aligned}$$

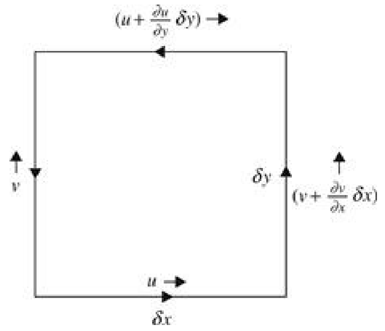

$(u+\frac{\partial u}{\partial y}\delta y)$
v
δy
$(v+\frac{\partial v}{\partial x}\delta x)$
u
δx

Figure 4.4 Relationship between circulation and vorticity for an area element in the horizontal plane.

Dividing through by the area $\delta A = \delta x\delta y$ gives

$$\frac{\delta C}{\delta A} = \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) \equiv \zeta$$

In more general terms the relationship between vorticity and circulation is given simply by Stokes's theorem applied to the velocity vector:

$$\oint \mathbf{U}\cdot d\mathbf{l} = \iint_A (\boldsymbol{\nabla}\times\mathbf{U})\cdot\mathbf{n}\,dA$$

Here A is the area enclosed by the contour and $\mathbf{n}$ is a unit normal to the area element dA (positive in the right sense). Thus, Stokes's theorem states that the circulation about any closed loop is equal to the integral of the normal component of vorticity over the area enclosed by the contour. Hence, for a finite area, circulation divided by area gives the *average* normal component of vorticity in the region. As a consequence, the vorticity of a fluid in solid-body rotation is just twice the angular velocity of rotation. Vorticity may thus be regarded as a measure of the local angular velocity of the fluid.

4.2.1 Vorticity in Natural Coordinates

Physical interpretation of vorticity is facilitated by considering the vertical component of vorticity in the natural coordinate system (see Section 3.2.1). If we compute the circulation about the infinitesimal contour shown in Figure 4.5, we obtain[3]

$$\delta C = V\left[\delta s + d(\delta s)\right] - \left(V + \frac{\partial V}{\partial n}\delta n\right)\delta s$$

However, from this figure, $d(\delta s) = \delta\beta\delta n$, where $\delta\beta$ is the angular change in the wind direction in the distance δs. Hence,

$$\delta C = \left(-\frac{\partial V}{\partial n} + V\frac{\delta\beta}{\delta s}\right)\delta n\,\delta s$$

or, in the limit $\delta n, \delta s \to 0$

$$\zeta = \lim_{\delta n,\delta s\to 0}\frac{\delta C}{(\delta n\,\delta s)} = -\frac{\partial V}{\partial n} + \frac{V}{R_s} \tag{4.9}$$

where R_s is the radius of curvature of the streamlines [Eq. (3.20)]. It is now apparent that the net vertical vorticity component is the result of the sum of two parts: (1) the rate of change of wind speed normal to the direction of flow $-\partial V/\partial n$, called the *shear* vorticity; and (2) the turning of the wind along a streamline V/R_s, called the *curvature* vorticity. Thus, even straight-line motion may have vorticity if the speed changes normal to the flow axis. For example, in the jet stream shown schematically in Figure 4.6a, there will be cyclonic relative vorticity north of the velocity maximum and anticyclonic relative vorticity to the south (Northern Hemisphere conditions), as is recognized easily when the turning of a small paddle wheel placed in the flow is considered. The lower of the two paddle wheels in Figure 4.6a will turn in a clockwise direction (anticyclonically) because the wind force on the blades north of its axis of rotation is stronger than the force on the blades to the south of the axis. The upper wheel will, of course, experience a counterclockwise (cyclonic) turning. Thus, the poleward and equatorward sides of a westerly jet stream are referred to as the cyclonic and anticyclonic shear sides, respectively.

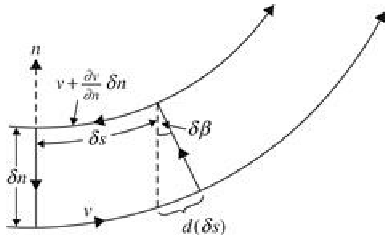


Figure 4.5 Circulation for an infinitesimal loop in the natural coordinate system.

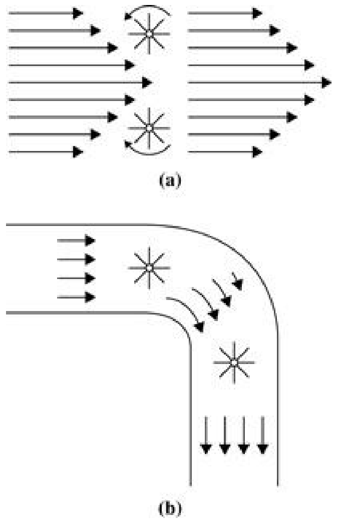


Figure 4.6 Two types of two-dimensional flow: (a) linear shear flow with vorticity and (b) curved flow with zero vorticity.

Conversely, curved flow may have zero vorticity provided that the shear vorticity is equal and opposite to the curvature vorticity. This is the case in the example shown in Figure 4.6b, where a frictionless fluid with zero relative vorticity upstream flows around a bend in a canal. The fluid along the inner boundary on the curve flows faster in just the right proportion so that the paddle wheel does not turn.

4.3 The Vorticity Equation

The previous section discussed kinematic properties of vorticity. This section addresses vorticity dynamics using the equations of motion to determine contributions to the time rate of change of vorticity.

4.3.1 Cartesian Coordinate Form

For motions of synoptic scale, the vorticity equation can be derived using the approximate horizontal momentum equations (2.24) and (2.25). We differentiate the zonal component equation with respect to y and the meridional component equation with respect to x:

$$\frac{\partial}{\partial y}\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+w\frac{\partial u}{\partial z}-fv=-\frac{1}{\rho}\frac{\partial p}{\partial x}\right) \tag{4.10}$$

$$\frac{\partial}{\partial x}\left(\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+w\frac{\partial v}{\partial z}+fu=-\frac{1}{\rho}\frac{\partial p}{\partial y}\right) \tag{4.11}$$

Subtracting (4.10) from (4.11) and recalling that $\zeta = \partial v/\partial x - \partial u/\partial y$, we obtain the vorticity equation

$$\begin{aligned}\frac{\partial \zeta}{\partial t} + u\frac{\partial \zeta}{\partial x} + v\frac{\partial \zeta}{\partial y} + w\frac{\partial \zeta}{\partial z} + (\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)\\ + \left(\frac{\partial w}{\partial x}\frac{\partial v}{\partial z} - \frac{\partial w}{\partial y}\frac{\partial u}{\partial z}\right) + v\frac{df}{dy} = \frac{1}{\rho^2}\left(\frac{\partial \rho}{\partial x}\frac{\partial p}{\partial y} - \frac{\partial \rho}{\partial y}\frac{\partial p}{\partial x}\right)\end{aligned} \tag{4.12}$$

Using the fact that the Coriolis parameter depends only on y so that $Df/Dt = v(df/dy)$, (4.12) may be rewritten in the form

$$\begin{aligned}\frac{D}{Dt}(\zeta + f) = -(\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)\\ - \left(\frac{\partial w}{\partial x}\frac{\partial v}{\partial z} - \frac{\partial w}{\partial y}\frac{\partial u}{\partial z}\right) + \frac{1}{\rho^2}\left(\frac{\partial \rho}{\partial x}\frac{\partial p}{\partial y} - \frac{\partial \rho}{\partial y}\frac{\partial p}{\partial x}\right)\end{aligned} \tag{4.13}$$

Equation (4.13) states that the rate of change of the absolute vorticity following the motion is given by the sum of the three terms on the right, called the divergence (or vortex stretching) term, the tilting or twisting term, and the solenoidal term, respectively.

The concentration or dilution of vorticity by the divergence field—the first term on the right in (4.13)—is the fluid analog of the change in angular velocity resulting from a change in the moment of inertia of a solid body when angular momentum is conserved. If the horizontal flow is divergent, the area enclosed by a chain of fluid parcels will increase with time, and if circulation is to be conserved, the average absolute vorticity of the enclosed fluid must decrease (i.e., the vorticity will be diluted). If, however, the flow is convergent, the area enclosed by a chain of fluid parcels will decrease with time and the vorticity will be concentrated. This mechanism for changing vorticity following the motion is very important in synoptic-scale disturbances.

The second term on the right in (4.13) represents vertical vorticity generated by the tilting of horizontally oriented components of vorticity into the vertical by a nonuniform vertical motion field. This mechanism is illustrated in Figure 4.7, which shows a region where the y component of velocity is increasing with height so that there is a component of shear vorticity oriented in the negative x direction as indicated by the double arrow. If at the same time there is a vertical motion field in which w decreases with increasing x, advection by the vertical motion will tend to tilt the vorticity vector initially oriented parallel to x so that it has a component in the vertical. Thus, if $\partial v/\partial z > 0$ and $\partial w/\partial x < 0$, there will be a generation of positive vertical vorticity.

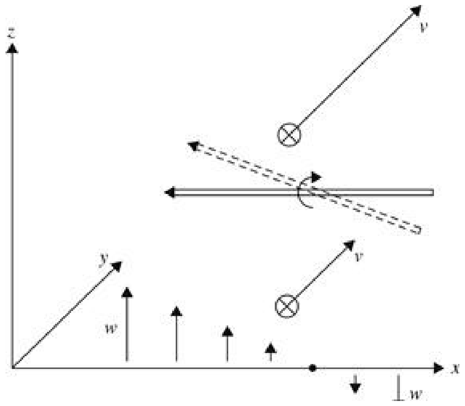


Figure 4.7 Vorticity generation by the tilting of a horizontal vorticity vector (*double arrow*).

Finally, the third term on the right in (4.13) is just the microscopic equivalent of the solenoidal term in the circulation theorem (4.5). To show this equivalence, we may apply Stokes's theorem to the solenoidal term to get

$$-\oint \alpha dp \equiv -\oint \alpha \boldsymbol{\nabla} p \cdot d\mathbf{l} = -\iint_A \boldsymbol{\nabla} \times (\alpha \boldsymbol{\nabla} p) \cdot \mathbf{k} dA$$

where A is the horizontal area bounded by the curve $\mathbf{l}$. Applying the vector identity $\boldsymbol{\nabla} \times (\alpha \boldsymbol{\nabla} p) \equiv \boldsymbol{\nabla} \alpha \times \boldsymbol{\nabla} p$, the equation becomes

$$-\oint \alpha dp = -\iint_A (\boldsymbol{\nabla} \alpha \times \boldsymbol{\nabla} p) \cdot \mathbf{k} dA$$

However, the solenoidal term in the vorticity equation can be written

$$-\left(\frac{\partial \alpha}{\partial x}\frac{\partial p}{\partial y} - \frac{\partial \alpha}{\partial y}\frac{\partial p}{\partial x}\right) = -(\boldsymbol{\nabla} \alpha \times \boldsymbol{\nabla} p) \cdot \mathbf{k}$$

Thus, the solenoidal term in the vorticity equation is just the limit of the solenoidal term in the circulation theorem divided by the area when the area goes to zero.

4.3.2 The Vorticity Equation in Isobaric Coordinates

A somewhat simpler form of the vorticity equation arises when the motion is referred to the isobaric coordinate system. This equation can be derived in vector form by operating on the momentum equation (3.2) with the vector operator $\mathbf{k} \cdot \nabla \times$, where ∇ now indicates the horizontal gradient on a surface of constant pressure. However, to facilitate this process it is desirable to first use the vector identity

$$(\mathbf{V} \cdot \boldsymbol{\nabla})\mathbf{V} = \boldsymbol{\nabla}\left(\frac{\mathbf{V} \cdot \mathbf{V}}{2}\right) + \zeta \, \mathbf{k} \times \mathbf{V} \tag{4.14}$$

where $\zeta = \mathbf{k} \cdot (\boldsymbol{\nabla} \times \mathbf{V})$, to rewrite (3.2) as

$$\frac{\partial \mathbf{V}}{\partial t} = -\boldsymbol{\nabla}\left(\frac{\mathbf{V} \cdot \mathbf{V}}{2} + \Phi\right) - (\zeta + f)\mathbf{k} \times \mathbf{V} - \omega \frac{\partial \mathbf{V}}{\partial p} \tag{4.15}$$

We now apply the operator $\mathbf{k} \cdot \boldsymbol{\nabla} \times$ to (4.15). Using the facts that for any scalar A, $\nabla \times \nabla A = 0$ and for any vectors **a**, **b**,

$$\nabla \times (\mathbf{a} \times \mathbf{b}) = (\nabla \cdot \mathbf{b})\mathbf{a} - (\mathbf{a} \cdot \nabla)\mathbf{b} - (\nabla \cdot \mathbf{a})\mathbf{b} + (\mathbf{b} \cdot \nabla)\mathbf{a} \tag{4.16}$$

we can eliminate the first term on the right and simplify the second term so that the resulting vorticity equation becomes

$$\frac{\partial \zeta}{\partial t} = -\mathbf{V}\cdot\boldsymbol{\nabla}(\zeta+f) - \omega\frac{\partial \zeta}{\partial p} - (\zeta+f)\,\boldsymbol{\nabla}\cdot\mathbf{V} + \mathbf{k}\cdot\left(\frac{\partial \mathbf{V}}{\partial p}\times\boldsymbol{\nabla}\omega\right) \tag{4.17}$$

Comparing (4.13) and (4.17), we see that in the isobaric system there is no vorticity generation by pressure-density solenoids. This difference arises because in the isobaric system, horizontal partial derivatives are computed with p held constant so that the vertical component of vorticity is $\zeta = (\partial v/\partial x - \partial u/\partial y)_p$, whereas in height coordinates it is $\zeta = (\partial v/\partial x - \partial u/\partial y)_z$. In practice the difference is generally unimportant because, as shown in the next section, the solenoidal term is usually sufficiently small so that it can be neglected for synoptic-scale motions.

4.3.3 Scale Analysis of the Vorticity Equation

In Section 2.4 the equations of motion were simplified for synoptic-scale motions by evaluating the order of magnitude of various terms. The same technique can be applied to the vorticity equation. Characteristic scales for the field variables based on typical observed magnitudes for synoptic-scale motions are chosen as follows:

$U \sim 10$ m s $^{-1}$	horizontal scale
$W \sim 1$ cm s $^{-1}$	vertical scale
$L \sim 10^6$ m	length scale
$H \sim 10^4$ m	depth scale
$\delta p \sim 10$ hPa	horizontal pressure scale
$\rho \sim 1$ kg m $^{-3}$	mean density
$\delta\rho/\rho \sim 10^{-2}$	fractional density fluctuation
$L/U \sim 10^5$ s	time scale
$f_0 \sim 10^{-4}$ s^{-1}	Coriolis parameter
$\beta \sim 10^{-11}$ m^{-1} s^{-1}	"beta” parameter

Again we have chosen an advective time scale because the vorticity pattern, like the pressure pattern, tends to move at a speed comparable to the horizontal wind speed. Using these scales to evaluate the magnitude of the terms in (4.12), we first note that

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \sim \frac{U}{L} \sim 10^{-5}\mathrm{s}^{-1}$$

where the inequality in this expression means less than or equal to in order of magnitude. Thus,

$$\zeta/f_0 \sim U/(f_0 L) \equiv \mathbf{Ro} \sim 10^{-1}$$

For midlatitude synoptic-scale systems, the relative vorticity is often small (order Rossby number) compared to the planetary vorticity. For such systems, ζ may be neglected compared to f in the divergence term in the vorticity equation

$$(\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \approx f\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)$$

This approximation does not apply near the center of intense cyclonic storms. In such systems $|\zeta/f| \sim 1$, and the relative vorticity should be retained.

The magnitudes of the various terms in (4.12) can now be estimated as

$$\begin{gathered}
\frac{\partial \zeta}{\partial t}, u\frac{\partial \zeta}{\partial x}, v\frac{\partial \zeta}{\partial y} \sim \frac{U^2}{L^2} \sim 10^{-10}\ \mathrm{s}^{-2} \\
w\frac{\partial \zeta}{\partial z} \sim \frac{WU}{HL} \sim 10^{-11}\ \mathrm{s}^{-2} \\
v\frac{df}{dy} \sim U\beta \sim 10^{-10}\ \mathrm{s}^{-2} \\
f\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \sim \frac{f_0 U}{L} \sim 10^{-9}\ \mathrm{s}^{-2} \\
\left(\frac{\partial w}{\partial x}\frac{\partial v}{\partial z} - \frac{\partial w}{\partial y}\frac{\partial u}{\partial z}\right) \sim \frac{WU}{HL} \sim 10^{-11}\ \mathrm{s}^{-2} \\
\frac{1}{\rho^2}\left(\frac{\partial \rho}{\partial x}\frac{\partial p}{\partial y} - \frac{\partial \rho}{\partial y}\frac{\partial p}{\partial x}\right) \sim \frac{\delta\rho\delta p}{\rho^2 L^2} \sim 10^{-11}\ \mathrm{s}^{-2}
\end{gathered}$$

The inequality is used in the last three terms because in each case it is possible that the two parts of the expression might partially cancel so that the actual magnitude would be less than indicated. In fact, this must be the case for the divergence term (the fourth in the list) because if $\partial u/\partial x$ and $\partial v/\partial y$ were not nearly equal and opposite, the divergence term would be an order of magnitude greater than any other term and the equation could not be satisfied. Therefore, scale analysis of the vorticity equation indicates that synoptic-scale motions must be quasi-nondivergent. The divergence term will be small enough to be balanced by the vorticity advection terms only if

$$\left|\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)\right| \sim 10^{-6}\ \mathrm{s}^{-1}$$

so that the horizontal divergence must be small compared to the vorticity in synoptic-scale systems. From the aforementioned scalings and the definition of the Rossby number, we see that

$$\left|\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)\Big/ f_0\right| \sim \mathrm{Ro}^2$$

and

$$\left|\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)\Big/ \zeta\right| \sim \mathrm{Ro}$$

Thus, the ratio of the horizontal divergence to the relative vorticity is the same magnitude as the ratio of relative vorticity to planetary vorticity.

Retaining only the terms of order $10^{-10}\ \mathrm{s}^{-2}$ in the vorticity equation yields the approximate form valid for synoptic-scale motions,

$$\frac{D_h(\zeta + f)}{Dt} = -f\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \tag{4.18}$$

where

$$\frac{D_h}{Dt} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}$$

As mentioned earlier, (4.18) is not accurate in intense cyclonic storms. For these the relative vorticity should be retained in the divergence term:

$$\frac{D_h(\zeta + f)}{Dt} = -(\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \tag{4.19}$$

Equation (4.18) states that the change of absolute vorticity following the horizontal motion on the synoptic scale is given approximately by the concentration or dilution of planetary vorticity caused by the convergence or divergence of the horizontal flow, respectively. In (4.19), however, it is the concentration or dilution of absolute vorticity that leads to changes in absolute vorticity following the motion.

The form of the vorticity equation given in (4.19) also indicates why cyclonic disturbances can be much more intense than anticyclones. For a fixed amplitude of convergence, relative vorticity will increase, and the factor $(\zeta + f)$ becomes larger, which leads to even higher rates of increase in the relative vorticity. For a fixed rate of divergence, however, relative vorticity will decrease, but when $\zeta \to -f$, the divergence term on the right approaches zero and the relative vorticity cannot become more negative no matter how strong the divergence. (This difference in the potential intensity of cyclones and anticyclones was discussed in Section 3.2.5 in connection with the gradient wind approximation.)

The approximate forms given in (4.18) and (4.19) do not remain valid, however, in the vicinity of atmospheric fronts. The horizontal scale of variation in frontal zones is only ~ 100 km, and the vertical velocity scale is ~ 10 cm s^{-1}. For these scales, vertical advection, tilting, and solenoidal terms all may become as large as the divergence term.

4.4 Potential Vorticity

We return now to circulation to show that Kelvin's theorem applies to baroclinic dynamics for certain integration contours, which has very deep implications for dynamic meteorology. Rather than arbitrary closed contours, we now restrict the possibilities to those that fall on isentropic surfaces, for which potential temperature is constant. We prove that for these contours, the solenoidal term is exactly zero.

With the aid of the ideal gas law (1.25), the definition of potential temperature (2.44) can be expressed as a relationship between pressure and density for a surface of constant θ:

$$\rho = p^{c_v/c_p}(R\theta)^{-1}(p_s)^{R/c_p}$$

Hence, on an isentropic surface, density is a function of pressure alone, and the solenoidal term in the circulation theorem (4.3) vanishes;

$$\oint \frac{dp}{\rho} \propto \oint dp^{(1-c_v/c_p)} = 0$$

For adiabatic and frictionless flow, then, the circulation computed for a closed chain of fluid parcels on a constant θ surface reduces to the same form as in a barotropic fluid; that is, it *satisfies Kelvin's circulation theorem* even though the fluid is baroclinic!

From Stokes's theorem, we may replace the circulation in favor of the component of vorticity normal to the surface

$$C_a = \oint \mathbf{U}_a \cdot d\mathbf{l} = \iint_A \omega_a \cdot \mathbf{n}\, dA \tag{4.20}$$

where **n** is a unit vector normal to the surface. Now we replace **n** and *dA* in favor of conserved quantities. Consider an infinitesimal cylinder (Figure 4.8) that is bounded above and below by constant potential temperature surfaces; for adiabatic flow, both the potential temperature and the mass of air in this cylinder, *dm*, are conserved. A leading-order Taylor approximation across the cylinder gives $d\theta \approx |\nabla\theta| dh$, where *dh* is the "height" of the cylinder. The mass of air in the cylinder is given by $dm = \rho dA dh$ so that

$$dA = \frac{dm}{\rho} \frac{|\nabla\theta|}{d\theta} \tag{4.21}$$

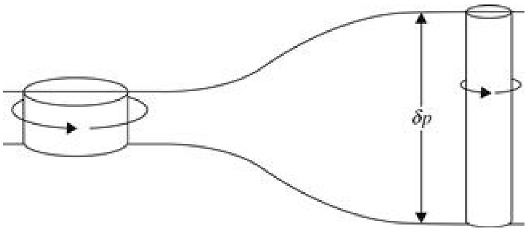


Figure 4.8 A cylindrical column of air moving adiabatically, conserving potential vorticity.

Taking the infinitesimal cylinder to be sufficiently small so that the vorticity normal to the surface is nearly constant,

$$\frac{DC_a}{Dt} \approx \frac{D}{Dt}[\boldsymbol{\omega}_a \cdot \mathbf{n}\, dA] = 0 \tag{4.22}$$

where

$$\mathbf{n} = \frac{\nabla\theta}{|\nabla\theta|} \tag{4.23}$$

Using (4.21) and (4.23) in (4.22), and noting that both *dm* and *dθ* are conserved following the motion, gives

$$\frac{D}{Dt}\left[\frac{\boldsymbol{\omega}_a \cdot \nabla\theta}{\rho}\right] = 0 \tag{4.24}$$

This celebrated equation, the **Ertel potential vorticity theorem**, is one of the most important theoretical results in dynamic meteorology. It says that the potential vorticity, or PV,

$$\Pi = \frac{\boldsymbol{\omega}_a \cdot \boldsymbol{\nabla}\theta}{\rho} \tag{4.25}$$

is conserved following the motion.

What makes this result so profound is that it links all of the basic physical conservation laws into a single expression. Although it may appear, as one studies a fluid, that the momentum field fluctuates independently of the thermodynamic fields, the potential vorticity places a powerful constraint on those fluctuations: They must evolve in a way that preserves the PV following the motion. Take, for example, the situation where the absolute vorticity vector has no horizontal components so that the PV is given by

$$\frac{1}{\rho}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} + f\right)\frac{\partial \theta}{\partial z} \tag{4.26}$$

In this case we see that the PV is the product of the absolute vorticity and the static stability so that if one increases, the other must decrease. This is illustrated in Figure 4.8: As the air parcel moves to the right, the static stability decreases, and as a result the vorticity must increase to conserve PV; the connections to angular momentum conservation and to vortex stretching are particularly clear.

The vertical contribution to the PV in (4.25) scales like $\frac{U\theta^*}{\rho H L} + \frac{f\theta^*}{H}$, where θ^* and ρ^* are typical potential temperature and density scale values; the second term represents the contribution from planetary rotation. The horizontal contribution—that is, from expanding (4.25)—scales like $\frac{U\theta^*}{\rho^* H L}$ (neglecting the small contribution from *w*), so that the ratio of vertical to horizontal contributions is

$$1 + \mathrm{Ro}^{-1} \tag{4.27}$$

On synoptic and larger scales, where the Rossby number is small (roughly 0.1), this analysis shows that the dominant contribution to the PV comes from the vertical direction. This serves to *explain why we focus on the vertical component of vorticity in dynamic meteorology*: Although the vorticity vector lies primarily in the horizontal direction, the gradient of potential temperature is mainly in the vertical and this promotes the importance of vertical vorticity. A leading-order approximation to the PV is

$$\Pi \approx \frac{f}{\rho}\frac{\partial \theta}{\partial z} \tag{4.28}$$

Picking typical values for the troposphere, we get an estimate of characteristic PV values of

$$\begin{aligned} \Pi_{trop} &\cong \left(\frac{10^{-4}\mathrm{s}^{-1}}{1\mathrm{kg\ m}^{-3}}\right)\left(5\ \mathrm{K\ km}^{-1}\right) \\ &= 0.5 \times 10^{-6}\,\mathrm{K\ m^2\ s^{-1}\ kg^{-1}} \equiv 0.5\,\mathrm{PVU} \end{aligned} \tag{4.29}$$

The definition “PVU” is a commonly adopted scaling of “PV Units” ($1\,\mathrm{PVU} = 10^{-6}\,\mathrm{K\ m^2\ s^{-1}\ kg^{-1}}$). In the lower stratosphere, where $\frac{\partial \theta}{\partial z}$ is an order of magnitude larger, characteristic values of PV are larger as well.

Because there is an abrupt jump in PV values at the troposphere PV provides a useful dynamical definition for that interface. In fact, for adiabatic and frictionless conditions, when both PV and potential temperature are conserved, maps of potential temperature on the “dynamical” tropopause (defined as a PV surface) provide a particularly useful summary of extratropical weather systems, because the potential temperature contours are simply advected by the wind on the tropopause. Moreover, compared to the alternative of PV contours on potential temperature surfaces, a single map suffices to describe the state of the tropopause.

An example comparing the structure of the dynamical tropopause with a conventional map of 500-hPa geopotential height is shown in Figure 4.9. Familiar features on the 500 hPa chart include a long-wave pattern with a ridge over western North America and a trough over the center of the continent, and short-wave troughs near New York and approaching the Pacific coastline near the Alaska panhandle. In addition to the wave-like features, cyclonic vortices are located southwest of California, over Hudson Bay, and near Baffin Island.

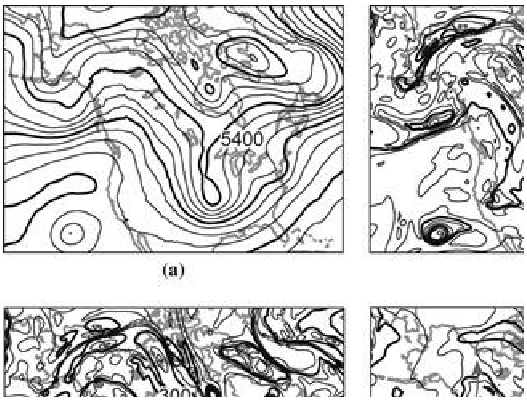


Figure 4.9 Comparison of 500-hPa geopotential height (a) and tropopause maps of pressure (b) potential temperature (c) and wind speed (d) on January 12, 2012. Geopotential height is shown every 60 m, potential temperature every 5 K, pressure every 50 hPa, and wind speed every 10 m s^{-1}. Every fourth contour is shown *bold* except every second contour for wind speed. *Gray arrows* in (d) illustrate the path of the main branches of jet streams.

Tropopause pressure (Figure 4.9b) reveals that troughs are regions where the tropopause is depressed to lower altitudes; in fact, the tropopause is below 500 hPa in several of the troughs and vortices. Therefore, troughs are associated with large values of stratospheric potential vorticity that have been locally lowered to elevations normally considered tropospheric. In ridges and subtropical latitudes, the tropopause is elevated, with pressure values of 200 to 250 hPa, and therefore the PV is anomalously low compared to the surroundings.

Tropopause potential temperature (Figure 4.9c) shows that most of the troughs apparent in the 500-hPa chart are in fact vortical near their core. We may infer this by the appearance of closed potential temperature contours, since, if potential temperature and potential vorticity are conserved, air is trapped within closed tropopause potential temperature contours. For this reason, these features are sometimes called material eddies, since the only means of moving the disturbance is to displace the material comprising it. This stands in contrasts to waves, which may propagate information without net transport of the medium. In addition to the disturbances, notice the long ribbons where the horizontal gradient of tropopause potential temperature is concentrated into fronts, separated by zones that are mixed and filled with small-scale noise. Along some of these fronts—for example, the base of the trough in the central United States—the pressure field reveals that the tropopause is essentially vertical.

Finally, since jet streams are located near the tropopause, this perspective is much more useful than isobaric surfaces, which require several maps to depict jet streams located at different elevations. We see that a midlatitude jet reaches the west coast of Canada before splitting into northern and southern branches; the southern branch also has a subtropical connection that is apparent near the southern edge of the domain (Figure 4.9d).

For completeness, we note that a form of the Ertel potential vorticity equation may be derived that includes a source of momentum, $\mathcal{F}$, such as frictional dissipation, and a source of entropy, $\mathcal{H}$, such as latent heating. The result is

$$\frac{D\Pi}{Dt} = \frac{\boldsymbol{\omega}_a}{\rho}\cdot\nabla\mathcal{H} + \frac{\nabla\theta}{\rho}\cdot\left(\nabla \times \frac{\mathcal{F}}{\rho}\right) \tag{4.30}$$

where Π is the Ertel PV defined by (4.25). The first right-side term is positive, where the vorticity vector points in the direction of a local maximum in the entropy source. This often occurs near extratropical cyclones, where clouds and precipitation are found in the lower and midtroposphere (Figure 4.10a).

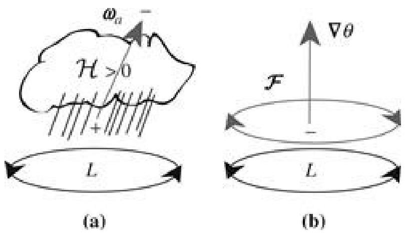


Figure 4.10 Cartoon illustrating the role of sources of (a) entropy and (b) momentum on changing the potential vorticity. (a) Depicts a precipitating cloud above a surface cyclone, where the vorticity vector, ω, points upward and to the right. Latent heating reaches a maximum in the cloud, which produces a positive (negative) PV tendency below (above) the cloud. (b) Depicts the role of surface friction (*gray arrow* opposite surface circulation). The curl of surface friction points in the opposite direction of the vorticity vector, producing a negative PV tendency.

In this case, near the surface the vertical component of vorticity points upward toward a maximum in latent heat release where condensation takes place; above the maximum in heating, potential vorticity decreases because the vorticity vector and gradient in heating point in opposite directions. The second right-side term is positive where the curl of the frictional force points in the same direction as the gradient in entropy. Again, appealing to extratropical cyclones, assuming that surface friction acts opposite to the motion, the curl of friction is a vector that points in the opposite direction from the gradient in entropy, which produces a negative tendency to the Ertel PV (Figure 4.10b).

4.5 Shallow Water Equations

In a homogeneous incompressible fluid, potential vorticity conservation takes a somewhat simpler form. To see this, we first derive the shallow water equations, which provide a very useful simplification that we will use later to isolate essential aspects of atmospheric dynamics. Beyond constant density, the shallow approximation implies that the depth of the fluid is small compared to the the horizontal scale of the features of interest. As a result of this small aspect ratio, we may use the hydrostatic approximation

$$\frac{\partial p}{\partial z} = \rho_0 g \tag{4.31}$$

where ρ_0 is a constant density. Integrating the hydrostatic equation over the depth of the fluid, $h(x,y)$, gives

$$p(z) = \rho_0 g(h - z) + p(h) \tag{4.32}$$

where $p(h)$ is the pressure at the top of the layer of shallow water due to the layer above, which we take to be a constant. Using (4.32) to replace pressure in the momentum equation gives

$$\frac{D_h \mathbf{V}}{Dt} = -g\boldsymbol{\nabla}_h h \; - \; f\mathbf{k} \times \mathbf{V} \tag{4.33}$$

We assume that **V** is initially a function of (x, y) only, and since h is a function of (x,y), (4.33) indicates that **V** will remain two-dimensional for all time.

Mass conservation for a constant density fluid has the simple form

$$\boldsymbol{\nabla}\cdot(u, v, w) = 0 \tag{4.34}$$

Incompressibility also simplifies the first law of thermodynamics, since the fluid can no longer do work. Therefore, if no heat is added, the temperature is constant following the motion (see 2.41). Water pressure is a function of density and temperature, $p = f(T,\rho)$, but following the motion incompressibility implies $dp = \frac{\partial f}{\partial T}dT + \frac{\partial f}{\partial \rho}d\rho = 0$, so that

$$\frac{Dp}{Dt} = 0 \tag{4.35}$$

Applying this fact to (4.32) yields

$$\frac{D_h h}{Dt} = w(h) \tag{4.36}$$

where $w = \frac{Dz}{Dt}$ is the vertical motion, which *is* a function of (x, y, z). Integrating (4.34) over depth h gives $w(h) = -h\boldsymbol{\nabla}_h\cdot\mathbf{V}$, which may be used to eliminate w from (4.36)

$$\frac{D_h h}{Dt} = -h\boldsymbol{\nabla}_h\cdot\mathbf{V} \tag{4.37}$$

The shallow water equations consist of (4.33) and (4.37), which describe the evolution of three unknowns (u,v,h). Following a derivation similar to that in Section 4.3.1, the shallow water vorticity equation may be derived from (4.33) and yields

$$\frac{D_h}{Dt}(\zeta + f) = -(\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \tag{4.38}$$

In the shallow water system, absolute vorticity increases following the motion through vortex stretching.

Shallow water potential vorticity conservation is obtained by replacing divergence on the right side of (4.38) using (4.37) to give

$$\frac{D_h}{Dt}\left[\frac{\zeta+f}{h}\right]=0 \tag{4.39}$$

Thus, the shallow water potential vorticity, $(\zeta+f)/h$, is given by the absolute vorticity divided by the fluid depth. Following the motion, if the absolute vorticity increases, then the depth of the fluid must as well. One may think of the inverse depth as an analog of static stability in shallow water by noting that the layer top and bottom surfaces are isentropic surfaces. When the layer depth, h, becomes smaller, the isentropes are closer together.

Since the potential vorticity depends on x and y, only (4.39) provides a dramatic simplification to the Ertel PV, which is very useful for gaining insight into large-scale dynamics. For example, consider westerly flow impinging on an infinitely long topographic barrier as shown in Figure 4.11. We suppose that flow upstream of the mountain barrier is uniform with $\zeta=0$. Each fluid column of depth h is confined between constant entropy surfaces θ_0 and $\theta_0+\delta\theta$ and remains between those surfaces as it crosses the barrier. Air that previously crossed the barrier leaves the upper surface higher near the mountain both upstream and downstream, for reasons that will be explored in Chapter 5.

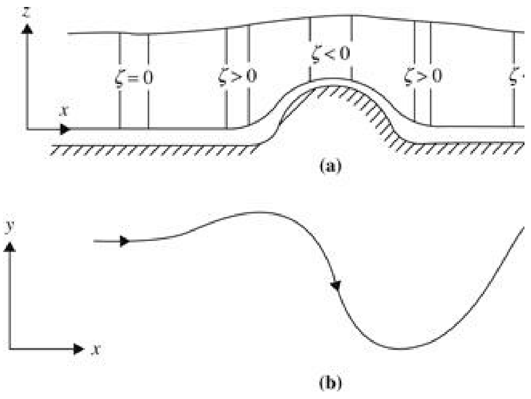


Figure 4.11 Schematic view of westerly flow over a topographic barrier: (a) the depth of a fluid column as a function of x and (b) the trajectory of a parcel in the (x,y) plane.

As fluid columns approach the topographic barrier, they stretch vertically, which causes ζ to increase in order to conserve potential vorticity, $(\zeta+f)/h$. Positive cyclonic vorticity is associated with cyclonic curvature in the flow, which causes a poleward drift so that f also increases; this in turn reduces the change in ζ required for potential vorticity conservation. As the column begins to cross the barrier, its vertical extent decreases, and the relative vorticity must then become negative. Thus, the air column will acquire anticyclonic vorticity and move southward.

When the air column has passed over the barrier and returned to its original depth, it will be south of its original latitude so that f will be smaller and the relative vorticity must be positive. Thus, the trajectory must have cyclonic curvature and the column will be deflected poleward. When the parcel returns again to its original latitude, it will still have a poleward velocity component and will continue poleward gradually, acquiring anticyclonic curvature until its direction is again reversed. The parcel will then move downstream, conserving potential vorticity by following a wave-like trajectory in the horizontal plane. Therefore, steady westerly flow over a large-scale ridge will result in an anticyclonic flow over the mountain, a cyclonic flow pattern to the east of the barrier, followed by a wavetrain downstream.

The situation for easterly flow impinging on a mountain barrier is quite different (Figure 4.12). Upstream stretching leads to a cyclonic turning of the flow, which results in an equatorward component of motion. As the column moves westward and equatorward over the barrier, its depth contracts and its absolute vorticity must then decrease so that potential vorticity can be conserved. This reduction in absolute vorticity arises both from development of anticyclonic relative vorticity and from a decrease in f due to the equatorward motion. The anticyclonic relative vorticity gradually turns the column so that when it reaches the top of the barrier, it is headed westward. As it continues westward down the barrier, conserving potential vorticity, the process is simply reversed, with the result that some distance downstream from the mountain barrier, the air column is again moving westward at its original latitude.

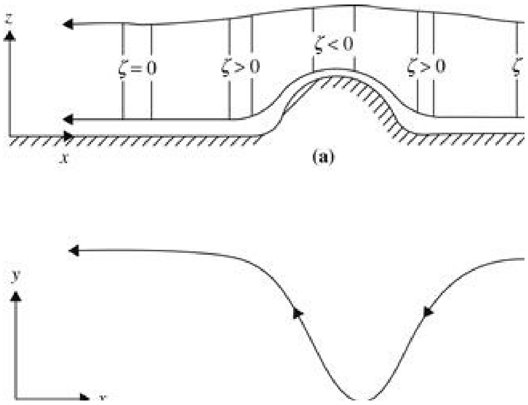


Figure 4.12 As in Figure 4.11, but for easterly flow.

Thus, the dependence of the Coriolis parameter on latitude creates a dramatic difference between westerly and easterly flow over large-scale topographic barriers. In the case of a westerly wind, the barrier generates a wave-like disturbance in the streamlines that extends far downstream. However, in the case of an easterly wind, the disturbance in the streamlines damps out away from the barrier.

4.5.1 Barotropic Potential Vorticity

Further simplification of potential vorticity is possible beyond the shallow water equations by making the barotropic assumption. Recall that a barotropic fluid is defined as one for which pressure depends only on density. For shallow water, (4.32) implies that in this case h and z are constant, so that $w = 0$ and the fluid is horizontally nondivergent

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0 \tag{4.40}$$

Potential vorticity conservation simplifies to

$$\frac{D_h}{Dt}(\zeta+f)=0 \tag{4.41}$$

which states that absolute vorticity is conserved following the motion. Equation (4.41) is called the barotropic vorticity equation and has been widely used in theoretical studies of large-scale dynamical meteorology.

For nondivergent horizontal motion, the flow field can be represented by a *streamfunction* $\psi(x,y)$ defined so that the velocity components are given as $u = -\partial\psi/\partial y,\ v = +\partial\psi/\partial x$. The vorticity is then given by

$$\zeta = \partial v/\partial x - \partial u/\partial y = \partial^2\psi/\partial x^2 + \partial^2\psi/\partial y^2 \equiv \nabla_h^2\psi$$

Thus, the velocity field and the vorticity can both be represented in terms of the variation of the single scalar field $\psi(x,y)$, and (4.41) can be written as a prognostic equation for vorticity in the form

$$\frac{\partial}{\partial t}\nabla^2\psi = -\mathbf{V}_\psi \cdot \nabla\left(\nabla^2\psi + f\right) \qquad (4.42)$$

where $\mathbf{V}_\psi \equiv \mathbf{k} \times \nabla\psi$ is a nondivergent horizontal wind. Equation (4.42) states that the local tendency of relative vorticity is given by the advection of absolute vorticity. This equation can be solved numerically to predict the evolution of the streamfunction and thus of the vorticity and wind fields. Because the flow in the midtroposphere is often nearly nondivergent on the synoptic scale, (4.42) provides a surprisingly good model for short-term forecasts of the synoptic-scale 500-hPa flow field.

Conservation of absolute vorticity following the motion provides a strong constraint on the flow, as can be shown by a simple example that again illustrates an asymmetry between westerly and easterly flow. Suppose that at a certain point (x_0, y_0) the flow is in the zonal direction and the relative vorticity vanishes so that $\eta(x_0, y_0) = f_0$. Then, if absolute vorticity is conserved, the motion at any point along a parcel trajectory that passes through (x_0, y_0) must satisfy $\zeta + f = f_0$. Because f increases toward the north, trajectories that curve northward in the downstream direction must have $\zeta = f_0 - f < 0$, whereas trajectories that curve southward must have $\zeta = f_0 - f > 0$. However, as indicated in Figure 4.13, if the flow is westerly, northward curvature downstream implies $\zeta > 0$, whereas southward curvature implies $\zeta < 0$. Thus, westerly zonal flow must remain purely zonal if absolute vorticity is to be conserved following the motion. The easterly flow case, also shown in Figure 4.13, is just the opposite. Northward and southward curvatures are associated with negative and positive relative vorticities, respectively. Hence, an easterly current can curve either to the north or to the south and still conserve absolute.

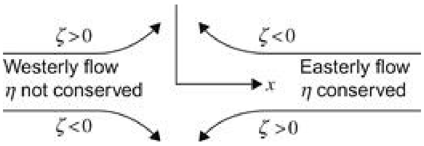


Figure 4.13 Absolute vorticity conservation for curved flow trajectories.

4.6 Ertel Potential Vorticity in Isentropic Coordinates

We consider here a more detailed treatment of the Ertel potential vorticity in isentropic coordinates, including nonconservative effects due to sources of momentum and entropy. We begin with a description of the basic conservation laws in isentropic coordinates.

4.6.1 Equations of Motion in Isentropic Coordinates

If the atmosphere is stably stratified so that potential temperature θ is a monotonically increasing function of height, θ may be used as an independent vertical coordinate. The vertical "velocity" in this coordinate system is just $\dot{\theta} \equiv D\theta/Dt$. Thus, adiabatic motions are two-dimensional when viewed in an isentropic coordinate frame. An infinitesimal control volume in isentropic coordinates with cross-sectional area δA and vertical extent $\delta\theta$ has a mass

$$\delta M = \rho\delta A\delta z = \delta A\left(-\frac{\delta p}{g}\right) = \frac{\delta A}{g}\left(-\frac{\partial p}{\partial\theta}\right)\delta\theta = \sigma\delta A\delta\theta \tag{4.43}$$

Here the "density" in (x, y, θ) space is defined as

$$\sigma \equiv -g^{-1}\partial p/\partial\theta \tag{4.44}$$

The horizontal momentum equation in isentropic coordinates may be obtained by transforming the isobaric form (4.15) to yield

$$\frac{\partial\mathbf{V}}{\partial t} + \boldsymbol{\nabla}_\theta\left(\frac{\mathbf{V}\cdot\mathbf{V}}{2} + \Psi\right) + (\zeta_\theta + f)\,\mathbf{k}\times\mathbf{V} = -\dot{\theta}\frac{\partial\mathbf{V}}{\partial\theta} + \mathbf{F}_r \tag{4.45}$$

where $\boldsymbol{\nabla}_\theta$ is the gradient on an isentropic surface, $\zeta_\theta \equiv \mathbf{k}\cdot\boldsymbol{\nabla}_\theta\times\mathbf{V}$ is the isentropic relative vorticity, and $\Psi \equiv c_pT + \Phi$ is the Montgomery streamfunction (see Problem 2.11 in Chapter 2). We have included a frictional term $\mathbf{F}_r$ on the right side, along with the diabatic vertical advection term. The continuity equation can be derived with the aid of (4.43) in a manner analogous to that used for the isobaric system in Section 3.1.2. The result is

$$\frac{\partial\sigma}{\partial t} + \boldsymbol{\nabla}_\theta\cdot(\sigma\mathbf{V}) = -\frac{\partial}{\partial\theta}\left(\sigma\dot{\theta}\right) \tag{4.46}$$

The Ψ and σ fields are linked through the pressure field by the hydrostatic equation, which in the isentropic system takes the form

$$\frac{\partial\Psi}{\partial\theta} = \Pi(p) \equiv c_p\left(\frac{p}{p_s}\right)^{R/c_p} = c_p\frac{T}{\theta} \tag{4.47}$$

where Π is called the *Exner function*. Equations (4.44) through (4.47) form a closed set for prediction of $\mathbf{V}$, σ, Ψ, and p, provided that $\dot{\theta}$ and $\mathbf{F}_r$ are known.

4.6.2 The Potential Vorticity Equation

If we take $\mathbf{k}\cdot\boldsymbol{\nabla}_\theta\times$ (4.45) and rearrange the resulting terms, we obtain this isentropic vorticity equation:

$$\frac{\tilde{D}}{Dt}(\zeta_\theta + f) + (\zeta_\theta + f)\,\boldsymbol{\nabla}_\theta \cdot \mathbf{V} = \mathbf{k} \cdot \boldsymbol{\nabla}_\theta \times \left(\mathbf{F}_r - \dot{\theta}\frac{\partial \mathbf{V}}{\partial \theta}\right) \tag{4.48}$$

where

$$\frac{\tilde{D}}{Dt} = \frac{\partial}{\partial t} + \mathbf{V} \cdot \boldsymbol{\nabla}_\theta$$

is the total derivative following the horizontal motion on an isentropic surface.

Noting that $\sigma^{-2}\partial\sigma/\partial t = -\partial\sigma^{-1}/\partial t$, we can rewrite (4.46) in the form

$$\frac{\tilde{D}}{Dt}\left(\sigma^{-1}\right) - \left(\sigma^{-1}\right)\nabla_\theta \cdot \mathbf{V} = \sigma^{-2}\frac{\partial}{\partial \theta}\left(\sigma\dot{\theta}\right) \tag{4.49}$$

Multiplying each term in (4.48) by σ^{-1} and in (4.49) by $(\zeta_\theta + f)$ and adding, we obtain the desired conservation law:

$$\frac{\tilde{D}\Pi}{Dt} = \frac{\partial \Pi}{\partial t} + \mathbf{V} \cdot \boldsymbol{\nabla}_\theta \Pi = \frac{\Pi}{\sigma}\frac{\partial}{\partial \theta}\left(\sigma\dot{\theta}\right) + \sigma^{-1}\mathbf{k} \cdot \boldsymbol{\nabla}_\theta \times \left(\mathbf{F}_r - \dot{\theta}\frac{\partial \mathbf{V}}{\partial \theta}\right) \tag{4.50}$$

where $\Pi \equiv (\zeta_\theta + f)/\sigma$ is the Ertel potential vorticity. If the diabatic and frictional terms on the right side of (4.50) can be evaluated, it is possible to determine the evolution of Π following the *horizontal* motion on an isentropic surface. When the diabatic and frictional terms are small, potential vorticity is approximately conserved following the motion on isentropic surfaces.

Weather disturbances that have sharp gradients in dynamical fields, such as jets and fronts, are associated with large anomalies in the Ertel potential vorticity. In the upper troposphere such anomalies tend to be advected rapidly under nearly adiabatic conditions. Thus, the potential vorticity anomaly patterns are conserved materially on isentropic surfaces. This material conservation property makes potential vorticity anomalies particularly useful in identifying and tracing the evolution of meteorological disturbances.

4.6.3 Integral Constraints on Isentropic Vorticity

The isentropic vorticity equation (4.48) can be written in the form

$$\frac{\partial \zeta_\theta}{\partial t} = -\boldsymbol{\nabla}_\theta \cdot \left[(\zeta_\theta + f)\mathbf{V}\right] + \mathbf{k} \cdot \boldsymbol{\nabla}_\theta \times \left(\mathbf{F}_r - \dot{\theta}\frac{\partial \mathbf{V}}{\partial \theta}\right) \tag{4.51}$$

Using the fact that any vector **A** satisfies the relationship

$$\mathbf{k} \cdot \left(\boldsymbol{\nabla}_\theta \times \mathbf{A}\right) = \boldsymbol{\nabla}_\theta \cdot \left(\mathbf{A} \times \mathbf{k}\right)$$

we can rewrite (4.51) in the form

$$\frac{\partial \zeta_\theta}{\partial t} = -\boldsymbol{\nabla}_\theta \cdot \left[(\zeta_\theta + f)\mathbf{V} - \left(\mathbf{F}_r - \dot{\theta}\frac{\partial \mathbf{V}}{\partial \theta}\right) \times \mathbf{k}\right] \tag{4.52}$$

Equation (4.52) expresses the remarkable fact that isentropic vorticity can only be changed by the divergence or convergence of the horizontal flux vector in brackets on the right side. The vorticity cannot be changed by vertical transfer across the isentropes. Furthermore, integration of (4.52) over the area of an isentropic surface and application of the divergence theorem (Appendix C.2) show that for an isentrope that does not intersect the surface of Earth, the global average of ζ_θ is constant. Furthermore, integration of ζ_θ over the sphere shows that the global average ζ_θ is exactly zero. Vorticity on such an isentrope is neither created nor destroyed; it is merely concentrated or diluted by horizontal fluxes along the isentropes.

Suggested References

cheson. Acheson, *Elementary Fluid Dynamics*, provides a good introduction to vorticity at a graduate level.

oskins et al. Hoskins et al. provide an advanced discussion of Ertel potential vorticity and its uses in diagnosis and prediction of synoptic-scale disturbances.

edlosky. Pedlosky, *Geophysical Fluid Dynamics*, provides a thorough treatment of circulation, vorticity, and potential vorticity in Chapter 2.

allis. Vallis, *Atmospheric and Oceanic Fluid Dynamics*, discusses vorticity and potential vorticity in Chapter 4.

illiams and Elder. Williams and Elder, *Fluid Physics for Oceanographers and Physicists*, provides an introduction to vorticity dynamics at an elementary level. This book also provides a good general introduction to fluid dynamics.

Problems

4.1. What is the circulation about a square of 1000 km on a side for an easterly (i.e., westward flowing) wind that decreases in magnitude toward the north at a rate of 10 m s^{-1} per 500 km? What is the mean relative vorticity in the square?

4.2. A cylindrical column of air at 30° N with radius 100 km expands to twice its original radius. If the air is initially at rest, what is the mean tangential velocity at the perimeter after expansion?

4.3. An air parcel at 30° N moves northward, conserving absolute vorticity. If its initial relative vorticity is $5 \times 10^{-5}\,\mathrm{s}^{-1}$, what is its relative vorticity upon reaching 90° N?

4.4. An air column at 60° N with $\zeta = 0$ initially stretches from the surface to a fixed tropopause at 10 km height. If the air column moves until it is over a mountain barrier 2.5 km high at 45° N, what is its absolute vorticity and relative vorticity as it passes the mountain top, assuming that the flow satisfies the barotropic potential vorticity equation?

4.5. Find the average vorticity within a cylindrical annulus of inner radius 200 km and outer radius 400 km if the tangential velocity distribution is given by $V = A/r$, where $A = 10^6\ \mathrm{m}^2\,\mathrm{s}^{-1}$ and r is in meters. What is the average vorticity within the inner circle of radius 200 km?

4.6. Show that the anomalous gradient wind cases discussed in Section 3.2.5 have negative absolute circulation in the Northern Hemisphere and thus have negative average absolute vorticity.

4.7. Compute the rate of change of circulation about a square in the (x,y) plane with corners at $(0,0), (0,L), (L,L),$ and $(L,0)$ if temperature increases eastward at a rate of 1° C per 200 km and presure increases northward at a rate of 1 hPa per 200 km. Let $L = 1000$ km and the pressure at the point (0,0) be 1000 hPa.

4.8. Verify the identity (4.14) by expanding the vectors in Cartesian components.

4.9. Derive a formula for the dependence of depth on radius for an incompressible fluid in solid-body rotation in a cylindrical tank with a flat bottom and free surface at the upper boundary. Let H be the depth at the center of the tank, Ω be the angular velocity of rotation of the tank, and a be the radius of the tank.

4.10. By how much does the relative vorticity change for a column of fluid in a rotating cylinder if the column is moved from the center of the tank to a distance 50 cm from the center? The tank is rotating at the rate of 20 revolutions per minute, the depth of the fluid at the center is 10 cm, and the fluid is initially in solid-body rotation.

4.11. A cyclonic vortex is in cyclostrophic balance with a tangential velocity profile given by the expression $V = V_0(r/r_0)^n$, where V_0 is the tangential velocity component at the distance r_0 from the vortex center. Compute the circulation about a streamline at radius r, the vorticity at radius r, and the pressure at radius r. (Let p_0 be the pressure at r_0 and assume that density is a constant.)

4.12. A westerly zonal flow at 45° is forced to rise adiabatically over a north–south-oriented mountain barrier. Before striking the mountain, the westerly wind increases linearly toward the south at a rate of 10 m s^{-1} per 1000 km. The crest of the mountain range is at 800 hPa and the tropopause, located at 300 hPa, remains undisturbed. What is the initial relative vorticity of the air? What is its relative vorticity when it reaches the crest if it is deflected 5° latitude toward the south during the forced ascent? If the current assumes a uniform speed of 20 m s^{-1} during its ascent to the crest, what is the radius of curvature of the streamlines at the crest?

4.13. A cylindrical vessel of radius a and constant depth H rotating at an angular velocity Ω about its vertical axis of symmetry is filled with a homogeneous, incompressible fluid that is initially at rest with respect to the vessel. A volume of fluid V is then withdrawn through a point sink at the center of the cylinder, thus creating a vortex. Neglecting friction, derive an expression for the resulting relative azimuthal velocity as a function of radius (i.e., the velocity in a coordinate system rotating with the tank). Assume that the motion is independent of depth and that $V \ll \pi a^2 H$. Also compute the relative vorticity and the relative circulation.

4.14.

(a) How far must a zonal ring of air initially at rest with respect to Earth's surface at 60° latitude and 100-km height be displaced latitudinally to acquire an easterly (east to west) component of 10 m s^{-1} with respect to Earth's surface?
(b) To what height must it be displaced vertically in order to acquire the same velocity? Assume a frictionless atmosphere.

4.15. The horizontal motion within a cylindrical annulus with permeable walls of inner radius 10 cm, outer radius 20 cm, and 10-cm depth is independent of height and azimuth and is represented by the expressions $u = 7 - 0.2r, v = 40 + 2r,$ where u and v are the radial and tangential velocity components in cm s^{-1}, positive outward and counterclockwise, respectively, and r is distance from the center of the annulus in centimeters. Assuming an incompressible fluid, find

(a) Circulation about the annular ring
(b) Average vorticity within the annular ring
(c) Average divergence within the annular ring
(d) Average vertical velocity at the top of the annulus if it is zero at the base

4.16. Prove that, as stated in Eq. (4.52), the globally averaged isentropic vorticity on an isentropic surface that does not intersect the ground must be zero. Show that the same result holds for the isobaric vorticity on an isobaric surface.

MATLAB Exercises

M4.1. Equation 4.41 showed that for nondivergent horizontal motion, the flow field can be represented by a *streamfunction* $\psi(x, y)$, and the vorticity is then given by $\zeta = \partial^2\psi/\partial x^2 + \partial^2\psi/\partial y^2 \equiv \nabla^2\psi$. Thus, if the vorticity is represented by a single sinusoidal wave distribution in both x and y, the streamfunction has the same spatial distribution as the vorticity and the opposite sign, as can be verified easily from the fact that the second derivative of a sine is proportional to minus the same sine function. An example is shown in the MATLAB script **vorticity_1.m**. However, when the vorticity pattern is localized in space, the space scales of the streamfunction and vorticity are much different. This latter situation is illustrated in the MATLAB script **vorticity_demo.m**, which shows the streamfunction corresponding to a point source of vorticity at $(x, y) = (0,0)$. For this problem you must modify the code in **vorticity_1.m** by specifying $\zeta(x, y) = \exp\left[-b\left(x^2 + y^2\right)\right]$ where b is a constant. Run the model for several values of b from $b = 1 \times 10^{-4}$ km^{-2} to 4×10^{-7} km^{-2}. Show in a table or line plot the dependence of the ratio of the horizontal scales on which vorticity and the streamfunction decay to one-half of their maximum values as a function of the parameter b. Note that for geostrophic motions (with constant Coriolis parameter), the streamfunction defined here is proportional to geopotential height. What can you conclude from this exercise about the information content of a 500-hPa height map versus that of a map of the 500-hPa vorticity field?

M4.2. The MATLAB scripts **geowinds_1.m** (to be used when the mapping toolbox is available) and **geowinds_2.m** (to be used if no mapping toolbox is available) contain contour plots showing the observed 500-hPa height and horizontal wind fields for November 10, 1998, in the North American sector. Also shown is a color plot of the magnitude of the 500-hPa wind with height contours superposed. Using centered difference formulas (see Section 13.2.1), compute the geostrophic wind components, the magnitude of the geostrophic wind, the relative vorticity, the vorticity of the geostrophic wind, and the vorticity minus the vorticity of the geostrophic wind. Following the models of Figures 4.1 and 4.2, superpose these fields on the maps of the 500-hPa height field. Explain the distribution and the sign of regions with large differences between vorticity and geostrophic vorticity in terms of the force balances that were studied in Chapter 3.

---

[1] Note that for a scalar, $D_a/Dt = D/Dt$ (i.e., the rate of change following the motion does not depend on the reference system). For a vector, however, this is not the case, as was shown in Section 2.1.1.

[2] Terms involving w are negligible by comparison for synoptic-scale weather systems.

[3] Recall that n is a coordinate in the horizontal plane perpendicular to the local flow direction with positive values to the left of an observer facing downstream.

Chapter 5

Atmospheric Oscillations

Linear Perturbation Theory

If one is interested in producing an accurate forecast of the circulation at some future time, a detailed numerical model based on the primitive equations and including processes such as latent heating, radiative transfer, and boundary layer drag should produce the best results. However, the inherent complexity of such a model generally precludes any simple interpretation of the physical processes that produce the predicted circulation. If we wish to gain physical insight into the fundamental nature of atmospheric motions, it is helpful to employ simplified models in which certain processes are omitted and compare the results with those of more complete models. It is difficult to gain an appreciation for the processes that produce the wave-like character observed in many meteorological disturbances through the study of numerical integrations alone. Therefore, analytical treatment of idealized representations of the atmosphere are valuable and are the main topic of this chapter.

First, we discuss the *perturbation method*, a simple technique that is useful for qualitative analysis of atmospheric waves. We then use this method to examine several types of waves in the atmosphere. In Chapters 6 and 7, the perturbation method is used to develop the quasi-geostrophic equations and to study the development of *synoptic-wave* disturbances, respectively.

5.1 The Perturbation Method

In the perturbation method, all field variables are divided into two parts: a *basic state* portion, which is usually assumed to be independent of time and longitude, and a perturbation portion, which is the local deviation of the field from the basic state. Thus, for example, if $\overline{u}$ designates a time and longitude-averaged zonal velocity, and u' is the deviation from that average, then the complete zonal velocity field is $u(x,t)=\overline{u}+u'(x,t)$. In that case, for example, the inertial acceleration $u\partial u/\partial x$ can be written

$$u\frac{\partial u}{\partial x}=\left(\overline{u}+u'\right)\frac{\partial}{\partial x}\left(\overline{u}+u'\right)=\overline{u}\frac{\partial u'}{\partial x}+u'\frac{\partial u'}{\partial x}$$

The basic assumptions of perturbation theory are that the basic state variables must themselves satisfy the governing equations when the perturbations are set to zero, and the perturbation fields must be small enough so that all terms in the governing equations that involve products of the perturbations can be neglected. Terms that depend linearly on perturbation fields drop by half when the perturbation amplitude is cut in half, but terms that are quadratic in the perturbation field become smaller by a factor of four. Therefore, the nonlinear terms become smaller faster than the linear terms and can be neglected in the small-amplitude limit. If $|u'/\overline{u}|\ll 1$, then

$$\left|\overline{u}\partial u'/\partial x\right|\gg\left|u'\partial u'/\partial x\right|$$

If terms that are products of the perturbation variables are neglected, the nonlinear governing equations are reduced to linear differential equations in the perturbation variables in which the basic state variables are specified coefficients. These equations can then be solved by standard methods to determine the character and structure of the perturbations in terms of the known basic state. For equations with constant coefficients, the solutions are sinusoidal or exponential in character. Solution of perturbation equations then determines such characteristics as the propagation speed, vertical structure, and conditions for growth or decay of the waves. The perturbation technique is especially useful in studying the stability of a given basic state flow with respect to small superposed perturbations. This application is the subject of Chapter 7.

5.2 Properties of Waves

Wave motions are oscillations in field variables (e.g., velocity and pressure) that propagate in space and time. In this chapter we are concerned with linear sinusoidal wave motions. Many of the mechanical properties of such waves are also features of a familiar system, the linear harmonic oscillator. An important property of the harmonic oscillator is that the period, or time required to execute a single oscillation, is independent of the amplitude of the oscillation. For most natural vibratory systems, this condition holds only for oscillations of sufficiently small amplitude. The classical example of such a system is the simple pendulum (Figure 5.1) consisting of a mass M suspended by a massless string of length l, free to perform small oscillations about the equilibrium position $\theta = 0$. The component of the gravity force parallel to the direction of motion is $-Mg \sin \theta$. Thus, the equation of motion for the mass M is

$$Ml\frac{d^2\theta}{dt^2} = -Mg\sin\theta$$

Now for small displacements, $\sin\theta \approx \theta$ so that the governing equation becomes

$$\frac{d^2\theta}{dt^2} + \nu^2\theta = 0 \tag{5.1}$$

where $\nu^2 \equiv g/l$. The harmonic oscillator equation (5.1) has the general solution

$$\theta = \theta_1 \cos\nu t + \theta_2 \sin\nu t = \theta_0 \cos(\nu t - \alpha)$$

where θ_1, θ_2, θ_0, and α are constants determined by the initial conditions (see Problem 5.1) and ν is the frequency of oscillation. The complete solution can thus be expressed in terms of an amplitude θ_0 and a phase $\phi(t) = \nu t - \alpha$. The phase varies linearly in time by a factor of 2π radians per wave period.

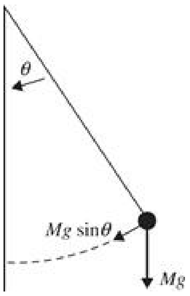


Figure 5.1 A simple pendulum.

Propagating waves can also be characterized by their amplitudes and phases. In a propagating wave, however, phase depends not only on time but on one or more space variables as well. Thus, for a one-dimensional wave propagating in the x direction, $\phi(x,t) = kx - \nu t - \alpha$. Here the *wave number*, k, is defined as 2π divided by the wavelength. For propagating waves the phase is constant for an observer moving at the *phase speed* $c \equiv \nu/k$. This may be verified by observing that if phase is to remain constant following the motion,

$$\frac{D\phi}{Dt} = \frac{D}{Dt}(kx - \nu t - \alpha) = k\frac{Dx}{Dt} - \nu = 0$$

Thus, $Dx/Dt = c = \nu/k$ for phase to be constant. For $\nu > 0$ and $k > 0$ we have $c > 0$. In that case, if $\alpha = 0$, $\phi = k(x - ct)$, so x must increase with increasing t for ϕ to remain constant. Phase then propagates in the positive direction, as illustrated for a sinusoidal wave in Figure 5.2.

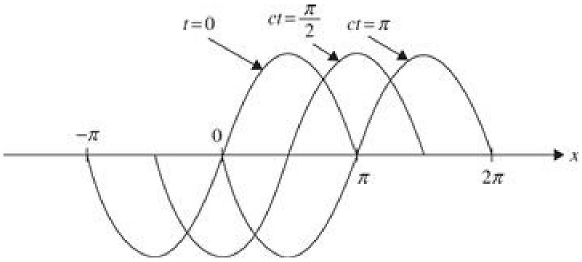


Figure 5.2 A sinusoidal wave traveling in the positive x direction at speed c. (Wave number is assumed to be unity.)

5.2.1 Fourier Series

The representation of a perturbation as a simple sinusoidal wave might seem an oversimplification, since disturbances in the atmosphere are never purely sinusoidal. It can be shown, however, that any reasonably well-behaved function of longitude can be represented in terms of a zonal mean plus a *Fourier* series of sinusoidal components:

$$f(x) = \sum_{s=1}^{\infty} (A_s \sin k_s x + B_s \cos k_s x) \tag{5.2}$$

where $k_s = 2\pi s/L$ is the *zonal* wave number (units m^{-1}), L is the distance around a latitude circle, and s, the *planetary* wave number, is an integer designating the number of waves around a latitude circle. The coefficients A_s are calculated by multiplying both sides of (5.2) by $\sin(2\pi nx/L)$, where n is an integer, and integrating around a latitude circle. Applying the orthogonality relationships

$$\int_0^L \sin\frac{2\pi sx}{L}\sin\frac{2\pi nx}{L}dx = \left\{\begin{array}{ll} 0, & s \neq n \\ L/2, & s = n \end{array}\right\}$$

we obtain

$$A_s = \frac{2}{L}\int_0^L f(x)\sin\frac{2\pi sx}{L}dx$$

In a similar fashion, multiplying both sides in (5.2) by $(2\pi nx/L)$ and integrating gives

$$B_s = \frac{2}{L}\int_0^L f(x)\cos\frac{2\pi sx}{L}dx$$

A_s and B_s are called the *Fourier coefficients*, and

$$f_s(x) = A_s \sin k_s x + B_s \cos k_s x \tag{5.3}$$

is called the *s*th *Fourier component* or *s*th harmonic of the function $f(x)$. If the Fourier coefficients are computed for, say, the longitudinal dependence of the (observed) geopotential perturbation, the largest-amplitude Fourier components will be those for which s is close to the observed number of troughs or ridges around a latitude circle. When only qualitative information is desired, it is usually sufficient to limit the analysis to a single typical Fourier component and assume that the behavior of the actual field will be similar to that of the component. The expression for a Fourier component may be written more compactly using complex exponential notation. According to the Euler formula,

$$\exp(i\phi) = \cos\phi + i\sin\phi$$

where $i \equiv (-1)^{1/2}$ is the imaginary unit. Thus, we can write

$$\begin{aligned} f_s(x) &= \mathrm{Re}[C_s \exp(ik_s x)] \\ &= \mathrm{Re}[C_s \cos k_s x + iC_s \sin k_s x] \end{aligned} \tag{5.4}$$

where Re[ ] denotes “real part of” and C_s is a complex coefficient. Comparing (5.3) and (5.4), we see that the two representations of $f_s(x)$ are identical, provided that

$$B_s = \mathrm{Re}[C_s] \quad \text{and} \quad A_s = -\mathrm{Im}[C_s]$$

where Im[ ] stands for “imaginary part of." This exponential notation will generally be used for applications of the perturbation theory that follows and also in Chapter 7.

5.2.2 Dispersion and Group Velocity

A fundamental property of linear oscillators is that the frequency of oscillation ν depends only on the physical characteristics of the oscillator, not on the motion itself. For propagating waves, however, ν generally depends on the wave number of the perturbation as well as the physical properties of the medium. Thus, because $c = \nu/k$, the phase speed also depends on the wave number except in the special case where $\nu \propto k$. For waves in which the phase speed varies with k, the various sinusoidal components of a disturbance originating at a given location are at a later time found in different places—that is, they are dispersed. Such waves are referred to as *dispersive*, and the formula that relates ν and k is called a *dispersion relationship*. Some types of waves, such as acoustic waves, have phase speeds that are independent of the wave number. In such *nondispersive* waves a spatially localized disturbance consisting of a number of Fourier wave components (a *wave group*) will preserve its shape as it propagates in space at the phase speed of the wave.

For dispersive waves, however, the shape of a wave group will not remain constant as the group propagates. Because the individual Fourier components of a wave group may either reinforce or cancel each other, depending on the relative phases of the components, the energy of the group will be concentrated in limited regions as illustrated in Figure 5.3. Furthermore, the group generally broadens in the course of time—that is, the energy is *dispersed*.

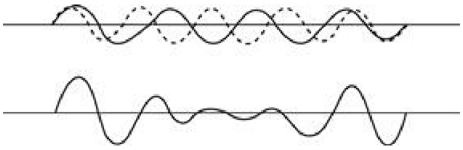


Figure 5.3 Wave groups formed from two sinusoidal components of slightly different wavelengths. For nondispersive waves, the pattern in the lower part of the diagram propagates without change of shape. For dispersive waves, the shape of the pattern changes in time.

When waves are dispersive, the speed of the wave group is generally different from the average phase speed of the individual Fourier components. Hence, as shown in Figure 5.4, individual wave components may move either more rapidly or more slowly than the wave group as the group propagates along. Surface waves in deep water (such as a ship wake) are characterized by dispersion in which individual wave crests move twice as fast as the wave group. In synoptic-scale atmospheric disturbances, however, the group velocity exceeds the phase velocity. The resulting downstream development of new disturbances will be discussed later.

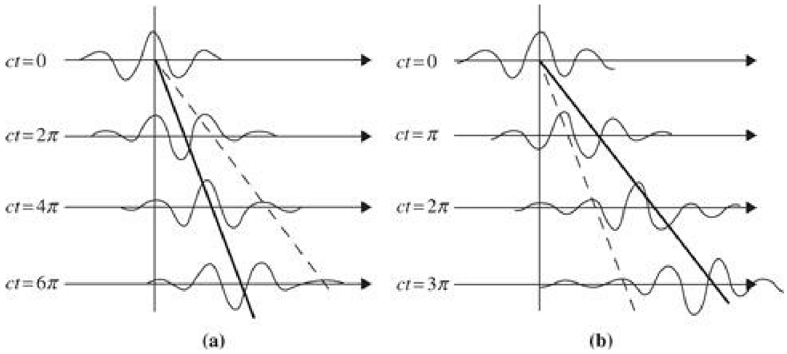


Figure 5.4 Schematic showing propagation of wave groups: (a) group velocity less than phase speed and (b) group velocity greater than phase speed. *Heavy lines* show group velocity, and *light lines* show phase speed.

An expression for *group velocity*, which is the velocity at which the observable disturbance (and thus the energy) propagates, can be derived as follows: We consider the superposition of two horizontally propagating waves of equal amplitude but slightly different wavelengths, with wave numbers and frequencies differing by $2\delta k$ and $2\delta\nu$, respectively. The total disturbance is thus

$$\Psi(x,t) = \exp\{i[(k+\delta k)x - (\nu+\delta\nu)t]\} + \exp\{i[(k-\delta k)x - (\nu-\delta\nu)t]\}$$

where for brevity the Re[ ] notation in (5.4) is omitted and it is understood that only the real part of the right side has physical meaning. Rearranging terms and applying the Euler formula gives

$$\begin{aligned}\Psi &= \left[e^{i(\delta kx - \delta\nu t)} + e^{-i(\delta kx - \delta\nu t)}\right]e^{i(kx-\nu t)} \\ &= 2\cos(\delta kx - \delta\nu t)e^{i(kx-\nu t)}\end{aligned} \tag{5.5}$$

The disturbance (5.5) is the product of a high-frequency *carrier wave* of wavelength $2\pi/k$ whose phase speed, ν/k, is the average for the two Fourier components, and a low-frequency *envelope* of wavelength $2\pi/\delta k$ that travels at the speed $\delta\nu/\delta k$. Thus, in the limit as $\delta k \to 0$, the horizontal velocity of the envelope, or *group velocity*, is just

$$c_{gx} = \partial\nu/\partial k$$

The wave energy thus propagates at the group velocity. This result applies generally to arbitrary wave envelopes provided that the wavelength of the wave group, $2\pi/\delta k$, is large compared to the wavelength of the dominant component, $2\pi/k$.

5.2.3 Wave Properties in Two and Three Dimensions

Developing a complete understanding of wave properties in two and three dimensions is sufficiently important to the remainder of this chapter and aspects of later chapters that we devote this section to a vector-oriented description of wave properties. We discuss two dimensions for ease of presentation, but the notation is sufficiently general that properties for three dimensions should be clear.

A two-dimensional plane wave in a scalar field, f, may be expressed as

$$f(x,y,t) = \mathrm{Re}\left\{A\,e^{i(kx+ly-\nu t)}\right\} = \mathrm{Re}\left\{A\,e^{i\phi}\right\} \tag{5.6}$$

Independent variables (x,y) and t represent space and time, respectively. k and l are the x and y wave numbers (m^{-1}), and ν is the frequency (s^{-1}). The wave is determined by *amplitude*, A (units of f), and *phase angle*, ϕ. Note that ϕ is a linear function of the independent variables, and it proves useful to consider the space and time properties of ϕ separately.

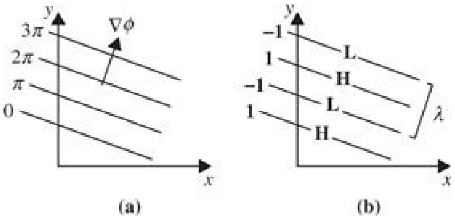
At any instant in time, $\phi = kx + ly + C$, C constant, so ϕ is uniform along lines of constant $kx + ly$. This implies $d\phi = \frac{\partial\phi}{\partial x}\delta x + \frac{\partial\phi}{\partial y}\delta y = 0$ for ϕ constant. Therefore, the slope of these lines is $\frac{\delta y}{\delta x}\big|_\phi = -k/l$. Constant values of

$$e^{i\phi} = e^{i(\phi+2\pi n)} \tag{5.7}$$

where n is an integer, define lines of constant *phase*, such as lows and highs (Figure 5.5). The *wave vector* is defined by

$$\mathbf{K} = \nabla\phi \tag{5.8}$$

$\mathcal{K} = |\mathbf{K}|$ is the *total wave number*, and therefore $\lambda = \dfrac{2\pi}{\mathcal{K}}$ is the *wavelength*—that is, the distance between lines of constant phase.



(a) (b)

Figure 5.5 Two-dimensional plane wave at a fixed time: (a) phase angle, ϕ, and (b) phase, $e^{i\phi}$. Wavelength is denoted λ. Note that if ν > 0, the wave travels in the direction of the wave vector, $\nabla\phi$.

At any fixed point in space, $\phi = C - \nu t$, *C* constant, so ϕ is a linear function of time. Therefore, frequency is defined as

$$\nu = -\frac{\partial\phi}{\partial t} \tag{5.9}$$

which is the rate at which lines of constant phase pass a fixed point in space. The wave *period* is $\frac{2\pi}{|\nu|}$, which is the length of time between points of constant phase (units: seconds). The wave *phase speed* (m s^{-1}) is determined by how fast lines of constant phase move along the wave vector

$$c = \frac{\nu}{\mathcal{K}} = -\frac{1}{|\nabla\phi|}\frac{\partial\phi}{\partial t} \tag{5.10}$$

This definition represents a generalization of phase speed from the one-dimensional example given in the previous section. In particular, for two and three dimensions, the phase speed is *not* the same as $c_x = \frac{\nu}{k}$ and $c_y = \frac{\nu}{l}$, which define the rate at which phase lines travel along the *x* and *y* coordinate axes, respectively. Furthermore, *c*, c_x, and c_y do not satisfy rules of vector addition, so $c^2 \neq c_x^2 + c_y^2$.

Note that, in general, ϕ and *A* may be complex numbers. If ϕ has an imaginary part, $\phi = \phi_r + i\phi_i$, then $e^{i\phi} = e^{i(\phi_r + i\phi_i)} = e^{i\phi_r}e^{-\phi_i} \equiv A^* e^{i\phi_r}$. ϕ_r is the wave phase angle as interpreted in the preceding, and $A^* = Ae^{-\phi_i}$ is a modified amplitude that depends on time and/or space. For example, if the frequency, ν, contributes the imaginary part, then the wave has *time-dependent* amplitude that grows or decays with time. Such waves are referred to as *unstable* to distinguish them from the *neutral* waves with fixed *A*. In Chapter 7 we seek waves that grow in time, and therefore will study conditions that contribute to imaginary frequency.

We return now to group velocity, which provides the speed and direction in which an observer must travel in order to follow waves of a certain wavelength and frequency. Rather than a pure plane wave, which exists over all space, we take the phase angle, ϕ in (5.6), to be a slowly varying function of space and time due to variable ν and **K**. Equations (5.8) and (5.9) define the wave vector and frequency respectively, even when these quantities vary in space and time. Together, (5.8) and (5.9) imply

$$\frac{\partial \mathbf{K}}{\partial t} + \nabla \nu = 0 \tag{5.11}$$

Frequency is defined by a dispersion relationship, so that ν is a function of **K**. As such, $\nabla \nu$ may be evaluated using the chain rule

$$\nabla \nu = (\nabla_k \nu \cdot \nabla)\mathbf{K} \tag{5.12}$$

where $\nabla_k \nu = \left(\frac{\partial \nu}{\partial k}; \frac{\partial \nu}{\partial l}\right)$ is the group velocity, $\mathbf{C}_g$. Therefore, (5.11) can be written as

$$\frac{\partial \mathbf{K}}{\partial t} + (\mathbf{C}_g \cdot \nabla)\mathbf{K} = 0 \tag{5.13}$$

Consequently, in a frame of reference moving with the group velocity, the wave vector is conserved; that is, we follow a group of waves with fixed wavelength and frequency.

5.2.4 A Wave Solution Strategy

Approximations to the full equations governing atmospheric dynamics will be solved for wave motions many times. Even though aspects of each individual case are different, a guide to the general approach to solving these problems is as follows:

1. **Choose a basic state** The basic state is a simplified representation of the atmosphere. Typically we wish to eliminate all complications except those essential to the motions of interest. For example, in this chapter we will often make use of basic states that are at rest and in hydrostatic balance. The basic state is like a musical instrument, and we wish to know what tones naturally emerge.
2. **Linearize the governing equations** Given a basic state, the governing equations can be linearized using the perturbation method of Section 5.1
3. **Assume wave solutions of the form in equation (5.6)** If the coefficients of the linearized equations do not depend on an independent variable, then the functions that depend on the independent variable may be represented as waves, $e^{i\phi}$, as in equation (5.6), provided that the independent variable is periodic. If, for example, the domain of interest is not periodic in the y direction, or the basic state varies in y, then we cannot assume wave solutions in that direction. Instead, a general functional form must be assumed in that direction, such as $F(y)e^{i(kx-\nu t)}$, which typically leads to an ordinary differential equation that must be solved for the structure function $F(y)$.
4. **Solve for the dispersion and polarization relationships** Substituting the wave form into the linearized governing equations typically results in a simpler closed set of algebraic equations or, in cases having nonconstant coefficients, with a set of ordinary differential equations. Normally the wave number is taken as the independent variable, and the frequency is determined as a function of the wave number; this gives the dispersion relationship. The wave amplitude is unconstrained, but the relationships between variables, the polarization relations, must be determined. Variables are *in phase* if they share the same form and sign (e.g., sin(x)) and *out of phase* if they have the same form and opposite sign (e.g., sin(x) and −sin(x)). If the variables are 90° out of phase, they are *in quadrature* (e.g., sin(x) and cos(x)).

5.3 Simple Wave Types

Waves in fluids result from the action of restoring forces on fluid parcels that have been displaced from their equilibrium positions. The restoring forces may be due to compressibility, gravity, rotation, or electromagnetic effects. This section considers the two simplest examples of linear waves in fluids: acoustic waves and shallow water gravity waves.

5.3.1 Acoustic or Sound Waves

Sound waves, or acoustic waves, are *longitudinal waves*. That is, they are waves in which the particle oscillations are parallel to the direction of propagation. Sound is propagated by the alternating adiabatic compression and expansion of the medium. As an example, Figure 5.6 shows a schematic section along a tube that has a diaphragm at its left end. If the diaphragm is set into vibration, the air adjacent to it will be alternately compressed and expanded as the diaphragm moves inward and outward. The resulting oscillating pressure gradient force will be balanced by an oscillating acceleration of the air in the adjoining region, which will cause an oscillating pressure oscillation further into the tube, and so on. The result of this continual adiabatic increase and decrease of pressure through alternating compression and rarefaction is, as shown in Figure 5.6, a sinusoidal pattern of pressure and velocity perturbations that propagates to the right down the tube. Individual air parcels do not, however, have a net rightward motion; they only oscillate back and forth, while the pressure pattern moves rightward at the speed of sound.

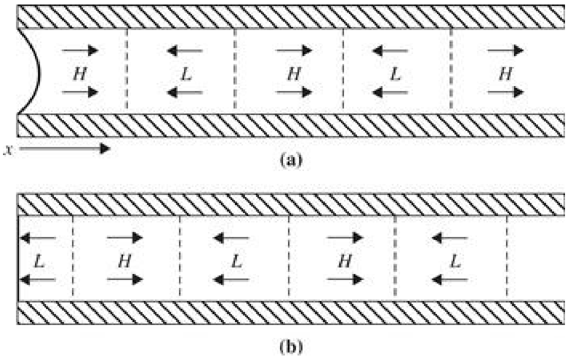


Figure 5.6 Schematic diagram illustrating the propagation of a sound wave in a tube with a flexible diaphragm at the left end. Labels *H* and *L* designate centers of high- and low-perturbation pressure. *Arrows* show velocity perturbations. The situation one-quarter period (b) later than in (a) for propagation in the positive *x* direction.

To introduce the perturbation method, we consider the problem illustrated by Figure 5.6—that is, one-dimensional sound waves propagating in a straight pipe parallel to the x axis. To exclude the possibility of *transverse* oscillations (i.e., oscillations in which the particle motion is at right angles to the direction of phase propagation), we assume at the outset that $v = w = 0$. In addition, we eliminate all dependence on y and z by assuming that $u = u(x, t)$. With these restrictions, the momentum equation, continuity equation, and thermodynamic energy equation for adiabatic motion are, respectively,

$$\frac{Du}{Dt}+\frac{1}{\rho}\frac{\partial p}{\partial x}=0 \tag{5.14}$$

$$\frac{D\rho}{Dt}+\rho\frac{\partial u}{\partial x}=0 \tag{5.15}$$

$$\frac{D\ln\theta}{Dt}=0 \tag{5.16}$$

where for this case $D/Dt = \partial/\partial t + u\partial/\partial x$. Recalling from (2.44) and the ideal gas law that potential temperature may be expressed as

$$\theta=(p/\rho R)(p_s/p)^{R/c_p}$$

where $p_s = 1000$ hPa, we may eliminate θ in (5.16) to give

$$\frac{1}{\gamma}\frac{D\ln p}{Dt}-\frac{D\ln\rho}{Dt}=0 \tag{5.17}$$

where $\gamma = c_p/c_v$. Eliminating ρ between (5.15) and (5.17) gives

$$\frac{1}{\gamma}\frac{D\ln p}{Dt}+\frac{\partial u}{\partial x}=0 \tag{5.18}$$

The dependent variables are now divided into constant basic state portions (denoted by overbars) and perturbation portions (denoted by primes):

$$\begin{aligned} u(x,t)&=\overline{u}+u'(x,t)\\ p(x,t)&=\overline{p}+p'(x,t)\\ \rho(x,t)&=\overline{\rho}+\rho'(x,t) \end{aligned} \tag{5.19}$$

Substituting (5.19) into (5.14) and (5.18) we obtain

$$\begin{aligned} &\frac{\partial}{\partial t}(\overline{u}+u')+(\overline{u}+u')\frac{\partial}{\partial x}(\overline{u}+u')+\frac{1}{(\overline{\rho}+\rho')}\frac{\partial}{\partial x}(\overline{p}+p')=0\\ &\frac{\partial}{\partial t}(\overline{p}+p')+(\overline{u}+u')\frac{\partial}{\partial x}(\overline{p}+p')+\gamma(\overline{p}+p')\frac{\partial}{\partial x}(\overline{u}+u')=0 \end{aligned}$$

We next observe that provided $|\rho'/\overline{\rho}|\ll 1$, we can use the binomial expansion to approximate the density term as

$$\frac{1}{(\overline{\rho}+\rho')}=\frac{1}{\overline{\rho}}\left(1+\frac{\rho'}{\overline{\rho}}\right)^{-1}\approx\frac{1}{\overline{\rho}}\left(1-\frac{\rho'}{\overline{\rho}}\right)$$

Neglecting products of the perturbation quantities and noting that the basic state fields are constants, we obtain the linear perturbation equations[1]

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)u'+\frac{1}{\overline{\rho}}\frac{\partial p'}{\partial x}=0 \tag{5.20}$$

$$\left(\frac{\partial}{\partial t}+\bar{u}\frac{\partial}{\partial x}\right)p'+\gamma\bar{p}\frac{\partial u'}{\partial x}=0 \tag{5.21}$$

Eliminating u' by operating on (5.21) with $(\partial/\partial t+\bar{u}\partial/\partial x)$ and substituting from (5.20), we get[2]

$$\left(\frac{\partial}{\partial t}+\bar{u}\frac{\partial}{\partial x}\right)^2 p'-\frac{\gamma\bar{p}}{\bar{\rho}}\frac{\partial^2 p'}{\partial x^2}=0 \tag{5.22}$$

which is a form of the standard *wave equation* familiar from electromagnetic theory. A simple solution representing a plane sinusoidal wave propagating in x is

$$p'=A\exp\left[ik(x-ct)\right] \tag{5.23}$$

where for brevity we omit the Re{ } notation, but it is to be understood that only the real part of (5.23) has physical significance. Substituting the assumed solution (5.23) into (5.22), we find that the phase speed c must satisfy

$$(-ikc+ik\bar{u})^2-(\gamma\bar{p}/\bar{\rho})(ik)^2=0$$

where we have canceled out the factor $A\exp\left[ik(x-ct)\right]$, which is common to both terms. Solving for c gives

$$c=\bar{u}\pm(\gamma\bar{p}/\bar{\rho})^{1/2}=\bar{u}\pm\left(\gamma R\bar{T}\right)^{1/2} \tag{5.24}$$

Therefore, (5.23) is a solution of (5.22), provided that the phase speed satisfies (5.24). According to (5.24), the speed of wave propagation relative to the zonal current is $c-\bar{u}=\pm c_s$, where $c_s\equiv\left(\gamma R\bar{T}\right)^{1/2}$ is called the *adiabatic speed of sound.*

The mean zonal velocity here plays only a role of *Doppler shifting* the sound wave so that the frequency relative to the ground corresponding to a given wave number k is

$$\nu=kc=k(\bar{u}\pm c_s)$$

Thus, in the presence of a wind, the frequency as heard by a fixed observer depends on the location of the observer relative to the source. If $\bar{u}>0$, the frequency of a stationary source will appear to be higher for an observer to the east (downstream) of the source $(c=\bar{u}+c_s)$ than for an observer to the west (upstream) of the source $c=\bar{u}-c_s$.

5.3.2 Shallow Water Waves

The second example of pure wave motion concerns the horizontally propagating oscillations known as shallow water waves. This analysis provides a clear introduction to the primary wave motions in the extratropical latitudes: Rossby waves and gravity waves.

We linearize the shallow water equations (derived in Section 4.5) for a resting basic state having water depth $\bar{h}$ and fixed ambient rotation given by a constant Coriolis parameter f. The linearized momentum (4.33) and mass continuity (4.37) equations are

$$\begin{aligned}\frac{\partial u'}{\partial t} &= -g\frac{\partial h'}{\partial x}+fv' \\ \frac{\partial v'}{\partial t} &= -g\frac{\partial h'}{\partial y}-fu' \\ \frac{\partial h'}{\partial t} &= -\bar{h}\left(\frac{\partial u'}{\partial x}+\frac{\partial v'}{\partial y}\right)\end{aligned} \tag{5.25}$$

Primes denote perturbation values—that is, departures from the state of rest. (5.25) represent a set of three coupled first-order partial differential equations in the unknowns, u', v', h' . One approach to solving these equations is to form and solve a single third-order partial differential equation. This is accomplished by taking $\partial/\partial t$ of the third equation in (5.25), which gives

$$\frac{\partial^2 h'}{\partial t^2} = -\bar{h}\left(\frac{\partial^2 u'}{\partial t \partial x}+\frac{\partial^2 v'}{\partial t \partial y}\right) \tag{5.26}$$

The terms on the right side of (5.26) may be replaced using the first two equations of (5.25):

$$\frac{\partial^2 h'}{\partial t^2} = \bar{h}\left(g\nabla_h^2 h' - f\zeta'\right) \tag{5.27}$$

Again take $\partial/\partial t$, which gives the third-order equation

$$\frac{\partial^3 h'}{\partial t^3} + \left(f^2 - g\bar{h}\nabla_h^2\right)\frac{\partial h'}{\partial t} = 0 \tag{5.28}$$

where $\partial\zeta'/\partial t$ is replaced using the linearized version of (4.38) (noting that f is taken constant here):

$$\frac{\partial \zeta'}{\partial t} = -f\left(\frac{\partial u'}{\partial x}+\frac{\partial v'}{\partial y}\right) \tag{5.29}$$

Assuming that the lateral boundaries are periodic and, since the coefficients f and $g\bar{h}$ are constant, we may assume wave solutions of the form

$$h' = \mathrm{Re}\left\{Ae^{i(kx+ly-\nu t)}\right\} \tag{5.30}$$

Using (5.30) in (5.28) gives a cubic polynomial for the frequency

$$\nu^3 - \nu\left[f^2 + g\bar{h}\left(k^2+l^2\right)\right] = 0 \tag{5.31}$$

This is the dispersion relationship for shallow water waves. Clearly $\nu = 0$ is a solution, and if $\nu \neq 0$, then

$$\nu^2 = f^2 + g\bar{h}\left(k^2+l^2\right) \tag{5.32}$$

For readers familiar with linear algebra, we note an alternative solution method. Taking solutions of the form (5.30) for h', u', and v', and substituting directly into (5.25), converts the set of partial differential equations to algebraic equations that may be written as

(5.33)

$$\mathbf{Ax} = 0$$

where **x** is the column vector of the unknowns $[u'v'h']^{\mathrm{T}}$, and

(5.34)

$$\mathbf{A} = \begin{bmatrix} -i\nu & -f & ikg \\ f & -i\nu & ilg \\ \overline{h}k & -\overline{h}l & -\nu \end{bmatrix}$$

A nontrivial solution to (5.33) is obtained only if **A** is not invertible. This is enforced by setting the determinant of **A** to zero, which gives (5.31).

Turning now to the structure of the waves, consider first the case where $\nu = 0$. Appealing to (5.25), we see that since the waves are stationary, the left sides are zero. Clearly, these waves depend on rotation, since when $f = 0$, the wave amplitude, A, must also be zero. For non zero f, (5.25) implies

(5.35)

$$v' = \frac{g}{f}\frac{\partial h}{\partial x}$$

$$u' = -\frac{g}{f}\frac{\partial h}{\partial y}$$

The stationary waves are in a state of geostrophic balance and are an example of **Rossby waves** in the limit of constant Coriolis parameter. The case of propagating Rossby waves, with nonconstant Coriolis parameter, will be explored in Section 5.7.

Solutions for $\nu \neq 0$ are called **inertia–gravity waves**, since particle oscillations depend on both gravitational and inertial forces. The inertial aspect will be discussed in Section 5.7 so that we may focus here on the gravitational aspect. In the limiting case $f = 0$, we have simply **gravity waves**, and from (5.32) the wave speed becomes

(5.36)

$$c = \sqrt{g\overline{h}}$$

Since the waves all move at the same speed, they are nondispersive. For a fluid $\overline{h} = 4$ km deep, the shallow water gravity wave speed is approximately 200 m s^{-1}. Thus, long waves on the ocean surface travel very rapidly. It should be emphasized again that this theory applies only to waves of wavelength much greater than $\overline{h}$. Such long waves may be produced by very large-scale disturbances such as earthquakes.[3]

In order to further expose wave structure and the physics of motion, we rotate coordinates such that the y direction is oriented along the phase lines (Figure 5.7). Since this yields the same interpretation as $l = 0$, we omit the prime notation on x and y. In this case, the general expressions for flow normal to and along phase lines, u' and v', respectively, are

(5.37)

$$u' = \frac{\nu k g}{\nu^2 - f^2} h'$$

$$v' = \frac{-ikgf}{\nu^2 - f^2} h'$$

From (4.34) and the third equation in (5.25),

$$\frac{\partial w'}{\partial z} = -\frac{\partial u'}{\partial x} = \frac{1}{\bar{h}}\frac{\partial h'}{\partial t} \tag{5.38}$$

Integrating in z gives the w' field

$$w'(z) = \frac{1}{\bar{h}}\frac{\partial h'}{\partial t}\int_0^z dz' = \frac{-ikcz}{\bar{h}}h' \tag{5.39}$$

Taking the real part of w' gives

$$\mathrm{Re}\{w'\} = -\frac{kcz}{\bar{h}}\mathrm{Re}\{ih'\} = \frac{kczA}{\bar{h}}\sin(\phi) \tag{5.40}$$

This shows that w' is in *quadrature* with h', so that w' will lead (lag) h' by 90° if the waves move to the right (left).

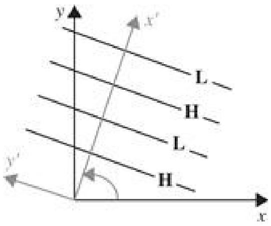


Figure 5.7 Coordinate rotation to align the y axis with wave fronts so that $l = 0$.

Setting $\nu = c = 0$ in (5.37) and (5.39) recovers the geostrophic flow of stationary Rossby waves: $u' = 0$ (no flow normal to phase lines), $v' = -ikg/f$, and $w = 0$. For the inertia gravity waves ($\nu \neq 0$), notice that the flow along phase lines is zero in the absence of ambient rotation. In that case, the physics of wave propagation is particularly clear. From (5.32) we see that $\nu^2 - f^2$ is strictly positive, and in the rotated coordinates, $u' = c/\bar{h}$. This shows that the u' field is either *in phase*, or *out of phase*, with the h' field depending on the sign of c. Waves that move to the right ($c > 0$) are shown in Figure 5.8. Convergence and divergence of the u field raise and lower the height of the fluid, respectively, moving the wave to the right in Figure 5.8. As expected from mass conservation, the w field exhibits rising motion near convergence, which increases linearly with height from zero at the bottom boundary. Changing the sign of c has no effect on h', but reverses the u and w fields so that the wave moves to the left.

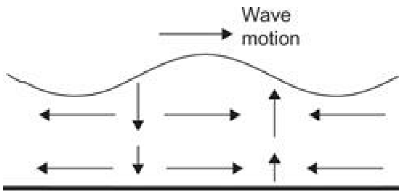


Figure 5.8 Depiction of shallow water gravity wave height and velocity fields for the case where the coordinates have been rotated as shown in Figure 5.7.

Finally, consider the vorticity and potential vorticity (PV) of the waves in the rotated coordinates. Since there is no variation in the *y* direction, the relative vorticity is simply

(5.41)

$$\zeta' = \frac{\partial v'}{\partial x} = \frac{k^2 g f}{\nu^2 - f^2} h'$$

A linearized version of the shallow water potential vorticity equation (4.39) is obtained from (5.29) and the third equation in (5.25), which yields

(5.42)

$$\frac{\partial}{\partial t}\left[\zeta' - \frac{f}{\bar{h}} h'\right] = 0$$

Linearization has removed all advection terms from the PV conservation equation, so that (5.42) states that the linearized PV

(5.43)

$$Q' = \zeta' - \frac{f}{\bar{h}} h'$$

is conserved *locally*, at every point in space. This is an extremely powerful constraint on the linear dynamics, which we will exploit in Section 5.6 when solving initial-vale problems for (5.25).

For shallow water Rossby waves, vorticity in the rotated coordinates becomes

(5.44)

$$\zeta'_{RW} = -\frac{k^2 g}{f} h'$$

which is out of phase with the height field and larger for shorter waves. Rossby wave PV is given by ζ reduced by $\frac{f}{\bar{h}} h'$, which has the same sign as ζ except for unrealistically long waves.

For shallow water inertia–gravity waves, vorticity in the rotated coordinates becomes

(5.45)

$$\zeta'_{GW} = \frac{f}{\bar{h}} h'$$

which is in phase with the height field, and nonzero provided there exists ambient rotation. From (5.45) and (5.43), it is clear that inertia–gravity wave PV is identically zero.

A summary of the essential differences between Rossby and inertia–gravity waves for rotating shallow water follows:

- Rossby waves are stationary for *f* constant, whereas inertia–gravity waves move very fast.
- Inertia–gravity waves have small relative vorticity but are strongly divergent, whereas Rossby waves have geostrophic winds with zero divergence and large relative vorticity.
- Inertia–gravity waves have zero potential vorticity, whereas Rossby waves generally have nonzero potential vorticity.

These properties generalize to waves in a stratified atmosphere, considered in the next two sections. They provide a useful framework for eliminating inertia–gravity waves from the full nonlinear equations, yielding the quasi-geostrophic equations, which we use to understand the dynamics of extratropical weather systems in Chapters 6 and 7.

5.4 Internal Gravity (Buoyancy) Waves

The nature of gravity wave propagation in the stratified atmosphere is considered now in the absence of ambient rotation. Atmospheric gravity waves can only exist when the atmosphere is stably stratified so that a fluid parcel displaced vertically will undergo buoyancy oscillations (see Section 2.7.3). Because the buoyancy force is the restoring force responsible for gravity waves, the term *buoyancy wave* is actually more appropriate as a name for these waves. However, in this text we will generally use the traditional name *gravity wave*.

In a fluid, such as the ocean, which is bounded both above and below, gravity waves propagate primarily in the horizontal plane, since vertically traveling waves are reflected from the boundaries to form standing waves. However, in a fluid that has no upper boundary, such as the atmosphere, gravity waves may propagate vertically as well as horizontally. In vertically propagating waves the phase is a function of height. Such waves are referred to as *internal* waves. Although internal gravity waves are not generally of great importance for synoptic-scale weather forecasting (and indeed are nonexistent in the filtered quasi-geostrophic models), they can be important in mesoscale motions. For example, they are responsible for the occurrence of mountain *lee waves*. They also are believed to be an important mechanism for transporting energy and momentum into the middle atmosphere, and are often associated with the formation of clear air turbulence (CAT).

5.4.1 Pure Internal Gravity Waves

For simplicity we neglect the Coriolis force and limit our discussion to two-dimensional internal gravity waves propagating in the x, z plane. An expression for the frequency of such waves can be obtained by modifying the parcel theory developed in Section 2.7.3.

Internal gravity waves are transverse waves in which the parcel oscillations are parallel to the phase lines as indicated in Figure 5.9. A parcel displaced a distance δs along a line tilted at an angle α to the vertical as shown in Figure 5.8 will undergo a vertical displacement $\delta z = \delta s \cos\alpha$. For such a parcel the *vertical* buoyancy force per unit mass is just $-N^2\delta z$, as was shown in (2.52). Thus, the component of the buoyancy force parallel to the tilted path along which the parcel oscillates is just

$$-N^2\delta z\cos\alpha = -N^2(\delta s\cos\alpha)\cos\alpha = -(N\cos\alpha)^2\delta s$$

The momentum equation for the parcel oscillation is then

$$\frac{d^2(\delta s)}{dt^2} = -(N\cos\alpha)^2\delta s \tag{5.46}$$

which has the general solution $\delta s = \exp[\pm i(N\cos\alpha)t]$. Thus, the parcels execute a simple harmonic oscillation at the frequency $\nu = N\cos\alpha$. This frequency depends only on the static stability (measured by the buoyancy frequency N) and the angle of the phase lines to the vertical.

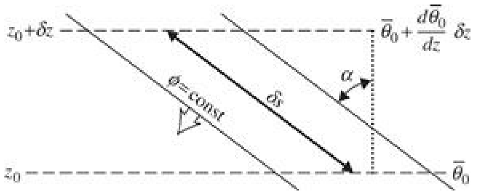


Figure 5.9 Parcel oscillation path (*heavy arrow*) for pure gravity waves with phase lines tilted at an angle α to the vertical.

The preceding heuristic derivation can be verified by considering the linearized equations for two-dimensional internal gravity waves. For simplicity, we employ the Boussinesq approximation of Section 2.8, in which density is treated as a constant except where it is coupled with gravity in the buoyancy term of the vertical momentum equation. Thus, in this approximation the atmosphere is considered to be incompressible, and local density variations are assumed to be small perturbations of the constant basic state density field. Because the vertical variation of the basic state density is neglected except where coupled with gravity, the Boussinesq approximation is only valid for motions in which the vertical scale is less than the atmospheric scale height $H(\approx 8\,\mathrm{km})$.

Neglecting the effects of rotation, the basic equations for two-dimensional motion of an incompressible atmosphere may be written as

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+w\frac{\partial u}{\partial z}+\frac{1}{\rho}\frac{\partial p}{\partial x}=0 \tag{5.47}$$

$$\frac{\partial w}{\partial t}+u\frac{\partial w}{\partial x}+w\frac{\partial w}{\partial z}+\frac{1}{\rho}\frac{\partial p}{\partial z}+g=0 \tag{5.48}$$

$$\frac{\partial u}{\partial x}+\frac{\partial w}{\partial z}=0 \tag{5.49}$$

$$\frac{\partial \theta}{\partial t}+u\frac{\partial \theta}{\partial x}+w\frac{\partial \theta}{\partial z}=0 \tag{5.50}$$

where the potential temperature θ is related to pressure and density by

$$\theta=\frac{p}{\rho R}\left(\frac{p_s}{p}\right)^{\kappa}$$

which, after taking logarithms on both sides, yields

$$\ln\theta=\gamma^{-1}\ln p-\ln\rho+\text{constant} \tag{5.51}$$

We now linearize (5.47) through (5.51) by letting

$$\begin{array}{ll} \rho=\rho_0+\rho' & u=\bar{u}+u' \\ p=\bar{p}(z)+p' & w=w' \\ \theta=\bar{\theta}(z)+\theta' & \end{array} \tag{5.52}$$

where the basic state zonal flow $\bar{u}$ and the density ρ_0 are both assumed to be constant. The basic state pressure field must satisfy the hydrostatic equation

$$d\bar{p}/dz=-\rho_0 g \tag{5.53}$$

while the basic state potential temperature must satisfy (5.51) so that

$$\ln\bar{\theta}=\gamma^{-1}\ln\bar{p}-\ln\rho_0+\text{constant} \tag{5.54}$$

The linearized equations are obtained by substituting from (5.52) into (5.47) through (5.51) and neglecting all terms that are products of the perturbation variables. Thus, for example, the last two terms in (5.48) are approximated as

$$\begin{aligned}\frac{1}{\rho}\frac{\partial p}{\partial z}+g&=\frac{1}{\rho_0+\rho'}\left(\frac{d\overline{p}}{dz}+\frac{\partial p'}{\partial z}\right)+g\\ &\approx\frac{1}{\rho_0}\frac{d\overline{p}}{dz}\left(1-\frac{\rho'}{\rho_0}\right)+\frac{1}{\rho_0}\frac{\partial p'}{\partial z}+g=\frac{1}{\rho_0}\frac{\partial p'}{\partial z}+\frac{\rho'}{\rho_0}g\end{aligned}\tag{5.55}$$

where (5.53) has been used to eliminate $\overline{p}$. The perturbation form of (5.51) is obtained by noting that

$$\begin{aligned}\ln\left[\overline{\theta}\left(1+\frac{\theta'}{\overline{\theta}}\right)\right]&=\gamma^{-1}\ln\left[\overline{p}\left(1+\frac{p'}{\overline{p}}\right)\right]\\ &\quad-\ln\left[\rho_0\left(1+\frac{\rho'}{\rho_0}\right)\right]+\text{const.}\end{aligned}\tag{5.56}$$

Now, recalling that $\ln(ab)=\ln(a)+\ln(b)$ and that $\ln(1+\varepsilon)\approx\varepsilon$ for any $\varepsilon \ll 1$, we find with the aid of (5.54) that (5.56) may be approximated by

$$\frac{\theta'}{\overline{\theta}}\approx\frac{1}{\gamma}\frac{p'}{\overline{p}}-\frac{\rho'}{\rho_0}$$

Solving for ρ' yields

$$\rho'\approx-\rho_0\frac{\theta'}{\overline{\theta}}+\frac{p'}{c_s^2}\tag{5.57}$$

where $c_s^2\equiv\overline{p}\gamma/\rho_0$ is the square of the speed of sound. For buoyancy wave motions $\left|\rho_0\theta'/\overline{\theta}\right|\gg\left|p'/c_s^2\right|$; that is, density fluctuations due to pressure changes are small compared with those due to temperature changes. Therefore, to a first approximation,

$$\theta'/\overline{\theta}=-\rho'/\rho_0\tag{5.58}$$

Using (5.55) and (5.58), the linearized version of the equation set (5.47) through (5.50), we can write as

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)u'+\frac{1}{\rho_0}\frac{\partial p'}{\partial x}=0\tag{5.59}$$

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)w'+\frac{1}{\rho_0}\frac{\partial p'}{\partial z}-\frac{\theta'}{\overline{\theta}}g=0\tag{5.60}$$

$$\frac{\partial u'}{\partial x}+\frac{\partial w'}{\partial z}=0\tag{5.61}$$

$$\left(\frac{\partial}{\partial t}+\bar{u}\frac{\partial}{\partial x}\right)\theta'+w'\frac{d\bar{\theta}}{dz}=0 \tag{5.62}$$

Subtracting ∂(5.59)/∂z from ∂(5.60)/∂x, we can eliminate p' to obtain

$$\left(\frac{\partial}{\partial t}+\bar{u}\frac{\partial}{\partial x}\right)\left(\frac{\partial w'}{\partial x}-\frac{\partial u'}{\partial z}\right)-\frac{g}{\bar{\theta}}\frac{\partial\theta'}{\partial x}=0 \tag{5.63}$$

which is just the y component of the vorticity equation. With the aid of (5.61) and (5.62), u' and θ' can be eliminated from (5.63) to yield a single equation for w':

$$\left(\frac{\partial}{\partial t}+\bar{u}\frac{\partial}{\partial x}\right)^2\left(\frac{\partial^2 w'}{\partial x^2}+\frac{\partial^2 w'}{\partial z^2}\right)+N^2\frac{\partial^2 w'}{\partial x^2}=0 \tag{5.64}$$

where $N^2 \equiv g\, d\ln\bar{\theta}/dz$ is the square of the buoyancy frequency, which is assumed to be constant.[4]

Equation (5.64) has harmonic wave solutions of the form

$$w'=\mathrm{Re}\left[\hat{w}\exp(i\phi)\right]=w_r\cos\phi-w_i\sin\phi \tag{5.65}$$

where $\hat{w}=w_r+iw_i$ is a complex amplitude with real part w_r and imaginary part w_i, and $\phi=kx+mz-\nu t$ is the phase, which is assumed to depend linearly on z as well as on x and t. Here the horizontal wave number k is real because the solution is always sinusoidal in x. The vertical wave number $m=m_r+im_i$ may, however, be complex, in which case m_r describes sinusoidal variation in z and m_i describes exponential decay or growth in z, depending on whether m_i is positive or negative. When m is real, the total wave number may be regarded as a vector $\boldsymbol{\kappa}\equiv(k,m)$, directed perpendicular to lines of constant phase, and in the direction of phase increase, whose components, $k = 2\pi/L_x$ and $m = 2\pi/L_z$, are inversely proportional to the horizontal and vertical wavelengths, respectively. Substitution of the assumed solution into (5.64) yields the dispersion relationship

$$(\nu-\bar{u}k)^2\left(k^2+m^2\right)-N^2k^2=0$$

so that

$$\hat{\nu}\equiv\nu-\bar{u}k=\pm Nk\Big/\left(k^2+m^2\right)^{1/2}=\pm Nk/|\boldsymbol{\kappa}| \tag{5.66}$$

where $\hat{\nu}$, the *intrinsic frequency*, is the frequency relative to the mean wind. Here, the plus sign is to be taken for eastward phase propagation and the minus sign for westward phase propagation, relative to the mean wind.

If we let $k > 0$ and $m < 0$, then lines of constant phase tilt eastward with increasing height as shown in Figure 5.10 (i.e., for $\phi=kx+mz$ to remain constant as x increases, z must also increase when $k > 0$ and $m < 0$). The choice of the positive root in (5.66) then corresponds to eastward and downward phase propagation relative to the mean flow, with horizontal and vertical phase speeds (relative to the mean flow) given by $c_x=\hat{\nu}/k$ and $c_z=\hat{\nu}/m$, respectively.[5] The components of the group velocity, c_{gx} and c_{gz}, however, are given by

$$c_{gx}=\frac{\partial \nu}{\partial k}=\bar{u}\pm\frac{Nm^2}{(k^2+m^2)^{3/2}} \tag{5.67}$$

$$c_{gz}=\frac{\partial \nu}{\partial m}=\pm\frac{(-Nkm)}{(k^2+m^2)^{3/2}} \tag{5.68}$$

where the upper or lower signs are chosen in the same way as in (5.66). Thus, the vertical component of group velocity has a sign opposite to that of the vertical phase speed relative to the mean flow (downward phase propagation implies upward energy propagation). Furthermore, it is easily shown from (5.45) that the group velocity vector is parallel to lines of constant phase. Internal gravity waves thus have the remarkable property that group velocity is perpendicular to the direction of phase propagation. Because energy propagates at the group velocity, this implies that energy propagates parallel to the wave crests and troughs, rather than perpendicular to them as in acoustic waves or shallow water gravity waves. In the atmosphere, internal gravity waves generated in the troposphere by cumulus convection, by flow over topography, and by other processes may propagate upward many scale heights into the middle atmosphere, even though individual fluid parcel oscillations may be confined to vertical distances much less than a kilometer.

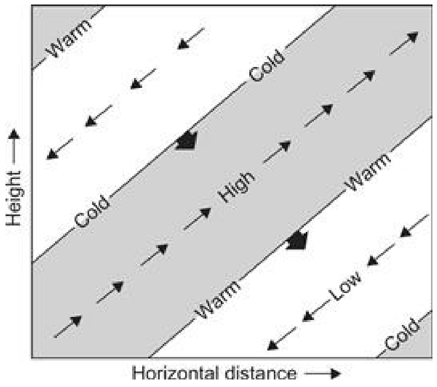


Figure 5.10 Idealized cross-section showing phases of pressure, temperature, and velocity perturbations for an internal gravity wave. *Thin arrows* indicate the perturbation velocity field; *blunt solid arrows*, the phase velocity. *Shading* shows regions of upward motion.

Referring again to Figure 5.10, it is evident that the angle of the phase lines to the local vertical is given by

$$\cos\alpha=L_z/(L_x^2+L_z^2)^{1/2}=\pm k/(k^2+m^2)^{1/2}=\pm k/|\mathbf{K}|$$

Thus, $\hat{\nu}=\pm N\cos\alpha$ (i.e., gravity wave frequencies must be less than the buoyancy frequency) in agreement with the heuristic parcel oscillation model (5.46). The tilt of phase lines for internal gravity waves depends only on the ratio of the intrinsic wave frequency to the buoyancy frequency and is independent of wavelength.

5.5 Linear Waves of A Rotating Stratified Atmosphere

Rotation is now considered by extending the analysis of linear waves in a stratified atmosphere. As in the shallow water analysis of Section 5.3.2, rotation permits the existence of stationary Rossby waves. Gravity waves with horizontal scales greater than a few hundred kilometers and periods greater than a few hours are hydrostatic, but they are influenced by the Coriolis effect and are characterized by parcel oscillations that are elliptical rather than straight lines as in the pure gravity wave case. This *elliptical polarization* can be understood qualitatively by observing that the Coriolis effect resists horizontal parcel displacements in a rotating fluid, but in a manner somewhat different from that in which the buoyancy force resists vertical parcel displacements in a statically stable atmosphere. In the latter case the resistive force is opposite to the direction of parcel displacement, whereas in the former it is at right angles to the horizontal parcel velocity.

5.5.1 Pure Inertial Oscillations

Section 3.2.3 showed that a parcel put into horizontal motion in a resting atmosphere with a constant Coriolis parameter executes a circular trajectory in an anticyclonic sense. A generalization of this type of inertial motion to the case with a geostrophic mean zonal flow can be derived using a parcel argument similar to that used for the buoyancy oscillation in Section 2.7.3.

If the basic state flow is assumed to be a zonally directed geostrophic wind u_g, and it is assumed that the parcel displacement does not perturb the pressure field, the approximate equations of motion become

$$\frac{Du}{Dt} = fv = f\frac{Dy}{Dt} \tag{5.69}$$

$$\frac{Dv}{Dt} = f\left(u_g - u\right) \tag{5.70}$$

We consider a parcel that is moving with the geostrophic basic state motion at a position $y = y_0$. If the parcel is displaced across stream by a distance δy, we can obtain its new zonal velocity from the integrated form of (5.49):

$$u\left(y_0 + \delta y\right) = u_g\left(y_0\right) + f\delta y \tag{5.71}$$

The geostrophic wind at $y_0 + \delta y$ can be approximated as

$$u_g\left(y_0 + \delta y\right) = u_g\left(y_0\right) + \frac{\partial u_g}{\partial y}\delta y \tag{5.72}$$

Using (5.71) and (5.72) to evaluate (5.50) at $y_0 + \delta y$ yields

$$\frac{Dv}{Dt} = \frac{D^2\delta y}{Dt^2} = -f\left(f - \frac{\partial u_g}{\partial y}\right)\delta y = -f\frac{\partial M}{\partial y}\delta y \tag{5.73}$$

where we have defined the *absolute momentum*, $M \equiv fy - u_g$.

This equation is mathematically of the same form as (2.52), the equation for the motion of a vertically displaced particle in a stratified atmosphere. Depending on the sign of the coefficient on the right side in (5.73), the parcel will either be forced to return to its original position or will accelerate further from that position. This coefficient thus determines the condition for *inertial* *instability*:

$$f\frac{\partial M}{\partial y}=f\left(f-\frac{\partial u_g}{\partial y}\right)\begin{cases}>0 & \text{stable}\\ =0 & \text{neutral}\\ <0 & \text{unstable}\end{cases} \tag{5.74}$$

Viewed in an inertial reference frame, instability results from an imbalance between the pressure gradient and inertial forces for a parcel displaced radially in an axisymmetric vortex. In the Northern Hemisphere, where f is positive, the flow is inertially stable provided that the absolute vorticity of the basic flow, $\partial M/\partial y$, is positive. In the Southern Hemisphere, however, inertial stability requires that the absolute vorticity be negative. Observations show that for extratropical synoptic-scale systems the flow is always inertially stable, although near neutrality often occurs on the anticyclonic shear side of upper-level jet streaks. The occurrence of inertial instability over a large area would immediately trigger inertially unstable motions, which would mix the fluid laterally just as convection mixes it vertically, and reduce the shear until the absolute vorticity times f was again positive. (This explains why anticyclonic shears cannot become arbitrarily large.)

5.5.2 Rossby and Inertia–Gravity Waves

When the flow is both inertially and gravitationally stable, parcel displacements are resisted by both rotation and buoyancy. The resulting oscillations are called *inertia–gravity waves*. The dispersion relation for such waves can be analyzed using a variant of the parcel method applied in Section 5.4 We consider parcel oscillations along a slantwise path in the (y, z) plane as shown in Figure 5.11. For a vertical displacement δz, the buoyancy force component parallel to the slope of the parcel oscillation is $-N^2\delta z\cos\alpha$, and for a meridional displacement δy the Coriolis (inertial) force component parallel to the slope of the parcel path is $-f^2\,\delta y\sin\alpha$, where we have assumed that the geostrophic basic flow is constant in latitude. Thus, the harmonic oscillator equation for the parcel (5.46) is modified to the form

$$\frac{D^2\delta s}{Dt^2}=-(f\sin\alpha)^2\,\delta s-(N\cos\alpha)^2\,\delta s \tag{5.75}$$

where δs is again the perturbation parcel displacement.

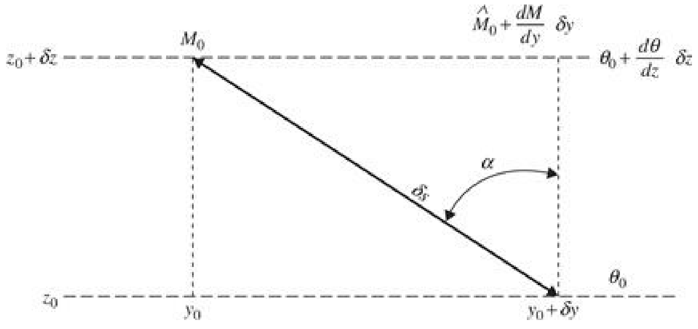


Figure 5.11 Parcel oscillation path in the meridional plane for an inertia–gravity wave. (See text for definition of symbols.)

The frequency now satisfies the dispersion relationship

$$\nu^2=N^2\cos^2\alpha+f^2\sin^2\alpha \tag{5.76}$$

Since in general $N^2 > f^2$, (5.76) indicates that inertia–gravity wave frequencies must lie in the range $f \le |\nu| \le N$. The frequency approaches N as the trajectory slope approaches the vertical, and approaches f as the trajectory slope approaches the horizontal. For typical midlatitude tropospheric conditions, inertia–gravity wave periods are in the approximate range of 12 min to 15 h. Rotational effects become important, however, only when the second term on the right in (5.76) is similar in magnitude to the first term. This requires that $\tan^2 \alpha \sim N^2/f^2 = 10^4$, in which case it is clear from (5.76) that $\nu \ll N$. Thus, only low-frequency gravity waves are modified significantly by Earth's rotation, and these have very small parcel trajectory slopes.

The heuristic parcel derivation can again be verified using the linearized dynamical equations. In this case, however, it is necessary to include rotation. The small parcel trajectory slopes of the relatively long period waves that are altered significantly by rotation imply that the horizontal scales are much greater than the vertical scales for these waves. Therefore, we may assume that the motions are in hydrostatic balance. If in addition we assume a motionless basic state, the linearized equations (5.59) through (5.62) are replaced by the set

$$\frac{\partial u'}{\partial t} - f v' + \frac{1}{\rho_0}\frac{\partial p'}{\partial x} = 0 \tag{5.77}$$

$$\frac{\partial v'}{\partial t} + f u' + \frac{1}{\rho_0}\frac{\partial p'}{\partial y} = 0 \tag{5.78}$$

$$\frac{1}{\rho_0}\frac{\partial p'}{\partial z} - \frac{\theta'}{\overline{\theta}} g = 0 \tag{5.79}$$

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z} = 0 \tag{5.80}$$

$$\frac{\partial \theta'}{\partial t} + w' \frac{d\overline{\theta}}{dz} = 0 \tag{5.81}$$

The hydrostatic relationship in (5.79) may be used to eliminate θ' in (5.81) to yield

$$\frac{\partial}{\partial t}\left(\frac{1}{\rho_0}\frac{\partial p'}{\partial z}\right) + N^2 w' = 0 \tag{5.82}$$

Letting

$$\begin{aligned}
u' &= \mathrm{Re}\left[\hat{u} \exp i(kx + ly + mz - \nu t)\right] \\
v' &= \mathrm{Re}\left[\hat{v} \exp i(kx + ly + mz - \nu t)\right] \\
w' &= \mathrm{Re}\left[\hat{w} \exp i(kx + ly + mz - \nu t)\right] \\
p'/\rho_0 &= \mathrm{Re}\left[\hat{p} \exp i(kx + ly + mz - \nu t)\right]
\end{aligned}$$

and substituting into (5.77), (5.78), and (5.82), we obtain

$$\hat{u} = \left(\nu^2 - f^2\right)^{-1}(\nu k + i l f)\,\hat{p} \tag{5.83}$$

$$\hat{v} = \left(\nu^2 - f^2\right)^{-1}(\nu l - i k f)\,\hat{p} \tag{5.84}$$

$$\hat{w} = -\left(\nu m/N^2\right)\hat{p} \tag{5.85}$$

which with the aid of (5.80) yields the dispersion relation

$$m^2\nu^3 - \left[N^2(k^2+l^2)+f^2m^2\right]\nu = 0 \tag{5.86}$$

As in the shallow water dispersion relationship (5.31), there exists a zero root corresponding to stationary Rossby waves. Time independence implies geostrophic flow from (5.77) and (5.78), and from (5.81) that $w' = 0$.

For $\nu \neq 0$, we have the classic dispersion relationship for hydrostatic inertia–gravity waves,

$$\nu^2 = f^2 + N^2\left(k^2+l^2\right)m^{-2} \tag{5.87}$$

Because hydrostatic waves must have $(k^2 + l^2)/m^2 \ll 1$, (5.87) indicates that for vertical propagation to be possible (m real) the frequency must satisfy the inequality $|f| < |\nu| \ll N$. Equation (5.87) is just the limit of (5.76) when we let

$$\sin^2\alpha \to 1,\ \cos^2\alpha = \left(k^2+l^2\right)/m^2$$

which is consistent with the hydrostatic approximation. Note also that (5.87) may be approximated exactly by the shallow water dispersion relationship (5.32) if one uses the "equivalent depth" of $\overline{h} = \frac{N^2}{gm^2}$ for the water depth.

If axes are chosen to make $l = 0$ (Figure 5.7), it may be shown (see Problem 5.14) that the ratio of the vertical to the horizontal components of group velocity is given by

$$\left|c_{gz}/c_{gx}\right| = |k/m| = \left(\nu^2 - f^2\right)^{1/2}/N \tag{5.88}$$

Thus, for fixed ν, inertia–gravity waves propagate more closely to the horizontal than pure internal gravity waves. However, as in the latter case the group velocity vector is again parallel to lines of constant phase.

Eliminating $\hat{p}$ between (5.83) and (5.84) for the case $l = 0$ yields the relationship $\hat{v} = -if\hat{u}/\nu$, from which it is easily verified that if $\hat{u}$ is real, the perturbation horizontal motions satisfy the relations

$$u' = \hat{u}\cos\left(kx+mz-\nu t\right),\quad v' = \hat{u}\left(f/\nu\right)\sin\left(kx+mz-\nu t\right) \tag{5.89}$$

so that the horizontal velocity vector rotates anticyclonically (that is, clockwise in the Northern Hemisphere) with time. As a result, parcels follow elliptical trajectories in a plane orthogonal to the wave number vector. Equations (5.89) also show that the horizontal velocity vector turns anticyclonically with height for waves with upward energy propagation (e.g., waves with $m < 0$ and $\nu < 0$). These characteristics are illustrated by the vertical cross-section shown in Figure 5.12. The anticyclonic turning of the horizontal wind with height and time is a primary method for identifying inertia–gravity oscillations in meteorological data.

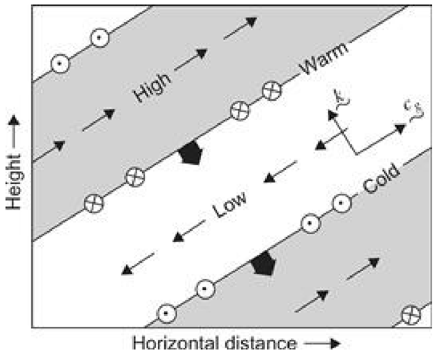


Figure 5.12 Vertical section in a plane containing the wave vector ***k*** showing the phase relationships among velocity, geopotential, and temperature fluctuations in an upward propagating inertia–gravity wave with $m < 0$, $\nu > 0$, and $f > 0$ (Northern Hemisphere). *Thin sloping lines* denote the surfaces of constant phase (perpendicular to the wave vector), and *thick arrows* show the direction of phase propagation. *Thin arrows* show the perturbation zonal and vertical velocity fields. Meridional wind perturbations are shown by *arrows* pointed into the page (northward) and out of the page (southward). Note that the perturbation wind vector turns clockwise (anticyclonically) with height.

After Andrews et al., 1987.

A linearized vorticity equation is derived from $\partial/\partial x$ (5.78) $-\partial/\partial y$ (5.77), which gives

$$\frac{\partial \zeta'}{\partial t} = -f\left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}\right) = f\frac{\partial w'}{\partial z} \tag{5.90}$$

From (5.81) $\partial w'/\partial z$ may be eliminated from (5.90) to give the linearized potential vorticity equation

$$\frac{\partial \Pi'}{\partial t} = 0 \tag{5.91}$$

This states that the linearized PV

$$\Pi' = \zeta' + f\frac{\partial}{\partial z}\left(\frac{\theta'}{d\bar{\theta}/dz}\right) \tag{5.92}$$

is conserved at every point in space, as was the case for linear shallow water waves. Polarization relationships (5.83) and (5.84) for the $l = 0$ case can be used to show that for Rossby waves

$$\zeta' = -\frac{k^2}{f}\hat{p} \tag{5.93}$$

and for inertia–gravity waves

$$\zeta' = \frac{fm^2}{N^2}\hat{p} \tag{5.94}$$

This result may be used to show (see Problem 5.16) that if the buoyancy frequency, N, is constant, inertia-gravity waves have zero linearized potential vorticity (5.92). Therefore, as in linearized shallow water, the linearized potential vorticity equation (5.91) effectively filters gravity waves from the dynamics. We will exploit this fact in the next section, where we show that geostrophic balance emerges in the long-time limit for arbitrary initial states. This result also motivates the derivation of the quasi-geostrophic equations in Chapter 6, which essentially generalizes (5.91) to include nonlinear advection of linearized PV (5.92).

5.6 Adjustment to Geostrophic Balance

This section investigates the process by which geostrophic balance is achieved from an unbalanced initial condition—that is, the *adjustment* process. For simplicity we utilize the linearized shallow water equations of Section 5.3.2; similar considerations apply to a continuously stratified atmosphere. The problem addressed here is the solution of the full linearized equations (5.25) in the limit of large time, starting from an arbitrary initial condition.

The key to solving this problem comes from (5.42), the linearized shallow water potential vorticity equation,

$$\frac{\partial Q'}{\partial t} = 0 \tag{5.95}$$

where Q' designates the perturbation potential vorticity. This implies the conservation relationship

$$Q'(x,y,t) = \zeta'/f - h'/\bar{h} = \text{Const.} \tag{5.96}$$

Thus, if we know the distribution of Q' at the initial time, we know Q' for all time

$$Q'(x,y,t) = Q'(x,y,0)$$

and the final adjusted state can be determined without solving the time-dependent problem. Note that, in general, u', v', and h' evolve in time, but only in such a way that Q' is constant at every point in space.

This problem was first solved by Rossby in the 1930s and is often referred to as the Rossby adjustment problem. As a simplified, albeit somewhat unrealistic, example of the adjustment process, we consider an idealized shallow water system on a rotating plane with initial conditions

$$u', v' = 0; \quad h' = -h_0 \operatorname{sgn}(x) \tag{5.97}$$

where $\operatorname{sgn}(x) = 1$ for $x > 0$ and $\operatorname{sgn}(x) = -1$ for $x < 0$. This corresponds to an initial step function in h' at $x = 0$, with the fluid motionless. Thus, from (5.96)

$$(\zeta'/f) - (h'/\bar{h}) = (h_0/\bar{h}) \operatorname{sgn}(x) \tag{5.98}$$

Using (5.98) to eliminate ζ' in (5.27) yields

$$\frac{\partial^2 h'}{\partial t^2} - c^2\left(\frac{\partial^2 h'}{\partial x^2} + \frac{\partial^2 h'}{\partial y^2}\right) + f^2 h' = -f^2 h_0 \operatorname{sgn}(x) \tag{5.99}$$

where $c^2 = g\bar{h}$.

Because initially h' is independent of y, it will remain so for all time. Thus, in the final steady state (5.99) becomes

$$-c^2\frac{d^2h'}{dx^2}+f^2h' = -f^2h_0\,\mathrm{sgn}(x) \tag{5.100}$$

which has the solution

$$\frac{h'}{h_0} = \left\{\begin{array}{ll} -1+\exp(-x/\lambda_R) & \text{for } x>0 \\ +1-\exp(+x/\lambda_R) & \text{for } x<0 \end{array}\right\} \tag{5.101}$$

where $\lambda_R \equiv f_0^{-1}\sqrt{g\bar{h}}$ is the Rossby *radius of deformation*. Hence, the radius of deformation may be interpreted as the horizontal length scale over which the height field adjusts during the approach to geostrophic equilibrium. For $|x| \gg \lambda_R$ the original h' remains unchanged. Substituting from (5.101) into (5.25) shows that the steady velocity field is geostrophic and nondivergent:

$$u' = 0, \quad \text{and} \quad v' = \frac{g}{f}\frac{\partial h'}{\partial x} = -\frac{gh_0}{f\lambda_R}\exp(-|x|/\lambda_R) \tag{5.102}$$

The steady-state solution (5.102) is shown in Figure 5.13.

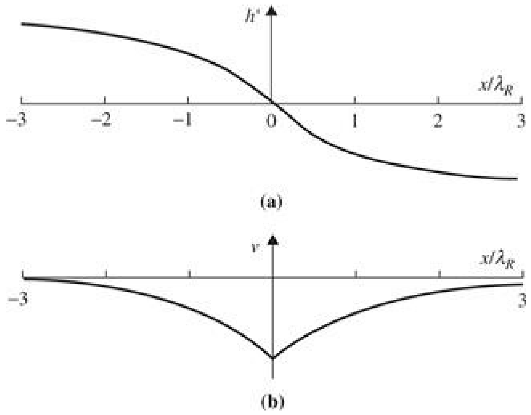


Figure 5.13 The geostrophic equilibrium solution corresponding to adjustment from the initial state defined in (5.97). (a) Final surface elevation profiles and (b) the geostrophic velocity profile in the final state.

After Gill, 1982.

Note that (5.102) could not be derived merely by setting $\partial/\partial t = 0$ in (5.25). That would yield geostrophic balance, and *any* distribution of h' would satisfy the equations:

$$fu' = -g\frac{\partial h'}{\partial y}, \quad fv' = g\frac{\partial h'}{\partial x}, \quad \frac{\partial u'}{\partial x}+\frac{\partial v'}{\partial y} = 0$$

Only by combining (5.25) to obtain the potential vorticity equation, and requiring the flow to satisfy potential vorticity conservation at all intermediate times, can the degeneracy of the geostrophic final state be eliminated. In other words, although any height field can satisfy the steady-state versions of (5.25), there is only one field that is consistent with a given initial state; this field can be found readily because it can be computed from the distribution of potential vorticity, which is conserved.

Although the final state can be computed without solving the time-dependent equation, if the evolution of the adjustment process is required, it is necessary to solve (5.99) subject to the initial conditions (5.97), which is beyond the scope of this discussion. We can, however, compute the amount of energy that is dispersed by gravity waves during the adjustment process. This only requires computing the energy change between initial and final states.

The potential energy per unit horizontal area is given by

$$\int_0^{h'} \rho g z dz = \rho g h'^2/2$$

Thus, the potential energy released per unit length in y during adjustment is

$$\begin{aligned} &\int_{-\infty}^{+\infty} \frac{\rho g h_0^2}{2} dx - \int_{-\infty}^{+\infty} \frac{\rho g h'^2}{2} dx \\ &= 2 \int_0^{+\infty} \frac{\rho g h_0^2}{2} \left[1 - \left(1 - e^{-x/\lambda_R} \right)^2 \right] dx \\ &= \frac{3}{2} \rho g h_0^2 \lambda_R \end{aligned} \tag{5.103}$$

In the nonrotating case $(\lambda_R \to \infty)$ all potential energy available initially is released (converted to kinetic energy) so that there is an infinite energy release. (Energy is radiated away in the form of gravity waves, leaving a flat free surface extending to $|x| \to \infty$ as $t \to \infty$.)

In the rotating case only the finite amount given in (5.103) is converted to kinetic energy, and only a portion of this kinetic energy is radiated away. The rest remains in the steady geostrophic circulation. The kinetic energy in the steady-state per unit length is

$$2 \int_0^{+\infty} \rho \bar{h} \frac{v'^2}{2} dx = \rho \bar{h} \left(\frac{g h_0}{f \lambda_R} \right)^2 \int_0^{+\infty} e^{-2x/\lambda_R} dx = \frac{1}{2} \rho g h_0^2 \lambda_R \tag{5.104}$$

Thus, in the rotating case a finite amount of potential energy is released, but only one-third of the potential energy released goes into the steady geostrophic flow. The remaining two-thirds is radiated away in the form of inertia–gravity waves. This simple analysis illustrates the following points

- It is difficult to extract the potential energy of a rotating fluid. Although there is an infinite reservoir of potential energy in this example (because h' is finite as $|x| \to \infty$), only a finite amount is converted before geostrophic balance is achieved.
- Conservation of potential vorticity allows one to determine the steady-state geostrophically adjusted velocity and height fields without carrying out a time integration.
- The length scale for the steady solution is the Rossby radius λ_R.

5.7 Rossby Waves

The wave type that is of most importance for large-scale meteorological processes is the *Rossby wave*, or *planetary wave*. In an inviscid barotropic fluid of constant depth (where the divergence of the horizontal velocity must vanish), the Rossby wave is an absolute vorticity-conserving motion that owes its existence to the variation of the Coriolis parameter with latitude. More generally, in a baroclinic atmosphere, the Rossby wave is a potential vorticity-conserving motion that owes its existence to the gradient of potential vorticity.

The variation of the Coriolis parameter with latitude can be approximated by expanding the latitudinal dependence of f in a Taylor series about a reference latitude ϕ_0 and retaining only the first two terms to yield

$$f = f_0 + \beta y \tag{5.105}$$

where $\beta \equiv (df/dy)_{\phi_0} = 2\Omega \cos \phi_0/a \text{ and } y = 0 \text{ at } \phi_0$. This approximation is usually referred to as the *midlatitude β-plane* approximation. Rossby wave propagation can be understood in a qualitative fashion by considering a closed chain of fluid parcels initially aligned along a circle of latitude. Recall that the absolute vorticity η is given by $\eta = \zeta + f$, where ζ is the relative vorticity and f is the Coriolis parameter. Assume that $\zeta = 0$ at time t_0. Now suppose that at t_1, δy is the meridional displacement of a fluid parcel from the original latitude. Then at t_1 we have

$$(\zeta + f)_{t_1} = f_{t_0}$$

or

$$\zeta_{t_1} = f_{t_0} - f_{t_1} = -\beta \delta y \tag{5.106}$$

where $\beta \equiv df/dy$ is the planetary vorticity gradient at the original latitude.

From (5.106) it is evident that if the chain of parcels is subject to a sinusoidal meridional displacement under absolute vorticity conservation, the resulting perturbation vorticity will be positive for a southward displacement and negative for a northward displacement.

This perturbation vorticity field will induce a meridional velocity field, which advects the chain of fluid parcels southward west of the vorticity maximum and northward west of the vorticity minimum, as indicated in Figure 5.14. Thus, the fluid parcels oscillate back and forth about their equilibrium latitude, and the pattern of vorticity maxima and minima propagates to the west. This westward propagating vorticity field constitutes a Rossby wave. Just as a positive vertical gradient of potential temperature resists vertical fluid displacements and provides the restoring force for gravity waves, the meridional gradient of absolute vorticity resists meridional displacements and provides the restoring mechanism for Rossby waves.

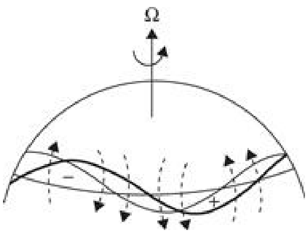


Figure 5.14 Perturbation vorticity field and induced velocity field (*dashed arrows*) for a meridionally displaced chain of fluid parcels. The *heavy wavy line* shows original perturbation position; the *light line* shows westward displacement of the pattern due to advection by the induced velocity.

The speed of westward propagation, c, can be computed for this simple example by letting $\delta y = a\sin[k(x-ct)]$, where a is the maximum northward displacement. Then $v = D(\delta y)/Dt = -kca\cos[k(x-ct)]$, and

$$\zeta = \partial v/\partial x = k^2 ca\sin[k(x-ct)]$$

Substitution for δy and ζ in (5.106) then yields

$$k^2 ca\sin[k(x-ct)] = -\beta a\sin[k(x-ct)]$$

or

$$c = -\beta/k^2 \tag{5.107}$$

Thus, the phase speed is westward relative to the mean flow and is inversely proportional to the square of the zonal wave number.

5.7.1 Free Barotropic Rossby Waves

The dispersion relationship for barotropic Rossby waves may be derived formally by finding wave-type solutions of the linearized barotropic vorticity equation. The barotropic vorticity equation (4.41) states that the vertical component of absolute vorticity is conserved following the horizontal motion. For a midlatitude β-plane this equation has the form

$$\left(\frac{\partial}{\partial t}+u\frac{\partial}{\partial x}+v\frac{\partial}{\partial y}\right)\zeta+\beta v=0 \tag{5.108}$$

We now assume that the motion consists of a constant basic state zonal velocity plus a small horizontal perturbation:

$$u=\overline{u}+u', \quad v=v', \quad \zeta=\partial v'/\partial x-\partial u'/\partial y=\zeta'$$

We define a perturbation streamfunction ψ' according to

$$u'=-\partial\psi'/\partial y, \quad v'=\partial\psi'/\partial x$$

from which $\zeta'=\nabla^2\psi'$. The perturbation form of (5.108) is then

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)\nabla^2\psi'+\beta\frac{\partial\psi'}{\partial x}=0 \tag{5.109}$$

where as usual we have neglected terms involving the products of perturbation quantities. We seek a solution of the form

$$\psi' = \mathrm{Re}[\Psi \exp(i\phi)]$$

where $\phi = kx + ly - \nu t$. Here k and l are wave numbers in the zonal and meridional directions, respectively. Substituting for ψ' in (5.109) gives

$$(-\nu + k\overline{u})\left(-k^2 - l^2\right) + k\beta = 0$$

which may immediately be solved for ν:

$$\nu = \overline{u}k - \beta k / K^2 \tag{5.110}$$

where $K^2 \equiv k^2 + l^2$ is the total horizontal wave number squared.

Recalling that $c = \nu/k$, we find that the zonal phase speed relative to the mean wind is

$$c - \overline{u} = -\beta / K^2 \tag{5.111}$$

which reduces to (5.109) when the mean wind vanishes and $l \rightarrow 0$. Thus, the Rossby wave zonal phase propagation is always *westward* relative to the mean zonal flow. Furthermore, the Rossby wave phase speed depends inversely on the square of the horizontal wave number. Therefore, Rossby waves are dispersive waves whose phase speeds increase rapidly with increasing wavelength.

For a typical midlatitude synoptic-scale disturbance, with similar meridional and zonal scales $(l \approx k)$ and zonal wavelength of order 6000 km, the Rossby wave speed relative to the zonal flow calculated from (5.111) is approximately -8 m s^{-1}. Because the mean zonal wind is generally westerly and greater than 8 m s^{-1}, synoptic-scale Rossby waves usually move eastward, but at a phase speed relative to the ground that is somewhat less than the mean zonal wind speed. For longer wavelengths the westward Rossby wave phase speed may be large enough to balance the eastward advection by the mean zonal wind so that the resulting disturbance is stationary relative to Earth’s surface. From (5.111) it is clear that the free Rossby wave solution becomes stationary when

$$K^2 = \beta / \overline{u} \equiv K_s^2 \tag{5.112}$$

The significance of this condition is discussed in the next subsection.

Unlike the phase speed, which is always westward relative to the mean flow, the zonal group velocity for a Rossby wave may be either eastward or westward relative to the mean flow, depending on the ratio of the zonal and meridional wave numbers (see Problem 5.20). Stationary Rossby modes (i.e., modes with $c = 0$) have zonal group velocities that are eastward relative to the ground. Synoptic-scale Rossby waves also tend to have zonal group velocities that are eastward relative to the ground. For synoptic waves, advection by the mean zonal wind is generally larger than the Rossby phase speed so that the phase speed is also eastward relative to the ground, but is slower than the zonal group velocity. As shown earlier in Figure 5.4b, this implies that new disturbances tend to develop downstream of existing disturbances, which is an important consideration for forecasting.

It is possible to carry out a less restrictive analysis of free planetary waves using the perturbation form of the full primitive equations. In that case the structure of the free modes depends critically on the boundary conditions at the surface and the upper boundary. The results of such an analysis are mathematically complicated, but qualitatively yield waves with horizontal dispersion properties similar to those in the shallow water model. It turns out that the free oscillations allowed in a hydrostatic gravitationally stable atmosphere consist of eastward- and westward-moving gravity waves that are slightly modified by Earth's rotation and westward-moving Rossby waves that are slightly modified by gravitational stability. These free oscillations are the normal modes of oscillation of the atmosphere. As such, they are continually excited by the various forces acting on the atmosphere. Planetary scale-free oscillations, although they can be detected by careful observational studies, generally have rather weak amplitudes. Presumably this is because the forcing is quite weak at the large phase speeds characteristic of most such waves. An exception is the 16-day period zonal wave number 1 normal mode, which can be quite strong in the winter stratosphere.

5.7.2 Forced Topographic Rossby Waves

Although free propagating Rossby modes are only rather weakly excited in the atmosphere, forced stationary Rossby modes are of primary importance for understanding the planetary-scale circulation pattern. Such modes may be forced by longitudinally dependent diabatic heating patterns or by flow over topography. Of particular importance for the Northern Hemisphere extratropical circulation are stationary Rossby modes forced by flow over the Rockies and the Himalayas. It is just the topographic Rossby wave that was described qualitatively in the discussion of streamline deflections in potential vorticity-conserving flows crossing mountain ranges in Section 4.3.

As the simplest possible dynamical model of topographic Rossby waves, we use a modified version of the the shallow potential vorticity equation (4.39). We assume that the upper boundary is at a fixed height *H*, and the lower boundary is at the variable height $h_T(x, y)$ where $|h_T| \ll H$. We also approximate ζ by the geostrophic value, ζ_g, and assume that $|\zeta_g| \ll f_0$. We can then approximate (4.39) by

$$H\left(\frac{\partial}{\partial t}+\mathbf{V}\cdot\nabla\right)(\zeta_g+f)=-f_0\frac{Dh_T}{Dt} \tag{5.113}$$

Linearizing and applying the midlatitude β-plane approximation yields

$$\left(\frac{\partial}{\partial t}+\bar{u}\frac{\partial}{\partial x}\right)\zeta_g'+\beta v_g'=-\frac{f_0}{H}\bar{u}\frac{\partial h_T}{\partial x} \tag{5.114}$$

We now examine solutions of (5.114) for the special case of a sinusoidal lower boundary. We specify the topography to have the form

$$h_T(x,y)=\mathrm{Re}\left[h_0\exp(ikx)\right]\cos ly \tag{5.115}$$

and represent the geostrophic wind and vorticity by the perturbation streamfunction

$$\psi(x,y)=\mathrm{Re}\left[\psi_0\exp(ikx)\right]\cos ly \tag{5.116}$$

Then (5.114) has a steady-state solution with complex amplitude given by

(5.117)

$$\psi_0 = f_0 h_0 / \left[H \left(K^2 - K_s^2 \right) \right]$$

The streamfunction is either exactly in phase (ridges over the mountains) or exactly out of phase (troughs over the mountains), with the topography depending on the sign of $K^2 - K_s^2$. For long waves ($K < K_s$) the topographic vorticity source in (5.114) is balanced primarily by the meridional advection of planetary vorticity (the β effect). For short waves ($K > K_s$) the source is balanced primarily by the zonal advection of relative vorticity.

The topographic wave solution (5.117) has the unrealistic characteristic that when the wave number exactly equals the critical wave number K_s, the amplitude goes to infinity. From (5.112) it is clear that this singularity occurs at the zonal wind speed for which the free Rossby mode becomes stationary. Thus, it may be thought of as a resonant response of the barotropic system.

Charney and Eliassen (1949) used the topographic Rossby wave model to explain the winter mean longitudinal distribution of 500-hPa heights in Northern Hemisphere midlatitudes. They removed the resonant singularity by including boundary layer drag in the form of Ekman pumping, which for the barotropic vorticity equation is simply a linear damping of the relative vorticity (see Section 8.3.4). The vorticity equation thus takes the form

(5.118)

$$\left(\frac{\partial}{\partial t} + \bar{u} \frac{\partial}{\partial x} \right) \zeta_g' + \beta v_g' + r \zeta_g' = -\frac{f_0}{H} \bar{u} \frac{\partial h_T}{\partial x}$$

where $r \equiv \tau_e^{-1}$ is the inverse of the spin-down time defined in (8.37).

For steady flow, (5.118) has a solution with complex amplitude

(5.119)

$$\psi_0 = f_0 h_0 / \left[H \left(K^2 - K_s^2 - i\varepsilon \right) \right]$$

where $\varepsilon \equiv r K^2 (k \bar{u})^{-1}$. Thus, boundary layer drag shifts the response's phase and removes the singularity at resonance. However, the amplitude is still a maximum for $K = K_s$ and the trough in the streamfunction occurs one-quarter cycle east of the mountain crest, in approximate agreement with observations.

By use of a Fourier expansion, (5.118) can be solved for realistic distributions of topography. The results for an x-dependence of h_T given by a smoothed version of Earth's topography at 45°N, a meridional wave number corresponding to a latitudinal half-wavelength of 35°, $\tau_e = 5$ days, $\bar{u} = 17\ \mathrm{m\,s^{-1}}$, $f_0 = 10^{-4}\ \mathrm{s^{-1}}$, and $H = 8$ km are shown in Figure 5.15. Despite its simplicity, the Charney–Eliassen model does a remarkable job of reproducing the observed 500-hPa stationary wave pattern in Northern Hemisphere midlatitudes.

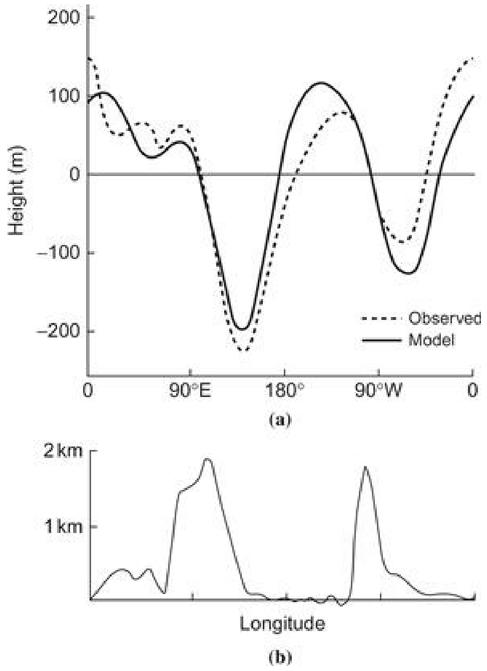


Figure 5.15 (a) Longitudinal variation of the disturbance in geopotential height ($\equiv f_0\,\Psi/g$) in the Charney–Eliassen model for the parameters given in the text (*solid line*) compared with the observed 500-hPa height perturbations at 45°N in January (*dashed line*). (b) Smoothed profile of topography at 45°N used in the computation.

After Held, 1983.

Suggested References

Chapman and Lindzen. Chapman and Lindzen, *Atmospheric Tides*: *Thermal and Gravitational*, is the classic reference on both observational and theoretical aspects of tides, a class of atmospheric motions for which the linear perturbations method has proved to be particularly successful.

Gill. Gill, *Atmosphere-Ocean Dynamics*, has a very complete treatment of gravity, inertia–gravity, and Rossby waves, with particular emphasis on observed oscillations in the oceans.

Hildebrand. Hildebrand, *Advanced Calculus for Applications*, is one of many standard textbooks that discuss the mathematical techniques used in this chapter, including the representation of functions in Fourier series and the general properties of the wave equation.

Nappo. Nappo, *An Introduction to Atmospheric Gravity Waves*, is an excellent introduction to the theory and observation of gravity waves in the atmosphere.

Scorer. Scorer, *Natural Aerodynamics*, contains an excellent qualitative discussion on many aspects of waves generated by barriers such as lee waves.

Problems

5.1. Show that the Fourier component $F(x) = \mathrm{Re}\left[C\exp(imx)\right]$ can be written as

$$F(x) = |C|\cos m(x + x_0)$$

where $x_0 = m^{-1}\sin^{-1}(C_i/|C|)$ and C_i stands for the imaginary part of C.

5.2. In the study of atmospheric wave motions, it is often necessary to consider the possibility of amplifying or decaying waves. In such a case we might assume that a solution has the form

$$\psi = A\cos(kx - \nu t - kx_0)\exp(\alpha t)$$

where A is the initial amplitude, α the amplification factor, and x_0 the initial phase. Show that this expression can be written more concisely as

$$\psi = \text{Re}\left[Be^{ik(x-ct)}\right]$$

where both B and c are complex constants. Determine the real and imaginary parts of B and c in terms of A, α, k, ν, and x_0.

5.3. Several of the wave types discussed in this chapter are governed by equations that are generalizations of the wave equation

$$\frac{\partial^2\psi}{\partial t^2} = c^2\frac{\partial^2\psi}{\partial x^2}$$

This equation can be shown to have solutions corresponding to waves of arbitrary profile moving at the speed c in both positive and negative x directions. We consider an arbitrary initial profile of the field ψ; $\psi = f(x)$ at $t = 0$. If the profile is translated in the positive x direction at speed c without change of shape, then $\psi = f(x')$, where x' is a coordinate moving at speed c so that $x = x' + ct$. Thus, in terms of the fixed coordinate x we can write $\psi = f(x - ct)$, corresponding to a profile that moves in the positive x direction at speed c without change of shape. Verify that $\psi = f(x - ct)$ is a solution for any arbitrary continuous profile $f(x - ct)$. [*Hint*: Let $x - ct = x'$ and differentiate f using the chain rule.]

5.4. Assuming that the pressure perturbation for a one-dimensional acoustic wave is given by (5.23), find the corresponding solutions of the zonal wind and density perturbations. Express the amplitude and phase for u' and ρ' in terms of the amplitude and phase of p'.

5.5. Show that for isothermal motion ($DT/Dt = 0$) the acoustic wave speed is given by $(gH)^{1/2}$, where $H = RT/g$ is the scale height.

5.6. In Section 5.3.1 the linearized equations for acoustic waves were developed for the special situation of one-dimensional propagation in a horizontal tube. Although this situation does not appear to be directly applicable to the atmosphere, there is a special atmospheric mode, the *Lamb wave*, which is a horizontally propagating acoustic mode with no vertical velocity perturbation ($w' = 0$). Such oscillations have been observed following violent explosions such as volcanic eruptions and atmospheric nuclear tests. Using (5.20) and (5.21), plus the linearized forms of the hydrostatic equation, and the continuity equation (5.15), derive the height dependence of the perturbation fields for the Lamb mode in an isothermal basic state atmosphere, assuming that the pressure perturbation at the lower boundary ($z = 0$) has the form (5.23). Determine the vertically integrated kinetic energy density per unit horizontal area for this mode.

5.7. If the surface height perturbation in a shallow water gravity wave is given by

$$h' = \text{Re}\left[Ae^{ik(x-ct)}\right]$$

find the corresponding velocity perturbation $u'(x,t)$. Sketch the phase relationship between h' and u' for an eastward propagating wave.

5.8. Assuming that the vertical velocity perturbation for a two-dimensional internal gravity wave is given by (5.65), obtain the corresponding solution for the u', p', and θ' fields. Use these results to verify the approximation

$$|\rho_0\theta'/\overline{\theta}| \gg |p'/c_s^2|$$

which was used in (5.58).

5.9. For the situation in the preceding problem, express the vertical flux of horizontal momentum, $\rho_0\overline{u'w'}$, in terms of the amplitude A of the vertical velocity perturbation. Hence, show that the momentum flux is positive for waves in which phase speed propagates eastward and downward.

5.10. Show that if (5.60) is replaced by the hydrostatic equation (i.e., the terms in w' are neglected), the resulting frequency equation for internal gravity waves is just the asymptotic limit of (5.66) for waves in which $|k| \ll |m|$.

5.11.

(a) Show that the intrinsic group velocity vector in two-dimensional internal gravity waves is parallel to lines of constant phase.

(b) Show that in the long-wave limit $(|k| \ll |m|)$ the magnitude of the zonal component of the group velocity equals the magnitude of the zonal phase speed so that energy propagates one wavelength per wave period.

5.12. Determine the perturbation horizontal and vertical velocity fields for stationary gravity waves forced by flow over sinusoidally varying topography given the following conditions: The height of the ground is $h = h_0 \cos kx$, where $h_0 = 50\,\mathrm{m}$ is a constant; $N = 2\times10^{-2}\,\mathrm{s}^{-1}$; $\overline{u} = 5\,\mathrm{m\,s}^{-1}$; and $k = 3\times10^{-3}\,\mathrm{m}^{-1}$. *Hint*: For small-amplitude topography $(h_0 k \ll 1)$ we can approximate the lower boundary condition by

$$w' = Dh/Dt = \overline{u}\partial h/\partial x \quad \text{at } z = 0$$

5.13. Verify the group velocity relationship for inertia–gravity waves that are given in (5.88).

5.14. Show that when $\overline{u} = 0$ the wave number vector $\boldsymbol{\kappa}$ for an internal inertia–gravity wave is perpendicular to the group velocity vector.

5.15. Derive an expression for the group velocity of a barotropic Rossby wave with dispersion relation (5.110). Show that for stationary waves the group velocity always has an eastward zonal component relative to Earth. Hence, Rossby wave energy propagation must be downstream of topographic sources.

5.16. Prove that the internal gravity waves of Section 5.5.2 have zero linearized potential vorticity in the case of constant buoyancy frequency, N.

MATLAB Exercises

M5.1.

(a) The MATLAB script **phase_demo.m** shows that the Fourier series $F(x) = A\sin(kx) + B\cos(kx)$ is equivalent to the form $F(x) = \mathrm{Re}\, C \exp(ikx)$ where A and B are real coefficients and C is a complex coefficient. Modify the MATLAB script to confirm that the expression $F(x) = |C|\cos k(x + x_0) = |C|\cos(kx + \alpha)$ represents the same Fourier series where $kx_0 \equiv \alpha = \sin^{-1}(C_i/|C|)$, C_i stands for the imaginary part of C, α is the "phase" defined in the MATLAB script. Plot your results as the third subplot in the script.

(b) By running the script for several input phase angles (e.g., 0, 30, 60, and 90°), determine the relationship between α and the location of the maximum of $F(x)$.

M5.2. In this problem you will examine the formation of a "wave envelope" for a combination of dispersive waves. The example is that of a deep water wave in which the group velocity is half of the phase velocity. The MATLAB script **grp_vel_3.m** has code to show the wave height field at four different times for a group composed of various numbers of waves with differing wave numbers and frequencies. Study the code and determine the period and wavelength of the carrier wave. Then run the script several times, varying the number of wave modes from 4 to 32. Determine the half-width of the envelope at time $t = 0$ (top line on the graph) as a function of the number of modes in the group. The half-width is here defined as two times the distance from the point of maximum amplitude $(x = 0)$ to the point along the envelope where the amplitude is half the maximum. You can estimate this from the graph using the **ginput** command to determine the distance. Use MATLAB to plot a curve of the half-width versus number of wave modes.

M5.3. The script **geost_adjust_1.m** together with the function **yprim_adj_1.m** illustrates one-dimensional geostrophic adjustment of the velocity field in a barotropic model for a sinusoidally varying initial height field. The equations are a simplification of (5.25) in the text for the case of no y dependence. Initially $u' = v' = 0,\ \phi' \equiv gh' = 9.8\cos(kx)$. The final balanced wind in this case will have only a meridional component. Run the program **geost_adjust_1.m** for the cases of latitude 30 and 60°, choosing a time value of at least 10 days. For each of these cases choose values of wavelength of 2000, 4000, 6000, and 8000 km (a total of eight runs). Construct a table showing the initial and final values of the fields u', v', ϕ' and the ratio of the final energy to the initial energy. You can read the values from the MATLAB graphs by using the **ginput** command or, for greater accuracy, add lines in the MATLAB code to print out the values needed. Modify the MATLAB script to determine the partition of final state energy per unit mass between the kinetic energy $(v'^2/2)$ and the potential energy $\phi'^2/(2gH)$. Compute the ratio of final kinetic energy to potential energy for each of your eight cases and show this in a table.

M5.4. The script **geost_adjust_2.m** together with the function **yprim_adj_2.m** extends Problem M5.3 by using Fourier expansion to examine the geostrophic adjustment for an isolated initial height disturbance of the form $h_0(x) = -h_m/\left[1 + (x/L)^2\right]$. The version given here uses 64 Fourier modes and employs a fast Fourier transform algorithm (FFT). (There are 128 modes in the FFT, but only one-half of them provide real information.) In this case you may run the model for only 5 days of integration time (it requires a lot of computation compared to the previous case). Choose an initial zonal scale of the disturbance of 500 km and run the model for latitudes (in degrees) of 15, 30, 45, 60, 75, and 90 (six runs). Study the animations for each case. Note that the zonal flow is entirely in a gravity wave mode that propagates away from the initial disturbance. The meridional flow has a propagating gravity wave component, but also a geostrophic part (cyclonic flow). Use the **ginput** command to estimate the zonal scale of the final geostrophic flow (v component) by measuring the distance from the negative velocity maximum just west of the center to the positive maximum just east of the center. Plot a curve showing the zonal scale as a function of latitude. Compare this scale with the Rossby radius of deformation defined in the text. (Note that $gH = 400\ \mathrm{m}^2\ \mathrm{s}^{-2}$ in this example.)

M5.5. This problem examines the variation of phase velocity for Rossby waves as the zonal wavelength is varied. Run the MATLAB program **rossby_1.m** with zonal wavelengths specified as 5000, 10,000 and 20,000 km. For each of these cases try different values of the mean zonal wind until you find the mean wind for which the Rossby wave is approximately stationary.

M5.6. The MATLAB script **rossby_2.m** shows an animation of the Rossby waves generated by a vorticity disturbance initially localized in the center of the domain, with mean wind zero. Run the script and note that the waves excited have westward phase speeds, but that disturbances develop on the eastward side of the original disturbance. By following the development of these disturbances, make a crude estimate of the characteristic wavelength and the group velocity for the disturbances appearing to the east of the original disturbance at time $t = 7.5$ days. (Wavelength can be estimated by using **ginput** to measure the distance between adjacent troughs.) Compare your estimate with the group velocity formula derived from (5.108). Can you think of a reason why your estimate for group velocity may differ from that given in the formula?

M5.7. The MATLAB script **rossby_3.m** gives the surface height and meridional velocity disturbances for topographic Rossby waves generated by flow over an isolated ridge. The program uses a Fourier series approach to the solution. Ekman damping with a 2-day damping time is included to minimize the effect of waves propagating into the mountain from upstream (but this cannot be entirely avoided). Run this program for input zonal mean winds from 10 to 100 m/s at 10-m/s intervals. For each run use **ginput** to estimate the scale of the leeside trough by measuring the zonal distance between the minimum and maximum in the meridional velocity. (This should be equal approximately to one-half the wavelength of the dominant disturbance.) Compare your results in each case with the zonal wavelength for resonance given by solving (5.112) to determine the resonant $L_x = 2\pi/k$, where $K^2 = k^2 + l^2$ and $l = \pi/8 \times 10^6$ in units of 1/m. (Do not expect exact agreement because the actual disturbance corresponds to the sum over many separate zonal wavelengths.)

---

[1] It is not necessary that the perturbation velocity be small compared to the mean velocity for linearization to be valid. It is only required that quadratic terms in the perturbation variables be small compared to the dominant linear terms in (5.20) and (5.21).

[2] Note that the squared differential operator in the first term expands in the usual way as

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)^2 = \frac{\partial^2}{\partial t^2} + 2\overline{u}\frac{\partial^2}{\partial t \partial x} + \overline{u}^2\frac{\partial^2}{\partial x^2}$$

[3] Long waves excited by underwater earthquakes or volcanic eruptions are called *tsunamis*.

[4] Strictly speaking, N^2 cannot be exactly constant if ρ_0 is constant. However, for shallow disturbances the variation of N^2 with height is unimportant.

[5] Note that phase speed is not a vector. The phase speed in the direction perpendicular to constant phase lines (i.e., the blunt arrows in Figure 5.10) is given by $\nu/(k^2 + m^2)^{1/2}$, which is not equal to $\left(c_x^2 + c_z^2\right)^{1/2}$.

Chapter 6

Quasi-geostrophic Analysis

A primary goal of dynamic meteorology is to interpret the observed structure of large-scale atmospheric motions in terms of the physical laws governing the motions. These laws, which express the conservation of momentum, mass, and energy, completely determine the relationships among the pressure, temperature, and velocity fields. As we saw in Chapter 2, these governing laws are quite complicated even when the hydrostatic approximation (which is valid for all large-scale meteorological systems) is applied. For extratropical synoptic-scale motions, however, the horizontal velocities are approximately geostrophic (see Section 2.4). Such motions, which are usually referred to as *quasi-geostrophic*, are often simpler to analyze than are tropical disturbances or planetary-scale disturbances. They are also the major systems of interest in traditional short-range weather forecasting, and are thus a reasonable starting point for dynamical analysis.

In addition to the observation that extratropical weather systems are nearly geostrophic and hydrostatic, we found in the linear wave analyses of Chapter 5 that the properties of Rossby waves resemble these weather systems, whereas those of inertia–gravity waves do not. Specifically, Rossby waves have relatively slow phase speeds and nonzero potential vorticity, whereas the inertia–gravity waves are fast moving and have little or no potential vorticity. We will exploit these properties here in order to filter the full equations to a simpler set, the quasi-geostrophic (QG) equations. Before developing this system of equations, it is useful to briefly summarize the observed structure of midlatitude synoptic systems and the mean circulations in which they are embedded. We then develop the QG equations using potential vorticity and potential temperature, which provides one framework ("*PV Thinking*") for understanding. A second framework for dynamics ("*w Thinking*") is based on the diagnosis of vertical motion through an *"omega equation,"* and will also be thoroughly explored.

6.1 The Observed Structure of Extratropical Circulations

Atmospheric circulation systems depicted on a synoptic chart rarely resemble the simple circular vortices discussed in Chapter 3. Rather, they are generally highly asymmetric in form, with the strongest winds and largest temperature gradients concentrated along narrow bands called *fronts*. Also, such systems generally are highly baroclinic; the amplitudes and phases of the geopotential and velocity perturbations both change substantially with height. Part of this complexity is due to the fact that these synoptic systems are not superposed on a uniform mean flow, but are embedded in a slowly varying planetary scale flow that is itself highly baroclinic. Furthermore, this planetary scale flow is influenced by *orography* (i.e., by large-scale terrain variations) and continent-ocean heating contrasts so that it is highly longitude dependent. Therefore, while viewing synoptic systems as disturbances superposed on a zonal flow that varies only with latitude and height is a useful first approximation in theoretical analyses (see, e.g., Chapter 7), more complete descriptions require the observed zonal asymmetry.

Zonally averaged cross-sections do provide some useful information on the gross structure of the planetary scale circulation in which synoptic-scale eddies are embedded. Figure 6.1 shows meridional cross-sections of the longitudinally averaged zonal wind and temperature for the solstice seasons of (a) December, January, and February (DJF) and (b) June, July, and August (JJA). These sections extend from approximately sea level (1000 hPa) to about 32 km altitude (10 hPa). Thus, the troposphere and lower stratosphere are shown. This chapter is concerned with the structure of the wind and temperature fields in the troposphere, and the stratosphere is discussed in Chapter 12.

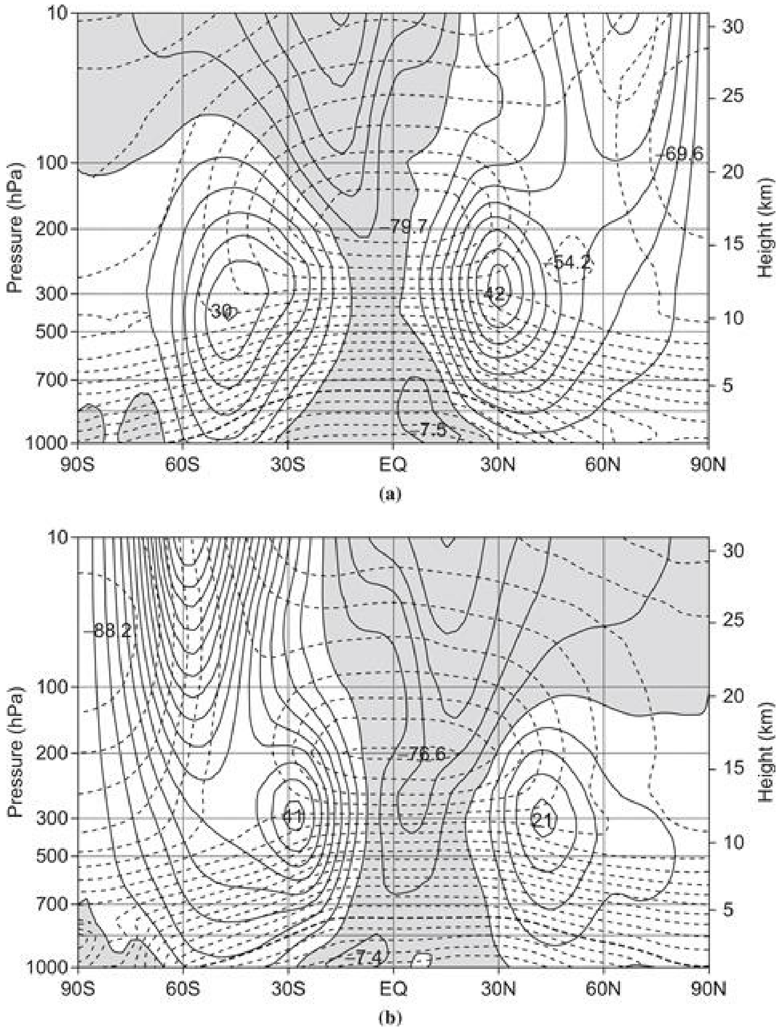


Figure 6.1 Meridional cross-sections of longitudinally and time-averaged zonal wind (*solid contours*, interval of 5ms^{-1}) and temperature (*dashed contours*, interval of 5 K) for December–February (a) and June–August (b). Easterly winds are *shaded* and 0° C isotherm is *darkened*. Wind maxima shown in m s^{-1}, temperature minima shown in ° C. (*Based on NCEP/NCAR reanalyses; after Wallace, 2003*.)

The average pole to equator temperature gradient in the Northern Hemisphere troposphere is much larger in winter than in summer. In the Southern Hemisphere the difference between summer and winter temperature distributions is smaller, due mainly to the large thermal inertia of the oceans, together with the greater fraction of the surface that is covered by oceans in the Southern Hemisphere. Since the mean zonal wind and temperature fields satisfy the thermal wind relationship to a high degree of accuracy, the seasonal cycle in zonal wind speeds is similar to that of the meridional temperature gradient. In the Northern Hemisphere the maximum zonal wind speed in the winter is twice as large as in the summer, whereas in the Southern Hemisphere the difference between winter and summer zonal wind maxima is much smaller. Furthermore, in both seasons the core of maximum zonal wind speed (called the mean jet stream axis) is located just below the *tropopause* (the boundary between the troposphere and stratosphere) at the latitude where the thermal wind integrated through the troposphere is a maximum. In both hemispheres, this is about 30° latitude during winter, but moves poleward to about 45° during summer.

The zonally averaged meridional cross-sections of Figure 6.1 are not representative of the mean wind structure at all longitudes, which can be seen in the distribution of the time-averaged zonal wind component for DJF on the 200-hPa surface in the Northern Hemisphere (Figure 6.2). It is clear that at some longitudes there are very large deviations of the time-mean zonal flow from its longitudinally averaged distribution. In particular, there are strong zonal wind maxima (jets) near 30° N just east of the Asian and North American continents and north of the Arabian peninsula; distinct minima occur in the eastern Pacific and eastern Atlantic. Synoptic-scale disturbances tend to develop preferentially in the regions of maximum time-mean zonal winds associated with the western Pacific and western Atlantic jets and to propagate downstream along *storm tracks* that approximately follow the jet axes.

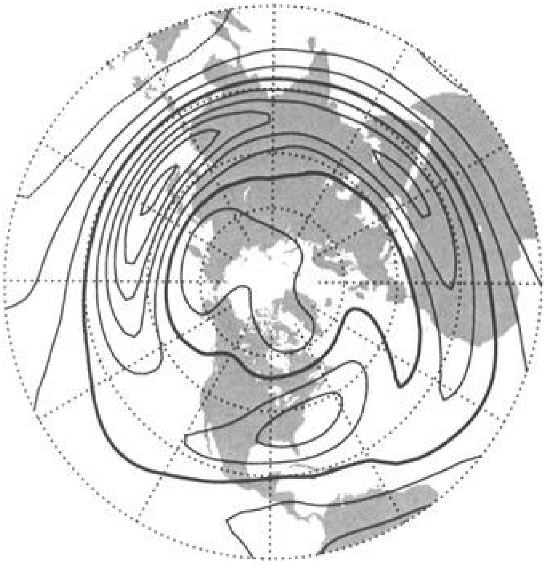


Figure 6.2 Mean zonal wind at the 200-hPa level for December through February, averaged for years 1958–1997. Contour interval 10 m s^{-1}(heavy contour, 20 m s^{-1}). (*Based on NCEP/NCAR reanalyses; after Wallace, 2003.*)

The large departure of the northern winter climatological jet stream from zonal symmetry can also be readily inferred from examination of Figure 6.3, which shows the mean 500-hPa geopotential contours for January and July in the Northern Hemisphere. Striking departures from zonal symmetry are apparent, which are linked to the distributions of continents and oceans. The most prominent asymmetries are the troughs in January to the east of the American and Asian continents. Referring to Figure 6.2, we see that the intense jet at 35° N and 140° E is a result of the semipermanent trough in that region. Thus, it is apparent that the mean flow in which synoptic systems are embedded should really be regarded as a longitude-dependent time-averaged flow. In northern summer, 500-hPa geopotential height increases much more over high latitudes as compared to the tropics due to warming in the polar region. As a result, the pole–equator temperature contrast is smaller than in winter, and the height contrast and jet stream are weaker and located at higher latitude.

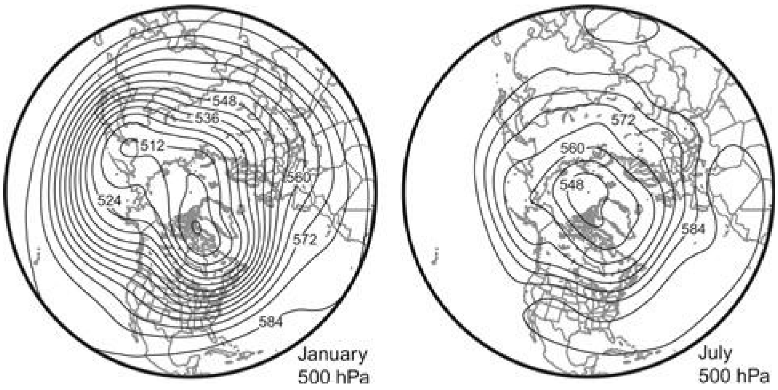


Figure 6.3 Mean 500-hPa geopotential height contours in January (*left*) and July (*right*) in the Northern Hemisphere (30-year average). Geopotential height contours are shown every 60 meters.

Adapted from NOAA/ESRL http://www.esrl.noaa.gov/psd/.

In addition to its longitudinal dependence, the planetary scale flow also varies from day to day due to its interactions with transient synoptic-scale disturbances. In fact, observations show that the transient planetary scale flow amplitude is comparable to that of the time-mean. As a result, monthly mean charts tend to smooth out the actual structure of the instantaneous jet stream since the position and intensity of the jet vary. Thus, at any time the planetary scale flow in the region of the jet stream has much greater baroclinicity than indicated on time-averaged charts. This point is illustrated schematically in Figure 6.4, which shows a latitude–height cross-section through an observed jet stream over North America. Figure 6.4a shows the zonal wind and potential temperature, whereas Figure 6.4b shows the potential temperature and Ertel potential vorticity. The 2-PVU contour of potential vorticity approximately marks the tropopause. As illustrated in Figure 6.4a, the axis of the jet stream tends to be located above a narrow sloping zone of strong potential temperature gradients called the *polar-frontal* zone. This is the zone that in general separates the cold air of polar origin from warm tropical air. The occurrence of an intense jet core above this zone of large-magnitude potential temperature gradients is, of course, not mere coincidence but rather a consequence of thermal wind balance.

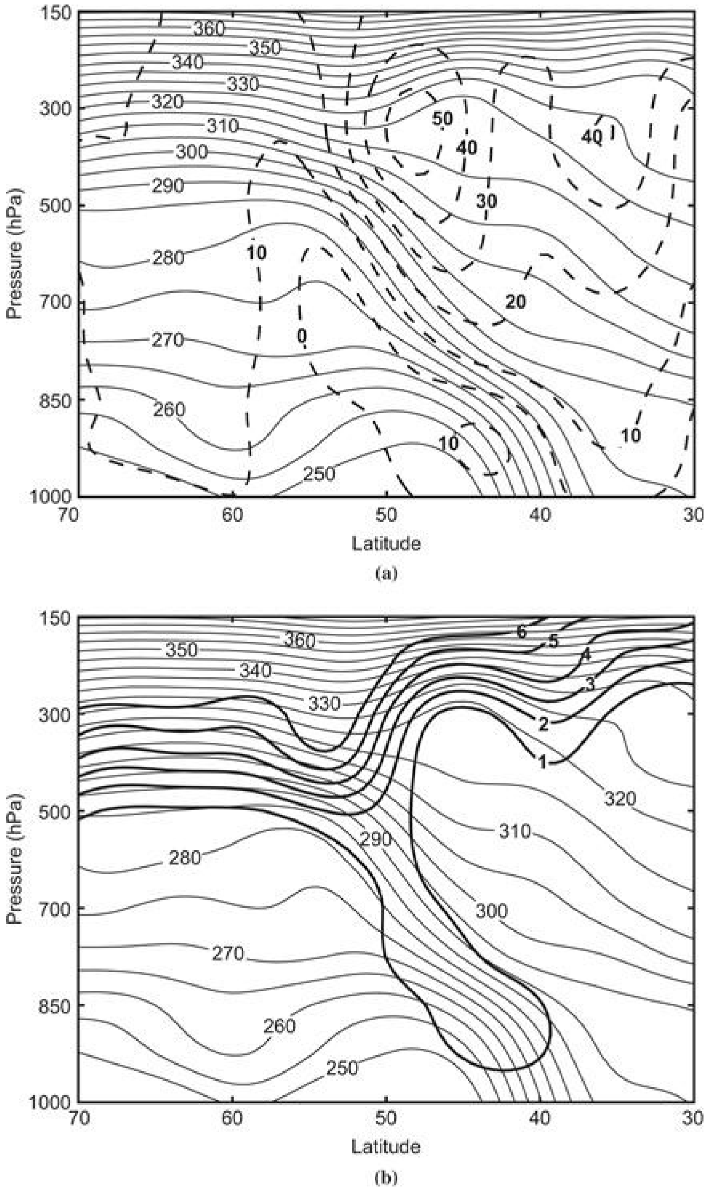


Figure 6.4 Latitude–height cross-sections through a cold front at 80W longitude on 2000 UTC January 14, 1999. (a) Potential temperature contours (*thin solid lines*, K) and zonal wind isotachs (*dashed lines*, m s^{-1}). (b) *Thin solid contours* as in (a), *heavy dashed contours*, Ertel potential vorticity labeled in PVU(1 − PVU = 10 $^{-6}$ K kg^{-1} m^2 s^{-1}).

The potential temperature contours in Figure 6.4 illustrate the strong static stability in the stratosphere. They also illustrate the fact that isentropes (constant θ surfaces) cross the tropopause in the vicinity of the jet so that air can move between the troposphere and the stratosphere without diabatic heating or cooling. The strong gradient of Ertel potential vorticity at the tropopause, however, provides a strong resistance to cross-tropopause flow along the isentropes. Note, however, that in the frontal region the potential vorticity surfaces are displaced substantially downward so that the frontal zone is characterized by a strong positive anomaly in potential vorticity associated with the strong relative vorticity on the poleward side of the jet and the strong static stability on the cold air side of the frontal zone.

It is a common observation in fluid dynamics that jets in which strong velocity shears occur may be unstable with respect to small perturbations. Any small disturbance introduced into an unstable jet will tend to amplify, drawing energy from the jet as it grows. For synoptic-scale systems in midlatitudes, the primary instability is called *baroclinic instability* because it depends on the meridional temperature gradient or, equivalently, from thermal-wind balance, on vertical wind shear. Athough horizontal temperature gradients maximize near frontal zones, baroclinic instability is not identical to frontal instability, as most baroclinic instability models describe only geostrophically scaled motions, whereas disturbances in the vicinity of strong frontal zones must be highly ageostrophic. As shown in Chapter 7, baroclinic disturbances may themselves act to intensify pre-existing temperature gradients and thus generate frontal zones.

Disturbances in the midlatitude storm tracks take the form of baroclinic waves and smaller-scale vortices. A statistical analysis of the meridional wind reveals that the baroclinic waves have a dominant wavelength of about 4000 km, and travel eastward at approximately 10 to 15 m s^{-1} (Figure 6.5). The vertical structure of the waves indicates a westward tilt with height in the geopotential height field and, consistent with hydrostatic balance, an eastward tilt with height in the temperature field (Figure 6.6). Vertical motion peaks in the middle troposphere, with rising motion between the trough and ridge and sinking motion west of the trough. It is these organized patterns of vertical motion that are largely responsible for clouds and precipitation in middle latitudes, and understanding the dynamical source for these circulations is a major subject of this chapter. Key aspects of an idealized developing baroclinic system are captured in the schematic vertical cross-section illustrated in Figure 6.7. Throughout the troposphere the trough and ridge axes tilt westward (or upstream) with height,[1] whereas the axes of warmest and coldest air are observed to have the opposite tilt. As will be shown later, the westward tilt of the troughs and ridges is necessary in order that the mean flow transfers potential energy to the developing wave. In the mature stage (not shown in Figure 6.7) the troughs at 500 and 1000 hPa are nearly in phase and, as a consequence, thermal advection and energy conversion are quite weak.

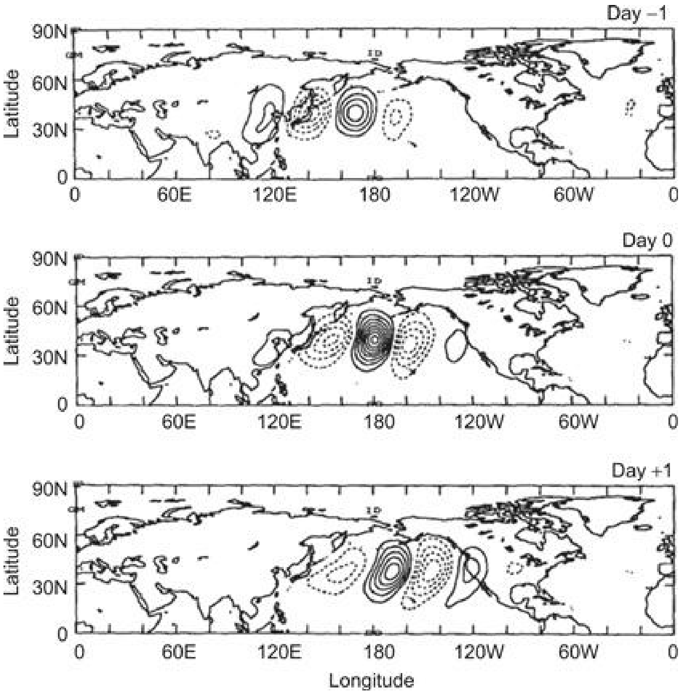


Figure 6.5 Linear regression of the meridional wind relative to the base point at 40° N, 180° W. Contours are shown every 2 m s^{-1} with negative values *dashed*. The top, middle, and bottom panels show lags of − 1, 0, and + 1 days, respectively, relative to the base point, which reveals the motion of the baroclinic waves across the North Pacific ocean.

After Chang, 1993. Copyright © American Meteorological Society. Reprinted with permission.

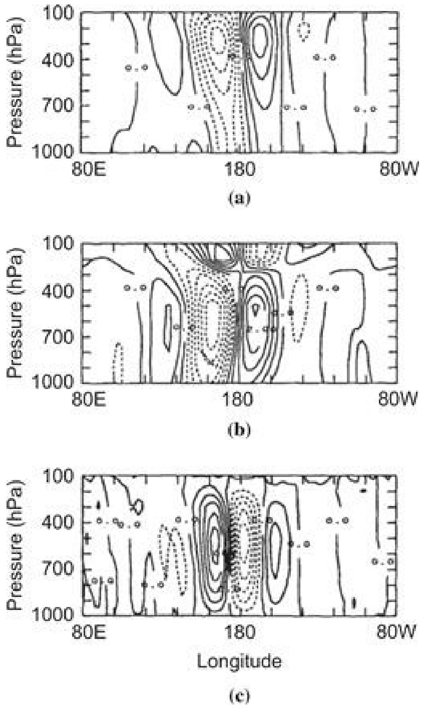


Figure 6.6 Linear regression is shown on the longitude–height plane of (a) geopotential height (b) temperature, and (c) vertical motion (ω) relative to the meridional wind at the base point located at 300 hPa, 40° N, and 180° W. Contours are shown every 20 m for geopotential height, 0.5 K for temperature, and 0.02 Pa s^{-1} for vertical motion, with negative values *dashed*. After Chang, 1993. Copyright © American Meteorological Society. Reprinted with permission. The vertical structure of baroclinic waves is revealed: westward tilt with height for geopotential height anomalies, eastward tilt with height for temperature, and rising motion eastward of the geopotential height trough axis.

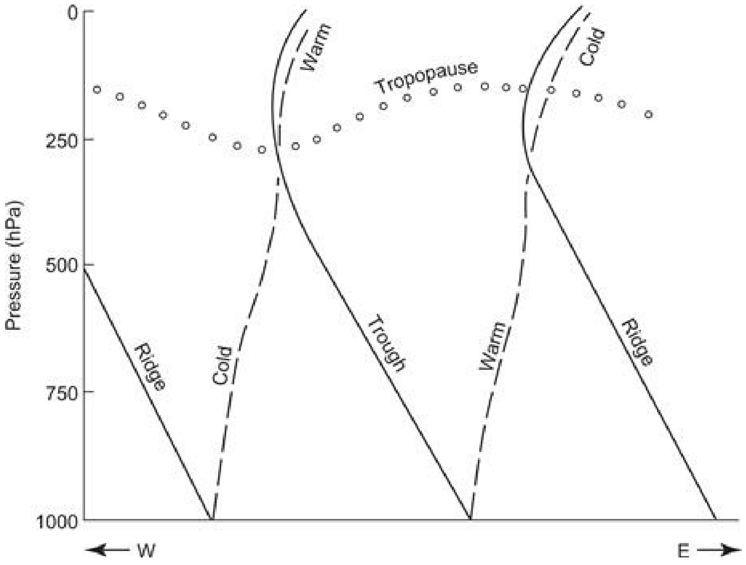


Figure 6.7 West–east cross-section through a developing baroclinic wave. *Solid lines* are trough and ridge axes; *dashed lines* are axes of temperature extrema; the chain of *open circles* denotes the tropopause.

This statistical description of baroclinic waves is manifest in individual realizations of weather disturbances in the form of extratropical cyclones. The stages in the development of a typical extratropical cyclone are shown schematically in Figure 6.8. In the stage of rapid development there is a cooperative interaction between the upper-level and surface flows; strong cold advection is seen to occur west of the trough at the surface, with weaker warm advection to the east. This pattern of thermal advection is a direct consequence of the fact that the trough at 500 hPa lags (lies to the west of) the surface trough so that the mean geostrophic wind in the 1000- to 500-hPa layer is directed across the 1000- to 500-hPa thickness lines toward larger thickness west of the surface trough and toward smaller thickness east of the surface trough.

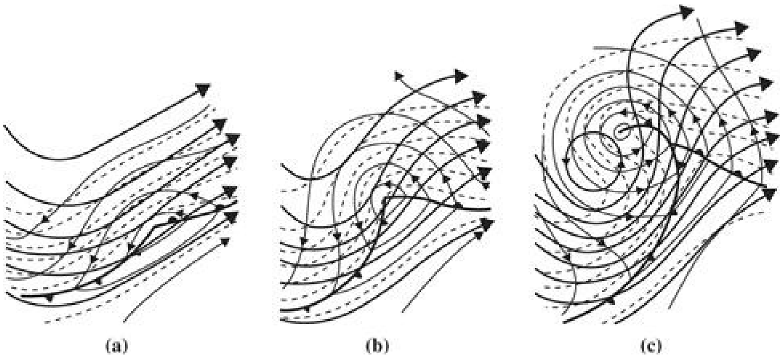


Figure 6.8 Schematic 500-hPa contours (*heavy solid lines*), 1000-hPa contours (*thin lines*), and 1000–500-hPa thickness (*dashed lines*) for a developing extratropical cyclone at three stages of development: (a) incipient development stage, (b) rapid development stage, and (c) occlusion stage.

After Palmèn and Newton, 1969.

6.2 Derivation of the Quasi-Geostrophic Equations

The main goal of this chapter is to show how the observed structure of midlatitude synoptic systems can be interpreted in terms of the constraints imposed on synoptic-scale motions by the dynamical equations. Specifically, we show that for motions that are nearly hydrostatic and geostrophic, the three-dimensional flow field is determined approximately by the pressure field. Given the small magnitude of vertical motion on these scales, this is a remarkable result. Here we use height, z, as the vertical coordinate, and make the Boussinesq approximation so that density is treated as a constant everywhere except for the buoyancy force. This approach has the advantage of simplicity and clarity compared to using pressure as a vertical coordinate, which involves a left-handed coordinate system and reverses the sign of vertical motion (negative is rising). Nevertheless, most weather and climate data are contained in pressure coordinates, so in Section 6.7 we present the isobaric-coordinate form of the equations. For completeness, the isobaric equations also allow for a linear variation of the Coriolis parameter (**β**-plane approximation), whereas the rest of this chapter employs the f-plane approximation (constant value); this simplifies the analysis and is a good approximation in midlatitudes, since ambient potential vorticity gradients near jet streams are much larger than those of f.

6.2.1 Preliminaries

The geostrophic wind, $\mathbf{V}_g = (u_g, v_g)$ is given by

$$\mathbf{V}_g = \frac{1}{\rho_0 f}\mathbf{k} \times \boldsymbol{\nabla}_h p = \frac{1}{\rho_0 f}\left(-\frac{\partial p}{\partial y}, \frac{\partial p}{\partial x}\right) \tag{6.1}$$

where ρ_0 and f are constants as a result of the Boussinesq and f-plane approximations, respectively. Recall that $\boldsymbol{\nabla}_h$ is the horizontal gradient operator $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$. The geostrophic relative vorticity, ζ_g, measures the rotational component of the geostrophic wind about the vertical axis,

$$\zeta_g = \frac{\partial v_g}{\partial x} - \frac{\partial u_g}{\partial y} = \frac{1}{\rho_0 f}\boldsymbol{\nabla}_h^2 p \tag{6.2}$$

where the second relationship derives from (6.1); the geostrophic relative vorticity is determined completely by the pressure field.

We will assume that the wind in all but one of the advection terms of the equations may be approximated by the geostrophic wind. The basis for this assumption derives from the following scaling arguments. The Eulerian time tendency and horizontal advection scale like

$$\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} \sim \frac{U}{L} \tag{6.3}$$

and, for $w \sim Ro\frac{H}{L}U$, vertical advection scales like

$$w\frac{\partial}{\partial z} \sim \frac{U}{L}Ro \tag{6.4}$$

where *Ro* is the Rossby number, *U/fL*, discussed in Section 2.4.2. Vertical advection may be neglected for small Rossby number, provided the vertical gradient of the advected quantity is not on the order of Ro^{-1} or larger. Moreover, since the Lagrangian tendency of horizontal momentum scales with the Rossby number, the ageostrophic wind does as well—see (2.24) and (2.25); as a result, the ageostrophic wind is small compared to the geostrophic wind. Therefore, for a small Rossby number, to good approximation we can we make the replacement

$$\frac{D}{Dt} \rightarrow \frac{D}{Dt_g} = \frac{\partial}{\partial t} + \mathbf{V}_g \cdot \boldsymbol{\nabla}_h = \frac{\partial}{\partial t} + u_g\frac{\partial}{\partial x} + v_g\frac{\partial}{\partial y} \tag{6.5}$$

which is simply the material derivative following the geostrophic wind.

In addition to a nearly geostrophic horizontal wind (“quasi-geostrophic”), we assume that the vertical distribution of pressure is nearly hydrostatic (“quasi-hydrostatic”). Recall that although we do not have a prognostic (forecast) equation for *w*, this does not mean that *w* is zero, but rather that it must be deduced diagnostically from the other variables. As was done in Section 2.4.3, we assume a *reference atmosphere* (denoted here by overbars) that is at rest and in exact hydrostatic balance:

$$\frac{\partial \overline{p}}{\partial z} = -\overline{\rho} g$$

Perturbations from the reference state are also in hydrostatic balance,

$$\frac{\partial p}{\partial z} = \frac{\rho_0}{\bar{\theta}}\theta g \tag{6.6}$$

where p and θ are the disturbance pressure and potential temperature, respectively, ρ_0 is a constant density (a result of the Boussinesq approximation), and $\bar{\theta}$ is a function of z only. The full value of potential temperature will be denoted by θ_{tot}.

Together, the assumptions of geostrophic and hydrostatic balance produce the thermal wind relation (see Section 3.4):

$$\mathbf{V}_{\mathrm{T}} = \frac{\partial \mathbf{V}_g}{\partial z} = \frac{g}{f\bar{\theta}}\mathbf{k}\times\nabla_h\theta \tag{6.7}$$

We notice that the disturbance potential temperature plays the same role as pressure does for the geostrophic wind: Wind flows along the contours with greatest magnitude near the largest gradients.

From equation (2.54), in the absence of strong diabatic heating, and for a small Rossby number, the first law of thermodynamics can be expressed as

$$\boxed{\frac{D\theta}{Dt_g} = -w\frac{d\bar{\theta}}{dz}} \tag{6.8}$$

which is the **quasi-geostrophic thermodynamic energy equation**. Equation (6.8) says that the perturbation potential temperature following the geostrophic motion changes due to vertical advection of the reference potential temperature. Note that the appearance of vertical advection in (6.8) reflects the large static stability of the reference atmosphere (on the order of $1/Ro$). To simplify the subsequent discussion, we take the static stability to be constant, which is a reasonable leading-order approximation for the troposphere. This implies that the buoyancy frequency (see Section 2.7.3),

$$N = \left[\frac{g}{\bar{\theta}}\frac{d\bar{\theta}}{dz}\right]^{1/2} \tag{6.9}$$

is also constant.

Finally, since many results that follow assume a familiarity with the gradient and Laplacian operations on scalar fields, we provide a brief review of these terms and their interpretation. Given a two-dimensional scalar function, ϕ, the gradient of ϕ, $\nabla_h\phi$, is a *vector* quantity that is directed toward larger values and orthogonal to lines of constant ϕ(Figure 6.9). The Laplacian is given by the divergence of the gradient of ϕ, $\nabla_h^2\phi = \nabla_h\cdot(\nabla_h\phi)$, which is a *scalar* quantity. Since the gradient points toward larger values, the Laplacian is positive and a local maximum near local minima in ϕ; conversely, the Laplacian is negative and a local minimum near local maxima in ϕ.

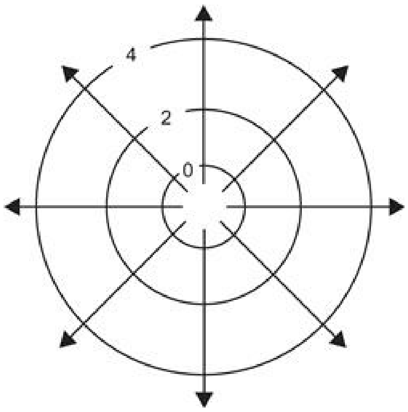


Figure 6.9 Illustration of the gradient operator for a scalar function, ϕ. The gradient of ϕ, a vector field, is illustrated by *arrows* that are directed toward larger values and orthogonal to *lines* of constant ϕ. The Laplacian is given by the divergence of the gradient, which is a local maximum near the local minimum in ϕ.

6.3 Potential vorticity derivation of the QG equations

Consider now conservation of potential vorticity, which will provide the basis for one perspective of quasi-geostrophic dynamics. In Sections 5.3.2 and 5.5.2, we discovered that fast gravity-wave motions have zero potential vorticity. The dynamics of extratropical weather systems are more closely related to Rossby waves, which are geostrophically balanced and have nonzero PV, so it proves useful to filter the gravity waves from the governing equations. This is accomplished by the linearized PV equation (5.91), but the linear dynamics are deficient in many aspects needed to understand weather systems, including interactions between these systems and the jet stream. These interactions come from advection terms in the PV equation, which are lacking in (5.91) and must be reintroduced. It turns out that this is easily accomplished. First, we use the perturbation method of Section 5.1 to linearize the Ertel PV around the reference atmosphere described in Section 6.2.1, while retaining the nonlinear advection terms. Second, we replace the wind by the geostrophic wind in both the PV and the advection terms by the reasoning leading to (6.5).

For the Boussinesq approximation, the Ertel PV conservation law (4.24) can be expressed as

$$\frac{D\Pi}{Dt} = \frac{D}{Dt}\left[\frac{\boldsymbol{\omega}_a \cdot \boldsymbol{\nabla}\theta_{\text{tot}}}{\rho_o}\right] = 0 \tag{6.10}$$

where

$$\boldsymbol{\omega}_a = (\eta, \xi, \zeta + f) \tag{6.11}$$

is the three-dimensional vorticity vector. Since $\boldsymbol{\nabla}\theta_{\text{tot}} = \left(\frac{\partial\theta}{\partial x}, \frac{\partial\theta}{\partial y}, \frac{\partial\theta}{\partial z} + \frac{d\bar{\theta}}{dz}\right)$, we immediately notice that the x and y contributions to the dot product in (6.10) involve nonlinear terms[2] and must be discarded, leaving

$$\frac{D}{Dt}\left[\frac{1}{\rho_o}(\zeta + f)\frac{\partial\theta_{\text{tot}}}{\partial z}\right] = 0 \tag{6.12}$$

Substituting $\theta_{\text{tot}} = \bar{\theta} + \theta$, and again discarding nonlinear terms gives

$$\frac{D}{Dt}\frac{1}{\rho_o}\left[(\zeta+f)\frac{d\bar{\theta}}{dz}+f\frac{\partial\theta}{\partial z}\right]=0 \tag{6.13}$$

Factoring the constant $\frac{d\bar{\theta}}{dz}$ yields

$$\frac{D}{Dt}\frac{1}{\rho_o}\frac{d\bar{\theta}}{dz}\left[\zeta+f+f\frac{\partial}{\partial z}\left(\frac{d\bar{\theta}}{dz}^{-1}\theta\right)\right]=0 \tag{6.14}$$

Although $\frac{d\bar{\theta}}{dz}$ is assumed constant, we introduce it within the vertical derivative since this general form holds even in the case where $\frac{d\bar{\theta}}{dz}$ is a function of z. Finally, approximating the wind geostrophically gives

$$\frac{D}{Dt_g}\frac{1}{\rho_o}\frac{d\bar{\theta}}{dz}\left[\zeta_g+f+f\frac{\partial}{\partial z}\left(\frac{d\bar{\theta}}{dz}^{-1}\theta\right)\right]=0 \tag{6.15}$$

or

$$\boxed{\frac{D}{Dt_g}\frac{1}{\rho_o}\frac{d\bar{\theta}}{dz}q=0} \tag{6.16}$$

where

$$\boxed{q=\zeta_g+f+f\frac{\partial}{\partial z}\left(\frac{d\bar{\theta}}{dz}^{-1}\theta\right)} \tag{6.17}$$

is the **quasi-geostrophic potential vorticity** (QG PV).

Equation (6.16) says that the QG PV is conserved following the geostrophic motion, analogous to Ertel PV conservation following the full flow. We note that the leading factor, $\frac{1}{\rho_o}\frac{d\bar{\theta}}{dz}$, is independent of time and merely scales the QG PV by a constant function of z. It is customary to discard this factor; however, in doing so the units change to those of vorticity (s^{-1}). Although, as a result of linearization, the QG PV involves a sum of vorticity and static stability compared to a product for the Ertel PV, the dynamical interpretation following an air parcel is qualitatively similar: Increases in vorticity must be accompanied by decreases in static stability in order to conserve PV.

Taking the vertical derivative of the quasi-geostrophic thermodynamic energy equation (6.8) and using the result to replace the static stability term in (6.15) gives

$$\boxed{\frac{D}{Dt_g}(\zeta_g+f)=f\frac{\partial w}{\partial z}} \tag{6.18}$$

which is the **quasi-geostrophic vorticity equation**. Following the geostrophic motion, the geostrophic absolute vorticity, ζ_g+f, changes due to stretching of planetary rotation by vertical motion. The fact that the contribution from stretching of relative vorticity is absent indicates that the quasi-geostrophic approximation is only formally valid when the relative vorticity is small compared to f. This is consistent with the small Rossby number approximation, since the ratio of the scaling for vorticity, *U/L*, to *f* is *U/*(*fL*). Fortunately, it turns out that quasi-geostrophic equations remain qualitatively useful for observed disturbance amplitudes despite this restrictive condition. The vertical motion in the quasi-geostrophic vorticity equation is linked to horizontal motion by the **quasi-geostrophic mass continuity equation**,

$$\boxed{\frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y} + \frac{\partial w}{\partial z} = 0} \tag{6.19}$$

where $\mathbf{V}_a = (u_a, v_a)$ is the ageostrophic wind. This result derives directly from the Boussinesq approximation (see 2.60) and the fact that the geostrophic wind is horizontally nondivergent on the *f*-plane.

We return now to the question of momentum conservation for the quasi-geostrophic approximation. If we approximate the momentum geostrophically in the momentum equations (e.g., 2.24 and 2.25), steady-state expressions result and there are no dynamics. Therefore, ageostrophic effects must be present in the momentum equation (hence, *quasi*-geostrophy), and this is one reason why we did not begin our analysis with momentum conservation. Approximating the material derivative geostrophically, the momentum equation can be written in general as

$$\frac{D\mathbf{V}_g}{Dt_g} = -\frac{1}{\rho_o}\boldsymbol{\nabla}_h(p_g + p_a) - f\mathbf{k}\times(\mathbf{V}_g + \mathbf{V}_a) \tag{6.20}$$

which allows for ageostrophic effects in both the pressure and wind fields. Forming a vorticity equation as in Chapter 4, by taking $\frac{\partial}{\partial x}$ of the v_g-momentum equation minus $\frac{\partial}{\partial y}$ of the u_g-momentum equation, gives

$$\frac{D\zeta_g}{Dt_g} = -f\left(\frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y}\right) = f\frac{\partial w}{\partial z} \tag{6.21}$$

where the last equality is accomplished by the quasi-geostrophic mass continuity equation. Noting that *f* is a constant allows it to be added to the material derivative, which recovers the quasi-geostrophic vorticity equation (6.18). This shows that (6.20) is consistent with the previously derived equations, but for reasons described subsequently, there is ambiguity in the individual terms in (6.20).

Ambiguity in the momentum equation involves two effects. First, since the curl of a gradient is zero, we may add an arbitrary field, $\tilde{p}_a$, to p_a in (6.20) and arrive at the same vorticity equation, (6.18). Second, we may add an arbitrary vector field, $\tilde{\mathbf{V}}_a$, to the ageostrophic wind in the momentum equation and arrive at the same vorticity equation, provided that $\boldsymbol{\nabla}_h\cdot\tilde{\mathbf{V}}_a = 0$. Thus, p_a and $\mathbf{V}_a$ are arbitrary fields, but they are related through the **quasi-geostrophic divergence equation**,

$$\nabla_h^2 p_a = \rho_0 f\zeta_a + 2J(u_g, v_g) \tag{6.22}$$

which is obtained from equation (6.20) by taking $\frac{\partial}{\partial x}$ of the u_g-momentum equation plus $\frac{\partial}{\partial y}$ of the v_g-momentum equation. Here, $J(u_g, v_g) = \frac{\partial u_g}{\partial x}\frac{\partial v_g}{\partial y} - \frac{\partial v_g}{\partial x}\frac{\partial u_g}{\partial y}$ is a Jacobian, and $\zeta_a = \frac{\partial v_a}{\partial x} - \frac{\partial u_a}{\partial y}$ is the ageostrophic vertical vorticity. Although (6.22) provides a diagnostic equation relating the ageostrophic pressure and the rotational part of the ageostrophic wind to the geostrophic wind, lacking additional information, one of the ageostrophic quantities must be eliminated in order to close the system. This choice is arbitrary, and therefore so are the terms on the right side of (6.20). The most common choice, $p_a = 0$, puts all ageostrophic effects into $\mathbf{V}_a$, giving the **quasi-geostrophic momentum equation**

(6.23)

$$\boxed{\frac{D\mathbf{V}_g}{Dt_g} = -f\mathbf{k} \times \mathbf{V}_a}$$

We emphasize that this arbitrary choice has no effect on the dynamics of vorticity or potential vorticity. The divergent part of the ageostrophic wind is well constrained through the continuity equation and will play an essential role in "*w*-Thinking." First we explore the "PV Thinking" view of quasi-geostrophic dynamics, which has no role for ageostrophic motion.

Section's Key Points

- The QG equations are defined by a small Rossby number and for a reference atmosphere that has strong stratification.
- The QG PV is a linearization of the Ertel PV consisting of a sum of contributions from the vertical component of vorticity and the disturbance static stability.
- Following the geostrophic motion, geostrophic vorticity changes due to stretching of ambient planetary rotation.
- Neglecting ageostrophic effects on pressure, following the geostrophic motion, geostrophic momentum changes due to Coriolis turning of the ageostrophic wind. The divergent ageostrophic wind links to vertical motion by the mass continuity equation, and the rotational ageostrophic wind is approximated from the geostrophic wind by the QG divergence equation.

6.4 Potential Vorticity Thinking

The fact that potential vorticity conservation represents all of the fundamental physical conservation laws in a single equation provides the basis for a very powerful and simple interpretation of the dynamics of weather systems. Before reviewing the two essential elements comprising "PV Thinking," inversion and conservation, we examine the scaling properties of the QG PV.

Using the geostrophic and hydrostatic relationships to replace the vorticity and potential temperature in the QG PV (6.17), and recalling that N is the buoyancy frequency, gives

(6.24)

$$q - f = \frac{1}{\rho_o f}\nabla_h^2 p + \frac{1}{\rho_o}\frac{\partial}{\partial z}\frac{f}{N^2}\frac{\partial p}{\partial z}$$

so that the QG PV is determined completely by the pressure field. This remarkable fact will allow us to make forecasts knowing only the pressure at the initial time, and to understand why and how the system evolves. Nondimensionalizing (6.24) using length scales L and H in the horizontal and vertical, respectively, and a geostrophic scaling[3] for the pressure field, $p \sim \rho_o U f L$ gives

(6.25)

$$q - f \sim \frac{U}{L}\left(\hat{\nabla}_h^2 \hat{p} + \frac{1}{B^2}\frac{\partial^2 \hat{p}}{\partial \hat{z}^2}\right)$$

where we have assumed that the buoyancy frequency, N, is a constant, hats denote nondimensional variables that are assumed to have values on the order of one, and B is the nondimensional Burger number. In analogy with vorticity, $q - f$ may be thought of as the "relative" QG PV, and we see that the natural scaling for this quantity is the same as for vertical vorticity, U/L. Including the scaling factor from (6.16) gives a scaling for QG PV, $\frac{U}{\rho_o L}\frac{d\bar{\theta}}{dz}$, in the same units as the Ertel PV.

The Burger number is a fundamental measure of the importance of stratification relative to rotation, and can be expressed in at least four useful forms:

$$B = \frac{NH}{fL} = \frac{L_R}{L} = \frac{H}{H_R} = \frac{Ro}{Fr} \tag{6.26}$$

The second equality shows that the Burger number is given by the ratio of the Rossby radius, $L_R = \frac{NH}{f}$, to the scaling length of the motion, L. Disturbances that are large compared to the Rossby radius have a small Burger number, which promotes the importance of vortex stretching. The third equality shows that the Burger number is also given by the ratio of the scaling depth, H, to the Rossby depth, $H_R = \frac{fL}{N}$. Rossby depth measures the vertical influence of disturbances, which increases with disturbance length scale and rotation, and for smaller static stability. Finally, the Burger number is also given by the ratio of the Rossby number, $Ro = \frac{U}{fL}$, to the Froude number, $Fr = \frac{U}{NH}$, both of which are non-dimensional. As we have seen, the Rossby number measures the importance of disturbance vorticity, $\frac{U}{L}$, relative to the ambient vorticity, f, and, in the present context, the Froude number[4] measures the importance of disturbance vertical shear, $\frac{U}{H}$, relative to the buoyancy frequency, N. In deriving the QG equations we have assumed strong ambient rotation ($\zeta \ll f$) and stratification (large $\frac{d\bar{\theta}}{dz}$), so that both Ro and Fr are small (much less than one), which implies a Burger number close to unity and equal importance for the two terms in the QG PV. This in turn implies that the natural disturbance length scales are $L \sim L_R$ and $H \sim H_R$, which provide useful guides in discussing the dynamics of weather features.

Section's Key Points

- The QG PV is determined completely by the pressure field.
- The relative importance of the vorticity and static stability contributions to the QG PV is measured by the Burger number.

6.4.1 PV Inversion, Induced Flow, and Piecewise PV Inversion

PV inversion is a kinematic diagnostic technique involving the recovery of other fields from the PV. For unity B, (6.25) can be written as

$$\hat{q} - Ro^{-1} = \hat{\nabla}^2 \hat{p} \tag{6.27}$$

The constant Ro^{-1} is due to the ambient rotation and stratification and, although large, has no dynamical significance since it merely changes the value of the PV everywhere. Therefore, features of interest in the nondimensional QG PV, $\hat{q}$, are determined by the Laplacian of the nondimensional pressure field. Areas of relatively low pressure are associated with large values of QG PV, and areas of high pressure with relatively small QG PV. QG PV inversion involves recovering the pressure from QG PV by solving

$$\hat{p} = \hat{\nabla}^{-2} \hat{q} \tag{6.28}$$

Since $\hat{\nabla}^2$ involves taking derivatives, we expect that $\hat{\nabla}^{-2}$ will involve taking antiderivatives, integrals, to recover the pressure field from the QG PV.

Returning to the full-dimensional form of the QG PV, multiplying (6.24) by f we find

$$f(q-f) = \frac{1}{\rho_o}\nabla_h^2 p + \frac{1}{\rho_o}\frac{\partial}{\partial z}\frac{f^2}{N^2}\frac{\partial p}{\partial z} = \mathsf{L}p \tag{6.29}$$

where

$$\mathsf{L} = \frac{1}{\rho_o}\nabla_h^2 + \frac{1}{\rho_o}\frac{\partial}{\partial z}\frac{f^2}{N^2}\frac{\partial}{\partial z} \tag{6.30}$$

is an operator that is approximately a three-dimensional Laplacian, “stretched” in the z direction relative to x and y by $\frac{f^2}{N^2}$; the nondimensionalized version discussed in (6.28) removes this stretching. The product fq is a useful combination because it is positive (negative) for cyclones (anticyclones) in both Northern and Southern hemispheres, providing a “hemispherically neutral” interpretation for PV Thinking.[5] Since L acts like a Laplacian, $f(q-f)$ will be a maximum where p is a minimum, which for cyclones implies $q > 0$ in the Northern Hemisphere and $q < 0$ in the Southern Hemisphere; the reverse applies to anticyclones.

As for the nondimensional case, symbolically we may define QG PV inversion by the operator that reverses the action of L,

$$p = \mathsf{L}^{-1}(fq) \tag{6.31}$$

which recovers the pressure field from the QG PV field (the constant from $L^{-1} f^2$ is taken to be zero). Given the pressure field, the geostrophic wind and the disturbance potential temperature are recovered from (6.1) and (6.6), respectively. From this point of view, the wind and potential temperature field are “induced," or caused, by the QG PV.

Boundary conditions must be supplied when inverting the potential vorticity. Assuming periodic conditions in the horizontal, boundary conditions on horizontal boundaries such as the surface include a specification of the pressure (Dirichlet condition), the vertical derivative of pressure (Neumann condition), or a linear combination of the pressure and its vertical derivative (Robin condition); “free-space” solutions apply in the absence of boundary conditions. From the hydrostatic equation (6.6), Neumann conditions involve a specification of the potential temperature field at the boundary. Assuming that the surface potential temperature field is independent of the potential vorticity provides a powerful approach to atmospheric dynamics because both the QG PV and θ are conserved following the motion. Since the conservation law for surface potential temperature (6.8 with $w = 0$) resembles that for QG PV (6.16), surface potential temperature may be thought of as QG PV at the surface. Mathematically, one can make this association precise by replacing the inhomogeneous Neumann boundary condition with a homogeneous one (setting $\theta = 0$ at $z = 0$), provided the QG PV is augmented by a “spike” contribution at $z = 0$ that is directly proportional to surface θ. This “PV-θ" perspective is particularly useful for understanding how PV anomalies in the troposphere, such as those associated with tropopause undulations, interact with surface features, such as surface cyclones.

To illustrate the concepts associated with PV inversion, we consider two canonical disturbances in this chapter. The first disturbance is modeled after a “short-wave” trough, which appears on upper-tropospheric weather maps as an isolated trough of low pressure (low geopotential height on pressure surfaces). This disturbance is associated with a localized blob of cyclonic QG PV, taken here to be in the middle troposphere with Gaussian structure and a maximum value of 2 PVU (Figure 6.10). The flow induced by the PV blob is largest near the "edge" of the disturbance, and decays with distance (Figure 6.10a), which illustrates the important nonlocal aspect of the inversion: The wind is nonzero more than 1000 km from the region having QG PV. This leads to the notion of "action at a distance," where a dynamical process may depend on the flow induced by a remote PV anomaly.

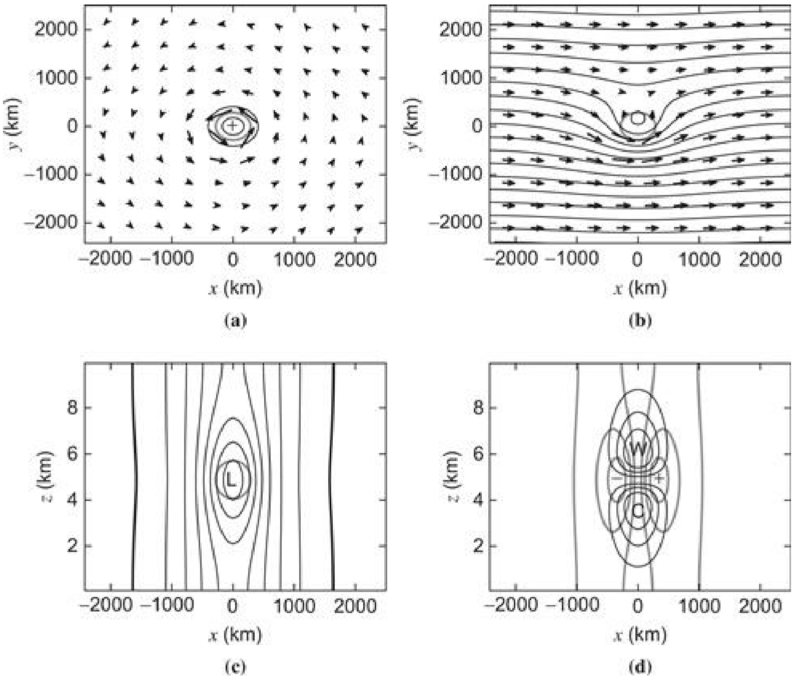


Figure 6.10 Inversion of an idealized localized blob of QG PV having a magnitude of 2 PVU embedded in a jet stream consisting of a westerly wind increasing linearly with height. (a) PV contours at altitude of 5 km are shown every 0.5 PVU; wind induced by the PV anomaly is shown by *arrows*, with a maximum wind speed of 22 m s^{-1}; (b) *pressure contours* every 42 hPa (equivalent to 60-m height contours on an isobaric surface) and full wind vectors (wind due to the PV anomaly plus the westerly jet), with a maximum wind speed of 36 m s^{-1}; (c) anomaly pressure every 21 hPa (equivalent to 30-m height contours) with *bold zero contour*; and (d) meridional wind every 6 m s^{-1} and potential-temperature anomalies every 2 K ("W" and "C" imply "warm" and "cold" values). In (a) through (c), the 1-PVU QG PV contour is denoted by the *thick gray line*.

The disturbance in Figure 6.10 is embedded within the simplest possible jet stream, with a zonal wind that is constant in x and y and if the zonal wind increases linearly upward from zero at the ground. Despite its simplicity, it is apparent from Figure 6.10b that this configuration is a rather good model for a short-wave trough in a westerly jet stream. The "stretched"-Laplacian solution is evident in the vertical cross-sections, which reveal that the nonlocal influence extends to the surface, with an area of low pressure due to the QG PV aloft (Figure 6.10c). Finally, in Figure 6.10d we see that the PV blob is associated with a local maximum in cyclonic vorticity, cyclonic wind flow around the PV blob, and a local maximum in static stability, with "warm" air above the PV blob and "cold" air below.

As mentioned previously, surface potential temperature plays the role of PV. Warm (cold) air at the surface is associated with low (high) pressure and, by thermal wind reasoning, cyclonic circulation that decays upward from the surface into the troposphere (Figure 6.11). This may be understood from the hydrostatic equation and the fact that, with zero QG PV, solutions to Laplace problems have extrema on the boundaries. A leading-order approximation to the tropopause is a rigid boundary similar to the surface, in which case cold (warm) air is associated with low (high) pressure and cyclonic (anticyclonic) circulation that decays downward from the tropopause.

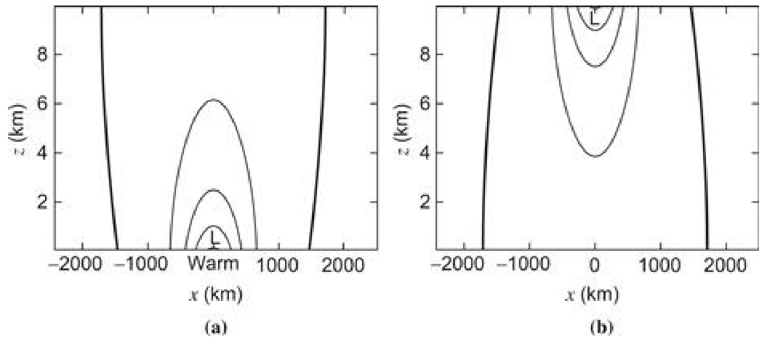


Figure 6.11 Pressure contours associated with (a) surface warm anomaly and (b) tropopause cold anomaly. Pressure is contoured every 21 hPa (equivalent to 30-m height contours) with *bold zero contour*. The potential temperature anomalies have an amplitude of 10 K and the same horizontal shape as the PV anomaly in Figure 6.10.

The notion of induced flow is most powerful when considering discrete portions of the atmosphere, and their interaction (i.e., dynamics). Since the QG PV operator, L, is linear, the atmosphere may be broken into components, the sum of which gives the full fields. This "piecewise PV inversion" framework is expressed mathematically as

$$p = \sum_{i=1}^{N} p_i = \sum_{i=1}^{N} \mathsf{L}^{-1}(f q_i) \tag{6.32}$$

To each PV element, q_i, one associates a pressure field, p_i, from which the wind and temperature may be recovered; adding up the pieces gives the full field for each variable (because geostrophic and hydrostatic balance are also linear operations).

To illustrate the concept of piecewise PV inversion we introduce the second canonical disturbance: a straight localized wind maximum within the westerly jet stream known as a "jet streak" (Figure 6.12). An elliptical dipole is used to model the jet streak, with a cyclonic QG PV anomaly poleward of an anticyclonic QG PV anomaly. The flow induced by this dipole produces a jet of strong wind between them (Figure 6.12a) and, adding in the simple linear westerly jet as in the previous example, we see a spatially localized burst of stronger wind within the broader westerlies (Figure 6.12b). Considering now the flow induced by only the positive and negative ellipses of QG PV, we see that the jet is present between the vortices because the contribution from each ellipse reinforces that from the other in this location. Moreover, the flow from each ellipse extends to the other vortex in a symmetrical fashion (Figure 6.12c, d). The dynamics of this system are simply understood by the component of flow from each ellipse advecting the QG PV, which reveals that (1) there is very little dynamics associated with each ellipses' "self" advection; and (2) the dynamics focus around the advection of the QG PV by the opposing ellipse, which advects the ellipses symmetrically downstream (the system moves in a straight line) at a speed much slower than the wind in the jet streak itself.

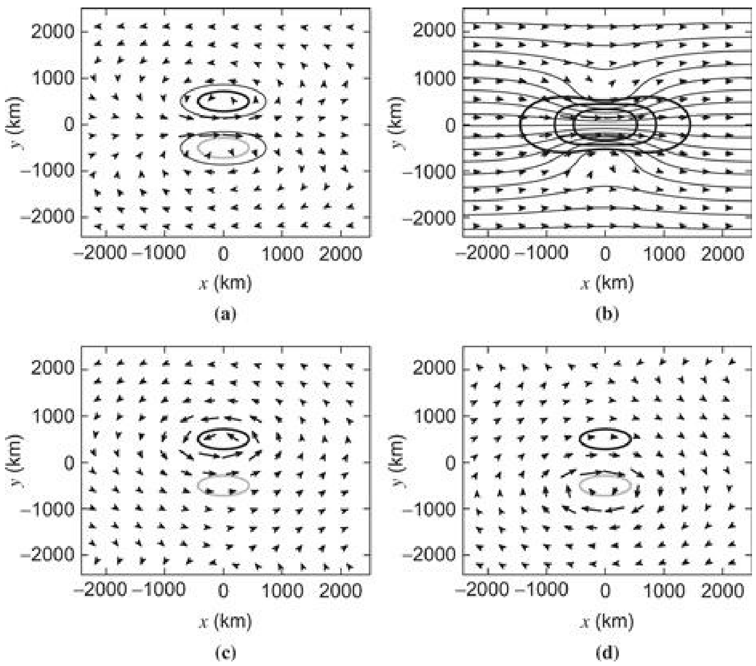


Figure 6.12 Idealized model of a "jet streak" as a dipole of QG PV having a magnitude of 2 PVU embedded in a jet stream consisting of a westerly wind increasing linearly with height. (a) PV contours at a height of 5 km are shown every 0.5 PVU; wind induced by the PV anomaly is shown by *arrows*, with a maximum wind speed of 22 m s^{-1}; (b) pressure contours at 5 km every 42 hPa (equivalent to 60-m geopotential height contours on an isobaric surface), full wind vectors (wind due to the PV dipole plus the westerly jet), and isotachs (20, 30, and 40 m s^{-1}) in *thick lines*; (c) wind induced by the positive anomaly; (d) wind induced by the negative PV anomaly. In (c) and (d) the −1- and 1-PVU QG PV contours are denoted by the *thick* gray and black lines, respectively.

Section's Key Points

- The pressure field is determined completely by the QG PV through an inversion relationship. Through the geostrophic and hydrostatic relations, the geostrophic wind and potential temperature fields are recovered from the QG PV as well.
- Warm air at the surface is analogous to cyclonic QG PV; cold air at the tropopause is analogous to cyclonic QG PV.
- QG PV inversion is defined by a Laplacian-like operator, so that local maxima in cyclonic QG PV yield local minima in pressure.
- Piecewise QG PV inversion involves breaking the QG PV field into components and associating with each component a pressure field by QG PV inversion; the sum of the pieces is equal to the total pressure field.

6.4.2 PV Conservation and the QG "Height Tendency" Equation

Given the ability to recover the geostrophic wind from the QG PV, QG prediction may now be explored from the PV Thinking perspective. Expanding the QG material derivative in (6.16) gives

$$\frac{\partial q}{\partial t} = -\mathbf{V}_g \cdot \boldsymbol{\nabla}_h q \tag{6.33}$$

which says that the time rate of change of QG PV at a point in space is determined entirely by the advection of the QG PV by the geostrophic wind. This provides the basis for a dynamical forecast model. Given an initial distribution of QG PV, one inverts this field for the pressure using (6.31), which then gives the geostrophic wind. Knowledge of the QG PV and geostrophic wind allows for a prediction of the future configuration of the QG PV from (6.33). At the future time, the QG PV can be inverted for the wind, and the process repeated indefinitely into the future.

To complete the picture, however, we need to specify boundary conditions for (6.33). At Earth's surface, which we approximate as a rigid horizontal surface, there is no vertical motion, and therefore the QG thermodynamic energy equation takes the form

$$\frac{\partial \theta}{\partial t} = -\mathbf{V}_g \cdot \boldsymbol{\nabla}_h \theta \tag{6.34}$$

Notice that this equation is identical to (6.33), where θ plays the role of q. Therefore, in addition to advancing q by (6.33), we also advance θ by (6.34), which provides boundary conditions for the QG PV inversion at the future time. For the upper boundary condition, another rigid horizontal boundary may be used to represent either a crude approximation to the tropopause or a higher level in the stratosphere.

This approach is an extremely powerful approximation to the evolution of extratropical weather systems, where one needs to evaluate the horizontal advection of only a single quantity to predict the future, compared to solving complicated equations for five variables in the full system (u,v,w,p,θ). Nevertheless, we do not measure QG PV directly, and a forecast of pressure is often more intuitive. We may obtain a pressure forecast equation from (6.33) by noting that $\frac{\partial}{\partial t} f(q-f) = \frac{\partial}{\partial t}(fq) = \mathsf{L}\frac{\partial p}{\partial t}$, so that

$$\mathsf{L}\frac{\partial p}{\partial t} = -\mathbf{V}_g \cdot \boldsymbol{\nabla}_h (fq) \tag{6.35}$$

Here, L is the quasi-Laplacian operator (6.30). Therefore, in order to obtain the pressure tendency, $\frac{\partial p}{\partial t}$, we apply the inversion operator L^{-1} to the QG PV advection term. We expect that near local maxima in QG PV advection, such as downwind of a QG PV maximum, the pressure tendency will be a local minimum (that is, the pressure will fall).

Boundary conditions for the pressure tendency are obtained using the hydrostatic equation (6.6) in (6.34) to yield

$$\frac{\partial}{\partial z}\frac{\partial p}{\partial t} = \frac{\rho_0 g}{\overline{\theta}}(-\mathbf{V}_g \cdot \boldsymbol{\nabla}_h \theta) \tag{6.36}$$

Warm-air advection at the surface implies that the pressure tendency increases upward from the surface so that, if the tendency is negative (pressure falls), it will be most negative on the boundary.

Replacing q on the right side of (6.35) using (6.17) gives

$$\mathsf{L}\frac{\partial p}{\partial t} = -\mathbf{V}_g \cdot \boldsymbol{\nabla}_h \zeta_g - \mathbf{V}_g \cdot \boldsymbol{\nabla}_h \left(f \frac{\partial}{\partial z} \frac{d\overline{\theta}}{dz}^{-1} \theta \right) \tag{6.37}$$

Using the thermal-wind equation (6.7) on the last term leaves

$$\boxed{\mathsf{L}\frac{\partial p}{\partial t} = -\mathbf{V}_g \cdot \boldsymbol{\nabla}_h (f\zeta_g) + f^2 \frac{\partial}{\partial z} \left[\frac{d\overline{\theta}}{dz}^{-1} (-\mathbf{V}_g \cdot \boldsymbol{\nabla}_h \theta) \right]} \tag{6.38}$$

which is known as the **quasi-geostrophic pressure tendency equation**, or the **height tendency equation** since, in pressure coordinates, geopotential height plays the role of pressure (see Section 6.7). Recalling the rough rule of thumb that ʟ reverses the sign of the quantity it operates upon, we see that pressure will decrease in time where there is (1) cyclonic geostrophic vorticity advection by the geostrophic wind, and (2) an upward increase in the geostrophic advection of potential temperature. An example of the first situation occurs downwind of upper-level troughs, since cyclonic vorticity is a local maximum in troughs. An example of the second situation occurs where cold air advection in the lower troposphere decreases with height, such as behind a surface cold front. Notice that if the atmosphere is very stable $\left(\text{large } \frac{d\bar{\theta}}{dz}\right)$, the pressure tendency is dominated by vorticity advection *and*, from (6.30), the response is local in the vertical, so that the dynamics become a set of nearly horizontal uncoupled layers.

For the PV blob in westerly shear (refer to Figure 6.10), solution of the pressure tendency equation shows, as expected, that the pressure falls downwind of the PV blob (Figure 6.13a). In the vertical, the pressure falls over a deep layer downwind of the upper-level feature, including at the surface (Figure 6.13b). The response is asymmetric in the vertical direction due to the boundary condition (6.36): Warm air advection to the right of the PV blob implies an upward increase in $\frac{\partial p}{\partial t}$ at the boundaries, so that pressure falls are a local minimum at the surface. From the PV perspective, recall that warm air on the surface acts like cyclonic QG PV, whereas warm air on the tropopause acts like anticyclonic PV; therefore, temperature advection on the surface reinforces the tendency from the PV blob, whereas there is cancellation on the tropopause.

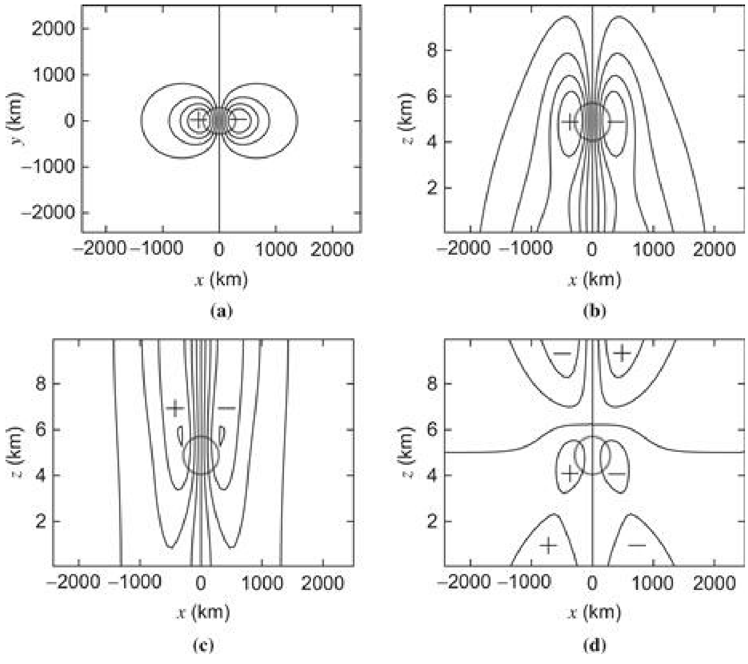


Figure 6.13 Pressure tendency associated with the PV blob in vertical shear shown in Figure 6.10. Pressure tendency: (a) plan view at 5-km altitude; (b) cross-section along $y = 0$; (c) cross-section along $y = 0$ of vorticity advection contribution; and (d) cross-section along $y = 0$ of temperature advection contribution. The sum of (c) and (d) equals (b). Pressure tendency *contours* are shown every 5 hPa $(\text{day})^{-1}$, and the 1-PVU QG PV contour by a *thick gray line*.

Examining the individual contributions from advection of vorticity and potential temperature, we see that, due to the increase of westerly winds with height, the magnitude of the pressure tendency from vorticity advection is largest above the PV blob and the pressure tendency from potential temperature advection changes sign between the surface and tropopause (Figure 6.13c, d) due in part to the boundary condition effect described previously.

Section’s Key Points

- Pressure falls near local maxima of cyclonic QG PV advection.
- Pressure falls near local maxima of cyclonic geostrophic vorticity PV advection.
- Pressure falls near locations having an increase with height of the geostrophic advection of potential temperature (cold-air advection decreasing with height or warm-air advection increasing with height).
- At the surface, geostrophic advection of potential temperature determines the sign of the vertical gradient of the pressure tendency (warm-air advection implies the pressure tendency increases with height).

6.5 Vertical Motion (*w*) Thinking

Vertical motion plays no role in PV Thinking because advection of potential vorticity is determined entirely by the geostrophic wind. Although temperature and momentum adjustments are implicit in the material conservation of QG PV, it is often helpful for understanding the dynamics of weather systems to consider these changes explicitly. In this framework the QG dynamics are expressed through the thermodynamic energy and vorticity equations

$$\frac{\partial \theta}{\partial t} = -\mathbf{V}_g \cdot \nabla_h \theta - w\frac{d\bar{\theta}}{dz} \tag{6.39}$$

and

$$\frac{\partial}{\partial t}(\zeta_g + f) = -\mathbf{V}_g \cdot \nabla_h (\zeta + f) + f\frac{\partial w}{\partial z} \tag{6.40}$$

respectively. Vertical motion controls the evolution of both θ and ζ_g. Therefore, we need an equation for *w*, which will not only mathematically close this framework but will also provide useful information on locations where clouds might form and conditional instability might be released.

To obtain the *w* equation we eliminate the time derivative between (6.39) and (6.40). First, we use (6.2) and (6.6) to express vorticity and potential temperature in terms of pressure:

$$\frac{\partial \theta}{\partial t} = \frac{\bar{\theta}}{\rho_0 g}\frac{\partial}{\partial z}\frac{\partial p}{\partial t} \quad \frac{\partial \zeta}{\partial t} = \frac{1}{\rho_0 f}\nabla_h^2 \frac{\partial p}{\partial t} \tag{6.41}$$

Time tendencies of pressure can then be eliminated by taking $\frac{g}{f\bar{\theta}}\nabla^2\frac{\partial \theta}{\partial t}$ and $\frac{\partial}{\partial z}\frac{\partial \zeta}{\partial t}$, and subtracting the resulting equations, which gives

$$\nabla_h^2 w + \frac{f^2}{N^2}\frac{\partial^2 w}{\partial z^2} = \frac{g}{\bar{\theta}N^2}\nabla_h^2(-\mathbf{V}_g \cdot \nabla_h \theta) - \frac{f}{N^2}\frac{\partial}{\partial z}(-\mathbf{V}_g \cdot \nabla_h \zeta) \tag{6.42}$$

This equation is more commonly known as the “traditional form” of the **quasi-geostrophic vertical motion** equation (or “omega” equation, where omega (ω) is the symbol for vertical motion in pressure coordinates). The left side involves again a quasi-Laplacian operator,

$$\tilde{\mathrm{L}} = \nabla_h^2 + \frac{f^2}{N^2}\frac{\partial^2}{\partial z^2} \tag{6.43}$$

and its interpretation is similar to that for L in QG PV inversion (6.30). Specifically, we see that, for a given forcing, the response in w will be mainly in the horizontal when $\frac{f^2}{N^2}$ is small—that is, for small planetary rotation and large static stability.

From (6.42) we see that upward motion ($w > 0$) is associated with local maxima in warm-air advection and an upward increase in the advection of geostrophic vorticity by the geostrophic wind. In both cases, the term on the right side of (6.42) is negative, which becomes positive when we "flip the sign" for the quasi-Laplacian, $\tilde{\mathrm{L}}$.

A problem with the traditional form of the omega equation is that there exists significant cancellation between the two right-side terms. To expose this cancellation and develop two different forms of the omega equation, we need to expand the right side of (6.42). This involves taking the gradient of vector products, for which vector notation is not well suited. Just as we use vector notation to simplify mathematical manipulations when scalar notation becomes cumbersome, it is often prudent to move to *indicial notation* for situations where vector notation becomes awkward. In many ways, indicial notation is simpler and less ambiguous than vector notation; we introduce only the basic elements needed here.

Indicial notation uses a subscript index to represent the components of a vector. For example, the vector wind U = (u, v, w) in vector notation may be written as u_i, where $u_1 = u, u_2 = v$, and $u_3 = w$ in indicial notation. Similarly, the vector coordinate directions (x, y, z) may be written simply as x_i. The key simplification of indicial notation is that *repeated indices represent sums*. For example, $a_i b_i = a_1 b_1 + a_2 b_2$ in two dimensions. Indices left over after summation, called dummy indices, are free to be changed if done so consistently. The dot product between two vectors can in general be written as

$$c = \mathbf{a} \cdot \mathbf{b} = a_i b_j \delta_{ij} = a_i b_i = a_j b_j \tag{6.44}$$

where δ_{ij} is the Kronecker delta, which takes the value one when $i = j$ and zero otherwise, leaving the familiar sum of products of vector components. Therefore, when expressing dot products it is convenient to match the subscripts of the effected quantities, which obviates the need for δ_{ij}. The gradient operator, ∇ in vector notation, takes the form $e_i \frac{\partial}{\partial x_i}$ for indicial notation, where e_i is a unit vector. This turns out to be useful for our purposes because all of the traditional rules of differential calculus, such as the product and chain rules, apply as usual. Advection combines the gradient operator and dot product, and appears for indicial notation as, for example,

$$U \cdot \boldsymbol{\nabla} \theta = u_i \frac{\partial \theta}{\partial x_i} \tag{6.45}$$

We may now write the traditional form of the omega equation using indicial notation as

$$\tilde{\mathrm{L}} w = \frac{g}{\bar{\theta} N^2} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_i} \left(-v_j \frac{\partial \theta}{\partial x_j} \right) - \frac{f}{N^2} \frac{\partial}{\partial x_3} \left(-v_j \frac{\partial \zeta}{\partial x_j} \right) \tag{6.46}$$

where we have dropped the subscript g on the wind and vorticity, and indices i and j take values 1 and 2 (x and y directions). Taking the innermost derivatives gives

$$\tilde{L}w = \frac{g}{\overline{\theta}N^2}\frac{\partial}{\partial x_i}\left(-\frac{\partial v_j}{\partial x_i}\frac{\partial \theta}{\partial x_j} - v_j\frac{\partial^2\theta}{\partial x_i \partial x_j}\right) - \frac{f}{N^2}\left(-\frac{\partial v_j}{\partial x_3}\frac{\partial \zeta}{\partial x_j} - v_j\frac{\partial}{\partial_j}\frac{\partial \zeta}{\partial x_3}\right) \tag{6.47}$$

From the definition of vorticity, $\zeta_g = \frac{1}{\rho_o f}\frac{\partial^2 p}{\partial x_i^2}$, and the hydrostatic balance, $\frac{1}{\rho_o}\frac{\partial p}{\partial x_3} = \frac{g}{\overline{\theta}}\theta$, we find $\frac{\partial \zeta}{\partial x_3} = \frac{g}{f\overline{\theta}}\frac{\partial^2\theta}{\partial x_i^2}$. Using this identity in (6.47), and taking the remaining derivative on the first term, gives

$$\tilde{L}w = \frac{g}{\overline{\theta}N^2}\left(-\frac{\partial^2 v_j}{\partial x_i^2}\frac{\partial \theta}{\partial x_j} - 2\frac{\partial v_j}{\partial x_i}\frac{\partial^2\theta}{\partial x_i \partial x_j} - v_j\frac{\partial}{\partial x_j}\frac{\partial^2\theta}{\partial x_i^2}\right) + \frac{f}{N^2}\left(\frac{\partial v_j}{\partial x_3}\frac{\partial \zeta}{\partial x_j} + v_j\frac{\partial}{\partial x_j}\frac{g}{f\overline{\theta}}\frac{\partial^2\theta}{\partial x_i^2}\right) \tag{6.48}$$

Cancellation is now apparent between the last term of each bracketed quantity. Moreover, it may be shown that (see Problem 6.3)

$$\frac{g}{\overline{\theta}N^2}\frac{\partial^2 v_j}{\partial x_i^2}\frac{\partial \theta}{\partial x_j} = \frac{f}{N^2}\frac{\partial v_j}{\partial x_3}\frac{\partial \zeta}{\partial x_j} \tag{6.49}$$

Two other forms of the omega equation are now immediately recovered. Using (6.49) to replace the term involving vorticity in (6.48) gives the compact form

$$\tilde{L}w = -2\frac{g}{N^2\overline{\theta}}\frac{\partial}{\partial x_i}\left(\frac{\partial v_j}{\partial x_i}\frac{\partial \theta}{\partial x_j}\right) \tag{6.50}$$

In vector notation this may be written as

$$\boxed{\tilde{L}w = 2\boldsymbol{\nabla}_h \cdot \mathbf{Q}} \tag{6.51}$$

where $\mathbf{Q} \equiv -\frac{g}{N^2\overline{\theta}}\frac{\partial v_j}{\partial x_i}\frac{\partial \theta}{\partial x_j}$ is the "**Q**-vector." Writing the **Q**-vector in component form,

$$\begin{aligned}\mathbf{Q} &= -\frac{g}{N^2\overline{\theta}}(Q_1, Q_2) \\ &= -\frac{g}{N^2\overline{\theta}}\left(\frac{\partial u_g}{\partial x}\frac{\partial \theta}{\partial x} + \frac{\partial v_g}{\partial x}\frac{\partial \theta}{\partial y}, \frac{\partial u_g}{\partial y}\frac{\partial \theta}{\partial x} + \frac{\partial v_g}{\partial y}\frac{\partial \theta}{\partial y}\right)\end{aligned} \tag{6.52}$$

we see that vertical motion is associated with products of horizontal wind gradients and horizontal potential temperature gradients. These effects involve changes in horizontal potential temperature gradients, which we shall explore more thoroughly in Chapter 9. Rising air is found in regions where the **Q**-vector converges, so that the righthand side of (6.51) is negative, and $\tilde{L}$ flips the sign.

The direction and magnitude of the **Q**-vector at a given point on a weather map can be estimated as follows. First note that the **Q**-vector can be written as

$$\mathbf{Q} = -\frac{g}{N^2\overline{\theta}}\left(\frac{\partial \mathbf{V}_g}{\partial x}\cdot\boldsymbol{\nabla}_h\theta, \frac{\partial \mathbf{V}_g}{\partial y}\cdot\boldsymbol{\nabla}_h\theta\right) \tag{6.53}$$

By referring the motion to a Cartesian coordinate system in which the *x* axis is parallel to the local isotherm with cold air on the left, (6.53) can be simplified to give

$$\mathbf{Q} = -\frac{g}{N^2\bar{\theta}}\left|\frac{\partial\theta}{\partial y}\right|\mathbf{k}\times\frac{\partial\mathbf{V}_g}{\partial x} \tag{6.54}$$

where we have used the fact that $\partial u_g/\partial x = -\partial v_g/\partial y$ and $\frac{\partial\theta}{\partial y} < 0$ in the specified coordinates. Thus, the **Q**-vector can be obtained by evaluating the vectorial change of $\mathbf{V}_g$ along the isotherm (with cold air on the left), rotating this change vector by 90° clockwise, and multiplying the resulting vector by $\frac{g}{N^2\bar{\theta}}\left|\frac{\partial\theta}{\partial y}\right|$. An example is shown in Figure 6.14, which shows an idealized pattern of cyclones and anticyclones in a slightly perturbed westerly thermal wind. Near the center of the low, the geostrophic wind change moving eastward along the isotherm is from northerly to southerly. Thus the geostrophic wind change vector points northward, and a 90° clockwise rotation produces a **Q**-vector parallel to the thermal wind. In the highs, by the same reasoning, the **Q**-vectors are antiparallel to the thermal wind. The pattern of $\nabla\cdot\mathbf{Q}$ thus yields descent in the region of cold-air advection west of the trough, and ascent in the warm-air advection region east of the trough.

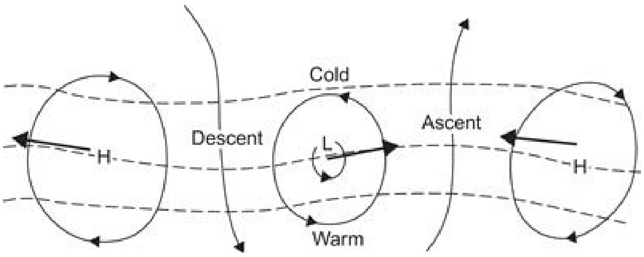


Figure 6.14 Q vectors (*bold arrows*) for an idealized pattern of isobars (*solid lines*) and isotherms (*dashed lines*) for a family of cyclones and anticyclones.

After Sanders and Hoskins, 1990.

Yet another form of the omega equation is obtained using (6.49) to replace the first term in (6.48) to yield

$$\boxed{\tilde{L}w = 2\frac{f}{N^2}\frac{\partial v_i}{\partial x_3}\frac{\partial\zeta}{\partial x_i} + D} \tag{6.55}$$

where $D = -2\frac{g}{\bar{\theta}N^2}\frac{\partial v_j}{\partial x_i}\frac{\partial^2\theta}{\partial x_i\partial x_j}$ is the “deformation term." Noting that $\partial v_i \partial x_3$ is the thermal wind, V_T, then (6.55) becomes, in vector notation, the deformation term is one of two terms in the divergence of the **Q**-vector. In vector notation,

$$\tilde{L}w = -\frac{2}{N^2}\left[-\mathbf{V}_\mathrm{T}\cdot\boldsymbol{\nabla}_h(f\zeta_g)\right] + D \tag{6.56}$$

is called the *Sutcliffe form* of the QG omega equation. Neglecting the deformation term, we see that rising air is located near regions of cyclonic vorticity advection by the thermal wind. Although this approximation neglects a potentially important source of vertical motion from the deformation term, advection of vorticity by the thermal wind is easily visualized on upper-level weather maps.

The vertical motion field for the PV blob in westerly shear is shown in Figure 6.15. Notice that **Q**-vectors converge (diverge) in the region of rising (sinking) motion downwind (upwind) of the PV blob. An east–west cross-section through the PV-blob reveals the ageostrophic circulation that links rising and sinking motion by mass continuity. Consider the local time tendency of potential temperature at the level of maximum QG PV. Since disturbance potential temperature is zero at this level, there is no advection of this quantity (see Figure 6.10d), whereas the disturbance advection of the basic-state potential temperature gradient (linear decrease with y) implies increasing (decreasing) values to the right (left) of the PV blob. Geostrophic and hydrostatic balance are "disrupted" by this geostrophic advection pattern, but they are "restored" by the ageostrophic circulation. Notice that the vertical motion pattern cools (warms) the atmosphere in the region of rising (sinking) motion, which opposes the tendency from geostrophic advection discussed previously. Moreover, Coriolis turning of the horizontal ageostrophic wind acts to increase (decrease) the v_g component of the wind above (below) the PV blob, which restores thermal wind balance with the new potential temperature distribution. In the quasi-geostrophic equations, this adjustment "process" is instantaneous, so that the atmosphere is always in quasi-geostrophic and quasi-hydrostatic balance. Linking back to the PV view of dynamics, one can say that the ageostrophic circulations that maintain geostrophic and hydrostatic balance are required if the QG PV is to be conserved following the geostrophic motion.

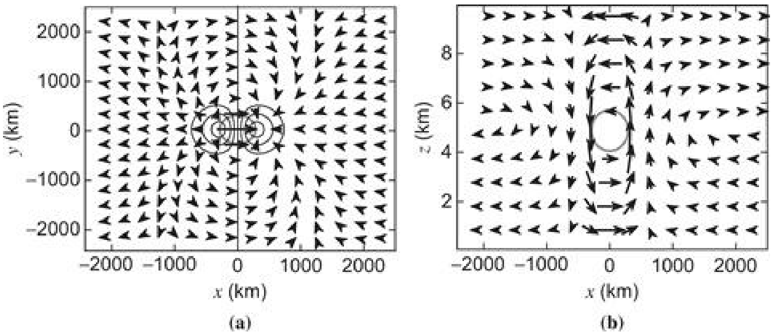


Figure 6.15 Vertical motion associated with the PV blob in vertical shear shown in Figure 6.10. (a) Q-vectors and contours of w every 2 cm s^{-1}, with rising air where Q-vectors converge; (b) vertical cross-section along $y = 0$ of the ageostrophic circulation (u_a, w). The 1-PVU QG PV contour is shown by a *thick gray line*.

Note that, in the absence of **Q**-vectors, one may still easily estimate the vertical motion pattern using the Sutcliffe form of the QG omega equation. The thermal wind, dominated by the westerly shear jet, points mostly toward positive x, so that the advection of geostrophic vorticity by the thermal wind is positive (negative) to the right (left) of the PV blob. As a result, one expects rising (sinking) motion to the right (left) of the blob. This technique can be extremely useful for quickly estimating regions of rising and sinking air from weather maps.

For the jet streak example, a four-cell pattern of vertical motion is apparent, with rising motion in the "right entrance" and "left exit" regions of the jet (Figure 6.16). Vertical cross sections through the jet entrance (Figure 6.17, left panel) and exit (Figure 6.17b) regions reveal transverse (across flow) ageostrophic circulations, which increase (decrease) zonal momentum in upper levels. These circulation patterns accelerate (decelerate) the wind into (out of) the jet, as well as increasing (decreasing) the vertical shear of the zonal wind in the jet entrance (exit) regions. Notice that, in the jet entrance region, relatively warm air rises while relatively cold air sinks, which converts potential energy to kinetic energy of the stronger winds in the jet streak; the reverse process occurs in the jet exit region.

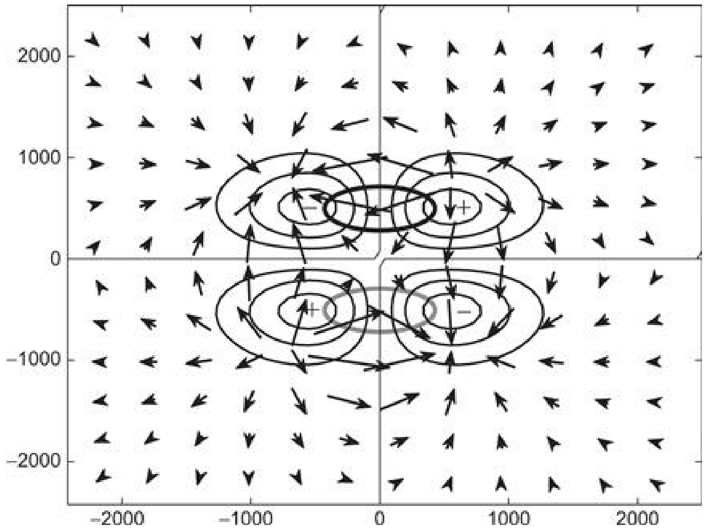
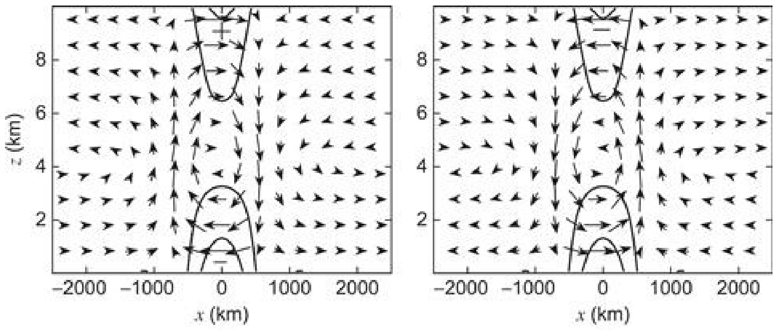


Figure 6.16 Vertical motion associated with the jet streak shown in Figure 6.12. *Contours* show vertical motion (w) at 5-km altitude every 1 cm s^{-1}, and *arrows* show the divergent ageostrophic wind. The -1- and 1-PVU QG PV contours are denoted by the *thick gray* and *black lines*, respectively.



(a)

(b)

Figure 6.17 Vertical cross-section of ageostrophic circulation (u_a, w) near the (a) jet entrance ($x = -700$ km) and (b) jet exit ($x = 700$ km) regions. *Contours* show the acceleration of the geostrophic zonal wind, $\frac{Du}{Dt}$, every 5 m s^{-1} per hour.

Section's Key Points

- The QG equations have a diagnostic equation for vertical motion.
- Air rises near local maxima in warm-air advection and an upward increase in the advection of geostrophic vorticity by the geostrophic wind.
- Air rises near local maxima of **Q**-vector convergence.
- Air rises near local maxima of cyclonic vorticity advection by the thermal wind.

6.6 Idealized Model of a Baroclinic Disturbance

We have developed two dynamically consistent frameworks for interpreting quasi-geostrophic dynamics. Here we apply these frameworks to understand intensifying extratropical cyclones, which is a process referred to as cyclogenesis (see Figures 6.7, 6.8, and accompanying text). Figure

6.18 illustrates the main elements involved in the development process. Nearly all developing cyclones are preceded by an upper-level disturbance that is due to a downward undulation of the tropopause. This downward undulation is associated with a blob of anomalous cyclonic PV, represented by the “+” sign in Figure 6.18. As discussed in connection with Figure 6.10, such a PV blob is associated with low-pressure, cyclonic winds, and warm (cold) air above (below) the blob. At the surface, a developing cyclone appears downshear of the PV blob as a region of relatively warm air at the surface. As discussed in connection with Figure 6.11, anomalously warm air at the surface behaves like cyclonic PV and is associated with low pressure and cyclonic winds that weaken upward. Taken together, these PV anomalies yield pressure fields that tilt westward with height and potential temperature fields that tilt eastward with height. A final important ingredient is the ambient flow, which takes the form of westerly winds that increase upward to a maximum at the tropopause.

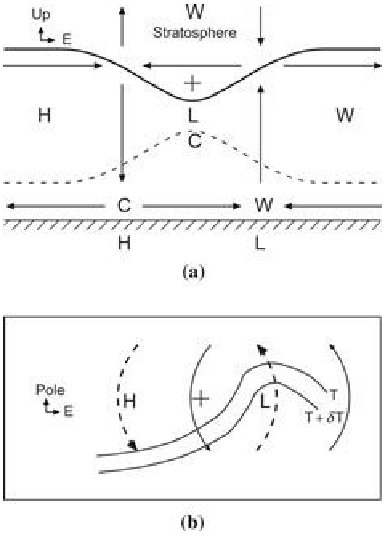


Figure 6.18 Schematic of quasi-geostrophic reasoning applied to extratropical cyclone development. (a) East–West vertical section showing the tropopause (*thick solid line*), the surface (*cross-hatched line*), a lower-tropospheric isentropic surface (*dashed line*), and the air circulations required to conserve potential vorticity for the typical case where westerly winds are increasing from the surface to the tropopause. Regions of relatively warm and cold air are given by “W” and “C," respectively, and regions of relatively low and high pressure are given by “L” and “H," respectively. (b) Horizontal plan view showing the surface isotherms (*solid lines*), surface wind associated with the upper-level PV disturbance (*thick dashed lines with arrows*), and surface wind associated with the surface cyclone (*thin solid lines with arrows*). In both (a) and (b), “+” shows the location of the PV disturbance due to the lowered tropopause.

From Hakim, 2002

From the PV-θ(kinematic) perspective, low pressure develops east of the upper-level PV blob because the induced flow from the blob acts to increase the amplitude of the surface warm anomaly through warm air advection (Figure 6.18b, *heavy dashed line*). Note that this mechanism requires a poleward decrease in the ambient potential temperature field and, by thermal wind, an upward increase in westerly winds as assumed before. Similarly, given a poleward increase of PV at the level of the upper-level PV blob, the amplitude of the anomaly will increase due to equatorward advection of PV by the flow induced from the surface warm anomaly (Figure 6.18b, *thin solid line*).

From a height tendency (dynamic) perspective, we know from Figure 6.13 that pressure falls east of the upper PV blob due to advection of cyclonic QG PV by the ambient westerly shear. Moreover, an upward decrease in the cold-air advection to the west of the developing surface low pressure results in intensification of the upper-level low pressure. One may understand the development of surface high pressure upshear from the PV blob by similar reasoning.

We may now add the contributions of clouds and precipitation to this dry description of cyclogenesis. In the region of rising motion we expect clouds to form due to latent heat release. From the *w* perspective, this added heat helps to intensify the developing cyclone by warming the column. From the PV-θ perspective, we must recognize that latent heating implies that PV is not conserved following air parcels, and from Section 4.4, we note that cyclonic PV is created below the level of maximum heating (see Figure 6.10a). This "new" PV is located above the surface cyclone, and as a result, produces a decrease in pressure and an increase in cyclonic circulation by PV inversion and superposition reasoning.

Finally, we note that the cyclogensis process is accelerated when the ambient static stability is reduced. This may be understood from the PV-θ perspective through an increase in the Rossby depth, which increases flow induced by the upper-level PV blob at the surface. From the *w* perspective, weaker static stability is also associated with a stronger and deeper ageostrophic circulation. Conversely, the cyclogenesis process is slowed for increasing ambient static stability. This is one reason why the main storm tracks for extratropical cyclones are found over the western margins of the Atlantic and Pacific oceans; the high heat capacity of the upper ocean allows relatively warm water in winter to heat the overlying lower atmosphere, weakening the static stability of the troposphere.

6.7 Isobaric Form of the QG Equations

Gridded weather data and weather maps are typically available in isobaric, rather than constant height, surfaces. The reasons for this are partly historical, but also that density becomes implicit in the governing equations when pressure is used as a vertical coordinate. Nevertheless, using pressure as the vertical coordinate yields a "left-handed" coordinate system, since pressure increases down rather than up, which results in a more cumbersome interpretation of the equations. Upward vertical motion is negative in the isobaric system ($\omega = Dp/Dt$ is the vertical motion). Following is a summary of the QG equations in isobaric coordinates presented without proof. Interpretation of these equations is the same as for the height-coordinate counterparts presented earlier in the chapter, except here additional terms appear due to the inclusion of the β effect (the Coriolis parameter varies linearly with latitude); these are often small in magnitude, but are included here for completeness.

Hydrostatic balance is given by

$$\frac{\partial \Phi}{\partial p} = -\alpha = -RT/p \tag{6.57}$$

The geostrophic wind is defined by

$$\mathbf{V}_g \equiv f_0^{-1}\mathbf{k} \times \nabla_h \Phi \tag{6.58}$$

where Φ is the geopotential, and the β-plane approximation applies to the Coriolis parameter,

$$f = f_0 + \beta y \tag{6.59}$$

where f_0 is a constant, and $\beta \equiv (df/dy)_{\phi_0} = 2\Omega\cos\phi_0/a \text{ and } y = 0 \text{ at } \phi_0$. The QG momentum equation takes the form

$$\frac{D_g \mathbf{V}_g}{Dt} = -f_0 \mathbf{k} \times \mathbf{V}_a - \beta y \mathbf{k} \times \mathbf{V}_g \tag{6.60}$$

and the QG mass continuity equation is

$$\frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y} + \frac{\partial \omega}{\partial p} = 0 \tag{6.61}$$

The adiabatic QG thermodynamic energy equation is

$$\left(\frac{\partial}{\partial t} + \mathbf{V}_g \cdot \boldsymbol{\nabla}_h\right) T - \left(\frac{\sigma p}{R}\right)\omega = 0 \tag{6.62}$$

where $\sigma \equiv -RT_0 p^{-1} d\ln\theta_0/dp$ measures the static stability of the reference atmosphere. The QG vorticity equation is unchanged aside from the **β** term,

$$\frac{D_g \zeta_g}{Dt} = -f_0\left(\frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y}\right) - \beta v_g \tag{6.63}$$

The vorticity–temperature-advection form of the QG height tendency equation ($\chi \equiv \partial\Phi/\partial t$),

$$\begin{aligned}\left[\nabla^2 + \frac{\partial}{\partial p}\left(\frac{f_0^2}{\sigma}\frac{\partial}{\partial p}\right)\right]\chi = &- f_0 \mathbf{V}_g \cdot \boldsymbol{\nabla}_h\left(\frac{1}{f_0}\nabla^2\Phi + f\right) \\ &- \frac{\partial}{\partial p}\left[-\frac{f_0^2}{\sigma}\mathbf{V}_g \cdot \boldsymbol{\nabla}_h\left(-\frac{\partial\Phi}{\partial p}\right)\right]\end{aligned} \tag{6.64}$$

includes advection of the planetary vorticity, as does the QG PV equation,

$$\left(\frac{\partial}{\partial t} + \mathbf{V}_g \cdot \boldsymbol{\nabla}_h\right) q = \frac{D_g q}{Dt} = 0 \tag{6.65}$$

where the QG PV is given by

$$q \equiv \frac{1}{f_0}\nabla^2\Phi + f + \frac{\partial}{\partial p}\left(\frac{f_0}{\sigma}\frac{\partial\Phi}{\partial p}\right) \tag{6.66}$$

The Sutcliffe form of the QG omega (vertical motion) equation includes advection of the planetary vorticity

$$\left(\nabla^2 + \frac{f_0^2}{\sigma}\frac{\partial^2}{\partial p^2}\right)\omega \approx \frac{f_0}{\sigma}\left[\frac{\partial \mathbf{V}_g}{\partial p} \cdot \boldsymbol{\nabla}_h\left(\frac{1}{f_0}\nabla^2\Phi + f\right)\right] \tag{6.67}$$

Finally, the **Q**-vector form of the QG omega (vertical motion) equation is

$$\sigma\nabla^2\omega + f_0^2\frac{\partial^2\omega}{\partial p^2} = -2\boldsymbol{\nabla}\cdot\mathbf{Q} + f_0\beta\frac{\partial v_g}{\partial p} \tag{6.68}$$

where the **Q**-vector is given by

$$\mathbf{Q} \equiv (Q_1, Q_2) = \left(-\frac{R}{p}\frac{\partial \mathbf{V}_g}{\partial x}\cdot\boldsymbol{\nabla}T, -\frac{R}{p}\frac{\partial \mathbf{V}_g}{\partial y}\cdot\boldsymbol{\nabla}T\right) \tag{6.69}$$

Suggested References

lackburn. Blackburn, *Interpretation of Ageostrophic Winds and Implications for Jetstream Maintenance*, discusses the differences between variable *f*(VF) and constant *f*(CF) ageostrophic motion.

luestein. Bluestein, *Synoptic-Dynamic Meteorology in Midlatitudes, Vol. II*, has a comprehensive treatment of midlatitude synoptic disturbances at the graduate level.

unningham and Keyser, 2006. Cunningham and Keyser (2006) provide the theoretical basis for understanding jet streaks from a PV perspective.

urran and Snellman, 1987. Durran and Snellman (1987) illustrate the application of both traditional and **Q**-vector forms of the omega equation in diagnosing the vertical motion of an observed system.

akim. Hakim, Cyclogenesis in *Encyclopedia of Atmospheric Sciences*.

ackmann. Lackmann, *Midlatitude Synoptic Meteorology: Dynamics, Analysis, and Forecasting*, applies the quasi-geostrophic equations to the problem of weather analysis and forecasting.

artin. Martin, *Mid-Latitude Atmospheric Dynamics: A First Course*, covers similar material as in this chapter, with an emphasis on the development of extratropical cyclones.

edlosky. Pedlosky, *Geophysical Fluid Dynamics, 2nd Edition*, presents a detailed formal derivation of the quasi-geostrophic system with applications to both the atmosphere and the oceans.

allis. Vallis, *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*, presents a comprehensive treatment of the dynamics of both the atmosphere and ocean, including the quasi-geostrophic equations.

allace and Hobbs. Wallace and Hobbs, *Atmospheric Science: An Introductory Survey*, has an excellent introductory-level description of the observed structure and evolution of midlatitude synoptic-scale disturbances.

Problems

6.1. Show that for a basic-state having linear shear, $\overline{U} = \lambda z, \lambda$ constant, the basic-state QG PV depends only on *z*, and that the disturbance QG PV obeys

$$\left(\frac{\partial}{\partial t} + \overline{U}\frac{\partial}{\partial x}\right)q' = 0$$

6.2. Show that the Boussinesq continuity equation,

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0$$

can be written as

$$\frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y} + \frac{\partial w}{\partial z} = 0$$

where the subscript *a* indicates ageostrophic wind.

6.3. Prove identity (6.49).

6.4. In this problem you will derive an equation for the QG streamfunction of the ageostrophic secondary circulation. You should assume constant *f* and $\frac{d\bar{\theta}}{dz}$, and that the geostrophic flow is two-dimensional (*y* –*z*); specifically: $u_g = u_g(z,t)$ only and $v_g = v_g(y,t)$ only.

(a) Starting with geostrophic and hydrostatic balance,

$$\mathbf{V}_g = \frac{1}{\rho_0 f}\mathbf{k} \times \boldsymbol{\nabla}_h p \qquad \frac{\partial p}{\partial z} = \frac{\rho_0 g}{\bar{\theta}}\theta$$

show that the maintenance of thermal wind balance requires

$$\frac{D}{Dt_g}\frac{\partial u_g}{\partial z} = -\frac{g}{\bar{\theta} f}\frac{D}{Dt_g}\frac{\partial \theta}{\partial y}$$

(b) Determine the left side of the result in (a) from the u_g momentum equation.

$$\frac{Du_g}{Dt_g} = f v_a$$

Interpret the result physically.

(c) Determine the right side of the result in (a) from the thermodynamic energy equation. Interpret the result physically.

(d) Now use your results from (b) and (c) in (a), using a streamfunction for the ageostrophic wind: $v_a = \frac{\partial \psi}{\partial z}$ and $w = -\frac{\partial \psi}{\partial y}$. Express your result with all streamfunction terms on the left side of the equation.

(e) Assume that the right side of your result in (d) reaches a local minimum in the midtroposphere. Sketch the ageostrophic circulation streamlines in a *y* –*z* cross-section along with arrows showing v_a and *w*.

6.5. Consider two spherical, cyclonic QG PV anomalies. One anomaly is located at the origin and has unity radius. The second anomaly is located at (*x*, *y*, *z*) = (2, 0, 0) and has radius and amplitude one-half the corresponding values of the other anomaly. Draw an (*x*,*y*) sketch estimating the trajectory traced out by these two anomalies assuming that their shape is preserved. Draw a second sketch showing the *w* field at a level above the anomalies.

6.6.

(a) Derive the dispersion relationship for 3D QG Rossby waves on the **β**-plane. Assume plane waves in the absence of mean flow and boundaries. The governing (disturbance PV) equation is modified by an additional term

$$\frac{Dq}{Dt_g} + v\beta = 0$$

(b) Derive the group velocity vector. Explain the result.

6.7. Derive the QG kinetic energy equation

$$\frac{\partial K}{\partial t_g} + \nabla \cdot \vec{S} = \frac{g}{\theta_0} w\theta$$

starting with the QG momentum equation

$$\frac{D\mathbf{V}_g}{Dt_g} = -f\mathbf{k} \times \vec{V}_a$$

6.8. We have considered simple PV distributions associated with points and spheres. Now consider an extremely complicated PV structure, which is completely contained within a cube having sides of length L. Without knowing the detailed structure of the PV, derive a formula relating the circulation on the lateral sides of the cube, and the mean θ on the top and bottom of the cube, to the mean PV in the cube. [*Hint*: First show that the QG PV can be expressed as the divergence of a vector field.]

MATLAB Exercises

M6.1. Examine the impact of disturbance horizontal scale using the idealized QG PV inversion routine **QG_PV_inversion.m**. Use the parallelpiped initial condition ($ipv = 1$) with an amplitude of 2 PVU. Make a plot of how the disturbance minimum pressure changes as a function of the size of the disturbance (controlled by the parameter *iw* in the routine **QG_initial_value.m**). Repeat the exercise using a Dirichlet boundary condition (zero surface pressure; set $idn = -1$) and compare your results to the Neumann case. Repeat the exercise again by moving the boundaries very far apart (set *ZH* to 10 in **grid_setup.m**). Compare the pressure of all three solutions at a fixed distance below the PV anomaly (e.g., 5.5 km below the anomaly). What can you conclude about the sensitivity to boundary conditions for the Neumann and Dirichlet cases?

M6.2. Set the disturbance in the idealized QG PV inversion routine **QG_PV_ inversion.m** to a jet streak ($ipv = 5$) with an amplitude of 2 PVU. Compute the wind speed and make a plot of the wind speed and wind vectors (using *quiver*). Make a blocking pattern by changing the sign on the QG PV amplitude variable *pvmag* so that the dipole pattern reverses. Make a plot of wind speed and wind vectors including a constant zonal wind so that the total zonal wind speed between the vortices is zero. Perform piecewise QG PV inversion on the cyclonic and anticyclone vortices in isolation and, including the constant zonal wind, use PV Thinking to explain why the blocking pattern does not move.

M6.3. The code **QG_model.m** is a numerical solver of the quasi-geostrophic equations in $PV - \theta$ form. Initialize the model with a “blob” of PV ($ipv = 4$) in a simple linear-shear jet ($ijet = 1$). Run the model for 96 hours. Use the plotting code **plot_QG_model.m** to create figures for the solution at various times. Diagnose the solution using **QG_diagnostics.m**. From the solutions of the height tendency equation, describe the evolution of the surface pressure field. From the diagnosed vertical motion field, diagnose the vortex stretching term in the vorticity equation, and use the result to describe the evolution of the surface geostrophic vorticity field. Finally, one might expect the PV blob to tilt over in the linear shear. Check to see if this happens, and explain what you find.

M6.4. For the solution in problem M6.3, solve for the rotational ageostrophic wind defined by (6.22) with $p_a = 0$. Since this component of the wind has no divergence, we may express it in terms of a streamfuntion

$$\mathbf{V}_a = \mathbf{k} \times \nabla \psi_a \tag{6.70}$$

Derive an expression for the ageostrophic vorticity from (6.70) that can be used in (6.22). Plot the rotational ageostrophic wind using *quiver* at the level of the blob along with the pressure field. The sum of the geostrophic and the total ageostrophic wind (rotational plus divergent) provides the QG estimate of the gradient wind. How do the results compare with your expectations based on Section 3.2.5?

M6.5. Initialize **QG_model.m** with a blocking configuration (dipole with high PV over low PV) for the PV described in Problem M6.2. Use **QG_diagnostics.m** to determine favored locations for cloud and precipitation. How does the *w* pattern compare with a jet streak? Add a constant wind to the linear shear using parameter *Unot*, and note that your value should be non-dimensionalized using the scaling parameter, *U*. Pick a value for *Unot* that makes the block stationary. Next, average the QG PV and boundary potential temperature fields for the last few output times, and use the result as the initial condition for a new experiment with the value of *Unot* that makes the pattern stationary. How does the result differ from the first experiment? Repeat the experiment by adding a small-amplitude random field to the initial condition. What does the result tell you about why blocks are so persistent in the atmosphere?

---

[1] As Figure 6.6 reveals, in reality, the phase tilts tend to be concentrated below the 700-hPa level.

[2] Products of the functions of the dependent variables—in this case, disturbance vorticity and potential temperature.

[3] The nondimensional geostrophic wind is then given in terms of the nondimensional pressure by $\hat{\mathbf{V}}_g = -\hat{k} \times \hat{p}$.

[4] By the thermal-wind relationship, vertical shear of the geostrophic wind is proportional to the horizontal temperature gradient, so the Froude number measures the slope of the isentropes.

[5] This relation derives from relative vorticity, so that cyclones have the same sign as the local planetary rotation, *f*.

Chapter 7

Baroclinic Development

Chapter 6 showed that the quasi-geostrophic equations can qualitatively account for the observed relationships among the vorticity, temperature, and vertical velocity fields in midlatitude synoptic-scale systems. The diagnostic approach used in that chapter provided useful insights into the structure of synoptic-scale systems; it also demonstrated the key role of potential vorticity in dynamical analysis. It did not, however, provide quantitative information on the origins, growth rates, and propagation speeds of such disturbances. This chapter shows how the linear perturbation analysis of Chapter 5 can be used to obtain such information from the quasi-geostrophic equations.

The development of synoptic-scale weather disturbances is often referred to as *cyclogenesis*, a term that emphasizes the role of relative vorticity in developing synoptic-scale systems. This chapter analyzes the processes that lead to cyclogenesis. Specifically, we discuss the role of dynamical instability of the mean flow in accounting for the growth of synoptic-scale disturbances. Moreover, even in the absence of such instability, we show that baroclinic disturbances may still amplify over finite periods of time if properly configured. Thus the quasi-geostrophic equations can, indeed, provide a reasonable theoretical basis for understanding the development of synoptic-scale storms, although as discussed in Section 9.2, ageostrophic effects must be included to model the development of fronts and subsynoptic-scale storms.

7.1 Hydrodynamic Instability

A zonal-mean flow field is said to be hydrodynamically unstable if a small disturbance introduced into the flow grows spontaneously, drawing energy from the mean flow. It is useful to divide fluid instabilities into two types: parcel instability and wave instability. The simplest example of a parcel instability is the convective overturning that occurs when a fluid parcel is displaced vertically in a statically unstable fluid (see Section 2.7.3). Another example is inertial instability, which occurs when a parcel is displaced radially in an axisymmetric vortex with negative absolute vorticity in the Northern Hemisphere or positive absolute vorticity in the Southern Hemisphere. This instability in discussed in Section 5.5.1. A more general type of parcel instability, called *symmetric instability*, may also be significant in weather disturbances; this is discussed in Section 9.3.

Most of the instabilities of importance in meteorology, however, are associated with wave propagation The wave instabilities important for synoptic-scale meteorology generally occur in the form of zonally asymmetric perturbations to a zonally symmetric basic flow field. In general the basic flow is a jet stream that has both horizontal and vertical mean-flow shears. *Barotropic instability* is a wave instability associated with the horizontal shear in a jet-like current. Barotropic instabilities grow by extracting kinetic energy from the mean-flow field. *Baroclinic instability*, however, is associated with vertical shear of the mean flow. Baroclinic instabilities grow by converting potential energy associated with the mean horizontal temperature gradient that must exist to provide thermal wind balance for the vertical shear in the basic state flow. In neither of these instability types does the parcel method provide a satisfactory stability criterion. A more rigorous approach is required in which a linearized version of the governing equations is analyzed to determine the structure and amplification rate for the various wave modes supported by the system.

As indicated in Problem 5.2 of Chapter 5, the traditional approach to instability analysis is to assume that a small perturbation consisting of a single Fourier wave mode of the form $\exp[ik(x - ct)]$ is introduced into the flow, and to determine the conditions for which the phase velocity c has an imaginary part. This technique, which is called the *normal modes* method, is applied in the next section to analyze the stability of a baroclinic current.

An alternative method of instability analysis is the *initial value* approach. This method is motivated by the recognition that in general the perturbations from which storms develop cannot be described as single normal mode disturbances, but may have a complex structure. The initial growth of such disturbances may strongly depend on the potential vorticity distribution in the initial disturbance. On the time scale of a day or two, such growth can be quite different from that of a normal mode of similar scale, although in the absence of nonlinear interactions the fastest-growing normal mode disturbance must eventually dominate.

A strong dependence of cyclogenesis on initial conditions occurs when a large-amplitude upper-level potential vorticity anomaly is advected into a region where there is a preexisting meridional temperature gradient at the surface. In that case, as shown schematically in Figure 7.1, the circulation induced by the upper-level anomaly (which extends downward, as discussed in Section 6.6) leads to temperature advection at the surface; this induces a potential vorticity anomaly near the surface, which in turn reinforces the upper-level anomaly. Under some conditions the surface and upper-level potential vorticity anomalies can become locked in phase so that the induced circulations produce a very rapid amplification of the anomaly pattern. Detailed discussion of the initial value approach to cyclogenesis is beyond the scope of this text. Here we concentrate primarily on the simplest normal mode instability models.

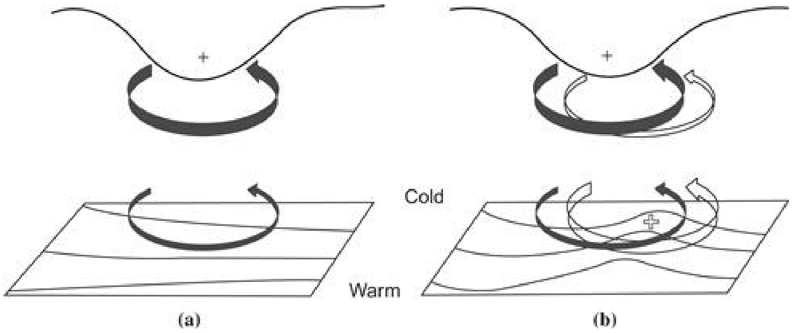


Figure 7.1 A schematic picture of cyclogenesis associated with the arrival of an upper-level potential vorticity perturbation over a lower-level baroclinic region. (a) Lower-level cyclonic vorticity induced by the upper-level potential vorticity anomaly. The circulation induced by the potential vorticity anomaly is shown by the *solid arrows*, and potential temperature contours are shown at the lower boundary. The advection of potential temperature by the induced lower-level circulation leads to a warm anomaly slightly east of the upper-level vorticity anomaly. This in turn will induce a cyclonic circulation as shown by the *open arrows* in (b). The induced upper-level circulation will reinforce the original upper-level anomaly and can lead to amplification of the disturbance.

After Hoskins et al., 1985.

7.2 Normal Mode Baroclinic Instability: A Two-Layer Model

Even for a highly idealized mean-flow profile, the mathematical treatment of baroclinic instability in a continuously stratified atmosphere is rather complicated. Before considering such a model, we first focus on the simplest model that can incorporate baroclinic processes. The atmosphere is represented by two discrete layers bounded by surfaces numbered 0, 2, and 4 (generally taken to be the 0-, 500-, and 1000-hPa surfaces, respectively) as shown in Figure 7.2. The quasi-geostrophic vorticity equation for the midlatitude β plane is applied at the 250- and 750-hPa levels, designated by 1 and 3 in Figure 7.2, whereas the thermodynamic energy equation is applied at the 500-hPa level, designated by 2 in Figure 7.2.

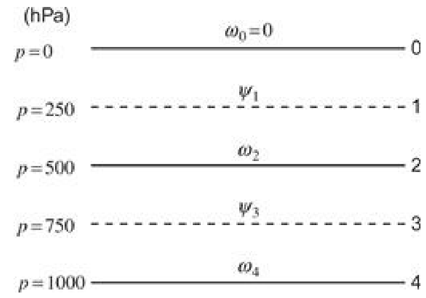


Figure 7.2 Arrangement of variables in the vertical for the two-level baroclinic model.

Before writing out the specific equations of the two-layer model, it is convenient to define a *geostrophic streamfunction*, $\psi \equiv \Phi/f_0$, for the isobaric form of the quasi-geostrophic equations. Then the geostrophic wind and the geostrophic vorticity can be expressed respectively as

$$\mathbf{V}_\psi = \mathbf{k} \times \boldsymbol{\nabla}_h \psi, \quad \zeta_g = \nabla_h^2 \psi \tag{7.1}$$

The quasi-geostrophic vorticity equation and the hydrostatic thermodynamic energy equation can then be written in terms of ψ and ω as

$$\frac{\partial}{\partial t}\nabla_h^2\psi + \mathbf{V}_\psi \cdot \boldsymbol{\nabla}_h\left(\nabla_h^2\psi\right) + \beta\frac{\partial\psi}{\partial x} = f_0\frac{\partial\omega}{\partial p} \tag{7.2}$$

$$\frac{\partial}{\partial t}\left(\frac{\partial\psi}{\partial p}\right) = -\mathbf{V}_\psi \cdot \boldsymbol{\nabla}_h\left(\frac{\partial\psi}{\partial p}\right) - \frac{\sigma}{f_0}\omega \tag{7.3}$$

We now apply the vorticity equation (7.2) at the two levels designated as 1 and 3, which are at the middle of the two layers. To do this we must estimate the divergence term $\partial\omega/\partial p$ at these levels using finite difference approximations to the vertical derivatives:

$$\left(\frac{\partial\omega}{\partial p}\right)_1 \approx \frac{\omega_2 - \omega_0}{\delta p}, \quad \left(\frac{\partial\omega}{\partial p}\right)_3 \approx \frac{\omega_4 - \omega_2}{\delta p} \tag{7.4}$$

where $\delta p = 500$ hPa is the pressure interval between levels 0 to 2 and 2 to 4, and subscript notation is used to designate the vertical level for each dependent variable. The resulting vorticity equations are

$$\frac{\partial}{\partial t}\nabla_h^2\psi_1 + \mathbf{V}_1 \cdot \nabla\left(\nabla_h^2\psi_1\right) + \beta\frac{\partial\psi_1}{\partial x} = \frac{f_0}{\delta p}\omega_2 \tag{7.5}$$

$$\frac{\partial}{\partial t}\nabla_h^2\psi_3 + \mathbf{V}_3 \cdot \boldsymbol{\nabla}_h\left(\nabla_h^2\psi_3\right) + \beta\frac{\partial \psi_3}{\partial x} = -\frac{f_0}{\delta p}\omega_2 \tag{7.6}$$

where we have used the fact that $\omega_0 = 0$, and assumed that $\omega_4 = 0$, which is approximately true for a level-lower boundary surface.

We next write the thermodynamic energy equation (7.3) at level 2. Here we must evaluate $\partial\psi/\partial p$ using the difference formula

$$\left(\partial\psi/\partial p\right)_2 \approx \left(\psi_3 - \psi_1\right)/\delta p$$

The result is

$$\frac{\partial}{\partial t}\left(\psi_1 - \psi_3\right) = -\mathbf{V}_2 \cdot \boldsymbol{\nabla}_h(\psi_1 - \psi_3) + \frac{\sigma\delta p}{f_0}\omega_2 \tag{7.7}$$

The first term on the right-hand side in (7.7) is the advection of the 250- to 750-hPa thickness by the wind at 500 hPa. However, ψ_2, the 500-hPa streamfunction, is not a predicted field in this model. Therefore, ψ_2 must be obtained by linearly interpolating between the 250- and 750-hPa levels:

$$\psi_2 = (\psi_1 + \psi_3)/2$$

If this interpolation formula is used, (7.5), (7.6), and (7.7) become a closed set of prediction equations in the variables ψ_1, ψ_3, and ω_2.

7.2.1 Linear Perturbation Analysis

To keep the analysis as simple as possible, we assume that the streamfunctions ψ_1 and ψ_3 consist of basic state parts that depend linearly on y alone, plus perturbations that depend only on x and t. Thus, we let

$$\begin{aligned}\psi_1 &= -U_1 y + \psi_1'(x,t)\\ \psi_3 &= -U_3 y + \psi_3'(x,t)\\ \omega_2 &= \omega_2'(x,t)\end{aligned} \tag{7.8}$$

The zonal velocities at levels 1 and 3 are then constants with the values U_1 and U_3, respectively. Hence, the perturbation field has meridional and vertical velocity components only.

Substituting from (7.8) into (7.5), (7.6), and (7.7) and linearizing yields the perturbation equations

$$\left(\frac{\partial}{\partial t} + U_1\frac{\partial}{\partial x}\right)\frac{\partial^2\psi_1'}{\partial x^2} + \beta\frac{\partial\psi_1'}{\partial x} = \frac{f_0}{\delta p}\omega_2' \tag{7.9}$$

$$\left(\frac{\partial}{\partial t} + U_3\frac{\partial}{\partial x}\right)\frac{\partial^2\psi_3'}{\partial x^2} + \beta\frac{\partial\psi_3'}{\partial x} = -\frac{f_0}{\delta p}\omega_2' \tag{7.10}$$

$$\left(\frac{\partial}{\partial t}+U_m\frac{\partial}{\partial x}\right)(\psi_1'-\psi_3')-U_T\frac{\partial}{\partial x}(\psi_1'+\psi_3')=\frac{\sigma\delta p}{f_0}\omega_2' \tag{7.11}$$

where we have linearly interpolated to express $\mathbf{V}_2$ in terms of ψ_1 and ψ_3, and have defined

$$U_m\equiv(U_1+U_3)/2,\quad U_T\equiv(U_1-U_3)/2$$

Thus, U_m and U_T are, respectively, the vertically averaged mean zonal wind and the mean thermal wind.

The dynamical properties of this system are more clearly expressed if (7.9), (7.10), and (7.11) are combined to eliminate ω_2'. We first note that (7.9) and (7.10) can be rewritten as

$$\left[\frac{\partial}{\partial t}+(U_m+U_T)\frac{\partial}{\partial x}\right]\frac{\partial^2\psi_1'}{\partial x^2}+\beta\frac{\partial\psi_1'}{\partial x}=\frac{f_0}{\delta p}\omega_2' \tag{7.12}$$

$$\left[\frac{\partial}{\partial t}+(U_m-U_T)\frac{\partial}{\partial x}\right]\frac{\partial^2\psi_3'}{\partial x^2}+\beta\frac{\partial\psi_3'}{\partial x}=-\frac{f_0}{\delta p}\omega_2' \tag{7.13}$$

We now define the barotropic and baroclinic perturbations as

$$\psi_m\equiv(\psi_1'+\psi_3')/2;\quad \psi_T\equiv(\psi_1'-\psi_3')/2 \tag{7.14}$$

Adding (7.12) and (7.13) and using the definitions in (7.14) yield

$$\left[\frac{\partial}{\partial t}+U_m\frac{\partial}{\partial x}\right]\frac{\partial^2\psi_m}{\partial x^2}+\beta\frac{\partial\psi_m}{\partial x}+U_T\frac{\partial}{\partial x}\left(\frac{\partial^2\psi_T}{\partial x^2}\right)=0 \tag{7.15}$$

while subtracting (7.13) from (7.12) and combining with (7.11) to eliminate ω_2' yield

$$\begin{aligned}\left[\frac{\partial}{\partial t}+U_m\frac{\partial}{\partial x}\right]\left(\frac{\partial^2\psi_T}{\partial x^2}-2\lambda^2\psi_T\right)+\beta\frac{\partial\psi_T}{\partial x}\\+U_T\frac{\partial}{\partial x}\left(\frac{\partial^2\psi_m}{\partial x^2}+2\lambda^2\psi_m\right)=0\end{aligned} \tag{7.16}$$

where $\lambda^2\equiv f_0^2/[\sigma(\delta p)^2]$. Equations (7.15) and (7.16) govern the evolution of the barotropic (vertically averaged) and baroclinic (thermal) perturbation vorticities, respectively.

As in Chapter 5, we assume that wave-like solutions exist of the form

$$\psi_m=Ae^{ik(x-ct)},\quad \psi_T=Be^{ik(x-ct)} \tag{7.17}$$

Substituting these assumed solutions into (7.15) and (7.16) and dividing through by the common exponential factor, we obtain a pair of simultaneous linear algebraic equations for the coefficients A, B:

$$ik\left[(c-U_m)k^2+\beta\right]A-ik^3U_TB=0 \tag{7.18}$$

$$ik\left[(c-U_m)\left(k^2+2\lambda^2\right)+\beta\right]B-ikU_T\left(k^2-2\lambda^2\right)A=0 \tag{7.19}$$

Because this set is homogeneous, nontrivial solutions will exist only if the determinant of the coefficients of A and B is zero. Thus, the phase speed c must satisfy the condition

$$\begin{vmatrix} (c-U_m)\,k^2+\beta & -k^2U_T \\ -U_T\left(k^2-2\lambda^2\right) & (c-U_m)\left(k^2+2\lambda^2\right)+\beta \end{vmatrix}=0$$

which gives a quadratic dispersion equation in c:

$$\begin{aligned}&\left(c-U_m\right)^2k^2(k^2+2\lambda^2)+2\left(c-U_m\right)\beta(k^2+\lambda^2)\\&\quad+\left[\beta^2+U_T^2k^2(2\lambda^2-k^2)\right]=0\end{aligned} \tag{7.20}$$

which is analogous to the linear wave dispersion equations developed in Chapter 5. The dispersion relationship in (7.20) yields for the phase speed

$$c=U_m-\frac{\beta\left(k^2+\lambda^2\right)}{k^2(k^2+2\lambda^2)}\pm\delta^{1/2} \tag{7.21}$$

where

$$\delta\equiv\frac{\beta^2\lambda^4}{k^4(k^2+2\lambda^2)^2}-\frac{U_T^2(2\lambda^2-k^2)}{(k^2+2\lambda^2)}$$

We have now shown that (7.17) is a solution for the system (7.15) and (7.16) only if the phase speed satisfies (7.21). Although (7.21) appears to be rather complicated, it is immediately apparent that if $\delta < 0$, the phase speed will have an imaginary part and the perturbations will amplify exponentially. Before discussing the general physical conditions required for exponential growth, it is useful to consider two special cases.

As the first special case we let $U_T = 0$ so that the basic state thermal wind vanishes and the mean flow is barotropic. The phase speeds in this case are

$$c_1=U_m-\beta k^{-2} \tag{7.22}$$

and

$$c_2=U_m-\beta\left(k^2+2\lambda^2\right)^{-1} \tag{7.23}$$

These are real quantities that correspond to the free (normal mode) oscillations for the two-level model with a barotropic basic state current. The phase speed c_1 is simply the dispersion relationship for a barotropic Rossby wave with no y dependence (see Section 5.7). Substituting from (7.22) for c in (7.18) and (7.19), we see that in this case $B = 0$ so that the perturbation is barotropic in structure. The expression of (7.23), however, may be interpreted as the phase speed for an internal baroclinic Rossby wave. Note that c_2 is a dispersion relationship analogous to the Rossby wave speed for a homogeneous ocean with a free surface, which was given in Problem 7.16.

However, in the two-level model, the factor $2\lambda^2$ appears in the denominator in place of the f_0^2/gH for the oceanic case.

In each of these cases there is vertical motion associated with the Rossby wave so that static stability modifies the wave speed. It is left as a problem for the reader to show that if c_2 is substituted into (7.18) and (7.19), the resulting fields of ψ_1 and ψ_3 are 180° out of phase so that the perturbation is baroclinic, although the basic state is barotropic. Furthermore, the ω_2' field is 1/4 cycle out of phase with the 250-hPa geopotential field, with the maximum upward motion occurring west of the 250-hPa trough.

This vertical motion pattern may be understood if we note that $c_2 - U_m < 0$, so that the disturbance pattern moves westward *relative to the mean wind*. Now, viewed in a coordinate system moving with the mean wind, the vorticity changes are due only to the planetary vorticity advection and the convergence terms, while the thickness changes must be caused solely by the adiabatic heating or cooling due to vertical motion. Thus, there must be rising motion west of the 250-hPa trough to produce the thickness changes required by the westward motion of the system.

Comparing (7.22) and (7.23), we see that the phase speed of the baroclinic mode is generally much less than that of the barotropic mode since for average midlatitude tropospheric conditions $\lambda^2 \approx 2 \times 10^{-12}\,\mathrm{m}^{-2}$, which is approximately equal to k^2 for a zonal wavelength of 4300 km.[1]

As the second special case, we assume that $\beta = 0$. This case corresponds, for example, to a laboratory situation in which the fluid is bounded above and below by rotating horizontal planes so that the gravity and rotation vectors are everywhere parallel. In such a situation

$$c = U_m \pm U_T \left(\frac{k^2 - 2\lambda^2}{k^2 + 2\lambda^2} \right)^{1/2} \tag{7.24}$$

For waves with zonal wave numbers satisfying $k^2 < 2\lambda^2$, (7.24) has an imaginary part. Thus, all waves longer than the critical wavelength $L_c = \sqrt{2}\pi/\lambda$ will amplify. From the definition of λ we can write

$$L_c = \delta p \pi (2\sigma)^{1/2} / f_0$$

For typical tropospheric conditions $(2\sigma)^{1/2} \approx 2 \times 10^{-3}\,\mathrm{N}^{-1}\,\mathrm{m}^3\,\mathrm{s}^{-1}$. Thus, with $\delta_p = 500$ hPa and $f_0 = 10^{-4}\,\mathrm{s}^{-1}$, we find that $L_c \approx 3000$ km. It is also clear from this formula that the critical wavelength for baroclinic instability increases with static stability. The role of static stability in stabilizing the shorter waves can be understood qualitatively as follows: For a sinusoidal perturbation, the relative vorticity, and thus the differential vorticity advection, increases with the square of the wave number. However, as shown in Chapter 6, a secondary vertical circulation is required to maintain hydrostatic temperature changes and geostrophic vorticity changes in the presence of differential vorticity advection. Thus, for a geopotential perturbation of fixed amplitude, the relative strength of the accompanying vertical circulation must increase as the wavelength of the disturbance decreases. Because static stability tends to resist vertical displacements, the shortest wavelengths will thus be stabilized.

It is also of interest that with $\beta = 0$ the critical wavelength for instability does not depend on the magnitude of the basic state thermal wind U_T. The growth rate, however, does depend on U_T. According to (7.17) the time dependence of the disturbance solution has the form $\exp(-ikct)$. Thus, the exponential growth rate is $\alpha = kc_i$, where c_i designates the imaginary part of the phase speed. In the present case

$$\alpha = kU_T\left(\frac{2\lambda^2-k^2}{2\lambda^2+k^2}\right)^{1/2} \tag{7.25}$$

so that the growth rate increases linearly with the mean thermal wind.

Returning to the general case where all terms are retained in (7.21), the stability criterion is understood most easily by computing the *neutral curve*, which connects all values of U_T and k for which $\delta = 0$ so that the flow is *marginally stable*. From (7.21), the condition $\delta = 0$ implies that

$$\frac{\beta^2\lambda^4}{k^4\left(2\lambda^2+k^2\right)} = U_T^2\left(2\lambda^2-k^2\right) \tag{7.26}$$

This complicated relationship between U_T and k can best be displayed by solving (7.26) for $k^4/2\lambda^4$, yielding

$$k^4/(2\lambda^4) = 1 \pm \left[1-\beta^2/(4\lambda^4U_T^2)\right]^{1/2}$$

In Figure 7.3 the nondimensional quantity $k^2/2\lambda^2$, which is proportional to the square of the zonal wave number, is plotted against the nondimensional parameter $2\lambda^2U_T/\beta$, which is proportional to the thermal wind. As indicated in Figure 7.3, the neutral curve separates the unstable region of the U_T, k plane from the stable region. It is clear that the inclusion of the β effect serves to stabilize the flow, since now unstable roots exist only for $|U_T| > \beta/(2\lambda^2)$. In addition, the minimum value of U_T required for unstable growth depends strongly on k. Thus, the β effect strongly stabilizes the long wave end of the wave spectrum ($k \to 0$). Again, the flow is always stable for waves shorter than the critical wavelength $L_c = \sqrt{2}\pi/\lambda$.

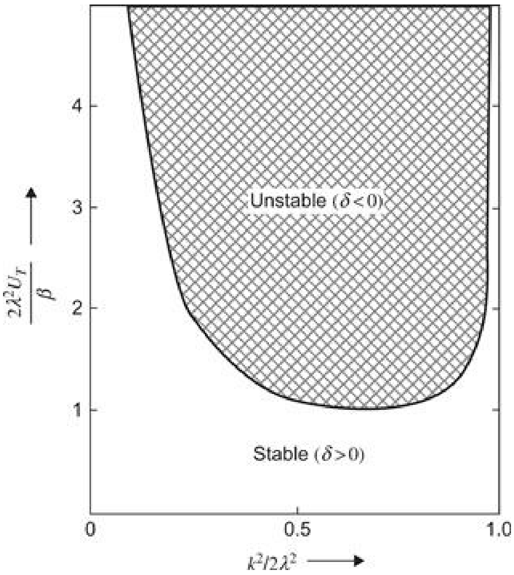


Figure 7.3 Neutral stability curve for the two-level baroclinic model.

This long-wave stabilization associated with the β effect is caused by the rapid westward propagation of long waves (i.e., Rossby wave propagation), which occurs only when the β effect is included in the model. It can be shown that baroclinically unstable waves always propagate at a speed that lies between maximum and minimum mean zonal wind speeds. Thus, for the two-level model in the usual midlatitude case where $U_1 > U_3 > 0$, the real part of the phase speed satisfies the inequality $U_3 < c_r < U_1$ for unstable waves. In a continuous atmosphere, this would imply that there must be a level where $U = c_r$. Such a level is called a *critical level* by theoreticians and a *steering level* by synopticians. For long waves and weak basic state wind shear, the solution given by (7.21) will have $c_r < U_3$, there is no steering level, and unstable growth cannot occur.

Differentiating (7.26) with respect to k and setting $dU_T/dk = 0$, we find that the minimum value of U_T for which unstable waves may exist occurs when $k^2 = \sqrt{2}\lambda^2$. This wave number corresponds to the wave of maximum instability. Wave numbers of observed disturbances should be close to the wave number of maximum instability, because if U_T were gradually raised from zero, the flow would first become unstable for perturbations of wave number $k = 2^{1/4}\lambda$. Those perturbations would then amplify and in the process remove energy from the mean thermal wind, thereby decreasing U_T and stabilizing the flow.

Under normal conditions of static stability the wavelength of maximum instability is about 4000 km, which is close to the average wavelength for midlatitude synoptic systems. Furthermore, the thermal wind required for marginal stability at this wavelength is only about $U_T \approx 4\ \mathrm{m\,s^{-1}}$, which implies a shear of $8\ \mathrm{m\,s^{-1}}$ between 250 and 750 hPa. Shears greater than this are certainly common in middle latitudes for the zonally averaged flow. Therefore, the observed behavior of midlatitude synoptic systems is consistent with the hypothesis that such systems can originate from infinitesimal perturbations of a baroclinically unstable basic current. Of course, in the real atmosphere many other factors may influence the development of synoptic systems—for example, due to lateral shear in the jet stream, nonlinear interactions of finite-amplitude perturbations, and the release of latent heat in precipitating systems. However, observational studies, laboratory simulations, and numerical models all suggest that baroclinic instability is a primary mechanism for synoptic-scale wave development in middle latitudes.

7.2.2 Vertical Motion in Baroclinic Waves

Since the two-level model is a special case of the quasi-geostrophic system, the physical mechanisms responsible for forcing vertical motion should be those discussed in Section 6.5. Thus, the forcing of vertical motion can be expressed in terms of the sum of the forcing by thermal advection (evaluated at level 2) plus the differential vorticity advection (evaluated as the difference between the vorticity advection at level 1 and that at level 3). Alternatively, the forcing of vertical motion can be expressed in terms of the divergence of the **Q** vector.

The **Q**-vector form of the omega equation for the two-level model can be derived simply from (6.68). We first estimate the second term on the left-hand side by finite differencing in p. Using (7.4) and again letting $\omega_0 = \omega_4 = 0$, we obtain

$$\frac{\partial^2 \omega}{\partial p^2} \approx \frac{(\partial \omega/\partial p)_3 - (\partial \omega/\partial p)_1}{\delta p} \approx -\frac{2\omega_2}{(\delta p)^2}$$

and observe that temperature in the two-level model is represented as

$$\frac{\partial^2\omega}{\partial p^2} \approx \frac{(\partial\omega/\partial p)_3 - (\partial\omega/\partial p)_1}{\delta p} \approx -\frac{2\omega_2}{(\delta p)^2}$$

Thus, (6.68) becomes

$$\sigma\left(\nabla_h^2 - 2\lambda^2\right)\omega_2 = -2\boldsymbol{\nabla}_h \cdot \mathbf{Q} \tag{7.27}$$

where

$$\mathbf{Q} = \frac{f_0}{\delta p}\left[-\frac{\partial \mathbf{V_2}}{\partial x} \cdot \boldsymbol{\nabla}_h\left(\psi_1 - \psi_3\right),\; -\frac{\partial \mathbf{V_2}}{\partial y} \cdot \boldsymbol{\nabla}_h\left(\psi_1 - \psi_3\right)\right]$$

In order to examine the forcing of vertical motion in baroclinically unstable waves, we linearize (7.27) by specifying the same basic state and perturbation variables as in (7.8). For this situation, in which the mean zonal wind and the perturbation streamfunctions are independent of *y*, the **Q** vector has only an *x* component:

$$Q_1 = \frac{f_0}{\delta p}\left[\frac{\partial^2\psi_2'}{\partial x^2}\left(U_1 - U_3\right)\right] = \frac{2f_0}{\delta p}U_T\zeta_2'$$

The pattern of the **Q** vector in this case is similar to that of Figure 6.14, with eastward-pointing **Q** centered at the trough and westward-pointing **Q** centered at the ridge. This is consistent with the fact that **Q** represents the change of temperature gradient forced by geostrophic motion alone. In this simple model the temperature gradient is entirely due to the vertical shear of the mean zonal wind [$U_T \propto -\partial T/\partial y$], and the shear of the perturbation meridional velocity tends to advect warm air poleward east of the 500-hPa trough and cold air equatorward west of the 500-hPa trough so that there is a tendency to produce a component of temperature gradient directed eastward at the trough.

The forcing of vertical motion by the **Q** vector in the linearized model is, from (7.27),

$$\left(\frac{\partial^2}{\partial x^2} - 2\lambda^2\right)\omega_2' = -\frac{4f_0}{\sigma\delta p}U_T\frac{\partial\zeta_2'}{\partial x} \tag{7.28}$$

Observing that

$$\left(\frac{\partial^2}{\partial x^2} - 2\lambda^2\right)\omega_2' \propto -\omega_2'$$

we may interpret (7.28) physically by noting that

$$w_2' \propto -\omega_2' \propto -U_T\frac{\partial\zeta_2'}{\partial x} \propto -v_2'\frac{\partial\overline{T}}{\partial y}$$

Thus, sinking motion is forced by negative advection of disturbance vorticity by the basic state thermal wind (or, alternatively, cold advection of the basic state thermal field by the perturbation meridional wind), while rising motion is forced by advection of the opposite sign.

We now have the information required to diagram the structure of a baroclinically unstable disturbance in the two-level model. The lower part of Figure 7.4 shows schematically the phase relationship between the geopotential field and the divergent secondary motion field for the usual midlatitude situation where $U_T > 0$. Linear interpolation has been used between levels so that the trough and ridge axes are straight lines tilted back toward the west with height. In this example the ψ_1 field lags the ψ_3 field by about 65° in phase so that the trough at 250 hPa lies 65° in phase west of the 750-hPa trough. At 500 hPa the perturbation thickness field lags the geopotential field by one-quarter wavelength as shown in the top part of Figure 7.4, and the thickness and vertical motion fields are in phase. Note that the temperature advection by the perturbation meridional wind is in phase with the 500-hPa thickness field so that the advection of the basic state temperature by the perturbation wind acts to intensify the perturbation thickness field. This tendency is also illustrated by the zonally oriented **Q** vectors shown at the 500-hPa level in Figure 7.4.

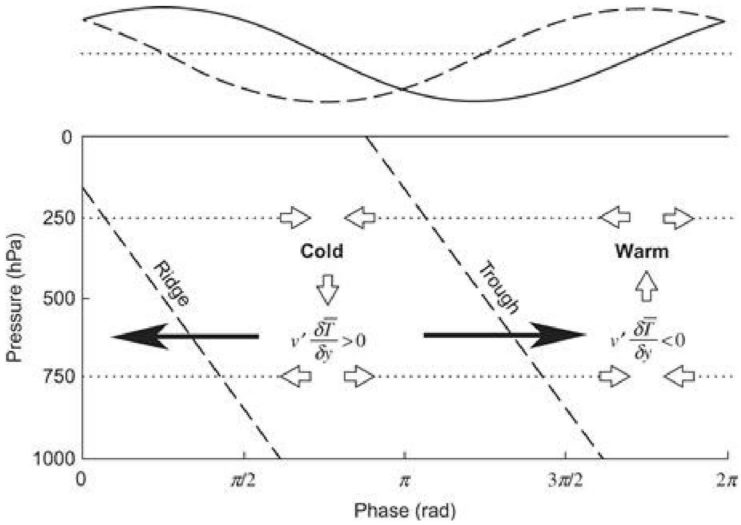

0
250
500
750
1000
Pressure (hPa)
Ridge
Trough
Cold
Warm
$v'\frac{\delta \overline{T}}{\delta y}>0$
$v'\frac{\delta \overline{T}}{\delta y}<0$
0
π/2
π
3π/2
2π
Phase (rad)

Figure 7.4 Structure of an unstable baroclinic wave in the two-level model. (*Top*) Relative phases of the 500-hPa perturbation geopotential (*solid line*) and temperature (*dashed line*). (*Bottom*) Vertical cross-section showing phases of geopotential, meridional temperature advection, ageostrophic circulation (*open arrows*), Q vectors (*solid arrows*), and temperature fields for an unstable baroclinic wave in the two-level model.

The pattern of vertical motion forced by the divergence of the **Q** vector, as shown in Figure 7.4, is associated with a divergence–convergence pattern that contributes a positive vorticity tendency near the 250-hPa trough and a negative vorticity tendency near the 750-hPa ridge, with opposite tendencies at the 250-hPa ridge and 750-hPa trough. Since in all cases these vorticity tendencies tend to increase the extreme values of vorticity at the troughs and ridges, this secondary circulation system will act to increase the strength of the disturbance.

The total vorticity change at each level is, of course, determined by the sum of vorticity advection and vortex stretching due to the divergent circulation. The relative contributions of these processes are indicated schematically in Figures 7.5 and 7.6, respectively. As can be seen in Figure 7.5, vorticity advection leads the vorticity field by one-quarter wavelength. Since in this case the basic state wind increases with height, the vorticity advection at 250 hPa is larger than that at 750 hPa. If no other processes influenced the vorticity field, the effect of this differential vorticity advection would be to move the upper-level trough and ridge pattern eastward more rapidly than the lower-level pattern. Thus, the westward tilt of the trough–ridge pattern would quickly be destroyed. The maintenance of this tilt in the presence of differential vorticity advection is due to the concentration of vorticity by vortex stretching associated with the divergent secondary circulation.

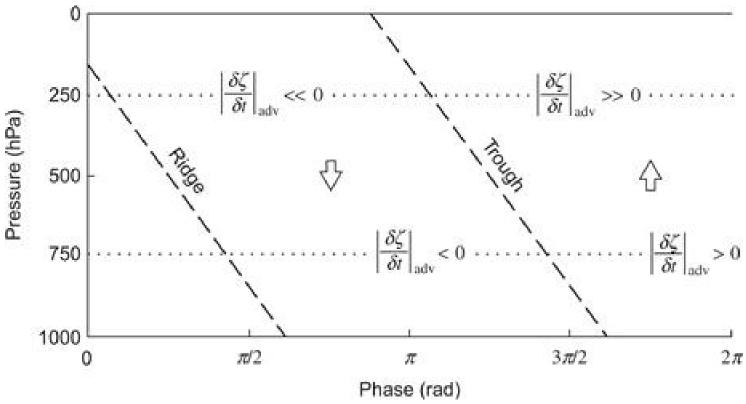

0
250
500
750
1000
Pressure (hPa)
$\left|\frac{\delta\zeta}{\delta t}\right|_{\mathrm{adv}} << 0$
$\left|\frac{\delta\zeta}{\delta t}\right|_{\mathrm{adv}} >> 0$
Ridge
Trough
$\left|\frac{\delta\zeta}{\delta t}\right|_{\mathrm{adv}} < 0$
$\left|\frac{\delta\zeta}{\delta t}\right|_{\mathrm{adv}} > 0$
0
π/2
π
3π/2
2π
Phase (rad)

Figure 7.5 Vertical cross-section showing the phase of vorticity change due to vorticity advection for an unstable baroclinic wave in the two-level model.

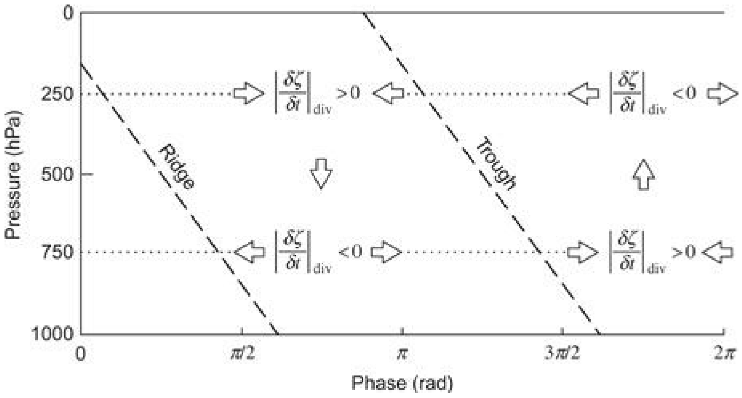

0
250
500
750
1000
Pressure (hPa)
$\left|\frac{\delta\zeta}{\delta t}\right|_{\mathrm{div}} >0$
$\left|\frac{\delta\zeta}{\delta t}\right|_{\mathrm{div}} <0$
Ridge
Trough
$\left|\frac{\delta\zeta}{\delta t}\right|_{\mathrm{div}} <0$
$\left|\frac{\delta\zeta}{\delta t}\right|_{\mathrm{div}} >0$
0
π/2
π
3π/2
2π
Phase (rad)

Figure 7.6 Vertical cross-section showing the phase of vorticity change due to divergence–convergence for an unstable baroclinic wave in the two-level model.

Referring to Figure 7.6, we see that concentration of vorticity by the divergence effect lags the vorticity field by about 65° at 250 hPa and leads the vorticity field by about 65° at 750 hPa. As a result, the net vorticity tendencies ahead of the vorticity maxima and minima are less than the advective tendencies at the upper level and greater than the advective tendencies at the lower level:

$$\left|\frac{\partial\zeta_1'}{\partial t}\right| < \left|\frac{\partial\zeta_1'}{\partial t}\right|_{\mathrm{adv}}, \quad \left|\frac{\partial\zeta_3'}{\partial t}\right| > \left|\frac{\partial\zeta_3'}{\partial t}\right|_{\mathrm{adv}}$$

Furthermore, vorticity concentration by the divergence effect will tend to amplify the vorticity perturbations in the troughs and ridges at both the 250- and 750-hPa levels as required for a growing disturbance.

7.3 The Energetics of Baroclinic Waves

The previous section showed that under suitable conditions a vertically sheared geostrophically balanced basic state flow is unstable to small wave-like perturbations with horizontal wavelengths in the range of observed synoptic-scale systems. Such baroclinically unstable perturbations will amplify exponentially by drawing energy from the mean flow. This section considers the energetics of linearized baroclinic disturbances and shows that the potential energy of the mean flow is the energy source for baroclinically unstable perturbations.

7.3.1 Available Potential Energy

Before discussing the energetics of baroclinic waves, it is necessary to consider the energy of the atmosphere from a more general point of view. For all practical purposes, the total energy of the atmosphere is the sum of internal energy, gravitational potential energy, and kinetic energy. However, it is not necessary to consider separately the variations of internal and gravitational potential energy because in a hydrostatic atmosphere these two forms of energy are proportional and may be combined into a single term called the *total potential energy*. The proportionality of internal and gravitational potential energy can be demonstrated by considering these forms of energy for a column of air of unit horizontal cross-section extending from the surface to the top of the atmosphere.

If we let dE_I be the internal energy in a vertical section of the column of height *dz*, then from the definition of internal energy—see Section (2.6):

$$dE_I = \rho c_v T dz$$

so that the internal energy for the entire column is

$$E_I = c_v \int_0^\infty \rho T dz \tag{7.29}$$

However, the gravitational potential energy for a slab of thickness *dz* at a height *z* is just

$$dE_P = \rho g z dz$$

so that the gravitational potential energy in the entire column is

$$E_P = \int_0^\infty \rho g z dz = -\int_{p_0}^0 z dp \tag{7.30}$$

where we have substituted from the hydrostatic equation to obtain the last integral in (7.30). Integrating (7.30) by parts and using the ideal gas law, we obtain

$$E_P = \int_0^\infty p dz = R \int_0^\infty \rho T dz \tag{7.31}$$

Comparing (7.29) and (7.31), we see that $c_v E_P = R E_I$. Thus, the total potential energy may be expressed as

$$E_P + E_I = (c_p/c_v)\,E_I = (c_p/R)\,E_P \tag{7.32}$$

Therefore, in a hydrostatic atmosphere the total potential energy can be obtained by computing either E_I or E_P alone.

The total potential energy is not a very useful measure of energy in the atmosphere because only a very small fraction of the total potential energy is available for conversion to kinetic energy in storms. To demonstrate qualitatively why most of the total potential energy is unavailable, we consider a simple model atmosphere that initially consists of two equal masses of dry air separated by a vertical partition as shown in Figure 7.7. The two air masses are at uniform potential temperatures θ_1 and θ_2, respectively, with $\theta_1 < \theta_2$. The ground-level pressure on each side of the partition is taken to be 1000 hPa. We wish to compute the maximum kinetic energy that can be realized by an adiabatic rearrangement of mass within the same volume when the partition is removed.

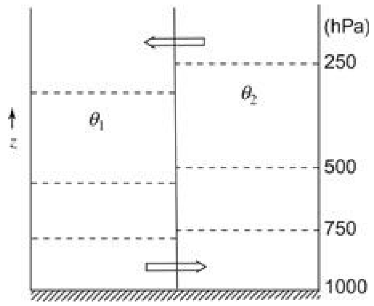


Figure 7.7 Two air masses of differing potential temperature separated by a vertical partition. *Dashed lines* indicate isobaric surfaces. *Arrows* show direction of motion when the partition is removed.

Now for an adiabatic process, total energy is conserved:

$$E_K + E_P + E_I = \text{constant}$$

where E_K denotes the kinetic energy. If the air masses are initially at rest, $E_K = 0$. Thus, if we let primed quantities denote the final state

$$E_K' + E_P' + E_I' = E_P + E_I$$

so that with the aid of (7.32) the kinetic energy realized by removal of the partition may be expressed as

$$E_K' = (c_p/c_v)\,(E_I - E_I')$$

Because θ is conserved for an adiabatic process, the two air masses cannot mix. It is clear that E_I' will be a minimum (designated by E_I'') when the masses are rearranged so that the air at potential temperature θ_1 lies entirely beneath the air at potential temperature θ_2, with the 500-hPa surface as the horizontal boundary between the two masses. In that case the total potential energy $(c_p/c_v)E_I''$ is not available for conversion to kinetic energy because no adiabatic process can further reduce E_I''.

The *available potential energy* (APE) can now be defined as the difference between the total potential energy of a closed system and the minimum total potential energy that could result from an adiabatic redistribution of mass. Thus, for the idealized model given earlier, the APE, which is designated by the symbol P, is

$$P = (c_p/c_v)\left(E_I - E_I''\right) \tag{7.33}$$

which is equivalent to the maximum kinetic energy that can be realized by an adiabatic process.

Lorenz (1960) showed that available potential energy is given approximately by the volume integral over the entire atmosphere of the variance of potential temperature on isobaric surfaces. Thus, letting $\bar{\theta}$ designate the average potential temperature for a given pressure surface and θ' the local deviation from the average, the average available potential energy per unit volume satisfies the proportionality

$$\overline{P} \propto V^{-1} \int \left(\overline{\theta'^2}/\bar{\theta}^2\right) dV$$

where V designates the total volume. For the quasi-geostrophic model, this proportionality is an exact measure of the available potential energy, as shown in the following subsection.

Observations indicate that for the atmosphere as a whole,

$$\overline{P}/\left[(c_p/c_v)\,\overline{E_I}\right] \sim 5 \times 10^{-3}, \quad \overline{K}/\overline{P} \sim 10^{-1}$$

Thus, only about 0.5% of the total potential energy of the atmosphere is available, and of the available portion only about 10% is actually converted to kinetic energy. From this point of view the atmosphere is a rather inefficient heat engine.

7.3.2 Energy Equations for the Two-Layer Model

In the two-layer model of Section 7.2, the perturbation temperature field is proportional to $\psi_1' - \psi_3'$, the 250- to 750-hPa thickness. Thus, in view of the discussion in the previous section, we anticipate that the available potential energy in this case will be proportional to $(\psi_1' - \psi_3')^2$. To show that this in fact must be the case, we derive the energy equations for the system in the following manner: We first multiply (7.9) by $-\psi_1$, (7.10) by $-\psi_3$, and (7.11) by $(\psi_1' - \psi_3')$. We then integrate the resulting equations over one wavelength of the perturbation in the zonal direction.

The resulting zonally averaged[2] terms are denoted by overbars as done previously in Chapter 5:

$$\overline{(\;)} = L^{-1} \int_0^L (\;)\, dx$$

where L is the wavelength of the perturbation. Thus, for the first term in (7.9) we have, after multiplying by $-\psi_1$, averaging and differentiating by parts:

$$-\overline{\psi_1'\frac{\partial}{\partial t}\left(\frac{\partial^2\psi_1'}{\partial x^2}\right)} = -\overline{\frac{\partial}{\partial x}\left[\psi_1'\frac{\partial}{\partial x}\left(\frac{\partial\psi_1'}{\partial t}\right)\right]} + \overline{\frac{\partial\psi_1'}{\partial x}\frac{\partial}{\partial t}\left(\frac{\partial\psi_1'}{\partial x}\right)}$$

The first term on the right side vanishes because it is the integral of a perfect differential in x over a complete cycle. The second term on the right can be rewritten in the form

$$\frac{1}{2}\frac{\partial}{\partial t}\overline{\left(\frac{\partial\psi_1'}{\partial x}\right)^2}$$

which is just the rate of change of the perturbation kinetic energy per unit mass averaged over a wavelength. Similarly, $-\psi_1'$ times the advection term on the left in (7.9) can be written after integration in x as

$$\begin{aligned} -U_1\overline{\psi_1'\frac{\partial^2}{\partial x^2}\left(\frac{\partial\psi_1'}{\partial x}\right)} &= -U_1\overline{\frac{\partial}{\partial x}\left[\psi_1'\frac{\partial}{\partial x}\left(\frac{\partial\psi_1'}{\partial x}\right)\right]} + U_1\overline{\frac{\partial\psi_1'}{\partial x}\frac{\partial^2\psi_1'}{\partial x^2}} \\ &= \frac{U_1}{2}\overline{\frac{\partial}{\partial x}\left(\frac{\partial\psi_1'}{\partial x}\right)^2} = 0 \end{aligned}$$

Thus, the advection of kinetic energy vanishes when integrated over a wavelength. Evaluating the various terms in (7.10) and (7.11) in the same manner after multiplying through by $-\psi_3'$ and $(\psi_1' - \psi_3')$, respectively, we obtain the following set of perturbation energy equations:

$$\frac{1}{2}\frac{\partial}{\partial t}\overline{\left(\frac{\partial\psi_1'}{\partial x}\right)^2} = -\frac{f_0}{\delta p}\overline{\omega_2'\psi_1'} \tag{7.34}$$

$$\frac{1}{2}\frac{\partial}{\partial t}\overline{\left(\frac{\partial\psi_3'}{\partial x}\right)^2} = +\frac{f_0}{\delta p}\overline{\omega_2'\psi_3'} \tag{7.35}$$

$$\frac{1}{2}\frac{\partial}{\partial t}\overline{(\psi_1' - \psi_3')^2} = U_T\overline{(\psi_1' - \psi_3')\frac{\partial}{\partial x}(\psi_1' + \psi_3')} + \frac{\sigma\delta p}{f_0}\overline{\omega_2'(\psi_1' - \psi_3')} \tag{7.36}$$

where as before $U_T \equiv (U_1 - U_3)/2$.

Defining the perturbation kinetic energy to be the sum of the kinetic energies of the 250- and 750-hPa levels,

$$K' \equiv (1/2)\left[\overline{(\partial\psi_1'/\partial x)^2} + \overline{(\partial\psi_3'/\partial x)^2}\right]$$

we find by adding (7.34) and (7.35) that

$$dK'/dt = -(f_0/\delta p)\overline{\omega_2'(\psi_1' - \psi_3')} = -(2f_0/\delta p)\overline{\omega_2'\psi_T} \tag{7.37}$$

Thus, the rate of change of perturbation kinetic energy is proportional to the correlation between perturbation thickness and vertical motion.

If we now define the perturbation available potential energy as

$$P' \equiv \lambda^2 \overline{(\psi_1' - \psi_3')^2}/2$$

we obtain from (7.36)

$$\begin{aligned} dP'/dt &= \lambda^2 U_T \overline{(\psi_1' - \psi_3')\partial(\psi_1' + \psi_3')/\partial x} + (f_0/\delta p)\overline{\omega_2'(\psi_1' - \psi_3')} \\ &= 4\lambda^2 U_T \overline{\psi_T \partial\psi_m/\partial x} + (2f_0/\delta p)\overline{\omega_2' \psi_T} \end{aligned} \tag{7.38}$$

The last term in (7.38) is just equal and opposite to the kinetic energy source term in (7.37). This term clearly must represent a conversion between potential and kinetic energy. If on average the vertical motion is positive $(\omega_2' < 0)$ where the thickness is greater than average $(\psi_1' - \psi_3' > 0)$ and vertical motion is negative where thickness is less than average, then

$$\overline{\omega_2'(\psi_1' - \psi_3')} = 2\overline{\omega_2' \psi_T} < 0$$

and the perturbation potential energy is being converted to kinetic energy. Physically, this correlation represents an overturning in which cold air aloft is replaced by warm air from below, a situation that clearly tends to lower the center of mass and thus the potential energy of the perturbation. However, the available potential energy and kinetic energy of a disturbance can still grow simultaneously, provided that the potential energy generation due to the first term in (7.38) exceeds the rate of potential energy conversion to kinetic energy.

The potential energy generation term in (7.38) depends on the correlation between the perturbation thickness ψ_T and the meridional velocity at 500 hPa, $\partial\psi_m/\partial x$. In order to understand the role of this term, it is helpful to consider a particular sinusoidal wave disturbance. Suppose that the barotropic and baroclinic parts of the disturbance can be written, respectively, as

$$\psi_m = A_m \cos k(x - ct) \quad \text{and} \quad \psi_T = A_T \cos k(x + x_0 - ct) \tag{7.39}$$

where x_0 designates the phase difference. Because ψ_m is proportional to the *500-hPa* geopotential and ψ_T is proportional to the 500-hPa temperature (or 250- to 750-hPa thickness), the phase angle kx_0 gives the phase difference between geopotential and temperature fields at 500 hPa. Furthermore, A_m and A_T are measures of the amplitudes of the 500-hPa disturbance geopotential and thickness fields, respectively. Using the expressions in (7.39), we obtain

$$\begin{aligned} \overline{\psi_T \frac{\partial\psi_m}{\partial x}} &= -\frac{k}{L}\int_0^L A_T A_m \cos k(x + x_0 - ct)\sin k(x - ct)\,dx \\ &= \frac{kA_T A_m \sin kx_0}{L}\int_0^L [\sin k(x - ct)]^2\,dx \\ &= (A_T A_m k \sin kx_0)/2 \end{aligned} \tag{7.40}$$

Substituting from (7.40) into (7.38), we see that for the usual midlatitude case of a westerly thermal wind ($U_T > 0$) the correlation in (7.40) must be positive if the perturbation potential energy is to increase. Thus, kx_0 must satisfy the inequality $0 < kx_0 < \pi$. Furthermore, the correlation will be a positive maximum for $kx_0 = \pi/2$—that is, when the temperature wave lags the geopotential wave by 90° in phase at 500hPa.

This case was shown schematically earlier in Figure 7.4. Clearly, when the temperature wave lags the geopotential by one-quarter cycle, the northward advection of warm air by the geostrophic wind east of the 500-hPa trough and the southward advection of cold air west of the 500-hPa trough are both maximized. As a result, cold advection is strong below the 250-hPa trough, and warm advection is strong below the 250-hPa ridge. In that case, as discussed previously in Section 6.4.2, the upper-level disturbance will intensify. It should also be noted here that if the temperature wave lags the geopotential wave, the trough and ridge axes will tilt westward with height, which, as mentioned in Section 6.1, is observed to be the case for amplifying midlatitude synoptic systems.

Referring again to Figure 7.4 and recalling the vertical motion pattern implied by the omega equation (7.28), we see that the signs of the two terms on the right in (7.38) cannot be the same. In the westward-tilting perturbation of Figure 7.4, the vertical motion must be downward in the cold air behind the trough at 500 hPa. Hence, the correlation between temperature and vertical velocity must be positive in this situation; that is,

$$\overline{\omega_2' \psi_T} < 0$$

Thus, for quasi-geostrophic perturbations, a westward tilt of the perturbation with height implies both that the horizontal temperature advection will increase the available potential energy of the perturbation and that the vertical circulation will convert perturbation available potential energy to perturbation kinetic energy. Conversely, an eastward tilt of the system with height would change the signs of both terms on the right in (7.38).

Although the signs of the potential energy generation term and the potential energy conversion term in (7.38) are always opposite for a developing baroclinic wave, it is only the potential energy generation rate that determines the growth of the total energy $P' + K'$ of the disturbance. This may be proved by adding (7.37) and (7.38) to obtain

$$d(P' + K')/dt = 4\lambda^2 U_T \overline{\psi_T \partial \psi_m / \partial x}$$

Provided the correlation between the meridional velocity and temperature is positive and $U_T > 0$, the total energy of the perturbation will increase. Note that the vertical circulation merely converts disturbance energy between the available potential and kinetic forms without affecting the total energy of the perturbation.

The rate of increase of the total energy of the perturbation depends on the magnitude of the basic state thermal wind U_T. This is, of course, proportional to the zonally averaged meridional temperature gradient. Because the generation of perturbation energy requires systematic poleward transport of warm air and equatorward transport of cold air, it is clear that baroclinically unstable disturbances tend to reduce the meridional temperature gradient and thus the available potential energy of the mean flow. This latter process cannot be described mathematically in terms of the linearized equations. However, from Figure 7.8 we can see qualitatively that parcels that move poleward and upward with slopes less than the slope of the zonal-mean potential temperature surface will become warmer than their surroundings, and vice versa for parcels moving downward and equatorward.

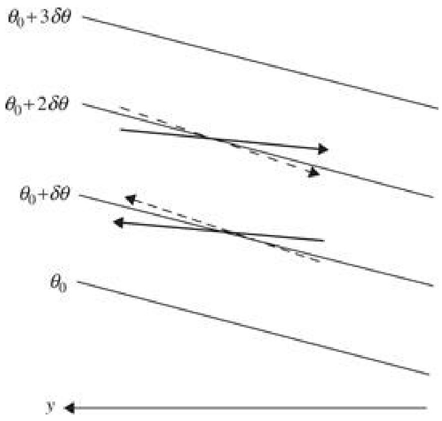


Figure 7.8 Slopes of parcel trajectories relative to the zonal-mean potential temperature surfaces for a baroclinically unstable disturbance (*solid arrows*) and for a baroclinically stable disturbance (*dashed arrows*).

For such parcels the correlations between disturbance meridional velocity and temperature and between disturbance vertical velocity and temperature will both be positive as required for baroclinically unstable disturbances. For parcels that have trajectory slopes greater than the mean potential temperature slope, however, both of these correlations will be negative. Such parcels must then convert disturbance kinetic energy to disturbance available potential energy, which is in turn converted to zonal-mean available potential energy. Therefore, in order that perturbations are able to extract potential energy from the mean flow, the perturbation parcel trajectories in the meridional plane must have slopes less than the slopes of the potential temperature surfaces, and a permanent rearrangement of air must take place for there to be a net heat transfer.

Since we have previously seen that poleward-moving air must rise and equatorward-moving air must sink, it is clear that the rate of energy generation can be greater for an atmosphere in which the meridional slope of the potential temperature surfaces is large. We can also see more clearly why baroclinic instability has a short-wave cutoff. As mentioned previously, the intensity of the vertical circulation must increase as the wavelength decreases. Thus, the slopes of the parcel trajectories must increase with decreasing wavelength, and for some critical wavelength the trajectory slopes will become greater than the slopes of the potential temperature surfaces. Unlike convective instability, where the most rapid amplification occurs for the smallest possible scales, baroclinic instability is most effective at an intermediate range of scales.

The energy flow for quasi-geostrophic perturbations is summarized in Figure 7.9 by means of a block diagram. In this type of energy diagram each block represents a reservoir of a particular type of energy, and arrows designate the direction of energy flow. The complete energy cycle cannot be derived in terms of linear perturbation theory but will be discussed qualitatively in Chapter 10.

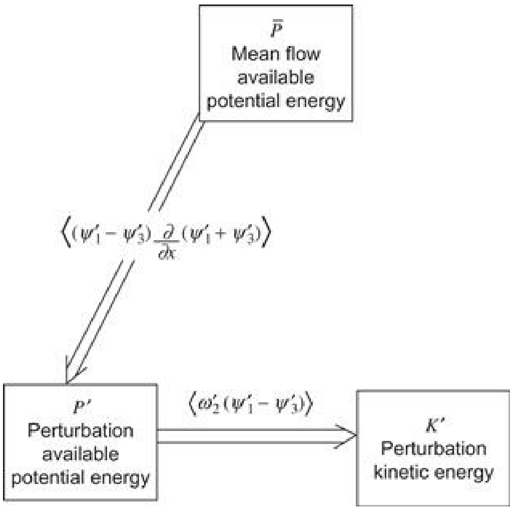

$\bar{P}$
Mean flow available potential energy
$\left\langle (\psi_1' - \psi_3') \frac{\partial}{\partial x} (\psi_1' + \psi_3') \right\rangle$
P'
Perturbation available potential energy
$\langle \omega_2' (\psi_1' - \psi_3') \rangle$
K'
Perturbation kinetic energy

Figure 7.9 Energy flow in an amplifying baroclinic wave.

7.4 Baroclinic Instability of a Continuously Stratified Atmosphere

In the two previous sections some basic aspects of baroclinic instability were elucidated in the context of a simple two-layer model. The dependence of growth rate on vertical shear and the existence of a short-wave cutoff were clearly demonstrated. The two-layer model, however, does have one severe constraint: It assumes that the altitude dependence of large-scale systems can be adequately represented with only two degrees of freedom in the vertical (i.e., the streamfunctions at the 250- and 750-hPa levels). Although most synoptic-scale systems in midlatitudes are observed to have vertical scales comparable to the depth of the troposphere, observed vertical structures do differ. Disturbances that are concentrated near the ground or near the tropopause can hardly be represented accurately in the two-layer model.

An analysis of the structure of baroclinic modes for realistic mean zonal wind profiles is quite complex, and indeed can only be done by numerical methods. However, without obtaining specific normal mode solutions, it is possible to obtain necessary conditions for baroclinic or barotropic instability from an integral theorem first developed by Rayleigh. This theorem, which is discussed in Section 7.4.2, also shows how baroclinic instability is intimately related to the mean meridional gradient of potential vorticity and the mean meridional temperature gradient at the surface.

If a number of simplifying assumptions are made, it is possible to pose the stability problem for a continuously stratified atmosphere in a fashion that leads to a second-order differential equation for the vertical structure that can be solved by standard methods. This problem was originally studied by the British meteorologist Eady (1949) and, although mathematically similar to the two-layer model, provides additional insights. It is developed in Section 7.4.3.

7.4.1 Log-Pressure Coordinates

Derivation of the Rayleigh theorem and the Eady stability model is facilitated if we transform from the standard isobaric coordinates to a vertical coordinate based on the logarithm of pressure. In the *log-pressure coordinates*, the vertical coordinate is defined as

$$z^* \equiv -H \ln (p/p_s) \tag{7.41}$$

where p_s is a standard reference pressure (usually taken to be 1000 hPa) and H is a standard scale height, $H \equiv RT_s/g$, with T_s a global average temperature. For the special case of an isothermal atmosphere at temperature T_s, z^* is exactly equal to geometric height, and the density profile is given by the reference density

$$\rho_0\left(z^*\right) = \rho_s \exp\left(-z^*/H\right)$$

where ρ_s is the density at $z^* = 0$.

For an atmosphere with a realistic temperature profile, z^* is only approximately equivalent to the height, but in the troposphere the difference is usually quite small. The vertical velocity in this coordinate system is

$$w^* \equiv Dz^*/dt$$

The horizontal momentum equation in the log-pressure system is the same as that in the isobaric system:

$$D\mathbf{V}/Dt + f\mathbf{k} \times \mathbf{V} = -\boldsymbol{\nabla}_h\Phi \tag{7.42}$$

However, the operator D/Dt is now defined as

$$D/Dt = \partial/\partial t + \mathbf{V} \cdot \boldsymbol{\nabla}_h + w^*\partial/\partial z^*$$

The hydrostatic equation $\partial\Phi/\partial p = -\alpha$ can be transformed to the log-pressure system by multiplying through by p and using the ideal gas law to get

$$\partial\Phi/\partial \ln p = -RT$$

which after dividing through by $-H$ and using (7.41) gives

$$\partial\Phi/\partial z^* = RT/\mathbf{H} \tag{7.43}$$

The log-pressure form of the continuity equation can be obtained by transforming from the isobaric coordinate form (3.5). We first note that

$$w^* \equiv -(H/p)\, Dp/Dt = -H\omega/p$$

so that

$$\frac{\partial \omega}{\partial p} = -\frac{\partial}{\partial p}\left(\frac{pw^*}{H}\right) = \frac{\partial w^*}{\partial z^*} - \frac{w^*}{H} = \frac{1}{\rho_0}\frac{\partial(\rho_0 w^*)}{\partial z^*}$$

Thus, in log-pressure coordinates the continuity equation becomes simply

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{1}{\rho_0}\frac{\partial(\rho_0 w^*)}{\partial z^*} = 0 \tag{7.44}$$

It is left as a problem for the reader to show that the first law of thermodynamics (3.6) can be expressed in log-pressure form as

$$\left(\frac{\partial}{\partial t}+\mathbf{V}\cdot\nabla_h\right)\frac{\partial\Phi}{\partial z^*}+w^*N^2=\frac{\kappa J}{H} \tag{7.45}$$

where

$$N^2\equiv(R/H)\left(\partial T/\partial z^*+\kappa T/H\right)$$

is the buoyancy frequency squared (see Section 2.7.3) and $\kappa\equiv R/c_p$. Unlike the static stability parameter, S_p, in the isobaric form of the thermodynamic equation (3.6), the parameter N^2 varies only weakly with height in the troposphere; it can be assumed to be constant without serious error. This is a major advantage of the log-pressure formulation.

The quasi-geostrophic potential vorticity equation has the same form as in the isobaric system, but with q defined as

$$q\equiv\nabla_h^2\psi+f+\frac{1}{\rho_0}\frac{\partial}{\partial z^*}\left(\varepsilon\rho_0\frac{\partial\psi}{\partial z^*}\right) \tag{7.46}$$

where $\varepsilon\equiv f_0^2/N^2$.

7.4.2 Baroclinic Instability: The Rayleigh Theorem

We now examine the stability problem for a continuously stratified atmosphere on the midlatitude β plane. The linearized form of the quasi-geostrophic potential vorticity equation can be expressed in log-pressure coordinates as

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)q'+\frac{\partial\overline{q}}{\partial y}\frac{\partial\psi'}{\partial x}=0 \tag{7.47}$$

where

$$q'\equiv\nabla_h^2\psi'+\frac{1}{\rho_0}\frac{\partial}{\partial z^*}\left(\varepsilon\rho_0\frac{\partial\psi'}{\partial z^*}\right) \tag{7.48}$$

and

$$\frac{\partial\overline{q}}{\partial y}=\beta-\frac{\partial^2\overline{u}}{\partial y^2}-\frac{1}{\rho_0}\frac{\partial}{\partial z^*}\left(\varepsilon\rho_0\frac{\partial\overline{u}}{\partial z^*}\right) \tag{7.49}$$

As in the two-layer model, boundary conditions are required at lower- and upper-boundary pressure surfaces. Assuming that the flow is adiabatic and that the vertical motion w^* vanishes at these boundaries, the linearized form of the thermodynamic energy equation (7.45) valid at horizontal boundary surfaces is simply

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)\frac{\partial\psi'}{\partial z^*}-\frac{\partial\psi'}{\partial x}\frac{\partial\overline{u}}{\partial z^*}=0 \tag{7.50}$$

The sidewall boundary conditions are

(7.51)

$$\partial \psi'/\partial x = 0, \quad \text{hence, } \psi' = 0 \quad \text{at} \quad y = \pm L$$

We now assume that the perturbation consists of a single zonal Fourier component propagating in the *x* direction:

(7.52)

$$\psi'(x, y, z, t) = \operatorname{Re}\left\{\Psi(y, z)\exp\left[ik(x - ct)\right]\right\}$$

where $\Psi(y, z) = \Psi_r + i\Psi_i$ is a complex amplitude, *k* is the zonal wave number, and $c = c_r + ic_i$ is a complex phase speed. Note that (7.52) can alternatively be expressed as

$$\psi'(x, y, z, t) = e^{kc_i t}\left[\Psi_r \cos k(x - c_r t) - \Psi_i \sin k(x - c_r t)\right]$$

Thus, the relative magnitudes of Ψ_r and Ψ_i determine the phase of the wave for any *y*, *z**.

Substituting from (7.52) into (7.47) and (7.50) yields

(7.53)

$$(\overline{u} - c)\left[\frac{\partial^2 \Psi}{\partial y^2} - k^2\Psi + \frac{1}{\rho_0}\frac{\partial}{\partial z^*}\left(\varepsilon\rho_0\frac{\partial \Psi}{\partial z^*}\right)\right] + \frac{\partial \overline{q}}{\partial y}\Psi = 0$$

and

(7.54)

$$(\overline{u} - c)\frac{\partial \Psi}{\partial z^*} - \frac{\partial \overline{u}}{\partial z^*}\Psi = 0 \quad \text{at } z^* = 0$$

If the upper boundary is taken to be a rigid lid at a finite height, as is sometimes done in theoretical studies, the condition (7.54) is appropriate at that boundary as well. Alternatively the upper boundary condition can be specified by requiring Ψ to remain finite as *z** → ∞.

Equation (7.53), together with its boundary conditions, constitutes a linear boundary value problem for $\Psi(y, z^*)$. It is generally not simple to obtain solutions to (7.53) for realistic mean zonal wind profiles. Nevertheless, we can obtain some useful information on stability properties simply by analyzing the energetics of the system.

Dividing (7.53) by $(\overline{u} - c)$ and separating the resulting equation into real and imaginary parts, we obtain

(7.55)

$$\frac{\partial^2 \Psi_r}{\partial y^2} + \frac{1}{\rho_0}\frac{\partial}{\partial z^*}\left(\varepsilon\rho_0\frac{\partial \Psi_r}{\partial z^*}\right) - \left[k^2 - \delta_r\left(\partial\overline{q}/\partial y\right)\right]\Psi_r - \delta_i\partial\overline{q}/\partial y\Psi_i = 0$$

(7.56)

$$\frac{\partial^2 \Psi_i}{\partial y^2} + \frac{1}{\rho_0}\frac{\partial}{\partial z^*}\left(\varepsilon\rho_0\frac{\partial \Psi_i}{\partial z^*}\right) - \left[k^2 - \delta_r\left(\partial\overline{q}/\partial y\right)\right]\Psi_i + \delta_i\partial\overline{q}/\partial y\Psi_r = 0$$

where

$$\delta_r = \frac{\bar{u} - c_r}{(\bar{u} - c_r)^2 + c_i^2}, \quad \text{and} \quad \delta_i = \frac{c_i}{(\bar{u} - c_r)^2 + c_i^2}$$

Similarly, dividing (7.54) through by $(\bar{u} - c)$ and separating into real and imaginary parts gives, for the boundary condition at $z^* = 0$,

$$\frac{\partial \Psi_r}{\partial z^*} + \frac{\partial \bar{u}}{\partial z^*} \left(\delta_i \Psi_i - \delta_r \Psi_r\right) = 0; \quad \frac{\partial \Psi_i}{\partial z^*} - \frac{\partial \bar{u}}{\partial z^*} \left(\delta_r \Psi_i + \delta_i \Psi_r\right) = 0 \tag{7.57}$$

Multiplying (7.55) by Ψ_i, (7.56) by Ψ_r, and subtracting the latter from the former yields

$$\begin{aligned} &\rho_0 \left[\Psi_i \frac{\partial^2 \Psi_r}{\partial y^2} - \Psi_r \frac{\partial^2 \Psi_i}{\partial y^2}\right] + \left[\Psi_i \frac{\partial}{\partial z^*}\left(\varepsilon \rho_0 \frac{\partial \Psi_r}{\partial z^*}\right)\right. \\ &\left. - \Psi_r \frac{\partial}{\partial z^*}\left(\varepsilon \rho_0 \frac{\partial \Psi_i}{\partial z^*}\right)\right] - \rho_0 \delta_i (\partial \bar{q}/\partial y) \left(\Psi_i^2 + \Psi_r^2\right) = 0 \end{aligned} \tag{7.58}$$

Using the chain rule of differentiation, (7.58) can be expressed in the form

$$\begin{aligned} &\rho_0 \frac{\partial}{\partial y}\left[\Psi_i \frac{\partial \Psi_r}{\partial y} - \Psi_r \frac{\partial \Psi_i}{\partial y}\right] + \frac{\partial}{\partial z^*}\left[\varepsilon \rho_0 \left(\Psi_i \frac{\partial \Psi_r}{\partial z^*} - \Psi_r \frac{\partial \Psi_i}{\partial z^*}\right)\right] \\ &- \rho_0 \delta_i (\partial \bar{q}/\partial y) \left(\Psi_i^2 + \Psi_r^2\right) = 0 \end{aligned} \tag{7.59}$$

The first term in brackets in (7.59) is a perfect differential in y; the second term is a perfect differential in z^*. Thus, if (7.59) is integrated over the y, z^* domain, the result can be expressed as

$$\begin{aligned} &\int_0^\infty \left[\Psi_i \frac{\partial \Psi_r}{\partial y} - \Psi_r \frac{\partial \Psi_i}{\partial y}\right]_{-L}^{+L} \rho_0 \, dz^* + \int_{-L}^{+L} \left[\varepsilon \rho_0 \left(\Psi_i \frac{\partial \Psi_r}{\partial z^*} - \Psi_r \frac{\partial \Psi_i}{\partial z^*}\right)\right]_0^\infty dy \\ &= \int_{-L}^{+L} \int_0^\infty \rho_0 \delta_i \frac{\partial \bar{q}}{\partial y} \left(\Psi_i^2 + \Psi_r^2\right) dy \, dz^* \end{aligned} \tag{7.60}$$

However, from (7.51) $\Psi_i = \Psi_r = 0$ at $y = \pm L$ so that the first integral in (7.60) vanishes. Furthermore, if Ψ remains finite as $z^* \rightarrow \infty$, the contribution to the second integral of (7.60) at the upper boundary vanishes. If we then use (7.57) to eliminate the vertical derivatives in this term at the lower boundary, (7.60) can be expressed as

$$c_i \left[\int_{-L}^{+L} \int_0^\infty \frac{\partial \bar{q}}{\partial y} \frac{\rho_0 |\Psi|^2}{|\bar{u} - c|^2} dy \, dz^* - \int_{-L}^{+L} \varepsilon \frac{\partial \bar{u}}{\partial z^*} \frac{\rho_0 |\Psi|^2}{|\bar{u} - c|^2}\bigg|_{z^*=0} dy\right] = 0 \tag{7.61}$$

where $|\Psi|^2 = \Psi_r^2 + \Psi_i^2$ is the disturbance amplitude squared.

Equation (7.61) has important implications for the stability of quasi-geostrophic perturbations. For unstable modes, c_i must be nonzero, and thus the quantity in square brackets in (7.61) must vanish. Because $|\Psi|^2 / |\bar{u} - c|^2$ is nonnegative, instability is possible only when $\partial \bar{u} / \partial z^*$ at the lower boundary and $(\partial \bar{q} / \partial y)$ in the whole domain satisfy certain constraints:

1. If $\partial \bar{u}/\partial z^*$ at $z^* = 0$ (which by thermal wind balance implies that the meridional temperature gradient vanishes at the boundary), the second integral in (7.61) vanishes. Thus, the first integral must also vanish for instability to occur. This can occur only if $\partial \bar{q}/\partial y$ changes sign within the domain (i.e., $\partial \bar{q}/\partial y = 0$ somewhere). This is referred to as the *Rayleigh necessary condition* and is another demonstration of the fundamental role played by potential vorticity. Because $\partial \bar{q}/\partial y$ is normally positive, it is clear that in the absence of temperature gradients at the lower boundary, a region of negative meridional potential vorticity gradients must exist in the interior for instability to be possible.
2. If $\partial \bar{q} / \partial y \geq 0$ everywhere, then it is necessary that $\partial \bar{u}/\partial z^* > 0$ somewhere at the lower boundary for $c_i > 0$.
3. If $\partial \bar{u}/\partial z^* < 0$ everywhere at $z^* = 0$, then it is necessary that $\partial \bar{q}/\partial y < 0$ somewhere for instability to occur. Thus, there is an asymmetry between westerly and easterly shear at the lower boundary, with the former more favorable for baroclinic instability.

The basic state potential vorticity gradient in (7.49) can be written in the form

$$\frac{\partial \bar{q}}{\partial y} = \beta - \frac{\partial^2 \bar{u}}{\partial y^2} + \frac{\varepsilon}{H}\frac{\partial \bar{u}}{\partial z^*} - \varepsilon \frac{\partial^2 \bar{u}}{\partial z^{*2}} - \frac{\partial \varepsilon}{\partial z^*}\frac{\partial \bar{u}}{\partial z^*}$$

Because β is positive everywhere, if ε is constant a negative basic state potential vorticity gradient can occur only for strong positive mean flow curvature (i.e., $\partial^2 \bar{u}/\partial y^2 \text{ or } \partial^2 \bar{u}/\partial z^{*2} \gg 0$) or strong negative vertical shear ($\partial \bar{u}/\partial z^* \ll 0$). Strong positive meridional curvature can occur at the core of an easterly jet or on the flanks of a westerly jet. Instability associated with such horizontal curvature is referred to as *barotropic instability*. The normal baroclinic instability in midlatitudes is associated with mean flows in which $\partial \bar{q}/\partial y > 0$ and $\partial \bar{u}/\partial z^* > 0$ at the ground. Hence, a mean meridional temperature gradient at the ground is essential for the existence of such instability. Baroclinic instability can also be excited at the tropopause because of the rapid decrease of ε with height if there is a sufficiently strong easterly mean wind shear to cause a local reversal in the mean potential vorticity gradient.

7.4.3 The Eady Stability Problem

This section analyzes the structures (eigenfunctions) and growth rates (eigenvalues) for unstable modes in the simplest possible model that satisfies the necessary conditions for instability in a continuous atmosphere given in the previous subsection. For simplicity, we make the following assumptions:

- Basic state density constant (Boussinesq approximation).
- *f*-plane geometry ($\beta = 0$).
- $\partial \bar{u}/\partial z^* = \Lambda = \text{constant}$.
- Rigid lids at $z^* = 0$ and H.

These conditions are only a crude model of the atmosphere, but provide a first approximation for study of the dependence of vertical structure on horizontal scale and stability. Despite the zero mean potential vorticity in the domain, the Eady model satisfies the necessary conditions for instability discussed in the previous subsection because vertical shear of the basic state mean flow at the upper boundary provides an additional term in (7.61) that is equal and opposite to the lower boundary integral.

Using the aforementioned approximations, the quasi-geostrophic potential vorticity equation is

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)q'=0 \tag{7.62}$$

where the perturbation potential vorticity is given by

$$q'=\nabla_h\psi'+\varepsilon\frac{\partial^2\psi}{\partial z^{*2}} \tag{7.63}$$

and $\varepsilon\equiv f_0^2/N^2$. The thermodynamic energy equation at the horizontal boundaries ($z^*=0,H$) is

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)T'=0 \tag{7.64}$$

Letting

$$\begin{aligned}\psi'(x,y,z^*,t)&=\Psi(z^*)\cos ly\exp\left[ik(x-ct)\right]\\ \overline{u}(z^*)&=\Lambda z^*\end{aligned} \tag{7.65}$$

where as in the previous subsection $\Psi(z^*)$ is a complex amplitude and *c* a complex phase speed, and substituting from (7.65) into (7.62) we find that

$$(\overline{u}-c)\,q'=0 \tag{7.66}$$

From (7.66) we see that either $q'=0$ or $\overline{u}-c=0$. Since $\overline{u}=\Lambda z^*$, $\overline{u}-c=0$ allows for nonzero q' only at the level where the wave speed matches the flow speed. Moreover, since $\overline{u}$ is strictly real, the wave speed in this case must be as well, so this particular solution branch describes singular neutral modes (spikes of PV at the steering level of the wave). They are useful for describing the evolution of potential vorticity disturbances in the Eady model, but they do not describe instability. Therefore, unstable modes must have $q'=0$, which from (7.63) implies that the vertical structure is given by the solution of the second-order differential equation

$$\frac{d^2\Psi}{dz^{*2}}-\alpha^2\Psi=0 \tag{7.67}$$

where $\alpha^2=(k^2+l^2)/\varepsilon$. A similar substitution into (7.64) yields the boundary conditions

$$(\Lambda z^*-c)\,d\Psi/dz^*-\Psi\Lambda=0,\quad \text{at } z^*=0,H \tag{7.68}$$

valid for rigid horizontal boundaries ($w^*=0$) at the surface ($z^*=0$) and the tropopause ($z^*=H$).

The general solution of (7.67) can be written in the form

$$\Psi(z^*)=A\sinh\alpha z^*+B\cosh\alpha z^* \tag{7.69}$$

Substituting from (7.69) into the boundary conditions (7.68) for $z^*=0$ and H yields a set of two linear homogeneous equations in the amplitude coefficients *A* and *B*:

$$
\begin{aligned}
-c\alpha A - B\Lambda &= 0 \\
\alpha(\Lambda H - c)\left(A\cosh\alpha H + B\sinh\alpha H\right) - \Lambda\left(A\sinh\alpha H + B\cosh\alpha H\right) &= 0
\end{aligned}
$$

As in the two-layer model, a nontrivial solution exists only if the determinant of the coefficients of *A* and *B* vanishes. Again, this leads to a quadratic equation in the phase speed *c*. The solution (see Problem 7.12) has the form

$$c = \frac{\Lambda H}{2} \pm \frac{\Lambda H}{2}\left[1 - \frac{4\cosh\alpha H}{\alpha H\sinh\alpha H} + \frac{4}{\alpha^2 H^2}\right]^{1/2} \tag{7.70}$$

Thus

$$c_i \neq 0 \quad \text{if} \quad 1 - \frac{4\cosh\alpha H}{\alpha H\sinh\alpha H} + \frac{4}{\alpha^2 H^2} < 0$$

and the flow is then baroclinically unstable. When the quantity in square brackets in (7.70) is equal to zero, the flow is said to be *neutrally stable*. This condition occurs for $\alpha = \alpha_c$ where

$$\alpha_c^2 H^2/4 - \alpha_c H\left(\tanh\alpha_c H\right)^{-1} + 1 = 0 \tag{7.71}$$

Using the identity

$$\tanh\alpha_c H = 2\tanh\left(\frac{\alpha_c H}{2}\right)\Big/\left[1 + \tanh^2\left(\frac{\alpha_c H}{2}\right)\right]$$

we can factor (7.71) to yield

$$\left[\frac{\alpha_c H}{2} - \tanh\left(\frac{\alpha_c H}{2}\right)\right]\left[\frac{\alpha_c H}{2} - \coth\left(\frac{\alpha_c H}{2}\right)\right] = 0 \tag{7.72}$$

Thus, the critical value of α is given by $\alpha_c H/2 = /\coth(\alpha_c H/2)$, which implies $\alpha_c H \cong 2.4$. Hence, instability requires $\alpha < \alpha_c$ or

$$\left(k^2 + l^2\right) < \left(\alpha_c^2 f_0^2/N^2\right) \approx 5.76/L_R^2$$

where $L_R \equiv NH/f_0 \approx 1000$ km is the Rossby *radius of deformation* for a continuously stratified fluid—compare to λ^{-1} defined just below equation (7.16). For waves with equal zonal and meridional wave numbers ($k = l$), the wavelength of maximum growth rate turns out to be

$$L_m = 2\sqrt{2}\pi L_R/(H\alpha_m) \cong 5500\,\text{km}$$

where α_m is the value of α for which kc_i is a maximum.

Substituting this value of α into the solution for the vertical structure of the streamfunction (7.69) and using the lower boundary condition to express the coefficient *B* in terms of *A*, we can determine the vertical structure of the most unstable mode. As shown in Figure 7.10, trough and ridge axes slope westward with height, in agreement with the requirements for extraction of available potential energy from the mean flow. The axes of the warmest and coldest air, however, tilt eastward with height, a result that could not be determined from the two-layer model where temperature was given at a single level. Furthermore, Figures 7.10a and 7.10b show that east of the upper-level trough axis, where the perturbation meridional velocity is positive, the vertical velocity is also positive. Thus, parcel motion is poleward and upward in the region where $\theta' > 0$. Conversely, west of the upper-level trough axis parcel motion is equatorward and downward where $\theta' < 0$. Both cases are thus consistent with the energy-converting parcel trajectory slopes shown earlier in Figure 7.8.

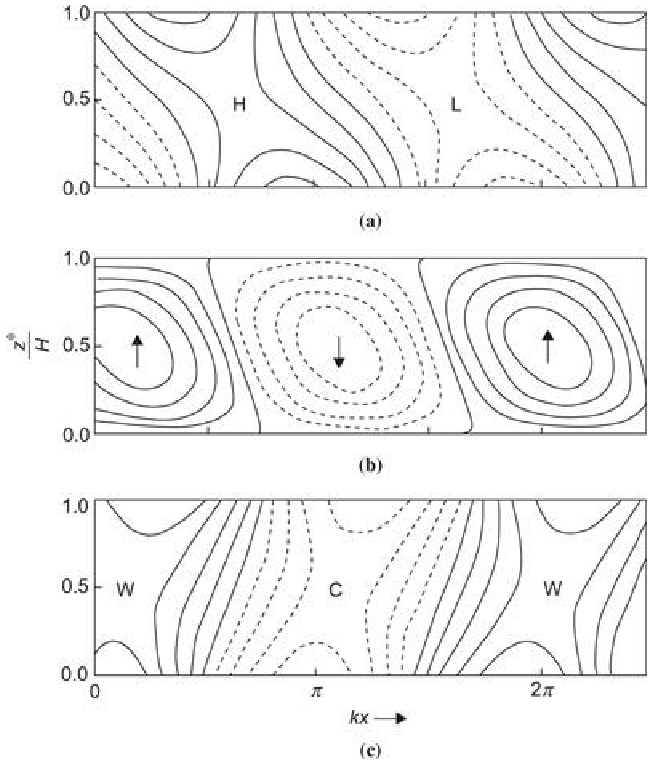


Figure 7.10 Properties of the most unstable Eady wave. (a) Contours of perturbation geopotential height; “H” and “L” designate ridge and trough axes, respectively. (b) Contours of vertical velocity; *up* and *down arrows* designate axes of maximum upward and downward motion, respectively. (c) Contours of perturbation temperature; “W” and “C” designate axes of warmest and coldest temperatures, respectively. In all panels 1 and 1/4 wavelengths are shown for clarity.

Although the Eady model yields unstable solutions that resemble the baroclinic waves that dominate the statistics of midlatitude eddy fields, individual cyclone events are highly localized and often develop at a faster rate than described by the growth rate of the most unstable mode. Both issues are resolved by considering initial value problems for finite-amplitude, *localized*, upper-level initial disturbances, as are often observed to precede the development of surface cyclones. Figure 7.11 shows the development of a surface cyclone from a localized initial disturbance. Note that, after 48 hours, the full nonlinear solution resembles a mature extratropical cyclone, with warm air moving poleward and cold air moving equatorward around an area of low pressure. The linear solution for only the unstable growing modes approximates the full nonlinear solution and shows that the development may be explained by the projection of the initial disturbance onto only the unstable modes. Nonlinear effects relate to the poleward drift of the surface cyclone because of the interaction of the surface and upper-level PV anomalies and details of the development of the surface cold and warm fronts. Although the unstable modes are sufficient to explain the development of surface cyclones, it turns out that *transient* development is also possible, even when all the modes are neutral.

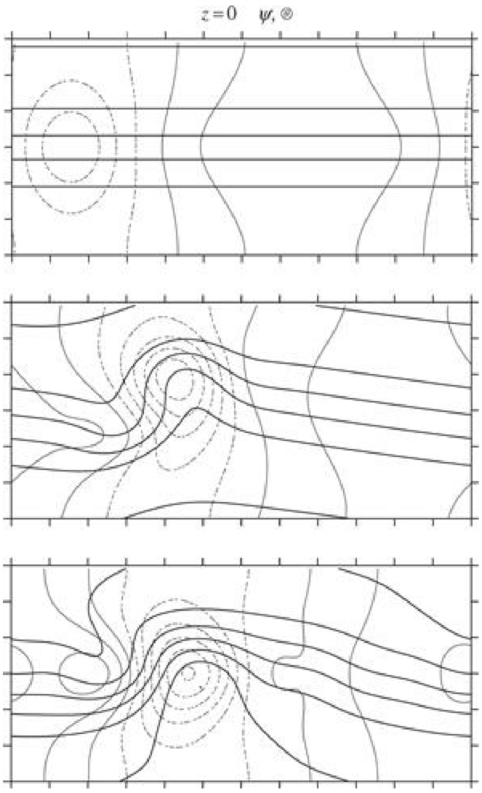


Figure 7.11 Development of a spatially localized disturbance in an idealized westerly jet stream. *Dashed lines* show pressure every 4 hPa and *solid lines* show potential temperature every 5 K. An initial surface disturbance (*top panel*) associated with an upper-level PV anomaly evolves into a well-developed extratropical cyclone 48 h later (*middle panel*). The linear solution for only the projection of the initial disturbance onto the unstable normal modes (*bottom panel*) captures most of the development of the full nonlinear solution.

Adapted from Hakim, 2000. Copyright © American Meteorological Society. Reprinted with permission.

7.5 Growth and Propagation of Neutral Modes

As suggested earlier, baroclinic wave disturbances with certain favorable initial configurations may amplify rapidly even in the absence of baroclinic instability. Although the theory for the optimal initial perturbations that give rise to rapid transient growth is beyond the scope of this book, the main ideas can be exposed mathematically and illustrated with an example from the two-layer model.

A central concept for growth involves a metric, which is needed to measure the amplitude of disturbances; we call this measure a *norm*. A familiar example is the magnitude of a scalar z, $|z|$, which gives the absolute value for real-valued z. For complex-valued $z = a+ib$, the magnitude gives the length of a vector on the complex plane, as defined by $|z| = (z^* z)^{1/2} = (a^2 + b^2)^{1/2}$, where the asterisk denotes a complex conjugate. For *functions*, we adopt a natural extension of this definition:

$$|f| = [\langle f, f\rangle]^{1/2} \tag{7.73}$$

where, for functions of a single independent variable, x,

$$\langle f, g\rangle = \int f^* g \, dx \tag{7.74}$$

is an *inner product* that gives a scalar result. In particular, for $g = f$, (7.74) reveals that $|f|^2$ gives the integral over f^*f, which is a continuous analog of the sum of squared scalar components for complex numbers discussed previously. We define *amplification* over a time interval as the ratio of the norm at the time of interest, t, to the norm at the initial time, t_0:

$$A = \frac{|f|_t}{|f|_{t_0}} \tag{7.75}$$

Consider the normal modes of the Eady model, which have the form (dropping the prime)

$$\psi = \Psi(z^*)\, e^{i\phi}\, e^{kc_i t} \tag{7.76}$$

where $\phi = k(x - c_r t)$ and c_r and c_i are the real and imaginary parts of the phase speed, respectively. The amplification of a single normal mode, as defined by (7.75), for $t_0 = 0$ is given by

$$\left[\frac{\iint \Psi^* e^{-i\phi} e^{kc_i t}\, \Psi e^{i\phi} e^{kc_i t}\, dx dz}{\iint \Psi^* e^{-ikx}\, \Psi e^{ikx}\, dx dz}\right]^{1/2} = \left[e^{2kc_i t} \frac{\iint \Psi^* \Psi \, dx dz}{\iint \Psi^* \Psi \, dx dz}\right]^{1/2} = e^{kc_i t} \tag{7.77}$$

For neutral modes $c_i = 0$ and $A = 1$; otherwise, amplitude grows or decays exponentially depending on the sign of c_i. Note that, because of the space and time separation in (7.76), any function of the spatial structure of the normal modes (e.g., potential vorticity) amplifies at the same rate given by (7.77). This means that, for a single mode, the exponential growth found in (7.77) is independent of the norm.

Consider now the sum of two *neutral* modes, $\psi_1 + \psi_2$. In this case, amplification is given by

$$\left[\frac{\langle \psi_1 + \psi_2, \psi_1 + \psi_2\rangle_t}{\langle \psi_1 + \psi_2, \psi_1 + \psi_2\rangle_{t_0}}\right]^{1/2} = \left[\frac{\langle \psi_1, \psi_1\rangle_t + \langle \psi_2, \psi_2\rangle_t + 2\langle \psi_1, \psi_2\rangle_t}{\langle \psi_1, \psi_1\rangle_{t_0} + \langle \psi_2, \psi_2\rangle_{t_0} + 2\langle \psi_1, \psi_2\rangle_{t_0}}\right]^{1/2} \tag{7.78}$$

where we have used the fact that $\langle f, g\rangle = \langle g, f\rangle^*$. Since the modes are neutral, $\langle \psi_1, \psi_1\rangle$ and $\langle \psi_2, \psi_2\rangle$ are constants, and amplification is determined by the term

$$\langle \psi_1, \psi_2\rangle_t = e^{ik(c_1 - c_2)t} \iint \Psi_1^* \Psi_2 \, dx dz \tag{7.79}$$

This shows that if the modes are *orthogonal in the chosen norm*—that is,

$$\iint \Psi_1^* \Psi_2 \, dx dz = 0 \tag{7.80}$$

then there is no amplification. On the other hand, if the modes are not orthogonal, then the amplitude fluctuates periodically proportional to the difference in the phase speeds of the modes. A physical interpretation is that, because the modes move relative to one another, there are times when the sum gives an upshear-tilted disturbance[3] and the amplitude of the sum increases. Eventually, the relationship changes so that the modes tilt downshear and the amplitude of the sum decreases. Therefore, the amplification effect is transient and depends on the chosen norm. In contrast, the unstable normal modes grow exponentially for all time, and measures of amplification do not depend on the norm.

7.5.1 Transient Growth of Neutral Waves

Consider now a specific example for the two-layer model of Section 7.2. If we neglect the β effect, let $U_m = 0$, and assume that $k^2 > 2\lambda^2$, the two-layer model has two oppositely propagating neutral solutions given by (7.24) with zonal phase speeds

$$c_1 = +U_T\mu; \quad c_2 = -U_T\mu \tag{7.81}$$

where

$$\mu = \left[\frac{k^2 - 2\lambda^2}{k^2 + 2\lambda^2}\right]^{1/2}$$

Then from (7.17) a disturbance consisting of these two modes can be expressed as

$$\psi_m = A_1 \exp\left[ik\left(x - c_1 t\right)\right] + A_2 \exp\left[ik\left(x - c_2 t\right)\right] \tag{7.82a}$$

$$\psi_T = B_1 \exp\left[ik\left(x - c_1 t\right)\right] + B_2 \exp\left[ik\left(x - c_2 t\right)\right] \tag{7.82b}$$

but from (7.18)

$$c_1 A_1 - U_T B_1 = 0; \quad c_2 A_2 - U_T B_2 = 0$$

So that with the aid of (7.81)

$$B_1 = \mu A_1; \quad B_2 = -\mu A_2 \tag{7.83}$$

For an initial disturbance confined entirely to the upper level, it is easily verified that $\psi_m = \psi_T$ ($\psi_1 = 2\psi_m$ and $\psi_3 = 0$). Thus, initially $A_1 + A_2 = B_1 + B_2$, and substituting from (7.83) gives

$$A_2 = -A_1\left[(1-\mu)/(1+\mu)\right]$$

Hence, if A_1 is real, the streamfunctions of (7.82a,b) can be expressed as

$$\begin{aligned}\psi_m(x,t) &= A_1\left[\cos\left[k\left(x-\mu U_T t\right)\right]-\frac{(1-\mu)}{(1+\mu)}\cos\left[k\left(x+\mu U_T t\right)\right]\right]\\ &= \frac{2\mu A_1}{(1+\mu)}\left[\cos kx\cos\left(k\mu U_T t\right)+\frac{1}{\mu}\sin kx\sin\left(k\mu U_T t\right)\right]\end{aligned}\tag{7.84a}$$

$$\begin{aligned}\psi_T(x,t) &= \mu A_1\left[\cos\left[k\left(x-\mu U_T t\right)\right]+\frac{(1-\mu)}{(1+\mu)}\cos\left[k\left(x+\mu U_T t\right)\right]\right]\\ &= \frac{2\mu A_1}{(1+\mu)}\left[\cos kx\cos\left(k\mu U_T t\right)+\mu\sin kx\sin\left(k\mu U_T t\right)\right]\end{aligned}\tag{7.84b}$$

The first forms on the right side in (7.84a,b) show that for small μ the barotropic mode initially consists of two waves of nearly equal amplitude that are 180° out of phase so that they nearly cancel, whereas the baroclinic mode initially consists of two very weak waves that are in phase.

As time advances, the two oppositely propagating barotropic modes begin to reinforce each other, leading to a maximum-amplitude trough 90° to the east of the initial trough at time $t_m = \pi/(2k\mu U_T)$. From (7.84) the baroclinic and barotropic modes at time t_m are then

$$\begin{aligned}(\psi_m)_{\max} &= 2A_1(1+\mu)^{-1}\sin kx\\ (\psi_T)_{\max} &= 2A_1\mu^2(1+\mu)^{-1}\sin kx\end{aligned}\tag{7.85a,b}$$

from which it is shown easily that

$$(\psi_1)_{\max} = 2A_1\left(1+\mu^2\right)(1+\mu)^{-1}\sin kx;\quad (\psi_3)_{\max} = 2A_1(1-\mu)\sin kx$$

so that for small μ the resulting disturbance is nearly barotropic. Thus, the initial disturbance not only amplifies but spreads in the vertical. The time for growth to maximum amplitude is inversely proportional to the basic state thermal wind; the maximum amplitude, however, depends only on the initial amplitude and the parameter μ.

Figure 7.12 shows the initial and maximum amplitudes of the barotropic and baroclinic streamfunctions for $f_0 = 10^{-4}\ \mathrm{s}^{-1}$, $\sigma = 2 \times 10^{-6}\ \mathrm{m}^2\ \mathrm{Pa}^{-2}\ \mathrm{s}^{-2}$, $U_T = 35\ \mathrm{m\,s}^{-1}$, and a zonal wavelength of 3000 km. In this case, the amplitude of the barotropic disturbance increases by a factor of 8 in about 48 h. Although the most unstable normal mode for these conditions amplifies by a similar amount in less than 1 day, if the initial upper-level neutral disturbance has a velocity amplitude of a few meters per second, its growth on the time scale of a few days may dominate over a normal mode instability that grows exponentially from a much smaller initial perturbation.

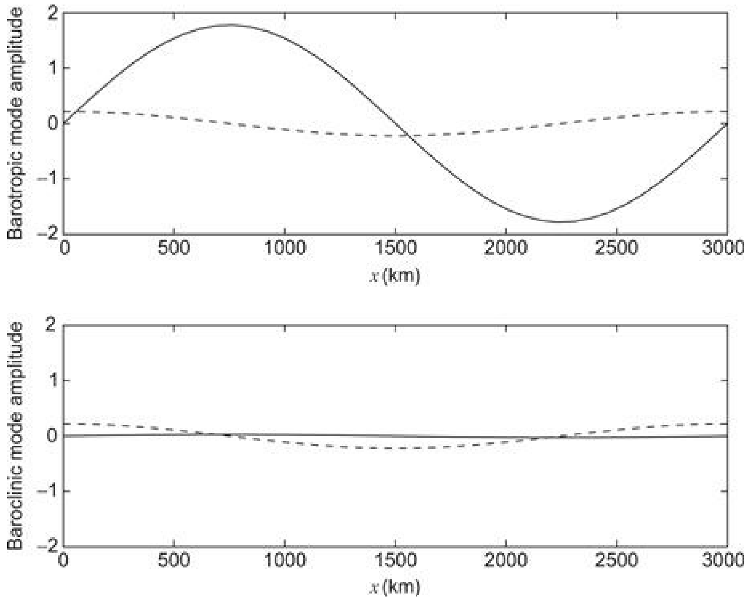


Figure 7.12 Zonal distribution of the amplitude of a transient neutral disturbance initially confined entirely to the upper layer in the two-layer model. *Dashed lines* show the initial distribution of the baroclinic and barotropic modes; *solid lines* show the distribution at the time of maximum amplification.

As with normal mode baroclinic instability, the energy source for the transient amplification of the neutral modes is conversion of the available potential energy of the mean flow into disturbance potential energy by meridional temperature advection, followed immediately by conversion to disturbance kinetic energy by the secondary vertical circulation. This secondary circulation has a vertical velocity field that lags the upper-level streamfunction field by 90° phase. Maximum upward motion and lower-level convergence thus occur to the west of the initial upper-level ridge, and maximum downward motion and lower-level divergence occur west of the initial upper-level trough. Vorticity tendencies associated with this convergence and divergence pattern partly balance the eastward vorticity advection at the upper level and also act to produce a lower-level trough to the east of the initial upper-level trough and lower-level ridge to the east of the initial upper-level ridge. The result is that a pattern of nearly barotropic troughs and ridges develops 90° to the east of the initial trough and ridge pattern. Unlike normal mode instability, the growth in this case does not continue indefinitely. Rather, for a given value of μ the maximum growth occurs at the time for which $t_m = \pi/(2kU_T\mu)$. As the zonal wave number approaches the short-wave instability cutoff, the total amplification increases, but the amplification time also increases.

7.5.2 Downstream Development

It was noted in Section 7.2.2 that under some circumstances the zonal group velocity may exceed the zonal phase speed so that the energy of a wave group propagates faster than the individual wave disturbances. This dispersion effect is demonstrated for barotropic Rossby waves in Problem 5.15. The β effect is not necessary, however, for the development of new disturbances to occur downstream of the original wave disturbances on the synoptic scale. The two-layer channel model with constant Coriolis parameter can be used to demonstrate the process of dispersion and downstream development of neutral waves in a simple context.

Expressed in terms of wave frequency, the dispersion relation of (7.24) becomes

$$\nu = kU_m \pm k\mu U_T \tag{7.86}$$

where μ is defined below (7.81). The corresponding group velocity is then

$$c_{gx} = \frac{\partial \nu}{\partial k} = U_m \pm U_T \mu \left(1 + \frac{4k^2\lambda^2}{k^4 - 4\lambda^4}\right) \tag{7.87}$$

Comparison of (7.87) with (7.81) shows that relative to the mean zonal flow the group velocity exceeds the phase velocity by the factor in parentheses on the right side of (7.87) for both eastward- and westward-directed neutral modes. For example, if $k^2 = 2\lambda^2(1+\sqrt{2})$, the group velocity equals twice the phase speed, which is the situation shown schematically earlier in Figure 7.4.

Observations of baroclinic waves in the midlatitude storm tracks show clearly the downstream development of wave energy. For example, over the North Pacific ocean, the baroclinic wave phase speed is about 10 m s^{-1}, while the group speed is about 30^{-1} (Figure 7.13). As a result, the impact of extratropical cyclones that develop near the Asian coastline may be felt along the West Coast of North America in three days, even though the initial disturbance may take five days to reach the middle Pacific during its decay phase.

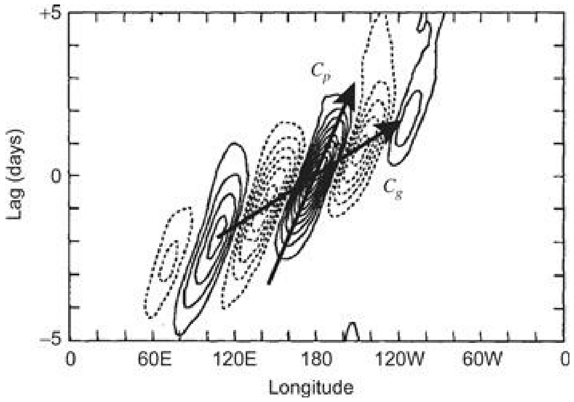


Figure 7.13 Time-series correlation of meridional wind as a function of longitude and time lag. Correlation is shown every 0.1 unit (i.e., a unitless number between −1 and 1), with the reference time series at 180. Results are averaged over 30 to 60° N latitude during winter.

From Chang, 1993. Copyright © American Meteorological Society. Reprinted with permission.

Suggested References

harney, 1947. Charney (1947) is the classic paper on baroclinic instability. The mathematical level is advanced, but Charney's paper contains an excellent qualitative summary of the main results that is very readable.

oskins et al., 1985. Hoskins, McIntyre, and Robertson (1985) discuss cyclogenesis and baroclinic instability from a potential vorticity perspective.

edlosky. Pedlosky, *Geophysical Fluid Dynamics* (2nd ed.), contains a thorough treatment of normal mode baroclinic instability at a mathematically advanced level.

ierrehumbert and Swanson, 1995. Pierrehumbert and Swanson (1995) review numerous aspects of baroclinic instability, including spatiotemporal development.

Vallis. Vallis, *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*, covers topics in baroclinic instability similar to those considered here.

Problems

7.1. Show using (7.25) that the maximum growth rate for baroclinic instability when $\beta = 0$ occurs for

$$k^2 = 2\lambda^2\left(\sqrt{2}-1\right)$$

How long does it take the most rapidly growing wave to amplify by a factor of e^1 if $\lambda^2 = 2 \times 10^{-12}\ \mathrm{m}^{-1}$ and $U_T = 20\mathrm{m\ s}^{-1}$?

7.2. Solve for ψ_3' and ω_2' in terms of ψ_1' for a baroclinic Rossby wave whose phase speed satisfies (7.23). Explain the phase relationship among $\psi_1', \psi_3', \omega_2'$ and in terms of the quasi-geostrophic theory. (Note that $U_T = 0$ in this case.)

7.3. For the case $U_1 = -U_3$ and $k^2 = \lambda^2$, solve for ψ_3' and ω_2' in terms of ψ_1' for marginally stable waves—that is, $\delta = 0$ in (7.21).

7.4. For the case $\beta = 0, k^2 = \lambda^2$, and $U_m = U_T$, solve for ψ_3' and ω_2' in terms of ψ_1'. Explain the phase relationships among ω_2', ψ_1', and ψ_3' in terms of the energetics of quasi-geostrophic waves for the amplifying wave.

7.5. Suppose that a baroclinic fluid is confined between two rigid horizontal lids in a rotating tank in which $\beta = 0$, but friction is present in the form of linear drag proportional to the velocity (i.e., $\mathbf{Fr} = -\mu\mathbf{V}$). Show that the two-level model perturbation vorticity equations in Cartesian coordinates can be written as

$$\left(\frac{\partial}{\partial t}+U_1\frac{\partial}{\partial x}+\mu\right)\frac{\partial^2\psi_1'}{\partial x^2}-\frac{f}{\delta p}\omega_2'=0$$
$$\left(\frac{\partial}{\partial t}+U_3\frac{\partial}{\partial x}+\mu\right)\frac{\partial^2\psi_3'}{\partial x^2}+\frac{f}{\delta p}\omega_2'=0$$

where perturbations are assumed in the form given in (7.8). Assuming solutions of the form (7.17), show that the phase speed satisfies a relationship similar to (7.21) with β replaced everywhere by $i\mu k$, and that as a result the condition for baroclinic instability becomes

$$U_T > \mu\left(2\lambda^2-k^2\right)^{-1/2}$$

7.6. For the case $\beta = 0$ determine the phase difference between the 250- and 750-hPa geopotential fields for the most unstable baroclinic wave (see Problem 7.1). Show that the 500-hPa geopotential and thickness fields (ψ_m, ψ_T) are 90° out of phase.

7.7. For the conditions of Problem 7.6, given that the amplitude of ψ_m is $A = 10^7\ \mathrm{m}^2\ \mathrm{s}^{-1}$, solve the system (7.18) and (7.19) to obtain B. Let $\lambda^2 = 2.0 \times 10^{-12}\ \mathrm{m}^{-2}$ and $U_T = 15\ \mathrm{m\ s}^{-1}$. Use your results to obtain expressions for ψ_1' and ψ_3'.

7.8. For the situation of Problem 7.7, compute ω_2' using (7.28).

7.9. Compute the total potential energy per unit cross-sectional area for an atmosphere with an adiabatic lapse rate given that the temperature and pressure at the ground are $p = 10^5$ Pa and $T = 300$ K, respectively.

7.10. Consider two air masses at the uniform potential temperatures $\theta_1 = 320\,\mathrm{K}$ and $\theta_2 = 340\,\mathrm{K}$ that are separated by a vertical partition as shown in Figure 7.7. Each air mass occupies a horizontal area of 10^4 m^2 and extends from the surface ($p_0 = 10^5$ Pa) to the top of the atmosphere. What is the available potential energy for this system? What fraction of the total potential energy is available in this case?

7.11. For the unstable baroclinic wave that satisfies the conditions given in Problems 7.7 and 7.8, compute the energy conversion terms in (7.37) and (7.38) and thus obtain the instantaneous rates of change of the perturbation kinetic and available potential energies.

7.12. Starting with (7.62) and (7.64), derive the phase speed c for the Eady wave given in (7.70).

7.13. Unstable baroclinic waves play an important role in the global heat budget by transferring heat poleward. Show that for the Eady wave solution the poleward heat flux averaged over a wavelength,

$$\overline{v'T'} = \frac{1}{L}\int_0^L v'T'dx$$

is independent of height and is positive for a growing wave. How does the magnitude of the heat flux at a given instant change if the mean wind shear is doubled?

7.14. Assuming that the coefficient A in (7.69) is real, obtain an expression for the geostrophic streamfunction $\psi'(x, y, z^*, t)$ for the most unstable mode in the Eady stability problem for the case $k = l$. Use this result to derive an expression for the corresponding vertical velocity w^* in terms of A.

7.15. For the neutral baroclinic wave disturbance in the two-layer model given by (7.85a,b), derive the corresponding ω'_2 field. Describe how the convergence and divergence fields associated with this secondary circulation influence the evolution of the disturbance.

7.16. For the situation of Problem 7.15, derive expressions for the conversion of zonal available potential energy to eddy available potential energy and the conversion of eddy available potential energy to eddy kinetic energy.

MATLAB Exercises

M7.1. By varying the input zonal wavelength in the MATLAB script **twolayer_model_1A.m**, find the shortest zonal wavelength for which the two-layer model is baroclinically unstable and the wavelength of maximum instability (i.e., most rapid growth rate) in the case where the basic state thermal wind is 15 m s^{-1}. The case given corresponds to the situation given in Section 7.2.1. The code of **twolayer_model_1B.m** differs only in that a finite meridional width is assumed with meridional wave number $m = \pi/(3000\,\mathrm{km})$. Compare the zonal wavelengths for short-wave cutoff and for maximum growth rate for these two cases.

M7.2. Use the MATLAB script **twolayer_model_2A.m** to examine the transient growth associated with a disturbance that is initially entirely confined to the 250-hPa level. Let the basic state thermal wind be 25m s^{-1}. By examining zonal wavelengths shorter than the instability cutoff determined in Exercise M7.1, find the zonal wavelength that gives the largest amplification of the disturbance. Repeat this exercise using the script **twolayer_model_2B.m**, which includes a meridional dependence with wave number $m=\pi/(3000\ \mathrm{km})$. Note that the program will terminate if you choose a wavelength that is long enough to be baroclinically unstable. Discuss the vertical structure of the transiently growing stable modes. What sort of situation in the real atmosphere does this solution crudely model?

M7.3. The MATLAB script **twolayer_model_3b.m** shows the propagation of a wave packet composed of the sum of nine eastward propagating neutral waves of the two-layer model, ranging in wave number from 0.6 to 1.6 k, where $k = 2\pi/L$ and $L = 1850$ km. Run the script and estimate from the animation the characteristic phase velocity and the characteristic group velocity for the cases of $U_T = 15$ and 30 m s^{-1}.

M7.4. Suppose that a meridional dependence is included in the eddy fields in the two-layer model so that (7.17) becomes

$$\psi_m = A\cos(my)\,e^{ik(x-ct)}; \quad \psi_T = B\cos(my)\,e^{ik(x-ct)}$$

where $m = 2\pi/L_y$ and $L_y = 3000\,\mathrm{km}$. Determine the zonal wavelength of the most unstable wave. Solve for the fields ψ_m, ψ_T, and ω_2' and plot (x, y) cross-sections of these fields using the MATLAB script **contour_sample.m** as a template. *Hint:* The solutions in this case are similar to those of Problems 7.6 through 7.8 but with k^2 replaced by $k^2 + m^2$ everywhere.

M7.5. The MATLAB script **eady_model_1.m** shows vertical and meridional cross-sections of the solution for the Eady model (Section 7.4.3) for the most unstable wave mode at time $t = 0$. Modify this script to plot the solution for the Eady wave corresponding to the neutral stability condition given in (7.71). Explain the vertical structure in this case in terms of quasi-geostrophic theory.

---

[1] The presence of the free internal Rossby wave should actually be regarded as a weakness of the two-level model. Lindzen et al. (1968) have shown that this mode does not correspond to any free oscillation of the real atmosphere. Rather, it is a spurious mode resulting from the use of the upper boundary condition $\omega = 0$ at $p = 0$, which formally turns out to be equivalent to putting a lid at the top of the atmosphere.

[2] A zonal average generally designates the average around an entire circle of latitude. However, for a disturbance consisting of a single sinusoidal wave of wave number $k = m/(a\cos\phi)$, where m is an integer, the average over a wavelength is identical to a zonal average.

[3] This interpretation is based on energy, which, as shown in Section 7.3.2, is proportional to ψ^2.

Chapter 8

The Planetary Boundary Layer

The *planetary boundary layer* is that portion of the atmosphere in which the flow field is strongly influenced directly by interaction with Earth's surface. Ultimately this interaction depends on molecular viscosity. It is, however, only within a few millimeters of the surface, where vertical shears are very intense, that molecular diffusion is comparable to other terms in the momentum equation. Outside this *viscous sublayer* molecular diffusion is not important in the boundary layer equations for the mean wind, although it is still important for small-scale turbulent eddies. However, viscosity still has an important indirect role; it causes the velocity to vanish at the surface. As a consequence of this *no-slip boundary condition*, even a fairly weak wind will cause a large-velocity shear near the surface, which continually leads to the development of turbulent eddies.

These turbulent motions have spatial and temporal variations at scales much smaller than those resolved by the meteorological observing network. Such shear-induced eddies, together with convective eddies caused by surface heating, are very effective in transferring momentum to the surface and transferring heat (latent and sensible) away from the surface at rates many orders of magnitude faster than can be done by molecular processes. The depth of the planetary boundary layer produced by this turbulent transport may range from as little as 30m in conditions of large static stability to more than 3km in highly convective conditions. For average midlatitude conditions the planetary boundary layer extends through the lowest kilometer of the atmosphere and thus contains about 10% of the mass of the atmosphere. Given the modest vertical displacement of air parcels, we apply the Boussinesq approximation of Section 2.8 in subsequent analysis.

The dynamical structure of the flow in the planetary boundary layer is not produced directly by viscosity. Rather, it is largely determined by the fact that the atmospheric flow is turbulent. In the *free atmosphere* (i.e., the region above the planetary boundary layer), this turbulence can be ignored in an approximate treatment of synoptic-scale motions, except perhaps in the vicinity of jet streams, fronts, and convective clouds. However, in the boundary layer the dynamical equations of the previous chapters must be modified to properly represent the effects of turbulence.

8.1 Atmospheric Turbulence

Turbulent flow contains irregular quasi-random motions spanning a continuous spectrum of spatial and temporal scales. Such eddies cause nearby air parcels to drift apart and thus mix properties such as momentum and potential temperature across the boundary layer. Unlike the large-scale rotational flows discussed in earlier chapters, which have depth scales that are small compared to their horizontal scales, the turbulent eddies of concern in the planetary boundary layer tend to have similar scales in the horizontal and vertical. The maximum eddy length scale is thus limited by the boundary layer depth to be about 10^3 m. The minimum length scale (10^{-3} m) is that of the smallest eddies that can exist in the presence of diffusion by molecular friction.

Even when observations are taken with very short temporal and spatial separations, a turbulent flow will always have scales that are unresolvable because they have frequencies greater than the observation frequency and spatial scales smaller than the scale of separation of the observations. Outside the boundary layer, in the free atmosphere, the problem of unresolved scales of motion is usually not a serious one for the diagnosis or forecasting of synoptic- and larger-scale circulations (although it is crucial for the mesoscale circulations discussed in Chapter 9). The eddies that contain the bulk of the energy in the free atmosphere are resolved by the synoptic network. However, in the boundary layer, unresolved turbulent eddies are of critical importance. Through their transport of heat and moisture away from the surface, they maintain the surface energy balance, and through their transport of momentum to the surface, they maintain the momentum balance. The latter process dramatically alters the momentum balance of the large-scale flow in the boundary layer so that geostrophic balance is no longer an adequate approximation to the large-scale wind field. It is this aspect of boundary layer dynamics that is of primary importance for dynamical meteorology.

8.1.1 Reynolds Averaging

In a turbulent fluid, a field variable such as velocity measured at a point generally fluctuates rapidly in time as eddies of various scales pass the point. In order that measurements be truly representative of the large-scale flow, it is necessary to average the flow over an interval of time long enough to average out small-scale eddy fluctuations, but still short enough to preserve the trends in the large-scale flow field. To do this we assume that the field variables can be separated into slowly varying mean fields and rapidly varying turbulent components. Following the scheme introduced by Reynolds, we then assume that for any field variables, w and θ, for example, the corresponding means are indicated by overbars and the fluctuating components by primes. Thus, $w = \overline{w} + w'$ and $\theta = \overline{\theta} + \theta'$.

By definition, the means of the fluctuating components vanish; the product of a deviation with a mean also vanishes when the time average is applied. Thus,

$$\overline{w'\overline{\theta}} = \overline{w'}\,\overline{\theta} = 0$$

where we have used the fact that a mean variable is constant over the period of averaging. The average of the product of deviation components (called the *covariance*) does not generally vanish. Thus, for example, if on average the turbulent vertical velocity is upward where the potential temperature deviation is positive, and downward where it is negative, the product $\overline{w'\theta'}$ is positive and the variables are said to be correlated positively.

These averaging rules imply that the average of the product of two variables will be the product of the average of the means plus the product of the average of the deviations:

$$\overline{w\theta} = \overline{(\overline{w} + w')(\overline{\theta} + \theta')} = \overline{w}\,\overline{\theta} + \overline{w'\theta'}$$

Before applying the Reynolds decomposition to the Boussinesq equations (2.56) through (2.59), it is convenient to rewrite the total derivatives in each equation in flux form. For example, the term on the left in (2.56) may be manipulated with the aid of the continuity equation (2.60) and the chain rule of differentiation to yield

$$\begin{aligned}\frac{Du}{Dt} &= \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} + u\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) \\ &= \frac{\partial u}{\partial t} + \frac{\partial u^2}{\partial x} + \frac{\partial uv}{\partial y} + \frac{\partial uw}{\partial z}\end{aligned} \tag{8.1}$$

Separating each dependent variable into mean and fluctuating parts, substituting into (8.1), and averaging then yields

$$\overline{\frac{Du}{Dt}} = \frac{\partial \bar{u}}{\partial t} + \frac{\partial}{\partial x}\left(\bar{u}\bar{u} + \overline{u'u'}\right) + \frac{\partial}{\partial y}\left(\bar{u}\bar{v} + \overline{u'v'}\right) + \frac{\partial}{\partial z}\left(\bar{u}\bar{w} + \overline{u'w'}\right) \tag{8.2}$$

Noting that the mean velocity fields satisfy the continuity equation (2.60), we can rewrite (8.2) as

$$\overline{\frac{Du}{Dt}} = \frac{\bar{D}\bar{u}}{Dt} + \frac{\partial}{\partial x}\left(\overline{u'u'}\right) + \frac{\partial}{\partial y}\left(\overline{u'v'}\right) + \frac{\partial}{\partial z}\left(\overline{u'w'}\right) \tag{8.3}$$

where

$$\frac{\bar{D}}{Dt} = \frac{\partial}{\partial t} + \bar{u}\frac{\partial}{\partial x} + \bar{v}\frac{\partial}{\partial y} + \bar{w}\frac{\partial}{\partial z}$$

is the rate of change following the mean motion. The mean equations thus have the form

$$\frac{\bar{D}\bar{u}}{Dt} = -\frac{1}{\rho_0}\frac{\partial \bar{p}}{\partial x} + f\bar{v} - \left[\frac{\partial \overline{u'u'}}{\partial x} + \frac{\partial \overline{u'v'}}{\partial y} + \frac{\partial \overline{u'w'}}{\partial z}\right] + \bar{F}_{rx} \tag{8.4}$$

$$\frac{\bar{D}\bar{v}}{Dt} = -\frac{1}{\rho_0}\frac{\partial \bar{p}}{\partial y} - f\bar{u} - \left[\frac{\partial \overline{u'v'}}{\partial x} + \frac{\partial \overline{v'v'}}{\partial y} + \frac{\partial \overline{v'w'}}{\partial z}\right] + \bar{F}_{ry} \tag{8.5}$$

$$\frac{\bar{D}\bar{w}}{Dt} = -\frac{1}{\rho_0}\frac{\partial \bar{p}}{\partial z} + g\frac{\bar{\theta}}{\theta_0} - \left[\frac{\partial \overline{u'w'}}{\partial x} + \frac{\partial \overline{v'w'}}{\partial y} + \frac{\partial \overline{w'w'}}{\partial z}\right] + \bar{F}_{rz} \tag{8.6}$$

$$\frac{\bar{D}\bar{\theta}}{Dt} = -\bar{w}\frac{d\theta_0}{dz} - \left[\frac{\partial \overline{u'\theta'}}{\partial x} + \frac{\partial \overline{v'\theta'}}{\partial y} + \frac{\partial \overline{w'\theta'}}{\partial z}\right] \tag{8.7}$$

$$\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{v}}{\partial y} + \frac{\partial \bar{w}}{\partial z} = 0 \tag{8.8}$$

The various covariance terms in square brackets in (8.4) through (8.7) represent turbulent fluxes. For example, $\overline{w'\theta'}$ is a vertical turbulent heat flux in kinematic form. Similarly $\overline{w'u'} = \overline{u'w'}$ is a vertical turbulent flux of zonal momentum. For many boundary layers the magnitudes of the turbulent flux divergence terms are of the same order as the other terms in (8.4) through (8.7). In such cases, it is not possible to neglect the turbulent flux terms even when only the mean flow is of direct interest. Outside the boundary layer the turbulent fluxes are often sufficiently weak so that the terms in square brackets in (8.4) through (8.7) can be neglected in the analysis of large-scale flows. This assumption was implicitly made in Chapters 3 and 4.

The complete equations for the mean flow (8.4) through (8.8), unlike the equations for the total flow (2.56) through (2.60), and the approximate equations of Chapters 3 and 4, are not a closed set since in addition to the five unknown mean variables $\bar{u}, \bar{v}, \bar{w}, \bar{\theta}, \bar{p}$, there are unknown turbulent fluxes. To solve these equations, *closure* assumptions must be made to approximate the unknown fluxes in terms of the five known mean state variables. Away from regions with horizontal inhomogeneities (e.g., shorelines, towns, forest edges), we can simplify by assuming that the turbulent fluxes are horizontally homogeneous so that it is possible to neglect the horizontal derivative terms in square brackets in comparison to the terms involving vertical differentiation.

8.2 Turbulent Kinetic Energy

Vortex stretching and twisting associated with turbulent eddies always tend to cause turbulent energy to flow toward the smallest scales, where it is dissipated by viscous diffusion. Thus, there must be continuing production of turbulence if the turbulent kinetic energy is to remain statistically steady. The primary source of boundary layer turbulence depends critically on the structure of the wind and temperature profiles near the surface. If the lapse rate is unstable, boundary layer turbulence is convectively generated. If it is stable, then instability associated with wind shear must be responsible for generating turbulence in the boundary layer. The comparative roles of these processes can best be understood by examining the budget for turbulent kinetic energy.

To investigate the production of turbulence, we subtract the component mean momentum equations (8.4) through (8.6) from the corresponding unaveraged equations (2.56), (2.57), and (2.58). We then multiply the results by u', v', w', respectively; add the resulting three equations; and average to obtain the turbulent kinetic energy equation. The complete statement of this equation is quite complicated, but its essence can be expressed symbolically as

$$\frac{\bar{D}(TKE)}{Dt} = MP + BPL + TR - \varepsilon \tag{8.9}$$

where $TKE \equiv (\overline{u'^2} + \overline{v'^2} + \overline{w'^2})/2$ is the turbulent kinetic energy per unit mass, *MP* is the mechanical production, *BPL* is the buoyant production or loss, *TR* designates redistribution by transport and pressure forces, and ε designates frictional dissipation. ε is always positive, reflecting the dissipation of the smallest scales of turbulence by molecular viscosity.

The buoyancy term in (8.9) represents a conversion of energy between mean flow potential energy and turbulent kinetic energy. It is positive for motions that lower the center of mass of the atmosphere and negative for motions that raise it. The buoyancy term has the form[1]

$$BPL \equiv \overline{w'\theta'}\,(g/\theta_0)$$

Positive buoyancy production occurs when there is heating at the surface so that an unstable temperature lapse rate (see Section 2.7.2) develops near the ground and spontaneous convective overturning can occur. As shown in the schematic of Figure 8.1, convective eddies have positively correlated vertical velocity and potential temperature fluctuations and hence provide a source of turbulent kinetic energy and positive heat flux. This is the dominant source in a convectively unstable boundary layer. For a statically stable atmosphere, *BPL* is negative, which tends to reduce or eliminate turbulence.

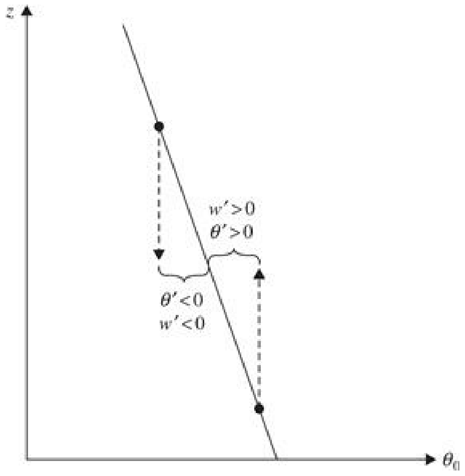


Figure 8.1 Correlation between vertical velocity and potential temperature perturbations for upward or downward parcel displacements when the mean potential temperature $\theta_0(z)$ decreases with height.

For both statically stable and unstable boundary layers, turbulence can be generated mechanically by dynamical instability due to wind shear. This process is represented by the mechanical production term in (8.9), which represents a conversion of energy between mean flow and turbulent fluctuations. This term is proportional to the shear in the mean flow and has the form

$$MP \equiv -\overline{u'w'}\frac{\partial \bar{u}}{\partial z} - \overline{v'w'}\frac{\partial \bar{v}}{\partial z} \tag{8.10}$$

MP is positive when the momentum flux is directed down the gradient of the mean momentum. Thus, if the mean vertical shear near the surface is westerly $(\partial \bar{u}/\partial z > 0)$, then $\overline{u'w'} < 0$ for *MP* > 0.

In a statically stable layer, turbulence can exist only if mechanical production is large enough to overcome the damping effects of stability and viscous dissipation. This condition is measured by a quantity called the *flux Richardson number*. It is defined as

$$Rf \equiv -BPL/MP$$

If the boundary layer is statically unstable, then Rf $<$ 0 and turbulence is sustained by convection. For stable conditions, Rf will be greater than zero. Observations suggest that only when Rf is less than about 0.25 (i.e., mechanical production exceeds buoyancy damping by a factor of 4) is the mechanical production intense enough to sustain turbulence in a stable layer. Since *MP* depends on the shear, it always becomes large close enough to the surface. However, as the static stability increases, the depth of the layer in which there is net production of turbulence shrinks. Thus, when there is a strong temperature inversion, such as produced by nocturnal radiative cooling, the boundary layer depth may be only a few decameters and vertical mixing is strongly suppressed.

8.3 Planetary Boundary Layer Momentum Equations

For the special case of horizontally homogeneous turbulence above the viscous sublayer, molecular viscosity and horizontal turbulent momentum flux divergence terms can be neglected. The mean flow horizontal momentum equations (8.4) and (8.5) then become

$$\frac{\bar{D}\bar{u}}{Dt} = -\frac{1}{\rho_0}\frac{\partial \bar{p}}{\partial x} + f\bar{v} - \frac{\partial \overline{u'w'}}{\partial z} \tag{8.11}$$

$$\frac{\bar{D}\bar{v}}{Dt} = -\frac{1}{\rho_0}\frac{\partial \bar{p}}{\partial y} - f\bar{u} - \frac{\partial \overline{v'w'}}{\partial z} \tag{8.12}$$

In general, (8.11) and (8.12) can only be solved for $\bar{u}$ and $\bar{v}$ if the vertical distribution of the turbulent momentum flux is known. Because this depends on the structure of the turbulence, no general solution is possible. Rather, a number of approximate semiempirical methods are used.

For midlatitude synoptic-scale motions, Section 2.4 showed that to a first approximation the inertial acceleration terms—the terms on the left in (8.11) and (8.12)—can be neglected compared to the Coriolis force and pressure gradient force terms. Outside the boundary layer, the resulting approximation is then simply geostrophic balance. In the boundary layer the inertial terms are still small compared to the Coriolis force and pressure gradient force terms, but the turbulent flux terms must be included. Thus, to a first approximation, planetary boundary layer equations express a three-way balance among the Coriolis force, the pressure gradient force, and the turbulent momentum flux divergence:

$$f\left(\bar{v} - \bar{v}_g\right) - \frac{\partial \overline{u'w'}}{\partial z} = 0 \tag{8.13}$$

$$-f\left(\bar{u} - \bar{u}_g\right) - \frac{\partial \overline{v'w'}}{\partial z} = 0 \tag{8.14}$$

where (2.23) is used to express the pressure gradient force in terms of geo-strophic velocity.

8.3.1 Well-Mixed Boundary Layer

If a convective boundary layer is topped by a stable layer, turbulent mixing can lead to formation of a well-mixed layer. Such boundary layers occur commonly over land during the day when surface heating is strong and over oceans when the air near the sea surface is colder than the surface water temperature. The tropical oceans typically have boundary layers of this type.

In a well-mixed boundary layer, the wind speed and potential temperature are nearly independent of height, as shown schematically in Figure 8.2, and to a first approximation it is possible to treat the layer as a slab in which the velocity and potential temperature profiles are constant with height and turbulent fluxes vary linearly with height. For simplicity, we assume that the turbulence vanishes at the top of the boundary layer. Observations indicate that the surface momentum flux can be represented by a *bulk aerodynamic formula*[2]:

$$\left(\overline{u'w'}\right)_s = -C_d\left|\overline{\mathbf{V}}\right|\bar{u}, \quad \text{and} \quad \left(\overline{v'w'}\right)_s = -C_d\left|\overline{\mathbf{V}}\right|\bar{v}$$

where C_d is a nondimensional *drag coefficient*, $|\overline{\mathbf{V}}| = h(\overline{u}^2 + \overline{v}^2)^{1/2}$, and the subscript s denotes surface values (referred to the standard anemometer height). Observations show that C_d is of order

1.5×10^{-3} over oceans, but may be several times as large over rough ground.

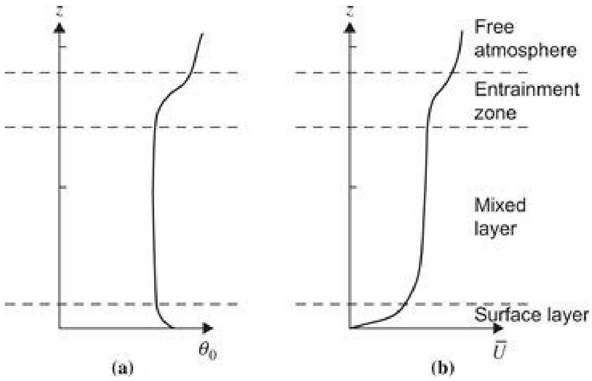


Figure 8.2 (a) Mean potential temperature, θ_0, and (b) mean zonal wind, U, profiles in a well-mixed boundary layer.

Adapted from Stull, 1988.

The approximate planetary boundary layer equations (8.13) and (8.14) can then be integrated from the surface to the top of the boundary layer at $z = h$ to give

$$f(\bar{v}-\bar{v}_g) = -\left(\overline{u'w'}\right)_s/h = C_d\left|\overline{\mathbf{V}}\right|\bar{u}/h \tag{8.15}$$

$$-f(\bar{u}-\bar{u}_g) = -\left(\overline{v'w'}\right)_s/h = C_d\left|\overline{\mathbf{V}}\right|\bar{v}/h \tag{8.16}$$

Without loss of generality we can choose axes such that $\bar{v}_g = 0$. Then (8.15) and (8.16) can be rewritten as

$$\bar{v} = \kappa_s\left|\overline{\mathbf{V}}\right|\bar{u}; \quad \bar{u} = \bar{u}_g - \kappa_s\left|\overline{\mathbf{V}}\right|\bar{v} \tag{8.17}$$

where $\kappa_s \equiv C_d/(fh)$. Thus, in the mixed layer the wind speed is less than the geostrophic speed and there is a component of motion directed toward lower pressure (i.e., to the left of the geostrophic wind in the Northern Hemisphere and to the right in the Southern Hemisphere) whose magnitude depends on κ_s. For example, if $\bar{u}_g = 10$ m s^{-1} and $\kappa_s = 0.05$ m^{-1} s, then $\bar{u} = 8.28$ m s^{-1}, $\bar{v} = 3.77$ m s^{-1}, and $\left|\overline{\mathbf{V}}\right| = 9.10$ m s^{-1} at all heights within this idealized slab mixed layer.

It is the work done by the flow toward lower pressure that balances the frictional dissipation at the surface. Because boundary layer turbulence tends to reduce wind speeds, the turbulent momentum flux terms are often referred to as boundary layer *friction*. It should be kept in mind, however, that the forces involved are due to turbulence, not molecular viscosity.

Qualitatively, the cross-isobar flow in the boundary layer can be understood as a direct result of the three-way balance among the pressure gradient force, the Coriolis force, and turbulent drag:

$$f\mathbf{k} \times \overline{\mathbf{V}} = -\frac{1}{\rho_0}\nabla\bar{p} - \frac{C_d}{h}\left|\overline{\mathbf{V}}\right|\overline{\mathbf{V}} \tag{8.18}$$

This balance is illustrated in Figure 8.3. Because the Coriolis force is always normal to the velocity and the turbulent drag is a retarding force, their sum can exactly balance the pressure gradient force only if the wind is directed toward lower pressure. Furthermore, it is easy to see that as the turbulent drag becomes increasingly dominant, the cross-isobar angle must increase.

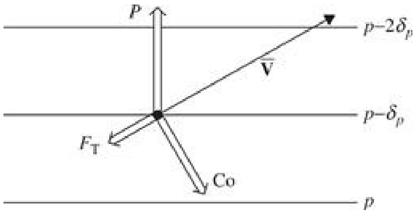


Figure 8.3 Balance of forces in the well-mixed planetary boundary layer: P designates the pressure gradient force; Co, the Coriolis force; and F_T, the turbulent drag.

8.3.2 The Flux–Gradient Theory

In neutrally or stably stratified boundary layers, the wind speed and direction vary significantly with height. The simple slab model is no longer appropriate; some means is needed to determine the vertical dependence of the turbulent momentum flux divergence in terms of mean variables in order to obtain closed equations for the boundary layer variables. The traditional approach to this closure problem is to assume that turbulent eddies act in a manner analogous to molecular diffusion so that the flux of a given field is proportional to the local gradient of the mean. In this case the turbulent flux terms in (8.13) and (8.14) are written as

$$\overline{u'w'} = -K_m\left(\frac{\partial \bar{u}}{\partial z}\right); \quad \overline{v'w'} = -K_m\left(\frac{\partial \bar{v}}{\partial z}\right)$$

and the potential temperature flux can be written as

$$\overline{\theta' w'} = -K_h\left(\frac{\partial \bar{\theta}}{\partial z}\right)$$

where $K_m(\mathrm{m}^2\mathrm{s}^{-1})$ is the *eddy viscosity* coefficient and K_h is the eddy diffusivity of heat. This closure scheme is often referred to as the K theory.

The K theory has many limitations. Unlike the molecular viscosity coefficient, eddy viscosities depend on the flow rather than the physical properties of the fluid and must be determined empirically for each situation. The simplest models have assumed that the eddy exchange coefficient is constant throughout the flow. This approximation may be adequate for estimating the small-scale diffusion of passive tracers in the free atmosphere. However, it is a very poor approximation in the boundary layer where the scales and intensities of typical turbulent eddies are strongly dependent on the distance to the surface as well as the static stability. Furthermore, in many cases the most energetic eddies have dimensions comparable to the boundary layer depth, and neither the momentum flux nor the heat flux is proportional to the local gradient of the mean. For example, in much of the mixed layer, heat fluxes are positive even though the mean stratification may be very close to neutral.

8.3.3 The Mixing Length Hypothesis

The simplest approach to determining a suitable model for the eddy diffusion coefficient in the boundary layer is based on the mixing length hypothesis introduced by the famous fluid dynamicist

Ludwig Prandtl. This hypothesis assumes that a parcel of fluid displaced vertically will carry the mean properties of its original level for a characteristic distance ξ' and then will mix with its surroundings just as an average molecule travels a mean free path before colliding and exchanging momentum with another molecule. By further analogy to the molecular mechanism, this displacement is postulated to create a turbulent fluctuation whose magnitude depends on ξ' and the gradient of the mean property. For example,

$$\theta' = -\xi'\frac{\partial\bar{\theta}}{\partial z}; \quad u' = -\xi'\frac{\partial\bar{u}}{\partial z}; \quad v' = -\xi'\frac{\partial\bar{v}}{\partial z}$$

where it must be understood that $\xi' > 0$ for upward parcel displacement and $\xi' < 0$ for downward parcel displacement. For a conservative property, such as potential temperature, this hypothesis is reasonable provided that the eddy scales are small compared to the mean flow scale or that the mean gradient is constant with height. However, the hypothesis is less justified in the case of velocity, as pressure gradient forces may cause substantial changes in the velocity during an eddy displacement.

Nevertheless, if we use the mixing length hypothesis, the vertical turbulent flux of zonal momentum can be written as

$$-\overline{u'w'} = \overline{w'\xi'}\frac{\partial\bar{u}}{\partial z} \tag{8.19}$$

with analogous expressions for the momentum flux in the meridional direction and the potential temperature flux. In order to estimate w' in terms of mean fields, we assume that the vertical stability of the atmosphere is nearly neutral so that buoyancy effects are small. The horizontal scale of the eddies should then be comparable to the vertical scale so that $|w'| \sim |\mathbf{V}'|$ and we can set

$$w' \approx \xi'\left|\frac{\partial\overline{\mathbf{V}}}{\partial z}\right|$$

where $\mathbf{V}'$ and $\overline{\mathbf{V}}$ designate the turbulent and mean parts of the horizontal velocity field, respectively. Here the absolute value of the mean velocity gradient is needed because if $\xi' > 0$, then $w > 0$ (i.e., upward parcel displacements are associated with upward eddy velocities). Thus, the momentum flux can be written

$$-\overline{u'w'} = \overline{\xi'^2}\left|\frac{\partial\overline{\mathbf{V}}}{\partial z}\right|\frac{\partial\bar{u}}{\partial z} = K_m\frac{\partial\bar{u}}{\partial z} \tag{8.20}$$

where the eddy viscosity is now defined as $K_m = \overline{\xi'^2}\left|\partial\overline{\mathbf{V}}/\partial z\right| = \overline{\ell^2}\left|\partial\overline{\mathbf{V}}/\partial z\right|$ and the mixing length,

$$\ell \equiv \left(\overline{\xi'^2}\right)^{1/2}$$

is the root mean square parcel displacement, which is a measure of average eddy size. This result suggests that larger eddies and greater shears induce greater turbulent mixing.

8.3.4 The Ekman Layer

If the flux–gradient approximation is used to represent the turbulent momentum flux divergence terms in (8.13) and (8.14), and the value of K_m is taken to be constant, we obtain the equations of the classical Ekman layer:

(8.21)

$$K_m \frac{\partial^2 u}{\partial z^2} + f\left(v - v_g\right) = 0$$

(8.22)

$$K_m \frac{\partial^2 v}{\partial z^2} - f\left(u - u_g\right) = 0$$

where we have omitted the overbars because all fields are Reynolds-averaged.

The Ekman layer equations (8.21) and (8.22) can be solved to determine the height dependence of the departure of the wind field in the boundary layer from geostrophic balance. In order to keep the analysis as simple as possible, we assume that these equations apply throughout the depth of the boundary layer. The boundary conditions on u and v then require that both horizontal velocity components vanish at the ground and approach their geostrophic valuesThen the general solution far from the ground:

(8.23)

$$\begin{aligned} &u = 0,\ v = 0 \text{ at } z = 0 \\ &u \to u_g,\ v \to v_g \text{ as } z \to \infty \end{aligned}$$

To solve (8.21) and (8.22), we combine them into a single equation by first multiplying (8.22) by $i = (-1)^{1/2}$ and then adding the result to (8.21) to obtain a second-order equation in the complex velocity, $(u + iv)$:

(8.24)

$$K_m \frac{\partial^2 (u + iv)}{\partial z^2} - if\left(u + iv\right) = -if\left(u_g + iv_g\right)$$

For simplicity, we assume that the geostrophic wind is independent of height and that the flow is oriented so that the geostrophic wind is in the zonal direction $(v_g = 0)$. Then the general solution of (8.24) is

(8.25)

$$(u + iv) = A\exp\left[\left(if/K_m\right)^{1/2} z\right] + B\exp\left[-\left(if/K_m\right)^{1/2} z\right] + u_g$$

It can be shown that $\sqrt{i} = (1 + i)/\sqrt{2}$. Using this relationship and applying the boundary conditions of (8.23), we find that for the Northern Hemisphere $(f > 0)$, $A = 0$ and $B = -u_g$. Thus,

$$u + iv = -u_g \exp\left[-\gamma\left(1 + i\right) z\right] + u_g$$

where $\gamma = \left(f/2K_m\right)^{1/2}$.

Applying the Euler formula $\exp(-i\theta) = \cos\theta - i\sin\theta$ and separating the real from the imaginary part, we obtain for the Northern Hemisphere

(8.26)

$$u = u_g\left(1 - e^{-\gamma z}\cos\gamma z\right), \quad v = u_g e^{-\gamma z}\sin\gamma z$$

This solution is the famous Ekman spiral named for the Swedish oceanographer Vagn Walfrid

Ekman, who first derived an analogous solution for the surface wind drift current in the ocean. The structure of the solution (8.26) is best illustrated by a hodograph as shown in Figure 8.4, where the zonal and meridional components of the wind are plotted as functions of height. The heavy solid curve traced out in Figure 8.4 connects all the points corresponding to values of u and v in (8.26) for values of γz increasing as one moves away from the origin along the spiral. Arrows show the velocities for various values of γz marked at the arrow points. When $z = \pi/\gamma$, the wind is parallel to and nearly equal to the geostrophic value. It is conventional to designate this level as the top of the Ekman layer and to define the layer depth as $De \equiv \pi/\gamma$.

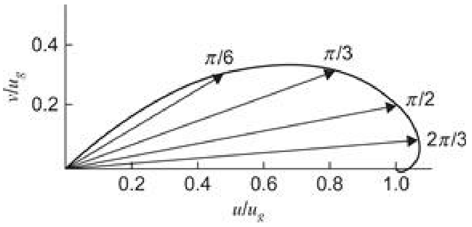


Figure 8.4 Hodograph of wind components in the Ekman spiral solution. *Arrows* show velocity vectors for several levels in the Ekman layer; the *spiral curve* traces out the velocity variation as a function of height. Points labeled on the spiral show the values of γz, which is a nondimensional measure of height.

Observations indicate that the wind approaches its geostrophic value at about 1 km above the surface. Letting $De = 1$ km and $f = 10^{-4}\mathrm{s}^{-1}$, the definition of γ implies that $K_m \approx 5\ \mathrm{m}^2\ \mathrm{s}^{-1}$. Referring back to (8.20), we see that for a characteristic boundary layer velocity shear of $|\delta \mathbf{V}/\delta z| \sim 5 \times 10^{-3}\ \mathrm{s}^{-1}$, this value of K_m implies a mixing length of about 30 m, which is small compared to the depth of the boundary layer, as it should be if the mixing length concept is to be useful.

Qualitatively, the most striking feature of the Ekman layer solution is that, like the mixed layer solution of Section 8.3.1, it has a boundary layer wind component directed toward lower pressure. As in the mixed layer case, this is a direct result of the three-way balance among the pressure gradient force, the Coriolis force, and the turbulent drag.

The ideal Ekman layer discussed here is rarely, if ever, observed in the atmospheric boundary layer. Partly this is because turbulent momentum fluxes are usually not simply proportional to the gradient of the mean momentum. However, even if the flux–gradient model were correct, it still would not be proper to assume a constant eddy viscosity coefficient, since in reality K_m must vary rapidly with height near the ground. Thus, the Ekman layer solution should not be carried all the way to the surface.

8.3.5 The Surface Layer

Some of the inadequacies of the Ekman layer model can be overcome if we distinguish a *surface layer* from the remainder of the planetary boundary layer. The surface layer, whose depth depends on stability, but is usually less than 10% of the total boundary layer depth, is maintained entirely by vertical momentum transfer by the turbulent eddies; it is not directly dependent on the Coriolis or pressure gradient forces. Analysis is facilitated by supposing that the wind close to the surface is directed parallel to the x axis. The kinematic turbulent momentum flux can then be expressed in terms of a *friction velocity*, u_*, which is defined as[3]

$$u_*^2 \equiv \left|\overline{(u'w')}_s\right|$$

Measurements indicate that the magnitude of the surface momentum flux is of the order 0.1 $\mathrm{m^2\,s^{-2}}$. Thus, the friction velocity is typically of the order 0.3 $\mathrm{m\,s^{-1}}$.

According to the scale analysis in Section 2.4, the Coriolis and pressure gradient force terms in (8.11) have magnitudes of about $10^{-3}\mathrm{m\,s^{-2}}$ in midlatitudes. The momentum flux divergence in the surface layer cannot exceed this magnitude or it would be unbalanced. Thus, it is necessary that

$$\frac{\delta\left(u_*^2\right)}{\delta z} \le 10^{-3}\,\mathrm{m\,s^{-2}}$$

For $\delta z = 10$ m, this implies that $\delta(u_*^2) \le 10^{-2}\,\mathrm{m^2\,s^{-2}}$ so that the change in the vertical momentum flux in the lowest 10 m of the atmosphere is less than 10% of the surface flux.

To a first approximation it is then permissible to assume that in the lowest several meters of the atmosphere the turbulent flux remains constant at its surface value, so with the aid of (8.20),

$$K_m \frac{\partial \bar{u}}{\partial z} = u_*^2 \tag{8.27}$$

where we have parameterized the surface momentum flux in terms of the eddy viscosity coefficient. In applying K_m in the Ekman layer solution, we assumed a constant value throughout the boundary layer. Near the surface, however, the vertical eddy scale is limited by the distance to the surface. Thus, a logical choice for the mixing length is $\ell = kz$, where k is a constant. In that case $K_m = (kz)^2\left|\partial\bar{u}/\partial z\right|$. Substituting this expression into (8.27) and taking the square root of the result gives

$$\partial\bar{u}/\partial z = u_*/(kz) \tag{8.28}$$

Integrating with respect to z yields the *logarithmic wind profile*:

$$\bar{u} = (u_*/k)\ln(z/z_0) \tag{8.29}$$

where z_0, the *roughness length*, is a constant of integration chosen so that $\bar{u} = 0$ at $z = z_0$. The constant k in (8.29) is a universal constant called *von Karman's constant*, which has an experimentally determined value of $k \approx 0.4$. The roughness length z_0 varies widely depending on the physical characteristics of the surface. For grassy fields, typical values are in the range of 1 to 4cm. Although a number of assumptions are required in the derivation of (8.29), many experimental programs have shown that the logarithmic profile provides a generally satisfactory fit to observed wind profiles in the surface layer.

8.3.6 The Modified Ekman Layer

As pointed out earlier, the Ekman layer solution is not applicable in the surface layer. A more satisfactory representation for the planetary boundary layer can be obtained by combining the logarithmic surface layer profile with the Ekman spiral. In this approach the eddy viscosity coefficient is again treated as a constant, but (8.24) is applied only to the region above the surface layer and the velocity and shear at the bottom of the Ekman layer are matched to those at the top of the surface layer. The resulting modified Ekman spiral provides a somewhat better fit to observations than the classical Ekman spiral. However, observed winds in the planetary boundary layer generally deviate substantially from the spiral pattern. Both transience and baroclinic effects (i.e., vertical shear of the geostrophic wind in the boundary layer) may cause deviations from the Ekman solution. However, even in steady-state barotropic situations with near neutral static stability, the Ekman spiral is seldom observed.

It turns out that the Ekman layer wind profile is generally unstable for a neutrally buoyant atmosphere. The circulations that develop as a result of this instability have horizontal and vertical scales comparable to the depth of the boundary layer. Thus, it is not possible to parameterize them by a simple flux–gradient relationship. However, these circulations do in general transport considerable momentum vertically. The net result is usually to decrease the angle between the boundary layer wind and the geostrophic wind from that characteristic of the Ekman spiral. A typical observed-wind hodograph is shown in Figure 8.5. Although the detailed structure is rather different from the Ekman spiral, the vertically integrated horizontal mass transport in the boundary layer is still directed toward lower pressure. As shown in the next section, this fact is of primary importance for synoptic- and larger-scale motions.

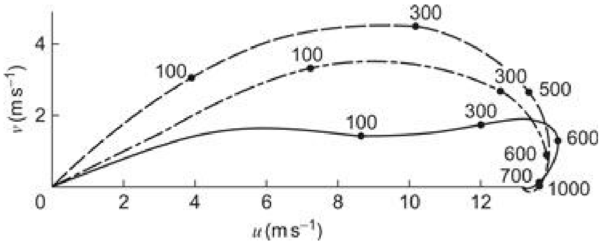


Figure 8.5 Mean wind hodograph for Jacksonville, Florida (≅30° N), April 4, 1968 (*solid line*) compared with the Ekman spiral (*dashed line*) and the modified Ekman spiral (*dash-dot line*) computed with *De* ≅ 1200 m. Heights are shown in meters.

Adapted from Brown, 1970. Copyright © American Meteorological Society. Reprinted with permission.

8.4 Secondary Circulations and Spin Down

Both the mixed layer solution (8.17) and the Ekman spiral solution (8.26) indicate that in the planetary boundary layer the horizontal wind has a component directed toward lower pressure. As suggested by Figure 8.6, this implies mass convergence in a cyclonic circulation and mass divergence in an anticyclonic circulation, which by mass continuity requires vertical motion out of and into the boundary layer, respectively. To estimate the magnitude of this induced vertical motion, we note that if $v_g = 0$, the cross-isobaric mass transport per unit area at any level in the boundary layer is given by $\rho_0 v$. The net mass transport for a column of unit width extending vertically through the entire layer is simply the vertical integral of $\rho_0 v$. For the mixed layer, this integral is simply $\rho_0 v h (\mathrm{kg\ m^{-1}\ s^{-1}})$, where h is the layer depth. For the Ekman spiral, it is given by

$$M = \int_0^{De} \rho_0 v\,dz = \int_0^{De} \rho_0 u_g \exp\left(-\pi z/De\right) \sin\left(\pi z/De\right) dz \tag{8.30}$$

where $De = \pi/\gamma$ is the Ekman layer depth defined in Section 8.3.4.

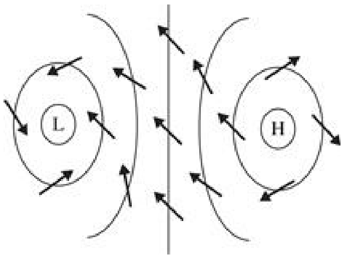


Figure 8.6 Schematic surface wind pattern (*arrows*) associated with high- and low-pressure centers in the Northern Hemisphere. Isobars are shown by thin lines, and "L" and "H" designate high- and low-pressure centers, respectively.

After Stull, 1988.

Integrating the mean continuity equation (8.8) through the depth of the boundary layer gives

$$w(De) = -\int_0^{De} \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) dz \tag{8.31}$$

where we have assumed that $w(0) = 0$. Assuming again that $v_g = 0$ so that u_g is independent of x, we find after substituting from (8.26) into (8.31) and comparing with (8.30) that the mass transport at the top of the Ekman layer is given by

$$\rho_0 w(De) = -\frac{\partial M}{\partial y} \tag{8.32}$$

Thus, the mass flux out of the boundary layer is equal to the convergence of the cross-isobar mass transport in the layer. Noting that $-\partial u_g/\partial y = \zeta_g$ is just the geostrophic vorticity in this case, we find after integrating (8.30) and substituting into (8.32) that[4]

$$w(De) = \zeta_g \left(\frac{1}{2\gamma}\right) = \zeta_g \left|\frac{K_m}{2f}\right|^{1/2} \left(\frac{f}{|f|}\right) \tag{8.33}$$

where we have neglected the variation of density with height in the boundary layer and have assumed that $1 + e^{-\pi} \approx 1$. Thus, we obtain the important result that the vertical velocity at the top of the boundary layer is proportional to the geostrophic vorticity. In this way the effect of boundary layer fluxes is communicated directly to the free atmosphere through a forced *secondary circulation* that usually dominates over turbulent mixing. This process is often referred to as *boundary layer pumping*. It only occurs in rotating fluids and is one of the fundamental distinctions between rotating and nonrotating flow. For a typical synoptic-scale system with $\zeta_g \sim 10^{-5}\ \mathrm{s}^{-1}$, $f \sim 10^{-4}\ \mathrm{s}^{-1}$, and $De \sim 1$ km, the vertical velocity given by (8.33) is of the order of a few millimeters per second.

An analogous boundary layer pumping is responsible for the decay of the circulation created when a cup of tea is stirred. Away from the bottom and sides of the cup there is an approximate balance between the radial pressure gradient and the centrifugal force of the spinning fluid. However, near the bottom, viscosity slows the motion and the centrifugal force is not sufficient to balance the radial pressure gradient. (Note that the radial pressure gradient is independent of depth, since water is an incompressible fluid.) Therefore, radial inflow takes place near the bottom of the cup. Because of this inflow, the tea leaves are always observed to cluster near the center at the bottom of the cup if the tea has been stirred. By continuity of mass, the radial inflow in the bottom boundary layer requires upward motion and a slow compensating outward radial flow throughout the remaining depth of the tea. This slow outward radial flow approximately conserves angular momentum, and by replacing high angular momentum fluid with low angular momentum fluid, it makes the vorticity in the cup spin down far more rapidly than could mere diffusion.

The characteristic time for the secondary circulation to spin down an atmospheric vortex is illustrated most easily in the case of a barotropic atmosphere. For synoptic-scale motions, equation (4.38) can be written approximately as

$$\frac{D\zeta_g}{Dt} = -f\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = f\frac{\partial w}{\partial z} \tag{8.34}$$

where we have neglected ζ_g compared to f in the divergence term and have also neglected the latitudinal variation of f. Recalling that the geostrophic vorticity is independent of height in a barotropic atmosphere, (8.34) can be integrated easily from the top of the Ekman layer ($z = De$) to the tropopause ($z = H$) to give

$$\frac{D\zeta_g}{Dt} = +f\left[\frac{w(H) - w(De)}{(H - De)}\right] \tag{8.35}$$

Substituting for $w(De)$ from (8.33), assuming that $w(H) = 0$ and that $H \gg De$, (8.35) may be written as

$$\frac{D\zeta_g}{Dt} = -\left|\frac{fK_m}{2H^2}\right|^{1/2}\zeta_g \tag{8.36}$$

This equation may be integrated in time to give

$$\zeta_g(t) = \zeta_g(0)\exp(-t/\tau_e) \tag{8.37}$$

where $\zeta_g(0)$ is the value of the geostrophic vorticity at time $t = 0$, and $\tau_e \equiv H\left|2/(fK_m)\right|^{1/2}$ is the time that it takes the vorticity to decrease to e^{-1} of its original value.

This *e-folding* time scale is referred to as the barotropic *spin-down time*. Taking typical values of the parameters as follows: $H \equiv 10$ km, $f = 10^{-4}$ s^{-1}, and $K_m = 10$ m^2 s^{-1}, we find that $\tau_e \approx 4$ days. Thus, for midlatitude synoptic-scale disturbances in a barotropic atmosphere, the characteristic spin-down time is a few days. This decay time scale should be compared to the time scale for ordinary viscous diffusion. For viscous diffusion the time scale can be estimated from scale analysis of the diffusion equation

$$\frac{\partial u}{\partial t} = K_m\frac{\partial^2 u}{\partial z^2} \tag{8.38}$$

If τ_d is the diffusive time scale and H is a characteristic vertical scale for diffusion, then from the diffusion equation

$$U/\tau_d \sim K_m U/H^2$$

so that $\tau_d \sim H^2/K_m$. For the preceding values of H and K_m, the diffusion time scale is thus about 100 days. Thus, in the absence of convective clouds the spin-down process is a far more effective mechanism than eddy diffusion for destroying vorticity in a rotating atmosphere. Cumulonimbus convection can produce rapid turbulent transports of heat and momentum through the entire depth of the troposphere. These must be considered together with boundary layer pumping for intense systems such as hurricanes.

Physically the spin-down process in the atmospheric case is similar to that described for the teacup, except that in synoptic-scale systems it is primarily the Coriolis force that balances the pressure gradient force away from the boundary, not the centrifugal force. Again, the role of the secondary circulation driven by forces resulting from boundary layer drag is to provide a slow radial flow in the interior that is superposed on the azimuthal circulation of the vortex above the boundary layer. This secondary circulation is directed outward in a cyclone so that the horizontal area enclosed by any chain of fluid particles gradually increases. Since the circulation is conserved, the azimuthal velocity at any distance from the vortex center must decrease in time or, from another point of view, the Coriolis force for the outward-flowing fluid is directed clockwise, and this force thus exerts a torque opposite to the direction of the circulation of the vortex. Figure 8.7 shows a qualitative sketch of the streamlines of this secondary flow.

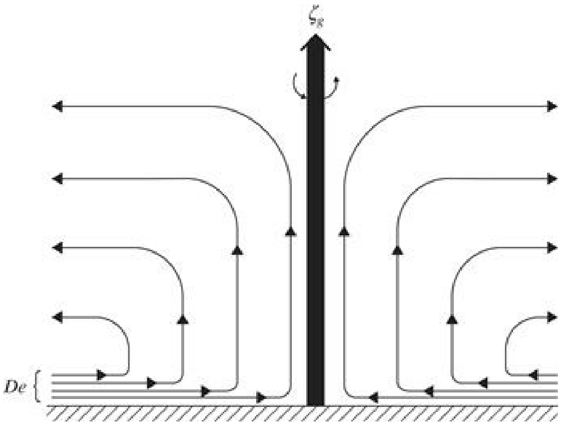


Figure 8.7 Streamlines of the secondary circulation forced by frictional convergence in the planetary boundary layer for a cyclonic vortex in a barotropic atmosphere. The circulation extends throughout the full depth of the vortex.

It should now be obvious exactly what is meant by the term *secondary circulation*. It is simply a circulation superposed on the primary circulation (in this case the azimuthal circulation of the vortex) by the physical constraints of the system. In the case of the boundary layer, viscosity is responsible for the presence of the secondary circulation. However, other processes, such as temperature advection and diabatic heating, may also lead to secondary circulations, as shown later.

The preceding discussion has concerned only the neutrally stratified barotropic atmosphere. An analysis for the more realistic case of a stably stratified baroclinic atmosphere is more complicated. However, qualitatively the effects of stratification may be easily understood. The buoyancy force (see Section 2.7.3) will act to suppress vertical motion, since air lifted vertically in a stable environment will be denser than the environmental air. As a result, the interior secondary circulation will decrease with altitude at a rate proportional to the static stability.

This vertically varying secondary flow, shown in Figure 8.8, will rather quickly spin down the vorticity at the top of the Ekman layer without appreciably affecting the higher levels. When the geostrophic vorticity at the top of the boundary layer is reduced to zero, the pumping action of the Ekman layer is eliminated. The result is a baroclinic vortex with a vertical shear of the azimuthal velocity that is just strong enough to bring ζ_g to zero at the top of the boundary layer. This vertical shear of the geostrophic wind requires a radial temperature gradient that is in fact produced during the spin-down phase by adiabatic cooling of the air forced out of the Ekman layer. Thus, the secondary circulation in the baroclinic atmosphere serves two purposes: (1) it changes the azimuthal velocity field of the vortex through the action of the Coriolis force, and (2) it changes the temperature distribution so that a thermal wind balance is always maintained between the vertical shear of the azimuthal velocity and the radial temperature gradient.

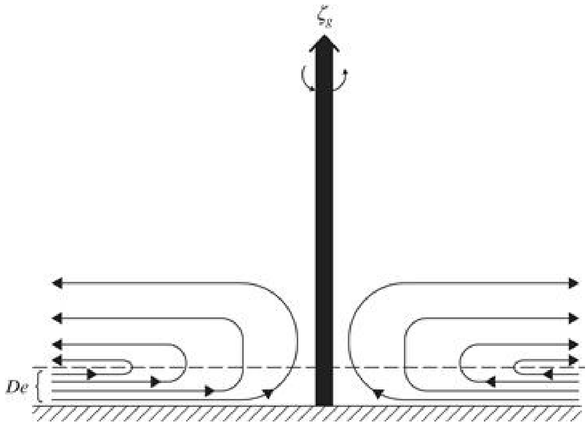


Figure 8.8 Streamlines of the secondary circulation forced by frictional convergence in the planetary boundary layer for a cyclonic vortex in a stably stratified baroclinic atmosphere. The circulation decays with height in the interior.

Suggested References

rya. Arya, *Introduction to Micrometeorology*, gives an excellent introduction to boundary layer dynamics and turbulence at the beginning undergraduate level.

arratt. Garratt, *The Atmospheric Boundary Layer*, is an excellent graduate-level introduction to the physics of the atmospheric boundary layer.

tull. Stull, *An Introduction to Boundary Layer Meteorology*, provides a comprehensive and very nicely illustrated treatment of all aspects of the subject at the beginning graduate level.

Problems

8.1. Verify by direct substitution that the Ekman spiral expression (8.26) is indeed a solution of the boundary layer equations (8.21) and (8.22) for the case $v_g = 0$.

8.2. Derive the Ekman spiral solution for the more general case where the geostrophic wind has both x and y components (u_g and v_g, respectively), which are independent of height.

8.3. Letting the Coriolis parameter and density be constants, show that (8.33) is correct for the more general Ekman spiral solution obtained in Problem 8.2.

8.4. For laminar flow in a rotating cylindrical vessel filled with water (molecular kinematic viscosity $\nu = 0.01\ \mathrm{cm^2\,s^{-1}}$), compute the depth of the Ekman layer and the spin-down time if the depth of the fluid is 30 cm and the rotation rate of the tank is 10 revolutions per minute. How small would the radius of the tank have to be in order for the time scale for viscous diffusion from the side walls to be comparable to the spin-down time?

8.5. Suppose that at 43°N the geostrophic wind is westerly at 15 m $\mathrm{s^{-1}}$. Compute the net cross-isobaric transport in the planetary boundary layer using both the mixed layer solution (8.17) and the Ekman layer solution (8.26). You may let $|\mathbf{V}| = u_g$ in equation (8.17), $h = De = 1$ km, $\kappa_s = 0.015\ \mathrm{m^{-1}s}$, and $\rho = 1\ \mathrm{kg\,m^{-3}}$.

8.6. Derive an expression for the wind-driven surface Ekman layer in the ocean. Assume that the wind stress τ_w is constant and directed along the x axis. The continuity of turbulent momentum flux at the air–sea interface ($z = 0$) requires that the wind stress divided by air density must equal the oceanic turbulent momentum flux at $z = 0$. Thus, if the flux–gradient theory is used, the boundary condition at the surface becomes

$$\rho_0 K \frac{\partial u}{\partial z} = \tau_w, \quad \rho_0 K \frac{\partial v}{\partial z} = 0, \quad \text{at } z = 0$$

where K is the eddy viscosity in the ocean (assumed constant). For a lower boundary condition, assume that $u, v \to 0$ as $z \to -\infty$. If $K = 10^{-3}\mathrm{m^2\,s^{-1}}$, what is the depth of the surface Ekman layer at 43°N latitude?

8.7. Show that the vertically integrated mass transport in the wind-driven oceanic surface Ekman layer is directed 90° to the right of the surface wind stress in the Northern Hemisphere. Explain this result physically.

8.8. A homogeneous barotropic ocean of depth H = 3 km has a zonally symmetric geostrophic jet whose profile is given by the expression

$$u_g = U \exp\left[-(y/L)^2\right]$$

where $U = 1\ \mathrm{ms^{-1}}$ and $L = 200$ km are constants. Compute the vertical velocity produced by convergence in the Ekman layer at the ocean bottom and show that the meridional profile of the secondary cross-stream motion forced in the interior is the same as the meridional profile of u_g. What are the maximum values of $\bar{v}$ in the interior and $\bar{w}$ if $K = 10^{-3}\mathrm{m^2\,s^{-1}}$ and $f = 10^{-4}\ \mathrm{s^{-1}}$ (assume that w and the eddy stress vanish at the surface).

8.9. Using the approximate zonally averaged momentum equation

$$\frac{\partial \bar{u}}{\partial t} \cong f\bar{v}$$

compute the spin-down time for the zonal jet in Problem 8.8.

8.10. Derive a formula for the vertical velocity at the top of the planetary boundary layer using the mixed layer expression (8.17). Assume that $|\mathbf{V}| = 5\,\mathrm{m\,s^{-1}}$ is independent of x and y and that $\bar{u}_g = \bar{u}_g(y)$. If $h = 1$ km and $\kappa_s = 0.05$, what value must K_m have if this result is to agree with the vertical velocity derived from the Ekman layer solution at 43° latitude with $De = 1$ km?

8.11. Show that $K_m = kzu_*$ in the surface layer.

MATLAB Exercises

M8.1. The MATLAB script **mixed_layer_wind1.m** uses a simple iterative technique to solve for u and v in (8.17) with u_g in the range 1–20 m s^{-1} for the case $v_g = 0$ and $\kappa_s = 0.05\ \mathrm{m}^{-1}\mathrm{s}$. If you run the script, you will observe that this iterative technique fails for u_g greater than 19 m s^{-1}. An alternative method, which works for a wide range of specified geostrophic winds, utilizes the natural coordinate system introduced in Section 3.2.1.

(a) Show that in the natural coordinate system the force balances parallel and perpendicular to the velocity vector in the mixed layer model (see Figure 8.3) can be expressed respectively as

$$f\kappa_s V^2 = f u_g \cos\beta \quad \text{and} \quad fV = f u_g \sin\beta$$

where it is assumed that the pressure gradient force is directed northward so that $fu_g = \left|\rho_0^{-1}\nabla p\right|$, and β designates the angle between the pressure gradient force vector and the mixed layer velocity, $\boldsymbol{V}$. (Other notation is as defined in Section 8.3.1.)

(b) Use MATLAB to solve for V and β given u_g in the range of 1 to 50m s^{-1}, and plot V and β versus u_g. *Hint*: Solve for V in the two equations in (a), and for each value of u_g, vary β until the two solutions for V agree.

M8.2. Suppose that the geopotential distribution at the top of the mixed layer can be expressed in the form $\Phi(x, y) = \Phi_0 - f_0 U_0 y + A \sin kx \sin ly$, where $\Phi_0 = 9800\,\mathrm{m}^2\,\mathrm{s}^{-2}$, $f_0 = 10^{-4}\mathrm{s}^{-1}$, $U_0 = 5\ \mathrm{m\,s}^{-1}$, $A = 1500\mathrm{m}^2\mathrm{s}^{-2}$, $k = \pi L^{-1}$, and, $l = \pi L^{-1}$, where $L = 6000$ km.

(a) Use the technique given in the demonstration script **mixed_layer_wind1.m** to determine the wind distribution in the mixed layer for this situation for the case where $\kappa_s = 0.05$.

(b) Using the formula obtained in Problem 8.10, compute the vertical velocity distribution at the top of the mixed layer for this distribution of geopotential when the depth of the mixed layer is 1km. (The MATLAB script **mixed_layer_wind_2.m** has a template that you can use to contour the vertical velocity and the vorticity fields.)

M8.3. For the geopotential distribution of Exercise M8.2, use the Ekman layer theory to derive the pattern of vertical velocity at the top of the boundary layer. Assume that $K_m = 10\ \mathrm{m}^2\ \mathrm{s}^{-1}$ and again use MATLAB to contour the fields of vorticity and vertical velocity. Explain why the vertical velocity patterns derived from the mixed layer and Ekman theories differ for this situation.

---

[1] In practice, buoyancy in the boundary layer is modified by the presence of water vapor, which has a density significantly lower than that of dry air. The potential temperature should be replaced by virtual potential temperature in (8.9) in order to include this effect. (See, for example, Curry and Webster, 1999, p. 67.)

[2] The turbulent momentum flux is often represented in terms of an "eddy stress" by defining, for example, $\tau_{ex} = \rho_o \overline{u'w'}$. We prefer to avoid this terminology to eliminate possible confusion with molecular friction.

[3] Thus, the surface eddy stress (see footnote 3) is equal to $\rho_0 u_*^2$.

[4] The ratio of the Coriolis parameter to its absolute value is included so that the formula will be valid in both hemispheres.

Chapter 9

Mesoscale Circulations

Earlier chapters focused primarily on the dynamics of synoptic- and planetary-scale circulations. Such large-scale motions are strongly influenced by the rotation of Earth so that outside the equatorial zone the Coriolis force dominates over inertia (i.e., the Rossby number is small). To first approximation, large-scale motions can be modeled by quasi-geostrophic theory (as shown in Chapter 6).

The study of quasi-geostrophic motions has been a central theme of dynamic meteorology for many years. Not all important circulations fit into the quasi-geostrophic classification, however. Some have Rossby numbers of order unity, and some are hardly influenced by Earth's rotation at all. Such circulations include a wide variety of phenomena. They all, however, are characterized by horizontal scales that are smaller than the synoptic scale (i.e., the *macroscale* of motion) but larger than the scale of an individual fair-weather cumulus cloud (i.e., the *microscale*). Hence, they can be classified conveniently as *mesoscale* circulations. Most severe weather is associated with mesoscale motion systems. Thus, understanding of the mesoscale is of both scientific and practical importance.

9.1 Energy Sources for Mesoscale Circulations

Mesoscale dynamics is generally defined to include the study of motion systems that have horizontal scales in the range of about 10 to 1000 km. It includes circulations ranging from thunderstorms and internal gravity waves at the small end of the scale to fronts and hurricanes at the large end. Given the diverse nature of mesoscale systems, it is not surprising that there is no single conceptual framework equivalent to the quasi-geostrophic theory that can provide a unified model for the dynamics of the mesoscale. Indeed, the dominant dynamical processes vary enormously depending on the type of mesoscale circulation system involved.

Possible sources of mesoscale disturbances include instabilities that occur intrinsically on the mesoscale, forcing by mesoscale thermal or topographic sources, nonlinear transfer of energy from either macroscale or microscale motions, and interaction of cloud physical and dynamical processes.

Although instabilities associated with the mean velocity or thermal structure of the atmosphere are a rich source of atmospheric disturbances, most instabilities have their maximum growth rates either on the large scale (baroclinic and most barotropic instability) or on the small scale (convection and Kelvin–Helmholtz instability). Only symmetric instability (to be discussed in Section 9.3) appears to be an intrinsically mesoscale instability.

Mountain waves created by flow over the individual peaks are generally regarded as small-scale phenomena. However, flow over large mountain ranges can produce orographic disturbances in the 10- to 100-km mesoscale range, whose characteristics depend on the mean wind and static stability profiles and the scale of the orography. Flow over mountain ranges, such as the Front Range of the Colorado Rockies, can under some conditions of mean flow and static stability lead to strong downslope wind storms.

Energy transfer from small scales to the mesoscale is a primary energy source for mesoscale convective systems. These may start as individual convective cells, which grow and combine to form thunderstorms; convective complexes such as squall lines and mesocyclones; and even hurricanes. Conversely, energy *transfer* from the large scale associated with temperature and vorticity advection in synoptic-scale circulations is responsible for the development of frontal circulations.

9.2 Fronts and Frontogenesis

In the discussion of baroclinic instability in Chapter 7, the mean thermal wind U_T was taken to be a constant independent of the y coordinate. That assumption was necessary to obtain a mathematically simple model that retained the basic instability mechanism. It was pointed out in Section 6.1, however, that baroclinicity is not distributed uniformly in the atmosphere. Rather, horizontal temperature gradients tend to be concentrated in baroclinic zones associated with tropospheric jet streams. Not surprisingly, the development of baroclinic waves is also concentrated in such regions.

We showed in Chapter 7 that the energetics of baroclinic waves require that they remove available potential energy from the mean flow. Thus, on average baroclinic wave development tends to weaken the meridional temperature gradient (i.e., reduce the mean thermal wind). The mean pole-to-equator temperature gradient is of course restored continually by differential solar heating, which maintains the time-averaged temperature gradient pattern. In addition, there are transient dynamical processes that produce zones with greatly enhanced temperature gradients within individual baroclinic eddies. Such zones, which are particularly intense at the surface, are referred to as *fronts*.

Processes that generate fronts are called *frontogenetic*. Frontogenesis usually occurs in association with developing baroclinic waves, which in turn are concentrated in the storm tracks associated with the time-mean jetstreams. Thus, even though on average baroclinic disturbances transport heat down the mean temperature gradient and tend to weaken the temperature difference between the polar and tropical regions, locally the flow associated with baroclinic disturbances may actually enhance the temperature gradient.

9.2.1 The Kinematics of Frontogenesis

A complete discussion of the dynamics of frontogenesis is beyond the scope of this text. A qualitative description of frontogenesis can be obtained, however, by considering the evolution of the temperature gradient when temperature is treated as a passive tracer in a specified horizontal flow field. Such an approach is referred to as *kinematic*; it considers the effects of advection on a field variable without reference to the underlying physical forces or to any influence of the advected tracer on the flow field.

The influence of a purely geostrophic flow on the temperature gradient was given in terms of the **Q** vector in (6.51). If for simplicity we focus on the meridional temperature gradient, then from (6.46) neglecting ageostrophic and diabatic effects gives

$$\frac{D_g}{Dt}\left(\frac{\partial T}{\partial y}\right) = -\left[\frac{\partial u_g}{\partial y}\frac{\partial T}{\partial x} - \frac{\partial u_g}{\partial x}\frac{\partial T}{\partial y}\right] \tag{9.1}$$

where we have used the fact that the geostrophic wind is nondivergent so that $\partial v_g/\partial y = -\partial u_g/\partial x$.

The two terms within the brackets on the right in (9.1) can be interpreted as the forcing of the meridional temperature gradient by horizontal shear deformation and stretching deformation, respectively.

Horizontal shear has two effects on a fluid parcel: It tends to rotate the parcel (due to shear vorticity) and to deform the parcel through stretching parallel to the shear vector (i.e., along the x axis in Figure 9.1a) and shrinking along the direction perpendicular to the shear vector. Thus, the x-directed temperature gradient in Figure 9.1a is both rotated into the positive y direction and intensified by the shear. Horizontal shear is an important frontogenetic mechanism in both cold and warm fronts. For example, in the schematic surface pressure chart of Figure 9.2, the geostrophic wind has a northerly component west of point B and a southerly component east of point B. The resulting cyclonic shear will tend to rotate the isotherms and to concentrate them along the line of maximum shear passing through B. (Note the strong cold advection northwest of B and the weak thermal advection southeast of B.)

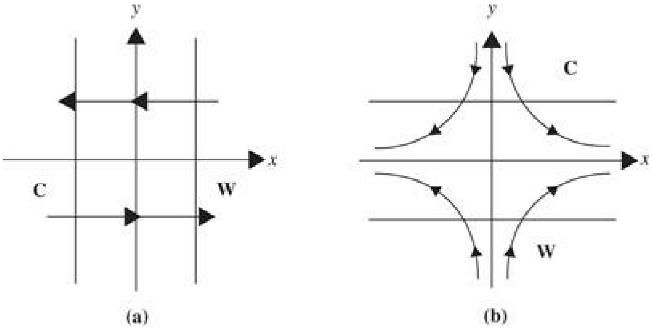


Figure 9.1 Frontogenetic flow configurations: (a) shows horizontal shearing deformation and (b) shows horizontal stretching deformation.

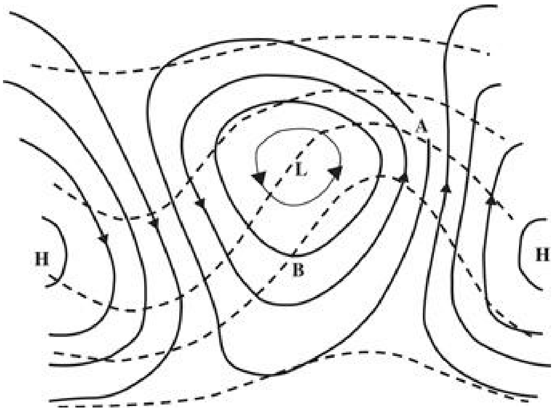


Figure 9.2 Schematic surface isobars (*solid lines*) and isotherms (*dashed lines*) for a baroclinic wave disturbance. *Arrows* show direction of geostrophic wind. Horizontal stretching deformation intensifies the temperature gradient at A, and horizontal shear deformation intensifies the gradient at B.

After Hoskins and Bretherton, 1972. Copyright © American Meteorological Society. Reprinted with permission.

Horizontal *stretching deformation* tends to advect the temperature field so that the isotherms become concentrated along the axis of *dilation* (the x axis in Figure 9.1b), provided that the initial temperature field has a finite gradient along the axis of *contraction* (the y axis in Figure 9.1b). That this effect is represented by the second term on the right in (9.1) can be verified by noting from Figure 9.1b that $\partial T/\partial y < 0$ and $\partial u_g/\partial x > 0$.

The velocity field shown in Figure 9.1b is a pure stretching deformation field. A pure deformation field is both irrotational and nondivergent. Thus, a parcel advected by a pure deformation field will merely have its shape changed in time, without any rotation or change in horizontal area. The deformation field of Figure 9.1b has a streamfunction given by $\psi = -Kxy$, where K is a constant. Such a field is characterized by the rate at which advection changes the shape of a horizontal area element. This can be illustrated by considering the rectangular element with sides δx and δy. The shape can be represented by the ratio $\delta x/\delta y$, and the fractional rate of change of shape can thus be expressed as

$$\frac{1}{(\delta x/\delta y)}\frac{D(\delta x/\delta y)}{Dt} = \frac{1}{\delta x}\frac{D\delta x}{Dt} - \frac{1}{\delta y}\frac{D\delta y}{Dt} = \frac{\delta u}{\delta x} - \frac{\delta v}{\delta y} \approx \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}$$

It is easily verified that the fractional rate of change of $\delta x/\delta y$ for the velocity field in Figure 9.1b equals $+2K$. Thus, a square parcel with sides parallel to the x and y axes would be deformed into a rectangle as the sides parallel to the x axis stretch and those parallel to the y axis contract in time at a constant rate.

Horizontal deformation at low levels is an important mechanism for the development of both cold and warm fronts. In the example of Figure 9.2 the flow near point A has $\partial u/\partial x > 0$ and $\partial v/\partial y < 0$ so that there is a stretching deformation field present with its axis of contraction nearly orthogonal to the isotherms. This deformation field leads to strong warm advection south of point A and a weak warm advection north of point A.

Although, as shown in Figure 9.2, the low-level flow in the vicinity of a developing warm front may resemble a pure deformation field, the total flow in the upper troposphere in baroclinic disturbances seldom resembles that of a pure deformation field due to the presence of strong mean westerlies. Rather, a combination of mean flow plus horizontal stretching deformation produces a *confluent* flow as shown in Figure 9.3. Such confluence acts to concentrate the cross-stream temperature gradient as parcels move downstream. Confluent regions are always present in the tropospheric jetstream due to the influence of quasi-stationary planetary-scale waves on the position and intensity of the jet. In fact, even a monthly mean 500-hPa chart (refer to Figure 6.3) reveals two regions of large-scale confluence immediately to the east of the continents of Asia and North America. Observationally, these two regions are known to be regions of intense baroclinic wave development and frontogenesis.

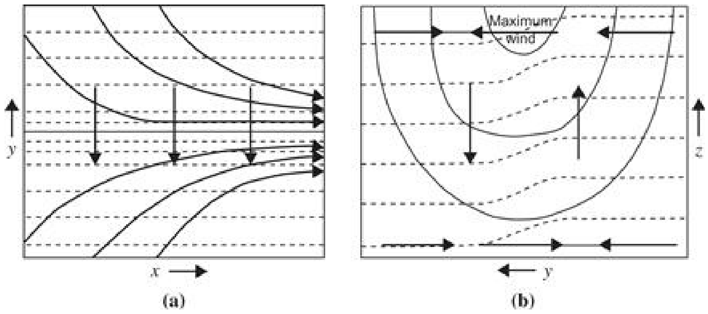


Figure 9.3 (a) Horizontal streamlines, isotherms, and Q-vectors in a frontogenetic confluence. (b) Vertical section across the confluence showing isotachs (*solid lines*), isotherms (*dashed lines*), and vertical and transverse motions (*arrows*).

After Sawyer, 1956. Copyright © The Royal Society. Reprinted with permission.

The mechanisms of horizontal shear and horizontal stretching deformation discussed previously operate to concentrate the pole–equator temperature gradient on the synoptic scale ($\sim$ 1000 km). The time scale on which these processes operate can be estimated with the aid of (9.1). Suppose that the geostrophic wind consists of a pure deformation field so that $u_g = Kx$, and $v_g = -Ky$, and let T be a function of y only. Then (9.1) simplifies to

$$\frac{D_g}{Dt}\left(\frac{\partial T}{\partial y}\right) = K\frac{\partial T}{\partial y}$$

so that following the geostrophic motion

$$\frac{\partial T}{\partial y} = e^{Kt}\left(\frac{\partial T}{\partial y}\right)_{t=0}$$

Typically, $K \sim 10^{-5}\ \mathrm{s}^{-1}$ so that the temperature gradient amplifies by a factor of 10 in about 3 days. This is much slower than observed rates of atmospheric frontogenesis.

Thus, geostrophic deformation fields alone cannot cause the rapid frontogenesis often observed in extratropical systems, in which the temperature gradient can become concentrated in a zone of $\sim$ 50 km width on a time scale of less than a day. This rapid reduction in scale is caused primarily by the frontogenetic character of the secondary circulation driven by the quasi-geostrophic synoptic-scale flow (moist processes may also be important in rapid frontogenesis).

The nature of the secondary flow may be deduced from the pattern of **Q** vectors illustrated in Figure 9.3a. As discussed in Section 6.5, the divergence of **Q** forces a secondary ageostrophic circulation. For the situation of Figure 9.3, this circulation is in the cross-frontal plane as illustrated in Figure 9.3b. Advection of the temperature field by this ageostrophic circulation tends to increase the horizontal temperature gradient at the surface on the warm side of the jet axis. Temperature advection by the upper-level secondary circulation, however, tends to concentrate the temperature gradient on the cold side of the jet axis. As a result, the frontal zone tends to slope toward the cold air side with height. The differential vertical motion associated with the ageostrophic circulation tends to weaken the front in the midtroposphere due to adiabatic temperature changes (adiabatic warming on the cold side of the front and adiabatic cooling on the warm side). For this reason fronts are most intense in the lower troposphere and near the tropopause.

The secondary circulation associated with frontogenesis is required to maintain the thermal wind balance between the along-front flow and the cross-front temperature gradient in the presence of advective processes that tend to destroy this balance. The concentration of the isotherms by the horizontal advection increases the cross-stream pressure gradient at upper levels, and thus requires an increase in the vertical shear of the along-jet flow in order to maintain thermal wind balance. The required upper-level acceleration of the jet is produced by the Coriolis force caused by the cross-jet ageostrophic wind, which develops in response to the increased cross-stream pressure gradient in the jetstream core. As the jet accelerates, cyclonic vorticity must be generated on the cold side of the jet axis and anticyclonic vorticity on the warm side. These vorticity changes require that the horizontal flow at the jetstream level be convergent on the cold side of the jet axis and divergent on the warm side of the jet axis. The resulting vertical circulation and low-level secondary ageostrophic motion required by mass continuity are indicated in Figure 9.3b.

9.2.2 Semigeostrophic Theory

To analyze the dynamics of the frontogenetic motion fields discussed in the previous subsection, it is convenient to use the Boussinesq approximation introduced in Section 2.8, in which density is replaced by a constant reference value ρ_0 except where it appears in the buoyancy force. This approximation simplifies the equations of motion without affecting the main features of the results. It is also useful to replace the total pressure and density fields with deviations from their standard atmosphere values. Thus, we let $\Phi(x, y, z, t) = (p - p_0)/\rho_0$ designate the pressure deviation normalized by density and $\Theta = \theta - \theta_0$ designate the potential temperature deviation, where $p_0(z)$ and $\theta_0(z)$ are the height-dependent standard atmosphere values of pressure and potential temperature, respectively.

With the preceding definitions, the horizontal momentum equations, thermodynamic energy equation, hydrostatic approximation, and continuity equation become

$$\frac{Du}{Dt} - fv + \frac{\partial \Phi}{\partial x} = 0 \tag{9.2}$$

$$\frac{Dv}{Dt} + fu + \frac{\partial \Phi}{\partial y} = 0 \tag{9.3}$$

$$\frac{D\Theta}{Dt} + w\frac{d\theta_0}{dz} = 0 \tag{9.4}$$

$$b \equiv \frac{g\Theta}{\theta_{00}} = \frac{\partial \Phi}{\partial z} \tag{9.5}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \tag{9.6}$$

where b is the *buoyancy*, θ_{00} is a constant reference value of potential temperature, and

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}$$

From the discussion of the previous subsection it should be clear that the horizontal scale of variations parallel to a front is much larger than the cross-frontal scale. This scale separation suggests that to a first approximation we can model fronts as two-dimensional structures. For convenience we choose a coordinate system in which the front is stationary and take the cross-frontal direction to be parallel to the y axis. Then $L_x \gg L_y$, where L_x and L_y designate the along-front and cross-front length scales. Similarly, $U \gg V$, where U and V, respectively, designate the along-front and cross-front velocity scales. Figure 9.4 shows these scales relative to the front.

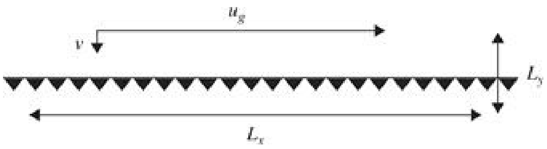


Figure 9.4 Velocity and length scales relative to a front parallel to the x axis.

Letting $U\sim$ 10 m s^{-1}, $V\sim$ 1 m s^{-1}, $L_x\sim$ 1000 km, and $L_y\sim$ 100 km, we find that it is possible to use the differing scales of the along-front and cross-front motion to simplify the dynamics. Assuming that $D/Dt \sim V/L_y$ (the cross-front advection time scale) and defining a Rossby number, Ro $\equiv V/fL_y \ll 1$, the magnitude of the ratios of the inertial and Coriolis terms in the x and y components of the momentum equation can be expressed as

$$\begin{aligned}\frac{|Du/Dt|}{|fv|} &\sim \frac{UV/L_y}{fV} \sim \mathrm{Ro}\left(\frac{U}{V}\right) \sim 1 \\ \frac{|Dv/Dt|}{|fu|} &\sim \frac{V^2/L_y}{fU} \sim \mathrm{Ro}\left(\frac{V}{U}\right) \sim 10^{-2}\end{aligned}$$

The along-front velocity is in geostrophic balance with the cross-front pressure gradient with error of order 1%, but geostrophy does not hold even approximately for the cross-front velocity. Therefore, if the geostrophic wind components are defined by

$$fu_g = -\partial\Phi/\partial y, \qquad fv_g = \partial\Phi/\partial x$$

and we separate the horizontal velocity field into geostrophic and ageostrophic parts, to a good approximation $u = u_g$, but $v = v_g + v_a$, where v_g and v_a are the same order of magnitude.

The x component of the horizontal momentum equation (9.2), the thermodynamic energy equation (9.4), and the continuity equation (9.6) for frontal scaling can thus be expressed as

$$\frac{Du_g}{Dt} - fv_a = 0 \tag{9.7}$$

$$\frac{Db}{Dt} + wN^2 = 0 \tag{9.8}$$

$$\frac{\partial v_a}{\partial y} + \frac{\partial w}{\partial z} = 0 \tag{9.9}$$

Here, (9.8) is obtained by using (9.5) to replace Θ by b in (9.4), and N is the buoyancy frequency defined in terms of potential temperature as

$$N^2 \equiv \frac{g}{\theta_{00}}\frac{\partial\theta_0}{\partial z}$$

Because the along-front velocity is in geostrophic balance, u_g and b are related by the thermal wind relationship:

$$f\frac{\partial u_g}{\partial z} = -\frac{\partial b}{\partial y} \tag{9.10}$$

Note that (9.7) and (9.8) differ from their quasi-geostrophic analogs; although zonal momentum is still approximated geostrophically, and advection parallel to the front is geostrophic, the advection of momentum and temperature across the front is due not only to the geostrophic wind but to the ageostrophic (v_a, w) circulation:

$$\frac{D}{Dt} = \frac{D_g}{Dt} + \left(v_a \frac{\partial}{\partial y} + w \frac{\partial}{\partial z} \right)$$

where D_g/Dt was defined in equation (6.5). Replacement of momentum by its geostrophic value in (9.7) is referred to as the *geostrophic momentum* approximation, and the set of prediction equations that result are called the *semigeostrophic* equations.[1]

9.2.3 Cross-Frontal Circulation

Equations (9.7) through (9.10) form a closed set that can be used to determine the cross-frontal ageostrophic circulation in terms of the distribution of zonal wind or temperature. Suppose that the large-scale geostrophic flow is acting to intensify the north–south temperature gradient through deformation as shown in Figure 9.3. As the temperature gradient increases, the vertical shear of the zonal wind must also increase to maintain geostrophic balance. This requires an increase of u_g in the upper troposphere, which must be produced by the Coriolis force associated with a cross-frontal ageostrophic circulation—see (9.7). The structure of this secondary circulation can be computed by deriving an equation analogous to the omega equation discussed in Section 6.5.

We first differentiate (9.8) with respect to y and use the chain rule to express the result as

$$\frac{D}{Dt}\left(\frac{\partial b}{\partial y}\right) = Q_2 - \frac{\partial v_a}{\partial y}\frac{\partial b}{\partial y} - \frac{\partial w}{\partial y}\left(N^2 + \frac{\partial b}{\partial z}\right) \tag{9.11}$$

where

$$Q_2 = -\frac{\partial u_g}{\partial y}\frac{\partial b}{\partial x} - \frac{\partial v_g}{\partial y}\frac{\partial b}{\partial y} \tag{9.12}$$

is just the y component of the **Q** vector discussed in Section 6.5, but expressed in the Boussinesq approximation.

Next we differentiate equation (9.7) with respect to z, again use the chain rule to rearrange terms, and use the thermal wind equation (9.10) to replace $\partial u_g/\partial z$ by $\partial b/\partial y$ on the right side. The result can then be written as

$$\frac{D}{Dt}\left(f\frac{\partial u_g}{\partial z}\right) = Q_2 + \frac{\partial v_a}{\partial z} f \left(f - \frac{\partial u_g}{\partial y}\right) + \frac{\partial w}{\partial z}\frac{\partial b}{\partial y} \tag{9.13}$$

Again, as shown in Section 6.5, the geostrophic forcing (given by Q_2) tends to destroy thermal wind balance by changing the temperature gradient and vertical shear parts of the thermal wind equation in equal but opposite senses. This tendency of geostrophic advection to destroy geostrophic balance is counteracted by the cross-frontal secondary circulation.

In this case the secondary circulation is a two-dimensional transverse circulation in the y,z plane. It can thus be represented in terms of a meridional streamfunction ψ_M defined so that

$$v_a = -\partial\psi_M/\partial z, \qquad w = \partial\psi_M/\partial y \tag{9.14}$$

which identically satisfies the continuity equation (9.9). Adding (9.11) and (9.13), and using the thermal wind balance (9.10) to eliminate the time derivative and (9.14) to eliminate v_a and w, we obtain the *Sawyer–Eliassen equation*:

$$N_s^2\frac{\partial^2\psi_M}{\partial y^2}+F^2\frac{\partial^2\psi_M}{\partial z^2}+2S^2\frac{\partial^2\psi_M}{\partial y\partial z}=2Q_2 \tag{9.15}$$

where

$$N_s^2\equiv N^2+\frac{\partial b}{\partial z},\quad F^2\equiv f\left(f-\frac{\partial u_g}{\partial y}\right)=f\frac{\partial M}{\partial y},\quad S^2\equiv-\frac{\partial b}{\partial y} \tag{9.16}$$

where *M* is the absolute momentum defined as

$$M=fy-u_g$$

Equation (9.15) can be compared with the quasi-geostrophic version obtained by neglecting advection by the ageostrophic circulation in (9.7) and (9.8). This has the form

$$N^2\frac{\partial^2\psi_M}{\partial y^2}+f^2\frac{\partial^2\psi_M}{\partial z^2}=2Q_2 \tag{9.17}$$

Thus, in the quasi-geostrophic case the coefficients in the differential operator on the left depend only on the standard atmosphere static stability, *N*, and the planetary vorticity, *f*, whereas in the semigeostrophic case they depend on the deviation of potential temperature from its standard profile through the N_s and *S* terms and the absolute vorticity through the *F* term.

An equation of the form (9.17), in which the coefficients of the derivatives on the left are positive, is referred to as an *elliptic boundary value problem*. It has a solution ψ_M that is uniquely determined by Q_2 plus the boundary conditions. For a situation such as that of Figure 9.1b, with both $\partial v_g/\partial y$ and $\partial b/\partial y$ negative, the forcing term Q_2 is negative in the frontal region. The streamfunction in that case describes a circulation symmetric about the *y* axis, with rising on the warm side and sinking on the cold side. The semigeostrophic case (9.15) is also an elliptic boundary value problem provided that $N_s^2F^2-S^4>0$. It can be shown (see Problem 9.1) that this condition requires that the Ertel potential vorticity be positive in the Northern Hemisphere and negative in the Southern Hemisphere, which is nearly always the case in the extratropics for an unsaturated atmosphere.

The spatial variation of the coefficients in (9.16) and the presence of the cross-derivative term produce a distortion of the secondary circulation, as shown in Figure 9.5. The frontal zone slopes toward the cold air side with height; there is an intensification of the cross-frontal flow near the surface in the region of large absolute vorticity on the warm air side of the front and a tilting of the circulation with height.

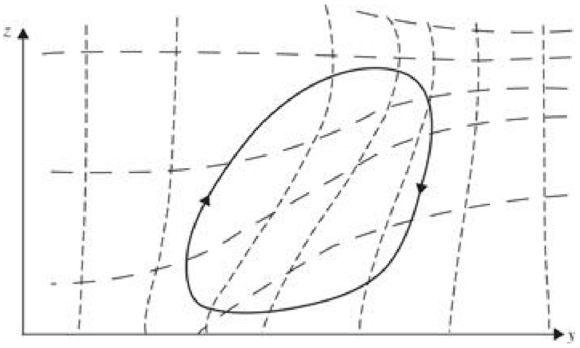


Figure 9.5 Relationship of the ageostrophic circulation (*heavy curve with arrows*) in two-dimensional frontogenesis to the potential temperature field (*long dashes*) and absolute momentum field (*short dashes*). Cold air is on the right and warm air is on the left. Note the tilt of the circulation toward the cold air side and the enhanced gradients of the absolute momentum and potential temperature fields in the frontal zone.

The influence of the ageostrophic circulation on the time scale for frontogenesis can be illustrated by comparing the processes included in quasi-geostrophic and semigeostrophic frontogenesis. For semigeostrophic advection there is a positive feedback that greatly reduces the time scale of frontogenesis compared to that in the quasi-geostrophic case. As the temperature contrast increases, Q_2 increases, and the secondary circulation must also increase so that the amplification *rate* of $|\partial T/\partial y|$ increases with $|\partial T/\partial y|$ rather than remaining constant as in the quasi-geostrophic case. Because of this feedback, in the absence of frictional effects, the semigeostrophic model can produce an infinite temperature gradient at the surface in less than half a day.

9.3 Symmetric Baroclinic Instability

Observations indicate that mesoscale bands of cloud and precipitation commonly occur in association with synoptic-scale systems. These are often aligned with fronts but are separate entities. In common with fronts, such features are generally associated with strong baroclinicity and have length scales parallel to the mean wind shear that are much larger than the scales across the wind shear. A plausible source for such features is a two-dimensional form of baroclinic instability known as *symmetric instability* (also referred to as slantwise convection).

For typical atmospheric conditions, buoyancy tends to stabilize air parcels against vertical displacements, and rotation tends to stabilize parcels with respect to horizontal displacements. Instability with respect to vertical displacements is referred to as hydrostatic (or, simply, static) instability (see Section 2.7.3). For an unsaturated atmosphere, static stability requires that the local buoyancy frequency squared be positive ($N_s^2 > 0$). Instability with respect to horizontal displacements, however, is referred to as inertial instability (see Section 5.5.1). The condition for inertial stability equation (5.74) requires that F^2 as defined in (9.16) be.

If parcels are displaced along slantwise paths rather than vertical or horizontal paths, it is possible under certain conditions for the displacements to be unstable even when the conditions for ordinary static and inertial stability are separately satisfied. Such instability can occur only in the presence of vertical shear of the mean horizontal wind and may be regarded as a special form of baroclinic instability in which the perturbations are independent of the coordinate parallel to the mean flow. Alternatively, as will be shown, symmetric instability may be regarded as isentropic inertial instability.

For convenience in deriving the conditions for symmetric instability, we use the Boussinesq equations of Section 2.8 and assume that the flow is independent of the x coordinate. The mean wind is directed along the x axis and is in thermal wind balance with the meridional temperature gradient:

$$f\partial u_g/\partial z = -\partial b/\partial y = -(g/\theta_{00})\,\partial\Theta/\partial y \tag{9.18}$$

where, as before, θ_{00} is a constant reference value of potential temperature.

Following Sections 2.7.3 and 5.5.1, we measure the stability with respect to vertical displacements by the distribution of total potential temperature $\theta = \theta_0 + \Theta$ and that with respect to horizontal displacements by the distribution of the absolute zonal momentum, $M \equiv fy - u_g$ (where $\partial M/\partial y = f - \partial u_g/\partial y$ is the zonal mean absolute vorticity).

For a barotropic flow, potential temperature surfaces are oriented horizontally and absolute momentum surfaces are oriented vertically in the meridional plane. When the mean flow is westerly and increases with height, however, the potential temperature and absolute momentum surfaces both slope upward toward the pole. The comparative strengths of the vertical and horizontal restoring forces in the midlatitude troposphere are given by the ratio $N_s^2/(f\partial M/\partial y)$. For typical tropospheric conditions, this ratio is $\sim 10^4$. Thus, parcel motion in the plane orthogonal to the mean flow will remain much closer to θ surfaces than to M surfaces. It is thus natural to utilize isentropic coordinates to analyze parcel displacements. The arguments of Section 5.5.1 still apply, provided that derivatives with respect to y are taken at constant θ. The stability of such motions thus depends on the relative slope of the θ and M surfaces. Ordinarily the M surfaces will slope more than the θ surfaces (see Figure 9.5), and parcel displacements are stable. However, when the θ surfaces slope more than the M surfaces so that

$$f(\partial M/\partial y)_\theta < 0 \tag{9.19}$$

the flow is unstable with respect to displacements along the θ surfaces. This situation may occur in regions of a very strong horizontal temperature gradient and weak vertical stability. The condition in (9.19) is similar to the criterion of (5.74) for inertial instability, except that the derivative of M is here taken along a sloping θ surface.

If (9.19) is multiplied by $-g(\partial\theta/\partial p)$, the criterion for symmetric instability can be expressed in terms of the distribution of Ertel potential vorticity (4.12) in the simple form

$$f\bar{P} < 0 \tag{9.20}$$

where $\bar{P}$ is the potential vorticity of the basic state geostrophic flow. Thus, if the initial state potential vorticity in the Northern Hemisphere is everywhere positive, then symmetric instability cannot develop through adiabatic motions, as potential vorticity is conserved following the motion and will always remain positive.

To demonstrate that (9.19) is the condition for symmetric instability, we consider the change in mean kinetic energy required for exchange of the tubes of fluid labeled 1 and 2 in Figure 9.6. (These tubes are located at y_1 and $y_2 = y_1 + \delta y$, respectively, and are assumed to extend infinitely along the x axis so that the problem is two-dimensional.) Because the tubes lie on the same potential temperature surface, they have the same available potential energy. Thus, a spontaneous exchange of parcels is possible if $\delta(\mathrm{KE})$, the kinetic energy of the zonal flow after exchange minus that in the initial state, is negative. Otherwise some external source of energy is required to furnish the kinetic energy of the meridional and vertical motions required for the exchange.

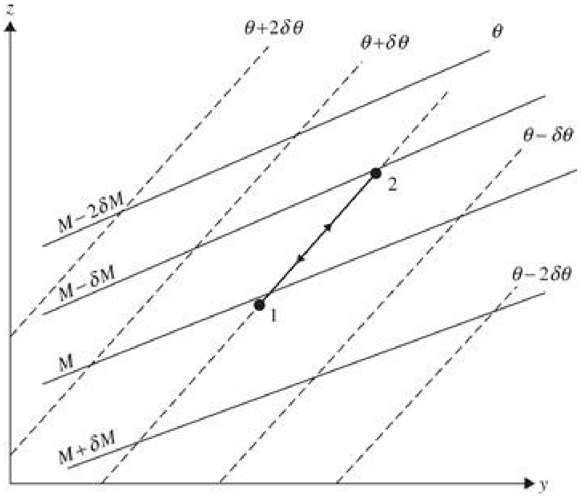


Figure 9.6 Cross-section showing isolines of absolute momentum and potential temperature for a symmetrically unstable basic state. Motion along the isentropic path between points labeled 1 and 2 is unstable, as *M* decreases with latitude along the path. See text for details.

Initially the motion of the tubes is parallel to the *x* axis and in geostrophic balance so that the absolute momentum for the two tubes is

$$\begin{aligned} M_1 &= fy_1 - u_1 = fy_1 - u_g(y_1, z) \\ M_2 &= fy_2 - u_2 = fy_1 + f\delta y - u_g\left(y_1 + \delta y, z + \delta z\right) \end{aligned} \tag{9.21}$$

Absolute momentum is conserved by the tubes so that after exchange the perturbation zonal velocities are given by

$$M_1' = fy_1 + f\delta y - u_1' = M_1, \quad M_2' = fy_1 - u_2' = M_2 \tag{9.22}$$

Eliminating M_1 and M_2 between (9.21) and (9.22) and solving for the disturbance zonal wind, we obtain

$$u_1' = f\delta y + u_1, \quad u_2' = -f\delta y + u_2$$

The difference in zonal kinetic energy between the final state and the initial state is given by

$$\begin{aligned} \delta(\mathrm{KE}) &= \frac{1}{2}\left(u_1'^2 + u_2'^2\right) - \frac{1}{2}\left(u_1^2 + u_2^2\right) \\ &= f\delta y\left(u_1 - u_2 + f\delta y\right) = f\delta y\left(M_2 - M_1\right) \end{aligned} \tag{9.23}$$

Thus, δ(KE) is negative, and unforced meridional motion may occur, provided that $f(M_2 - M_1) < 0$. This is equivalent to the condition (9.19), as the tubes lie along the same θ surface.

To estimate the likelihood that conditions for symmetric instability may be satisfied, it is useful to express the stability criterion in terms of a mean-flow *Richardson number*. To do this we first note that the slope of an *M* surface can be estimated from noting that on an *M* surface

$$\delta M = \frac{\partial M}{\partial y}\delta y + \frac{\partial M}{\partial z}\delta z = 0$$

so that the ratio of δz to δy at constant M is

$$\left(\frac{\delta z}{\delta y}\right)_M = \left(-\frac{\partial M}{\partial y}\right) \Big/ \left(\frac{\partial M}{\partial z}\right) = \left(f - \frac{\partial u_g}{\partial y}\right) \Big/ \left(\frac{\partial u_g}{\partial z}\right) \tag{9.24}$$

Similarly, the slope of a potential temperature surface is

$$\left(\frac{\delta z}{\delta y}\right)_\theta = \left(-\frac{\partial \theta}{\partial y}\right) \Big/ \left(\frac{\partial \theta}{\partial z}\right) = \left(f\frac{\partial u_g}{\partial z}\right) \Big/ \left(\frac{g}{\theta_{00}}\frac{\partial \theta}{\partial z}\right) \tag{9.25}$$

where we have used the thermal wind relationship to express the meridional temperature gradient in terms of the vertical shear of the zonal wind. The ratio of (9.24) to (9.25) is simply

$$\left(\frac{\delta z}{\delta y}\right)_M \Big/ \left(\frac{\delta z}{\delta y}\right)_\theta = f\left(f - \frac{\partial u_g}{\partial y}\right)\left(\frac{g}{\theta_{00}}\frac{\partial \theta}{\partial z}\right) \Big/ \left[f^2\left(\frac{\partial u_g}{\partial z}\right)^2\right] = \frac{F^2 N_s^2}{S^4}$$

where the notation in the last term is defined in (9.16).

Recalling that symmetric instability requires that the slopes of the θ surfaces exceed those of the *M* surfaces, the necessary condition for instability of geostrophic flow parallel to the *x* axis becomes

$$\left(\frac{\delta z}{\delta y}\right)_M \Big/ \left(\frac{\delta z}{\delta y}\right)_\theta = f\left(f - \frac{\partial u_g}{\partial y}\right)\mathrm{Ri}/f^2 = \frac{F^2 N_s^2}{S^4} < 1 \tag{9.26}$$

where the mean-flow *Richardson number* Ri is defined as

$$\mathrm{Ri} \equiv \left(\frac{g}{\theta_{00}}\frac{\partial \theta}{\partial z}\right) \Big/ \left(\frac{\partial u_g}{\partial z}\right)^2$$

Thus, if the relative vorticity of the mean flow vanishes $(\partial u_g/\partial y = 0)$, Ri < 1 is required for instability.

The condition (9.26) can be related to (9.20) by observing that (9.26) requires that $F^2 N_s^2 - S^4 < 0$ for symmetric instability. As is to be shown in Problem 9.1,

$$F^2 N_s^2 - S^4 = (\rho f g\,/\theta_{00})\,\overline{P} \tag{9.27}$$

Because the large-scale potential vorticity, $\bar{P}$, is normally positive in the Northern Hemisphere and negative in the Southern Hemisphere, (9.27) is ordinarily positive in both hemispheres; thus, the condition for symmetric instability is rarely satisfied. If the atmosphere is saturated, however, the relevant static stability condition involves the lapse rate of the equivalent potential temperature, and neutral conditions with respect to symmetric instability may easily occur (see Section 9.5).

Finally, it is worth noting that the condition for stability with respect to symmetric displacements, $F^2 N_s^2 - S^4 > 0$, is identical to the condition that the Sawyer–Eliassen equation (9.15) be an elliptic boundary value problem. Thus, when the flow is stable with respect to symmetric baroclinic perturbations, a nonzero *forced* transverse circulation governed by (9.15) will exist when there is nonzero forcing, Q_2—see (9.12). *Free* transverse oscillations, however, may occur in the absence of forcing. These require including the horizontal acceleration term in the *y* component of the momentum equation. The resulting equation for the transverse circulation has the form (see Appendix F):

$$\frac{\partial^2}{\partial t^2}\left(\frac{\partial^2 \psi_M}{\partial z^2}\right) + N_s^2 \frac{\partial^2 \psi_M}{\partial y^2} + F^2 \frac{\partial^2 \psi_M}{\partial z^2} + 2S^2 \frac{\partial^2 \psi_M}{\partial y \partial z} = 0 \tag{9.28}$$

which should be compared with (9.15). When $F^2 N_s^2 - S^4 > 0$, solutions of (9.28) correspond to stable oscillations, while for $F^2 N_s^2 - S^4 < 0$, solutions are exponentially growing, corresponding to symmetric baroclinic instability.

9.4 Mountain Waves

When stably stratified air is forced to flow over varying surface topography, buoyancy oscillations are created. First, consider the case with constant wind and stratification for sinusoidally varying topography, and then allow for localized mountain barriers and vertical variations in static stability and wind.

9.4.1 Waves over Sinusoidal Topography

When air with mean wind speed $\overline{u}$ is forced to flow over a sinusoidal pattern of ridges under statically stable conditions, individual air parcels are alternately displaced upward and downward from their equilibrium levels and will thus undergo buoyancy oscillations as they move across the ridges, as shown in Figure 9.7. In this case there are solutions in the form of waves that are stationary relative to the ground. For such stationary waves, w' depends only on (*x*,*z*) and the equation governing their motion, (5.64), simplifies to

$$\left(\frac{\partial^2 w'}{\partial x^2} + \frac{\partial^2 w'}{\partial z^2}\right) + \frac{N^2}{\overline{u}^2} w' = 0 \tag{9.29}$$

Substituting

$$w' = \mathrm{Re}\, \hat{w} \exp^{i(kx+mz)} \tag{9.30}$$

into (9.29) yields

$$m^2 = \frac{N^2}{\overline{u}^2} - k^2 \tag{9.31}$$

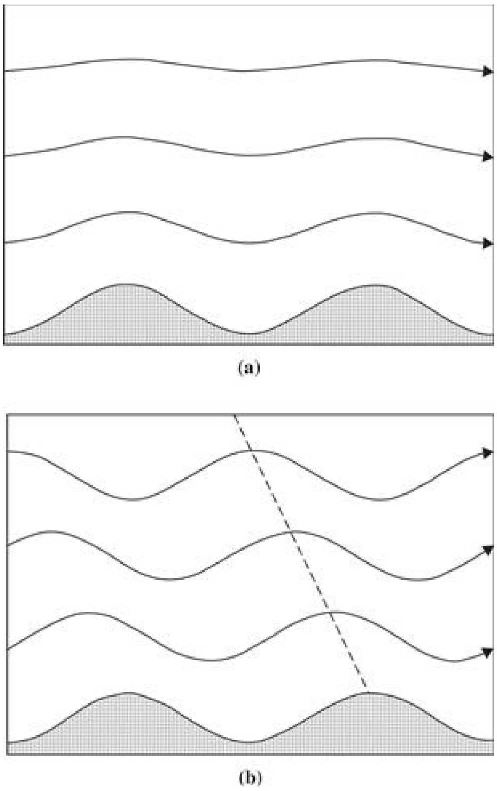


Figure 9.7 Streamlines in steady flow over an infinite series of sinusoidal ridges for the narrow ridge case (a) and broad ridge case (b). The *dashed line* in (b) shows the phase of maximum upward displacement.

From Durran, 1990. Copyright © American Meteorological Society. Reprinted with permission.

For given values of *N*, *k*, and $\overline{u}$, (9.31) determines the vertical structure. Then if $|\overline{u}| < N/k$, (9.31) shows that $m^2 > 0$ (i.e., *m* must be real) and solutions of (9.29) have the form of vertically propagating waves. When $m^2 < 0, m = im_i$ is imaginary and the solution to (9.29) will have the form of vertically trapped waves:

$$w' = \hat{w} \exp(ikx) \exp(-m_i z)$$

Thus, vertical propagation is possible only when $|\overline{u}k|$, the magnitude of the frequency relative to the mean flow, is less than the buoyancy frequency.

Stable stratification, wide ridges, and comparatively weak zonal flow provide favorable conditions for the formation of vertically propagating topographic waves (*m* real). Because the energy source for these waves is at the ground, they must transport energy upward. Hence, the phase speed relative to the mean zonal flow must have a downward component. Thus, if $\overline{u} > 0$, lines of constant phase must tilt westward with height. When *m* is imaginary, however, the solution (9.30) has exponential behavior in the vertical with an exponential decay height of μ^{-1}, where $\mu = |m|$. Boundedness as $z \to \infty$ requires that we choose the solution with exponential decay away from the lower boundary.

To contrast the character of the solutions for real and imaginary *m*, we consider a specific example in which there is westerly mean flow over topography with a height profile given by

$$h(x) = h_M \cos kx$$

where h_M is the amplitude of the topography. Then, because the flow at the lower boundary must be parallel to the boundary, the vertical velocity perturbation at the boundary is given by the rate at which the boundary height changes following the motion:

$$w'(x,0) = (Dh/Dt)_{z=0} \approx \overline{u}\partial h/\partial x = -\overline{u}kh_M \sin kx$$

and the solution of (9.29) that satisfies this condition can be written

$$w(x,z) = \left\{ \begin{array}{ll} -\overline{u}h_M k e^{-\mu z}\sin kx, & \overline{u}k > N \\ -\overline{u}h_M k \sin(kx+mz), & \overline{u}k < N \end{array} \right\} \tag{9.32}$$

For fixed mean wind and buoyancy frequency, the character of the solution depends only on the horizontal scale of the topography. The two cases of (9.32) may be regarded as narrow ridge and wide ridge cases, respectively, since for specified values of $\overline{u}$ and N, the character of the solution is determined by the zonal wave number k. The streamline patterns corresponding to these cases for westerly flow are illustrated in Figure 9.7. In the narrow ridge case (Figure 9.7a), the maximum upward displacement occurs at the ridge tops, and the amplitude of the disturbance decays with height. In the wide ridge case (Figure 9.7b), the line of maximum upward displacement tilts back toward the west ($m > 0$), and amplitude is independent of height consistent with an internal gravity wave propagating westward relative to the mean flow.

Alternatively, for fixed zonal wave number and buoyancy frequency, the solution depends only on the speed of the mean zonal wind. As indicated in (9.32), only for mean zonal wind magnitudes less than the critical value N/k will vertical wave propagation occur.

Equation (9.29) was obtained for conditions of constant basic state flow. In reality, both the zonal wind $\overline{u}$ and the stability parameter N generally vary with height, and ridges are usually isolated rather than periodic. A wide variety of responses are possible depending on the shape of the terrain and on wind and stability profiles. Under certain conditions, large-amplitude waves can be formed, which may generate severe downslope surface winds and zones of strong clear air turbulence. Such circulations are discussed further in Section 9.4.4.

9.4.2 Flow over Isolated Ridges

Just as flow over a periodic series of sinusoidal ridges can be represented by a single sinusoidal function flow—that is, by a single Fourier harmonic flow—over an isolated ridge can be approximated by the sum of a number of Fourier components (see Section 5.2.1). Thus, any zonally varying topography can be represented by a Fourier series of the form

$$h_M(x) = \sum_{s=1}^{\infty} \mathrm{Re}\left[h_s \exp(ik_s x)\right] \tag{9.33}$$

where h_s is the amplitude of the s th Fourier component of the topography. We can then express the solution to the wave equation (9.29) as the sum of Fourier components:

$$w(x,z) = \sum_{s=1}^{\infty} \mathrm{Re}\left\{W_s \exp\left[i(k_s x + m_s z)\right]\right\} \tag{9.34}$$

where $W_s = ik_s\overline{u}h_s$, and $m_s^2 = N^2/\overline{u}^2 - k_s^2$.

Individual Fourier modes will yield vertically propagating or vertically decaying contributions to the total solution (9.34) depending on whether m_s is real or imaginary. This in turn depends on whether k_s^2 is less than or greater than $N^2/\bar{u}^2$. Thus, each Fourier mode behaves just as the solution (9.30) for periodic sinusoidal topography behaves. For a narrow ridge, Fourier components with wave numbers greater than $N/\bar{u}$ dominate in (9.33), and the resulting disturbance decays with height. For a broad ridge, components with wave numbers less than $N/\bar{u}$ dominate, and the disturbance propagates vertically. In the wide mountain limit where $m_s^2 \approx N^2/\bar{u}^2$, the flow is periodic in the vertical with a vertical wavelength of $2\pi m_s^{-1}$, and phase lines tilt upstream with height as shown in Figure 9.8.

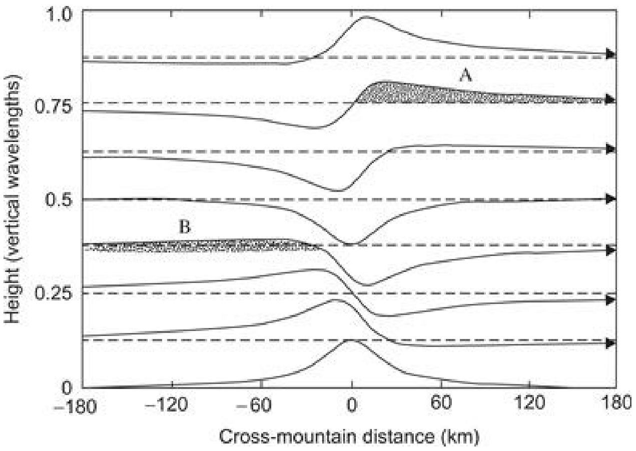


Figure 9.8 Streamlines of flow over a broad isolated ridge showing upstream phase tilt with height. The pattern is periodic in height, and one vertical wavelength is shown. Orographic clouds may form in the shaded areas where streamlines are displaced upward from equilibrium either upstream or downstream of the ridge if sufficient moisture is present.

After Durran, 1990. Copyright © American Meteorological Society. Reprinted with permission.

Vertically propagating gravity waves generated by flow over broad topography can produce clouds both upstream and downstream of the topography depending on variations of the moisture distribution with altitude. In the example shown in Figure 9.7, the positions labeled A and B indicate regions where streamlines are displaced upward downstream and upstream of the ridge, respectively. If sufficient moisture is present, orographic clouds may then form in region A or B as suggested by the shading in Figure 9.8.

9.4.3 Lee Waves

If $\bar{u}$ and N are allowed to vary in height, then (9.29) must be replaced by

$$\left(\frac{\partial^2 w'}{\partial x^2} + \frac{\partial^2 w'}{\partial z^2}\right) + l^2 w' = 0 \tag{9.35}$$

where the *Scorer parameter*, *l*, is defined as

$$l^2 = N^2/\overline{u}^2 - \overline{u}^{-1} d^2\overline{u}/dz^2$$

and the condition for vertical propagation becomes $k_s^2 < l^2$.

If the mean cross-mountain wind speed increases strongly with height, or if there is a low-level stable layer so that N decreases strongly with height, there may be a layer near the surface in which vertically propagating waves are permitted, which is topped by a layer in which the disturbance decays in the vertical. In that case vertically propagating waves in the lower layer are reflected when they reach the upper layer. Under some circumstances the waves may be repeatedly reflected from the upper layer and the surface downstream of the mountain, leading to a series of "trapped" lee waves as shown in Figure 9.9.

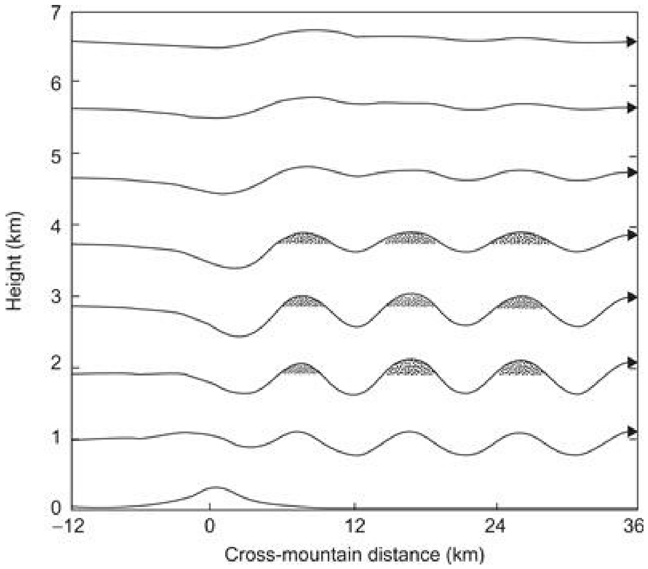


Figure 9.9 Streamlines for trapped lee waves generated by flow over topography with vertical variation of the Scorer parameter. *Shading* shows locations where lee wave clouds may form.

After Durran, 1990. Copyright © American Meteorological Society. Reprinted with permission.

Vertical variations in the Scorer parameter can also modify the amplitude of waves that are sufficiently long to propagate vertically through the entire troposphere. Amplitude enhancement leading to wave breaking and turbulent mixing can occur if there is a *critical level* where the mean flow goes to zero $(l \rightarrow \infty)$.

9.4.4 Downslope Windstorms

Strong downslope surface winds are occasionally observed along the lee slopes of mountain ranges. Although partial reflection of vertically propagating linear gravity waves may produce enhanced surface winds under some conditions, it appears that nonlinear processes are essential to account for observed windstorms associated with stable flow over topography.

To demonstrate the role of nonlinearity, we assume that the troposphere has a stable lower layer of undisturbed depth h topped by a weakly stable upper layer, and assume that the lower layer behaves as a barotropic fluid with a free surface $h(x,t)$. We assume that disturbances have zonal wavelengths much greater than the layer depth. The motion of the lower layer may then be described by the shallow water equations of Section 4.5, but with the lower boundary condition replaced by

$$w(x, h_M) = Dh_M/Dt = u\partial h_M/\partial x$$

where h_M again denotes the height of the topography.

We first examine the linear behavior of this model by considering steady flow over small-amplitude topography. Neglecting ambient rotation, the linearized shallow water equations in steady state become

$$\overline{u}\frac{\partial u'}{\partial x} + g\frac{\partial h'}{\partial x} = 0 \tag{9.36}$$

$$\overline{u}\frac{\partial \left(h' - h_M\right)}{\partial x} + H\frac{\partial u'}{\partial x} = 0 \tag{9.37}$$

Here, $h' = h - H$, where H is the mean height of the interface and $h' - h_M$ is the deviation from H of the thickness of the layer.

The solutions for (9.36) and (9.37) can be expressed as

$$h' = -\frac{h_M\left(\overline{u}^2/c^2\right)}{\left(1 - \overline{u}^2/c^2\right)}, \quad u' = \frac{h_M}{H}\left(\frac{\overline{u}}{1 - \overline{u}^2/c^2}\right) \tag{9.38}$$

where $c^2 \equiv gH$ is the shallow water gravity wave speed. The characteristics of the disturbance fields h' and u' depend on the magnitude of the mean-flow *Froude number*, defined by the relation $\mathrm{Fr}^2 = \overline{u}^2/c^2$. When $\mathrm{Fr} < 1$, the flow is referred to as *subcritical*. In subcritical flow, the shallow water gravity wave speed is greater than the mean-flow speed, and the disturbance height and wind fields are out of phase. The interface height disturbance is negative, and the velocity disturbance is positive over the topographic barrier as shown in Figure 9.10a. When $\mathrm{Fr} > 1$, the flow is referred to as *supercritical*. In supercritical flow the mean flow exceeds the shallow water gravity wave speed. Gravity waves cannot play a role in establishing the steady-state adjustment between height and velocity disturbances because such waves are swept downstream from the ridge by the mean flow. In this case the fluid thickens and slows as it ascends over the barrier (Figure 9.10b). It is also clear from (9.38) that for Fr ~ 1 the perturbations are no longer small and the linear solution breaks down.

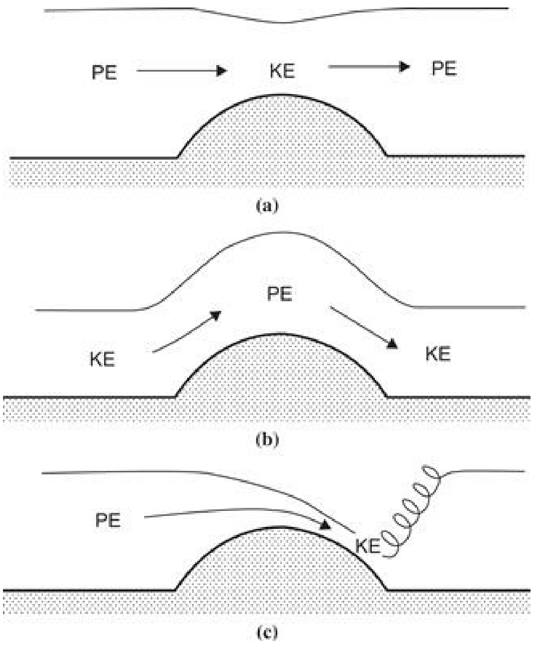


Figure 9.10 Flow over an obstacle for a barotropic fluid with free surface. (a) Subcritical flow (Fr > 1 everywhere). (b) Supercritical flow (Fr < 1 everywhere). (c) Supercritical flow on lee slope with adjustment to subcritical flow at hydraulic jump near base of obstacle.

After Durran, 1990. Copyright © American Meteorological Society. Reprinted with permission.

The nonlinear equations corresponding to (9.36) and (9.37) can be expressed as

$$u\frac{\partial u}{\partial x}+g\frac{\partial h}{\partial x}=0 \tag{9.39}$$

$$\frac{\partial}{\partial x}\left[u\left(h-h_M\right)\right]=0 \tag{9.40}$$

Equation (9.39) may be integrated immediately to show that the sum of kinetic and potential energy, $u^2/2+gh$, is constant following the motion. Thus, energy conservation requires that if u increases, h must decrease, and vice versa. In addition, (9.40) shows that the mass flux, $u(h-h_M)$, must also be conserved. The direction of the exchange between kinetic and potential energy in flow over a ridge is determined by the necessity that both (9.39) and (9.40) be satisfied.

Multiplying (9.39) by u and eliminating $\partial h/\partial x$ with the aid of (9.40) gives

$$\left(1-\mathrm{Fr}^2\right)\frac{\partial u}{\partial x}=\frac{ug}{c^2}\frac{\partial h_M}{\partial x} \tag{9.41}$$

where the shallow water gravity wave speed, c, and the Froude number are now defined using the local thickness and velocity of the fluid:

$$c^2\equiv g\left(h-h_M\right); \quad \mathrm{Fr}^2\equiv u^2/c^2$$

From (9.41) it is clear that the flow will accelerate on the upslope side of the ridge ($\partial u/\partial x>0$, where $\partial h_M/\partial x>0$) if the Froude number is less than unity, but will decelerate if the Froude number is greater than unity.

As a subcritical flow ascends the upslope side of a topographic barrier, Fr will tend to increase both from the increase in u and the decrease in c. If $\mathrm{Fr} = 1$ at the crest, then from (9.41) the flow will become supercritical and continue to accelerate as it descends the lee side until it adjusts back to the ambient subcritical conditions in a turbulent hydraulic jump as illustrated in Figure 9.10c. In this case very high velocities can occur along the lee slope, as potential energy is converted into kinetic energy during the entire period that a fluid column traverses the barrier. Although conditions in the continuously stratified atmosphere are clearly more complex than in the shallow water hydraulic model, numerical simulations have demonstrated that the hydraulic model provides a reasonable conceptual model for the primary processes occurring in downslope windstorms.

9.5 Cumulus Convection

Mesoscale storms associated with cumulus convection represent a large fraction of all of the meteorologically important mesoscale circulations. Before considering such systems, it is necessary to examine a few of the essential thermodynamic and dynamical aspects of individual cumulus clouds. The subject of cumulus convection is extremely complex to address theoretically. Much of this difficulty stems from the fact that the cumulus clouds have a complex internal structure. They are generally composed of a number of short-lived individual rising towers, which are produced by elements of ascending buoyant air called *thermals*. Rising thermals *entrain* environmental air and thus modify the cloud air through mixing.

Thermals are nonhydrostatic, nonsteady, and highly turbulent. The buoyancy of an individual thermal (i.e., the difference between its density and the density of the environment) depends on a number of factors, including the environmental lapse rate, the rate of dilution by entrainment, and the drag by the weight of liquid water in cloud droplets. A detailed discussion of the dynamics of thermal convection is beyond the scope of this text. This section utilizes a simple one-dimensional cloud model and focuses primarily on the thermodynamic aspects of moist convection. Convective storm dynamics is considered in Section 9.6.

9.5.1 Convective Available Potential Energy

Development of convective storms depends on the presence of environmental conditions favorable for the occurrence of deep convection. Several indices have been developed to measure the susceptibility of a given temperature and moisture profile to the occurrence of deep convection. A particularly useful measure is the *convective available potential energy*. Convective available potential energy (CAPE) provides a measure of the maximum possible kinetic energy that a statically unstable parcel can acquire (neglecting effects of water vapor and condensed water on the buoyancy), assuming that the parcel ascends without mixing with the environment and adjusts instantaneously to the local environmental pressure.

The momentum equation for such a parcel is (2.51), which can be rewritten following the vertical motion of the parcel as

$$\frac{Dw}{Dt} = \frac{Dz}{Dt}\frac{Dw}{Dz} = w\frac{Dw}{Dz} = b' \tag{9.42}$$

where $b'(z)$ is again the buoyancy given by

$$b' = g\frac{(\rho_{env} - \rho_{parcel})}{\rho_{parcel}} = g\frac{(T_{parcel} - T_{env})}{T_{env}} \tag{9.43}$$

and T_{env} designates the temperature of the environment. If (9.42) is integrated vertically from the level of free convection, z_{LFC}, to the level of neutral buoyancy, z_{LNB}, following the motion of the parcel the result is

$$\frac{w_{\max}^2}{2} = \int_{z_{LFC}}^{z_{LNB}} g\left(\frac{T_{parcel} - T_{env}}{T_{env}}\right) dz \equiv B \tag{9.44}$$

Here, B is the maximum kinetic energy per unit mass that a buoyant parcel can obtain by ascending from a state of rest at the level of free convection to the level of neutral buoyancy near the tropopause (see Figure 2.8). This is an overestimate of the actual kinetic energy realized by a nonentraining parcel, as the negative buoyancy contribution of liquid water reduces the effective buoyancy, especially in the tropics.

In a typical tropical oceanic sounding, parcel temperature excesses of 1 to 2 K may occur over a depth of 10 to 12 km. A typical value of CAPE is then $B \approx 500\ \mathrm{m^2\,s^{-2}}$. In severe storm conditions in the Midwest of North America, however, parcel temperature excesses can be 7 to 10 K (refer to Figure 2.8) and $B \approx$ 2000 to 3000 $\mathrm{m^2\,s^{-2}}$. Observed updrafts in the latter case (up to 50 $\mathrm{m\,s^{-1}}$) are much stronger than in the former case (5–10 $\mathrm{m\,s^{-1}}$). The small value of CAPE in the mean tropical environment is the major reason that updraft velocities in tropical cumulonimbus are observed to be much smaller than those in midlatitude thunderstorms.

9.5.2 Entrainment

In the previous subsection it was assumed that convective cells rise without mixing with environmental air so that they maintain constant θ_e during their rise. In reality, however, rising saturated air parcels tend to be diluted by entraining, or mixing in, some of the relatively dry environmental air. If the air in the environment is unsaturated, some of the liquid water in the rising parcel must be evaporated to maintain saturation in the convective cell as air from the environment is entrained. The evaporative cooling caused by entrainment will reduce the buoyancy of the convective parcel (i.e., lower its θ_e). Thus, the equivalent potential temperature in an entraining convection cell will decrease with height rather than remaining constant. Similar considerations hold for any other conservable variable[2] in which environmental values differ from cloud values; entrainment will modify the in-cloud vertical profiles.

Denoting as A the amount of an arbitrary conservable variable per unit mass of air, the vertical dependence of A in an entraining convective cell can be estimated by assuming that to a first approximation the cell can be modeled as a steady-state jet as shown in Figure 9.11. Thus, in a time increment δt, a mass m of saturated cloud air with an amount of the arbitrary variable given by mA_{cld} mixes with a mass δm of entrained environmental air, which has an amount of the arbitrary variable given by $\delta m A_{env}$. The change in the value of A within the cloud, δA_{cld}, is then given by the mass balance relationship

$$(m + \delta m)\left(A_{cld} + \delta A_{cld}\right) = mA_{cld} + \delta m A_{env} + \left(DA_{cld}/Dt\right)_S m\delta t \tag{9.45}$$

where $(DA_{cld}/Dt)_S$ designates the rate of change of A_{cld} due to sources and sinks unrelated to entrainment. Dividing through by δt in (9.45), neglecting the second-order term, and rearranging yields

$$\frac{\delta A_{cld}}{\delta t} = \left(\frac{DA_{cld}}{Dt}\right)_S - \frac{1}{m}\frac{\delta m}{\delta t}\left(A_{cld} - A_{env}\right) \tag{9.46}$$

Noting that in the time increment δt an ascending parcel rises a distance $\delta z = w\delta t$, where w is the speed of ascent of the parcel, we can eliminate δt in (9.46) to obtain an equation for the vertical dependence of A_{cld} in a continuously entraining convective cell:

$$w\frac{dA_{cld}}{dz} = \left(\frac{DA_{cld}}{Dt}\right)_S - w\lambda\left(A_{cld} - A_{env}\right) \tag{9.47}$$

where we have defined the *entrainment rate*; $\lambda \equiv d\ln m/dz$.

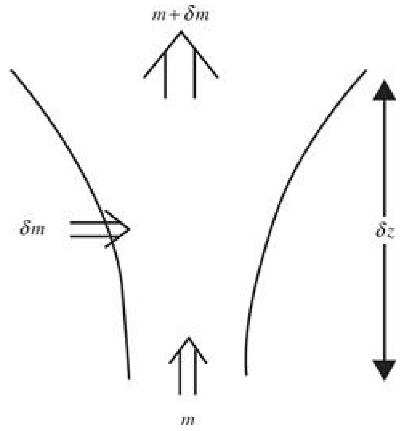


Figure 9.11 An entraining jet model of cumulus convection. See text for an explanation.

Letting $A_{cld} = \ln\theta_e$ and noting that θ_e is conserved in the absence of entrainment, (9.47) yields

$$\begin{aligned}\left(\frac{d\ln\theta_e}{dz}\right)_{cld} &= -\lambda\left[(\ln\theta_e)_{cld} - (\ln\theta_e)_{env}\right] \\ &\approx -\lambda\left[\frac{L_c}{c_pT}\left(q_s - q_{env}\right) + \ln\left(\frac{T_{cld}}{T_{env}}\right)\right]\end{aligned} \tag{9.48}$$

where (2.70) is used to obtain the latter form of the right side. Thus, an entraining convective cell is less buoyant than a nonentraining cell. Letting $A_{cld} = w$ in (9.47), applying (9.42), and neglecting the pressure contribution to buoyancy, we find that the height dependence of kinetic energy per unit mass is given by

$$\frac{d}{dz}\left(\frac{w^2}{2}\right) = g\left(\frac{T_{cld} - T_{env}}{T_{env}}\right) - \lambda w^2 \tag{9.49}$$

An entraining cell will undergo less acceleration than a nonentraining cell not only because the buoyancy is reduced, but also because of the drag exerted by mass entrainment.

Equations (9.48) and (9.49), together with suitable relations for the cloud moisture variables, can be used to determine the vertical profile of cloud variables. Such one-dimensional cloud models have been very popular in the past. Unfortunately, observed cloud properties, such as maximum cloud depth and cloud water concentration, cannot simultaneously be satisfactorily predicted by this type of model. In reality, pressure perturbations generated by the convective cells are important in the momentum budget, and entrainment does not occur instantaneously but in a sporadic manner that allows some cells to rise through most of the troposphere with very little dilution. Although more sophisticated one-dimensional models can partly overcome these deficiencies, some of the most important aspects of thunderstorm dynamics (e.g., the influence of vertical shear of the environmental wind) can only be included in multidimensional models.

9.6 Convective Storms

Convective storms can take a large variety of forms. These range in scale from isolated thunderstorms involving a single convective cloud (or *cell*) to mesoscale convective complexes consisting of ensembles of multicelled thunderstorms. Here we distinguish three primary types: the single-cell, the multicell, and the supercell storm. As shown in the previous section, convective available potential energy measures whether thermodynamical conditions are favorable for the development of cumulus convection. CAPE, therefore, provides a guide to the strength of convection. It does not, however, provide any notion of the most likely type of mesoscale organization. It turns out, as suggested earlier, that storm type also depends on the vertical shear of the lower tropospheric environment.

When vertical shear is weak ($< 10\ \mathrm{ms}^{-1}$ below 4 km), single cell storms occur. These tend to be short-lived (~ 30 min) and move with the mean wind in the lowest 8km. When there is moderate vertical shear (~ 10–$20\ \mathrm{ms}^{-1}$ below 4km), multicell storms arise in which individual cells have lifetimes of ~ 30 min, but the storm lifetime may be many hours. In multicell storms the downdrafts induced by the evaporation of precipitation form a dome of cold outflowing air at the surface. New cells tend to develop along the *gust front* where the cold outflow lifts conditionally unstable surface air. When vertical shear is large ($> 20\ \mathrm{ms}^{-1}$ below 4 km), the strong tilting of convective cells tends to delay storm development, even in a thermodynamically favorable environment, so that an hour or more may be required for the initial cell to completely develop. This development may be followed by a split into two storms, which move to the left and right of the mean wind. Usually the left-moving storm dies rapidly, while the right-moving storm slowly evolves into a rotating circulation with a single updraft core and trailing downdrafts, as discussed in the next section. Such supercell storms often produce heavy rain, hail, and damaging tornadoes. Ensembles of multicell or supercell storms are often organized along lines referred to as *squall lines*, which may move in a different direction than the individual thunderstorms.

9.6.1 Development of Rotation in Supercell Thunderstorms

The supercell thunderstorm is of particular dynamical interest because of its tendency to develop a rotating mesocyclone from an initially nonrotating environment. The dominance of cyclonic rotation in such systems might suggest that the Coriolis force plays a role in supercell dynamics. However, it can be shown readily that the rotation of Earth is not relevant to the development of rotation in supercell storms.

Although a quantitative treatment of supercell dynamics requires that the density stratification of the atmosphere be taken into account, for purposes of understanding the processes that lead to the development of rotation in such systems and to the dominance of the right-moving cell, it is sufficient to use the Boussinesq approximation. The Euler momentum equation and continuity equation may then be expressed as

$$\frac{D\mathbf{U}}{Dt} = \frac{\partial \mathbf{U}}{\partial t} + (\mathbf{U} \cdot \boldsymbol{\nabla})\mathbf{U} = -\frac{1}{\rho_0}\boldsymbol{\nabla} p + b\mathbf{k}$$

$$\boldsymbol{\nabla} \cdot \mathbf{U} = 0$$

Here, $\mathbf{U} \equiv \mathbf{V} + \mathbf{k}w$ is the three-dimensional velocity, ∇ is the three-dimensional del operator, ρ_0 is the constant basic state density, p is the deviation of pressure from its horizontal mean, and $b \equiv -g\rho'/\rho_0$ is the total buoyancy.

It is convenient to rewrite the momentum equation using the vector identity

$$(\mathbf{U} \cdot \boldsymbol{\nabla})\mathbf{U} = \boldsymbol{\nabla}\left(\frac{\mathbf{U} \cdot \mathbf{U}}{2}\right) - \mathbf{U} \times (\boldsymbol{\nabla} \times \mathbf{U})$$

to obtain

$$\frac{\partial \mathbf{U}}{\partial t} = -\boldsymbol{\nabla}\left(\frac{p}{\rho_0} + \frac{\mathbf{U} \cdot \mathbf{U}}{2}\right) + \mathbf{U} \times \boldsymbol{\omega} + b\mathbf{k} \tag{9.50}$$

Taking $\nabla \times$ (9.50) and recalling that the curl of the gradient vanishes, we obtain the three-dimensional vorticity equation

$$\frac{\partial \boldsymbol{\omega}}{\partial t} = \boldsymbol{\nabla} \times (\mathbf{U} \times \boldsymbol{\omega}) + \boldsymbol{\nabla} \times (b\mathbf{k}) \tag{9.51}$$

Letting $\zeta = \mathbf{k} \cdot \boldsymbol{\omega}$ be the vertical component of vorticity and taking $\mathbf{k} \cdot$ (9.51), we obtain an equation for the tendency of ζ in a nonrotating reference frame:

$$\frac{\partial \zeta}{\partial t} = \mathbf{k} \cdot \nabla \times (\mathbf{U} \times \boldsymbol{\omega}) \tag{9.52}$$

Note that buoyancy only affects the horizontal vorticity components.

We now consider a flow consisting of a single convective updraft embedded in a basic state westerly flow that depends on z alone. Linearizing about this basic state by letting

$$\boldsymbol{\omega} = \mathbf{j}d\overline{u}/dz + \boldsymbol{\omega}'(x, y, z, t), \quad \mathbf{U} = \mathbf{i}\overline{u} + \mathbf{U}'(x, y, z, t)$$

and noting that the linearized form of the right side in (9.52) becomes

$$\mathbf{k} \cdot \boldsymbol{\nabla} \times (\mathbf{U} \times \boldsymbol{\omega}) = -\mathbf{k} \cdot \boldsymbol{\nabla} \times (\mathbf{i}w'd\overline{u}/dz + \mathbf{j}\,\overline{u}\zeta')$$

we find that the linearized vorticity tendency is

$$\frac{\partial \zeta'}{\partial t} = -\overline{u}\frac{\partial \zeta'}{\partial x} + \frac{\partial w'}{\partial y}\frac{d\overline{u}}{dz} \tag{9.53}$$

The first term on the right in (9.53) is just the advection by the basic state flow. The second term represents tilting of horizontal shear vorticity into the vertical by differential vertical motion.

Because $d\overline{u}/dz$ is positive, the vorticity tendency due to this tilting will be positive to the south of the updraft core and negative to the north of the updraft core. As a result, a counterrotating vortex pair is established with cyclonic rotation to the south and anticyclonic rotation to the north of the initial updraft, as shown in Figure 9.12a. Eventually, the development of negative buoyancy due to precipitation loading generates an upper-level downdraft at the position of the initial updraft, and the storm splits as shown in Figure 9.12b. New updraft cores are established centered on the counterrotating vortex pair.

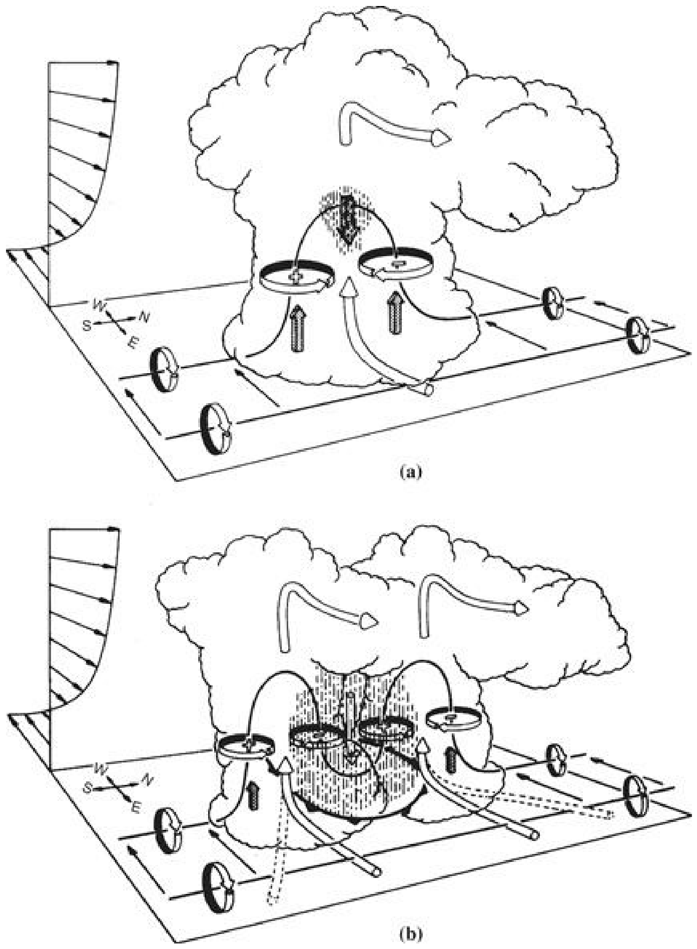


Figure 9.12 Development of rotation and splitting in a supercell storm with westerly mean wind shear (shown by storm relative wind *arrows* in the *upper left* corner of each panel). *Cylindrical arrows* show the direction of cloud relative air flow. *Heavy solid lines* show vortex lines with a sense of rotation shown by *circular arrows*. *Plus and minus signs* indicate cyclonic and anticyclonic rotation caused by vortex tube tilting. *Shaded arrows* represent updraft and downdraft growth. *Vertical dashed lines* denote regions of precipitation. (a) In the initial stage, the environmental shear vorticity is tilted and stretched into the vertical as it is swept into the updraft. (b) In the splitting stage, downdraft forms between the new updraft cells. The *arced line* with triangles in the lower middle of (b) at the surface indicates downdraft outflow at the surface.

After Klemp, 1987. Used with permission from Annual Reviews.

To understand the generation of updrafts in the vortices on the flanks of the storm, we examine the perturbation pressure field. A diagnostic equation for the disturbance pressure is obtained by taking $\nabla\cdot$ (9.53) to yield

$$\nabla^2\left(\frac{p}{\rho_0}\right) = -\nabla^2\left(\frac{\mathbf{U}\cdot\mathbf{U}}{2}\right) + \boldsymbol{\nabla}\cdot(\mathbf{U}\times\boldsymbol{\omega}) + \frac{\partial b}{\partial z} \tag{9.54}$$

The first two terms on the right in (9.54) represent dynamical forcing, whereas the last term represents buoyancy forcing. Observations and numerical models suggest that the buoyancy forcing in (9.54) produces pressure perturbations that tend partly to compensate for the buoyancy force in the vertical momentum equation. Dynamically forced pressure perturbations, however, may generate substantial vertical accelerations.

To compute the dynamical contribution to the disturbance pressure gradient force in either the right or left vortex, we use cylindrical coordinates (r, λ, z) centered on the axis of rotation of either vortex, and assume that to a first approximation the azimuthal velocity v_λ (positive for cyclonic flow) is independent of λ. In this system the storm relative horizontal motion and vertical component of vorticity are given approximately by

$$\mathbf{U}' \approx \mathbf{j}_\lambda v_\lambda, \quad \mathbf{k}\cdot\boldsymbol{\omega}' = \zeta' \approx r^{-1}\partial(rv_\lambda)/\partial r$$

where $\mathbf{j}_\lambda$ is the unit vector in the azimuthal direction (positive counterclockwise) and r is the distance from the axis of the vortex. Letting $\mathbf{i}_\lambda$ be the unit vector in the radial direction, we then have

$$\mathbf{U}\times\boldsymbol{\omega} \approx \mathbf{i}_\lambda \frac{v_\lambda}{r}\frac{\partial}{\partial r}(rv_\lambda)$$

Assuming that the vertical scale is much larger than the radial scale, the Laplacian in cylindrical coordinates can be approximated by

$$\nabla^2 \approx \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)$$

Thus, from (9.54) the dynamical component of the pressure perturbation in the vortices (designated p_{dyn}) can be expressed as

$$\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{r}{\rho_0}\frac{\partial p_{dyn}}{\partial r}\right) \approx -\frac{1}{r}\frac{\partial}{\partial r}\left[r\frac{\partial\left(v_\lambda^2/2\right)}{\partial r}\right] + \frac{1}{r}\frac{\partial}{\partial r}\left[v_\lambda\frac{\partial(rv_\lambda)}{\partial r}\right] = \frac{1}{r}\frac{\partial v_\lambda^2}{\partial r} \tag{9.55}$$

Integrating (9.55) with respect to r, we obtain the equation of cyclostrophic balance (see Section 3.2.4):

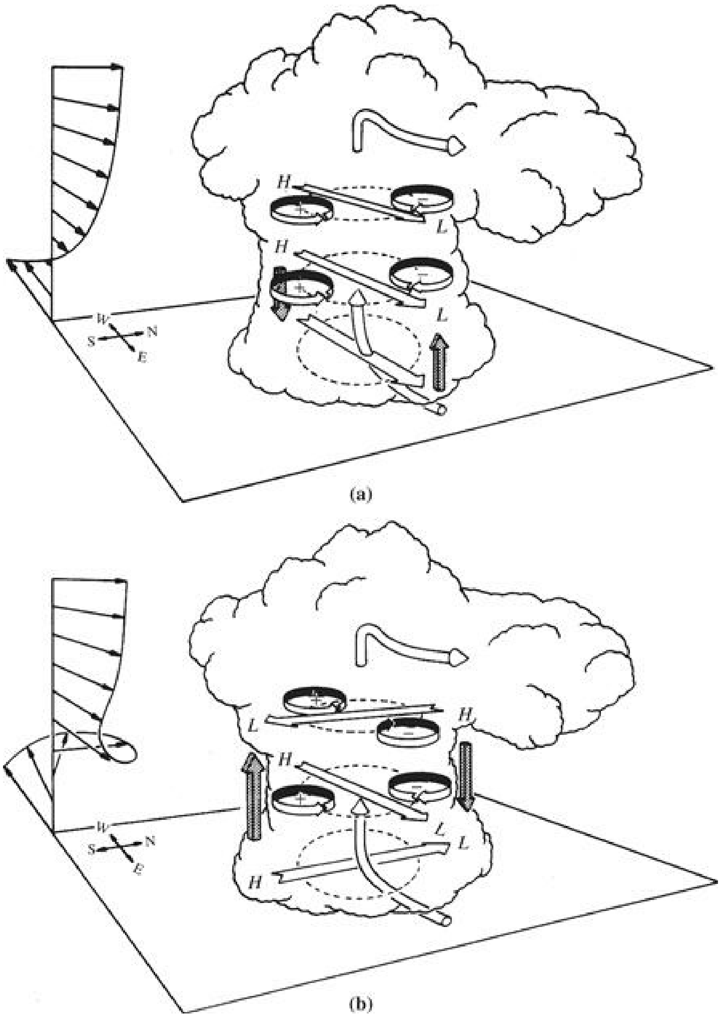
$$\rho_0^{-1}\partial p_{dyn}/\partial r \approx v_\lambda^2/r \tag{9.56}$$

Hence, there is a pressure minimum at the vortex center irrespective of whether the rotation is cyclonic or anticyclonic. The strong midtropospheric rotation induced by vortex tube twisting and stretching creates a “centrifugal pump” effect, which causes a negative dynamical pressure perturbation centered in the vortices in the midtroposphere. This in turn produces an upward-directed dynamical contribution to the vertical component of the pressure gradient force, and thus provides an upward acceleration, which produces updrafts in the cores of the counterrotating vortices as depicted in Figure 9.12. These updrafts are separated by a downdraft that leads to a splitting of the storm and the development of two new centers of convection that move to the right and left of the original storm (Figure 9.12b).

As discussed in this section, the tilting and stretching of horizontal vorticity associated with the vertical shear of the basic state wind can account for the development of mesoscale rotating supercells. This process does not, however, appear to be able to produce the large vorticities observed in the tornadoes that often accompany supercell thunderstorms. Numerical simulations suggest that these tend to involve tilting and stretching of especially strong horizontal vorticity produced by horizontal gradients in buoyancy that occur near the surface along the *gust front* where negatively buoyant outdrafts produced by convective downdrafts meet moist, warm boundary layer air.

9.6.2 The Right-Moving Storm

When the environmental wind shear is unidirectional, as in the case discussed previously (see also Figure 9.13a), the anticyclonic (left-moving) and cyclonic (right-moving) updraft cores are equally favored. In most severe storms in the central United States, however, the mean flow turns anticyclonically with height; this directional shear in the environment favors the right-moving storm center, while inhibiting the left-moving center. Thus, right-moving storms are observed far more than left-moving storms.



(a)

(b)

Figure 9.13 Pressure and vertical vorticity perturbations produced by interaction of the updraft with environmental wind shear in a supercell storm. (a) Wind shear does not change direction with height. (b) Wind shear turns clockwise with height. *Broad open arrows* designate the shear vectors. *H* and *L* designate high and low dynamical pressure perturbations, respectively. *Shaded arrows* show resulting disturbance vertical pressure gradients.

After Klemp, 1987. Used with permission from Annual Reviews.

The dominance of the right-moving storm can be understood qualitatively by again considering the dynamical pressure perturbations. We define the basic state wind shear vector $\overline{\mathbf{S}} \equiv \partial\overline{\mathbf{V}}/\partial z$, which is assumed to turn clockwise with height. Noting that the basic state horizontal vorticity in this case is

$$\overline{\omega} = \mathbf{k} \times \overline{\mathbf{S}} = -\mathbf{i}\partial\overline{v}/\partial z + \mathbf{j}\partial\overline{u}/\partial z$$

we see that there is a contribution to the dynamic pressure in (9.54) of the form

$$\boldsymbol{\nabla} \cdot (\mathbf{U}' \times \overline{\omega}) \approx -\boldsymbol{\nabla} \cdot (w'\overline{\mathbf{S}})$$

From (9.54) the sign of the pressure perturbation due to this effect may be determined by noting that

$$\nabla^2 p_{dyn} \sim -p_{dyn} \sim -\frac{\partial}{\partial x}(w'S_x) - \frac{\partial}{\partial y}(w'S_y) \tag{9.57}$$

which shows that there is a positive dynamical pressure perturbation upshear of the cell and a negative perturbation pressure downshear (analogous to the positive pressure perturbation upwind and negative perturbation downwind of an obstacle). The resulting pattern of dynamical pressure perturbations is shown in Figure 9.13. In the case of unidirectional shear (Figure 9.13a), the induced pressure pattern favors updraft growth on the leading edge of the storm. However, when the shear vector rotates clockwise with height as in Figure 9.13b, (9.57) shows that a dynamical pressure disturbance pattern is induced in which there is an upward-directed vertical pressure gradient force on the flank of the cyclonically rotating cell and a downward-directed pressure gradient force on the flank of the anticyclonic cell. Thus, in the presence of clockwise rotation of the environmental shear, stronger updrafts are favored in the right-moving cyclonic vortex to the south of the initial updraft.

9.7 Hurricanes

Hurricanes, which are also referred to as tropical cyclones and typhoons in some parts of the world, are intense vortical storms that develop over the tropical oceans in regions of very warm surface water. Although the winds and convective clouds observed in hurricanes are not truly axisymmetric about the vortex center, the fundamental aspects of hurricane dynamics can be modeled by idealizing the hurricane as an axisymmetric vortex. Typically hurricanes have radial scales of several hundred kilometers, similar to those of some midlatitude synoptic systems. However, the horizontal scale of the region of intense convection and strong winds in a hurricane is typically only about 100 km in radius. Thus, it is reasonable to classify hurricanes as mesoscale systems.

Unlike the rotating convective storms treated in the previous section, the hurricane vortex cannot be understood without including the rotation of Earth in the vorticity balance. The rapid rotation observed in hurricanes is produced by concentration of the vertical component of absolute vorticity by vortex stretching, not by tilting horizontal vorticity into the vertical. Maximum tangential wind speeds in these storms range typically from 50 to 100 m s^{-1}. For such high velocities and relatively small scales, the centrifugal force term cannot be neglected compared to the Coriolis force. Thus, to a first approximation, the azimuthal velocity in a steady-state hurricane is in gradient wind balance with the radial pressure gradient force. Hydrostatic balance also holds on the hurricane scale, which implies that the vertical shear of the azimuthal velocity is a function of the radial temperature gradient.

The hurricane *primary circulation* consists of cyclonic flow around a center of low pressure, with the maximum wind located roughly 15 to 50 km from the center and about 1 to 1.5 km altitude, near the top of the boundary layer (Figure 9.14). The hurricane *secondary circulation* consists of radial inflow at low levels, which rises in the eyewall while conserving equivalent potential temperature, and then moves radially away from the storm at upper levels in the outflow layer before gradually subsiding in the surrounding environment. This secondary circulation is driven by radial gradients of latent heating in the eyewall and vertical gradients of momentum dissipation in the boundary layer; it may be understood through a generalization of (9.15) to include these effects. Within the eyewall the θ_e and angular momentum surfaces coincide, so parcel ascent in the eyewall is neutral with respect to conditional symmetric instability and thus does not require external forcing. The eyewall surrounds a central eye of radius 15 to 50 km where subsidence occurs, resulting in relatively warm air that is cloud-free above the boundary layer.

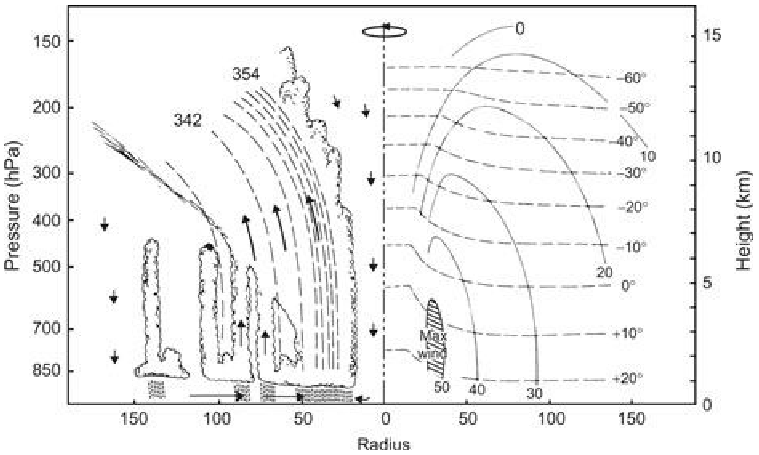


Figure 9.14 Schematic illustration of hurricane structure. *Thin solid curves* show the azimuthal wind speeds, which peak in the lower troposphere. Mainly *horizontal dashed lines* give the temperature field, which reflects the decrease with height in the troposphere. Mainly *vertical thin dashed lines* show equivalent potential temperature, which is well mixed in the eyewall. *Arrows* denote the secondary circulation, with radial inflow at low levels, ascent in the eyewall, and subsidence in the environment and inside the eye.

Adapted from Wallace and Hobbs, 2006.

9.7.1 Dynamics of Mature Hurricanes

Because of the fact that tropical cyclones are primarily axisymmetric and have well-defined centers, angular momentum is a useful measure of storm intensity. Angular momentum about the vertical axis is defined by

$$m = rv + fr^2 \tag{9.58}$$

The azimuthal gradient wind, *V*, in pressure coordinates is determined by

$$\frac{\partial \Phi}{\partial r} = \frac{V^2}{r} + fV \tag{9.59}$$

and hydrostatic balance in terms of the specific volume, α, is determined by

$$\frac{\partial \Phi}{\partial p} = -\alpha \tag{9.60}$$

A thermal wind equation is obtained by equating $\partial/\partial p$ (9.59) and $\partial/\partial r$ (9.60), giving

$$-\frac{\partial \alpha}{\partial r} = \frac{2m}{r^3}\frac{\partial m}{\partial p} \tag{9.61}$$

Although this form of the thermal wind equation is different than the one discussed in Chapter 3, the general interpretation is similar: Radial gradients in density are linked to vertical gradients in angular momentum. Specific volume may be eliminated from (9.61) in favor of entropy using so-called Maxwell relations[3]:

$$\begin{aligned} \left.\frac{\partial \alpha}{\partial r}\right|_p &= \frac{\partial \alpha}{\partial s}\frac{\partial s}{\partial r} \\ \left.\frac{\partial \alpha}{\partial s}\right|_p &= \left.\frac{\partial T}{\partial p}\right|_s \end{aligned} \tag{9.62}$$

Since entropy and angular momentum contours are, to good approximation, congruent in the eyewall, we take entropy to be a function of angular momentum alone. This implies that radial entropy gradients are proportional to radial angular momentum gradients:

$$\left.\frac{\partial s}{\partial r}\right|_p = \frac{\partial s}{\partial m}\frac{\partial m}{\partial r} \tag{9.63}$$

so that (9.61) becomes

$$\frac{2m}{r^3}\frac{\partial m}{\partial p} = -\left.\frac{\partial T}{\partial p}\right|_s \frac{\partial s}{\partial m}\frac{\partial m}{\partial r} \tag{9.64}$$

Alternatively, noting that

$$\frac{\partial m/\partial p}{\partial m/\partial r} = -\left.\frac{\partial r}{\partial p}\right|_m \tag{9.65}$$

thermal wind can be expressed in terms of the slope of angular momentum (and entropy) surfaces

$$\left.\frac{2m}{r^3}\frac{\partial r}{\partial p}\right|_m = -\left.\frac{\partial T}{\partial p}\right|_s \frac{\partial s}{\partial m} \tag{9.66}$$

Although (9.66) applies at each level in the eyewall, it proves useful to integrate this equation along the angular momentum surface that extends from the point of maximum wind above the boundary layer, **b**, to a point in the outflow, **t**, (Figure 9.15). Integration over the interval yields

$$\frac{1}{r_b^2} - \frac{1}{r_t^2} = \frac{1}{m}\frac{\partial s}{\partial m}(T_t - T_b) \tag{9.67}$$

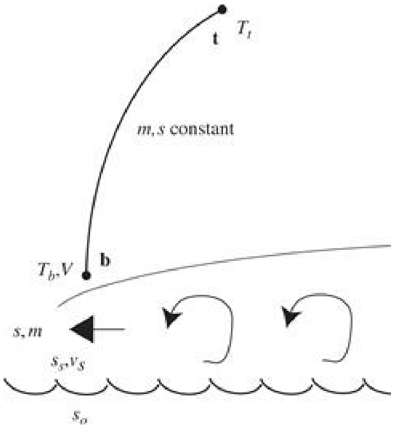


Figure 9.15 Essential elements of the potential intensity theory. The *bold curve* with endpoints b and t denotes the constant entropy and angular momentum surface along which the thermal wind equation is integrated. The curve originates at the top of the boundary layer, which is denoted by the *sloping gray curve*. Within the boundary layer, radial advection of entropy and angular momentum, denoted by the *bold arrow* (positive for angular momentum and negative for entropy) is balanced by vertical flux divergence due to turbulence, denoted by the *curved arrows*. At the sea surface, there exists an entropy difference between values in the atmosphere, s_s and ocean, s_o. The surface wind speed is given by v_s.

Assuming that the outflow radius is much larger than the radius of maximum wind, the second left-side term may be neglected, which reveals that angular momentum depends on the radius of maximum wind, r_b, as anticipated by (9.58), the temperature difference between point of maximum wind and the outflow, and the relationship between entropy and angular momentum. This last relationship is determined by the following closure scheme for the boundary layer. If the dominant balance in the entropy and angular momentum equations in the boundary layer is between radial advection and turbulent mixing, then

$$\begin{aligned} -u\frac{\partial m}{\partial r} &\approx \frac{\partial \tau_m}{\partial z} \\ -u\frac{\partial s}{\partial r} &\approx \frac{\partial \tau_s}{\partial z} \end{aligned} \tag{9.68}$$

where τ_m and τ_s are the vertical turbulent fluxes (see Section 8.3) of angular momentum and entropy, respectively. Dividing these equations indicates

$$\left.\frac{\partial s}{\partial m}\right|_r = \left.\frac{\partial \tau_s}{\partial \tau_m}\right|_z \tag{9.69}$$

which indicates that the relationship between entropy and angular momentum is controlled by the relationship between the turbulent fluxes. We now assume that these turbulent fluxes are independent of depth in the boundary layer, so they may be determined from surface fluxes, which are given by aerodynamic formulae (Section 8.3.1):

$$\begin{aligned} \tau_{s0} &= C_S|\mathbf{V}|_0(s_s - s_o) \\ \tau_{m0} &= C_D|\mathbf{V}|_0(m_s - m_o) = C_D|\mathbf{V}|_0 r v_s \end{aligned} \tag{9.70}$$

Here, $|\mathbf{V}_0|$ is the surface wind speed, and C_S and C_D are the surface exchange coefficients for entropy and angular momentum, respectively. The surface value of entropy (angular momentum) for the atmosphere is given by s_s (m_s), and analogously, for the surface of the ocean, we have s_o and m_o. In the last equality for the angular momentum flux we assumed that the azimuthal speed of the ocean is small compared to the atmosphere, v_s. Note that no assumption has been made about the azimuthal wind in the boundary layer; in particular, the wind is not assumed to be in gradient wind balance. Combining equations (9.70) gives

$$\tau_{s0} = \left[\frac{C_S(s_s - s_o)}{C_D r v_s}\right] \tau_{m0} \tag{9.71}$$

which from (9.69) immediately recovers

$$\left.\frac{\partial s}{\partial m}\right|_r = \frac{C_S(s_s - s_o)}{C_D r v_s} \tag{9.72}$$

Armed with the relationship between entropy and angular momentum, we return to (9.67) and use (9.72) assuming large r_t to yield

$$\frac{m v_s}{r_b} = \frac{C_S}{C_D}(s_o - s_s)(T_b - T_t) \tag{9.73}$$

Using (9.58) indicates that the left side of (9.73) is equal to

$$V v_s + 0.5 f r_b v_s \tag{9.74}$$

For typical hurricane parameters the second term may be neglected within approximately 5% error. Finally, we use a "surface wind reduction factor" to relate the gradient wind at the point of maximum wind, V, to the windspeed at the ocean surface (not a gradient wind), $v_s = aV$; observations indicate that the value of a is often around 0.8 to 0.9. These substitutions yield the "potential intensity" formula for a steady-state storm (Emanuel, 1988)

$$\boxed{V^2 = a^{-1}\frac{C_S}{C_D}(s_o - s_s)(T_b - T_t)} \tag{9.75}$$

Aside from the factor a, which has about a 5% effect, we see that steady-state hurricane intensity depends on the square root of the ratio of the surface exchange coefficients, the surface entropy difference between ocean and atmosphere, and the temperature difference between the top of the boundary layer and the outflow. The ratio of surface exchange coefficients is not known exactly, but is thought to be around 0.7. The surface entropy difference is due to a combination of moisture and thermal effects, with a net flux of both quantities from the ocean to the atmosphere. Therefore, although it may seem counterintuitive, (9.75) indicates that the potential intensity *in steady state* is higher when the atmosphere is colder and drier; the physical reason is that such situations promote the development of larger radial gradients of entropy. Although T_b applies at the top of the boundary layer, it is commonly approximated by the sea-surface temperature, which allows one to create global maps estimating the potential intensity of storms that form in these areas.

Since (9.75) estimates steady-state intensity, it does not apply to short-term intensity fluctuations. As such, one may expect to find storms that occasionally exceed (9.75), even when all assumptions are well satisfied; nevertheless, the time average over these transient periods should be limited by (9.75). Of course, if the assumptions behind (9.75) fail, then the estimate will as well. One aspect where this may occur fairly often is in the gradient wind assumption at the point of maximum wind. Specifically, vertical advection of azimuthal momentum in the boundary layer introduces an extra term in the momentum equation not considered by gradient wind balance. Below the level of maximum wind, the contribution from this term is of the same sign as the pressure gradient force (Figure 9.16). In steady state, the centrifugal and Coriolis forces required to balance these effects are larger than in the absence of vertical advection. As a result, the azimuthal wind tends to be supergradient near the top of the boundary layer, and this effect can become large if vertical advection is also large, which happens when the radial gradient of radial wind is large. As air parcels approach the level of maximum wind, vertical advection approaches zero, resulting in a net outward force that decelerates the air parcel. As a result, air tends to flow radially outward just above the boundary layer, particularly in cases when the supergradient effect is large.

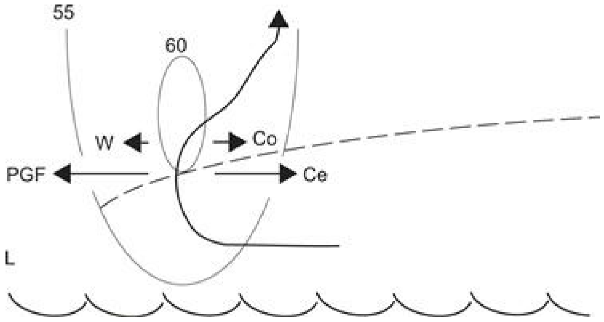


Figure 9.16 Illustration of unbalanced contributions to the azimuthal wind following an air parcel from the boundary layer into the free atmosphere (*thick arrow*) near the radius of maximum wind (isotachs of azimuthal wind speed given by *gray contours* in m/s). A large pressure gradient force is directed inward toward low pressure in the eye ("L"). If vertical motion is small, then the vertical advection of horizontal azimuthal momentum ("W") may be neglected, which results in a balance between the pressure gradient force and the sum of the Coriolis ("Co") and centrifugal ("Ce") forces—that is, gradient wind balance. If vertical advection is large, a stronger azimuthal wind is required to achieve balance, which yields winds stronger than that for gradient balance. In this situation, when parcels approach the level maximum wind, vertical advection approaches zero, resulting in a net outward force that decelerates the wind.

9.7.2 Hurricane Development

The origin of tropical cyclones is still a matter of active research. Although there are many tropical disturbances each year, only rarely does one develop into a hurricane. Thus, the development of a hurricane requires rather special conditions. As shown in Chapter 7, linear stability theory is quite successful in accounting for the development of extratropical baroclinic disturbances. As such, this approach provides a reasonable starting point for theoretical investigations of hurricane development, where one seeks the conditions under which unstable amplification of an initial disturbance can occur.

In the tropics, however, the only well-documented linear instability is conditional instability. This instability has its maximum growth rate on the scale of an individual cumulus cloud. Therefore, it cannot explain the synoptic-scale organization of the motion. Observations indicate, moreover, that the mean tropical atmosphere is not saturated, even in the planetary boundary layer. Thus, a parcel must undergo a considerable amount of forced ascent before it reaches its LFC and becomes positively buoyant. Such forced parcel ascent can only be caused by small-scale motions, such as turbulent plumes in the boundary layer. The efficacy of boundary layer turbulence in producing parcel ascent to the LFC clearly depends on the temperature and humidity of the boundary layer environment. In the tropics it is difficult to initialize deep convection unless the boundary layer is brought toward saturation and destabilized, which may occur if there is large-scale (or mesoscale) ascent in the boundary layer. Thus, convection tends to be concentrated in regions of large-scale low-level convergence. This concentration does not arise because the large-scale convergence directly “forces” the convection, but rather because it preconditions the environment to be favorable to parcel ascent to the LFC.

Cumulus convection and the large-scale environmental motion are thus often viewed as cooperatively interacting. In this viewpoint, diabatic heating due to latent heat released by cumulus clouds produces a large-scale (or mesoscale) cyclonic disturbance; this disturbance, in turn, through boundary layer pumping, drives the low-level moisture convergence necessary to maintain an environment favorable to the development of cumulus convection. There have been attempts to formalize these ideas into a linear stability theory; this is often referred to as *conditional instability of the second kind* (CISK), which attributes hurricane growth to the organized interaction between the cumulus-scale and the large-scale moisture convergence. This interaction process is indicated schematically in Figure 9.17a. The CISK model for hurricane development has not been successful because there is little evidence that such interaction leads to a growth rate maximum on the observed scale of hurricanes.

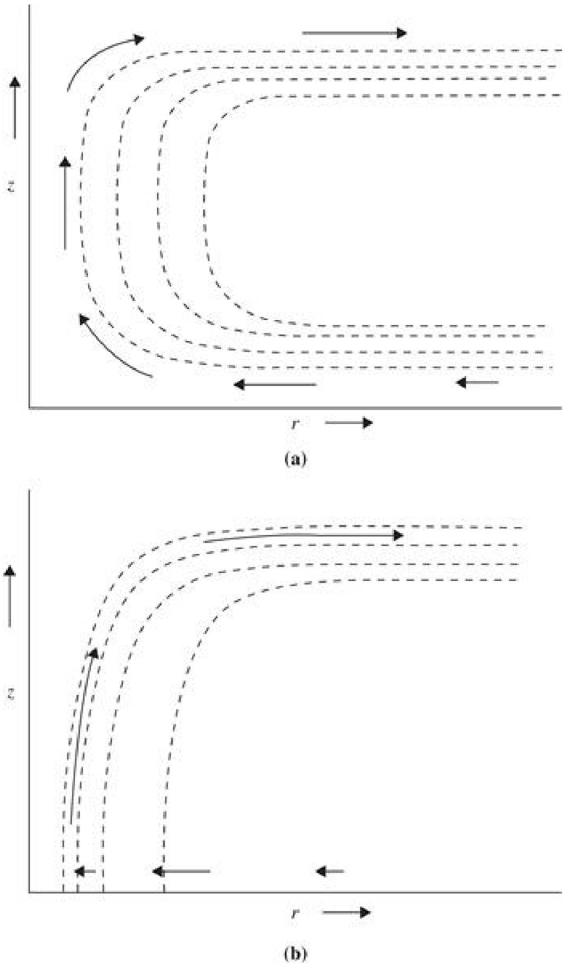


Figure 9.17 Meridional cross-sections showing the relationship between surfaces of constant saturation θ_e (*dashed contours*; values decreasing as r increases) and meridional circulation (*arrows*) in CISK (a) and WISHE (b) theories for hurricane development. (a) Frictionally induced boundary layer convergence moistens the environment and destabilizes it through layer ascent. This enables small-scale plumes to reach their levels of free convection easily and to produce cumulonimbus clouds. Diabatic heating due to the resulting precipitation drives large-scale circulation and thus maintains the large-scale convergence. (b) The saturation θ_e is tied to the θ_e of the boundary layer. The warm core occurs because enhanced surface fluxes of latent heat increase θ_e there.

After Emanuel, 2000.

A different view of the stability of the tropical atmosphere, referred to as *wind-induced surface heat exchange* (WISHE), is based on air–sea interactions. According to the WISHE view, illustrated schematically in Figure 9.17b, the potential energy for hurricanes arises from the thermodynamic disequilibrium between the atmosphere and the underlying ocean, which is quantified through the entropy difference in (9.75). The efficacy of air–sea interaction in providing potential energy to balance frictional dissipation depends on the rate of transfer of latent heat from the ocean to the atmosphere. This is a function of surface wind speed; strong surface winds, which produce a rough sea surface, can increase the evaporation rate greatly. Thus, hurricane development is enhanced by the presence of a finite-amplitude initiating disturbance, such as an equatorial wave, to provide the winds required to produce strong evaporation. Given a suitable initial disturbance, a feedback may occur in which an increase in inward-spiraling surface winds increases the rate of moisture transfer from the ocean, which increases the intensity of the convection, which further increases the secondary circulation. While it appears that observed cases of tropical cyclogenesis require finite-amplitude initial disturbances, the reason is likely that such disturbances require a shorter period of time to reach maturity. As initial disturbance amplitude decreases, the time required to reach maturity increases, as does the probability of adverse environmental factors disrupting the process.

Many environmental factors may slow or hinder the development process, but the most important are vertical wind shear and dry air. Vertical shear “tilts” the convection so that the heating is not vertically aligned. Moreover, shear produces a dipole of vertical motion about the nascent disturbance, which disturbs the circulation from axisymmetry and results in dry air in the midtroposphere on the upshear side of the storm. Dry air may be entrained into the convection, resulting in strong downdrafts that transport low-entropy air to the surface, reducing conditional instability. Thus, storm development can be substantially inhibited if dry air is continually transported into the midtroposphere above the developing storm. As a result, developing systems that have a deep layer with closed streamlines in a framework moving with the disturbance are favored, since the air is trapped within the disturbance, and convection may moisten the troposphere through detrainment without being replaced by drier environmental air.

Suggested References

urran, 1990. Durran (1990) reviews the dynamics of mountain waves and downslope windstorms.

liassen, 1990. Eliassen (1990) discusses the secondary circulation within frontal zones.

manuel. Emanuel’s *Atmospheric Convection* is a graduate-level text covering the subject of convection, including symmetric instability, in great detail.

manuel, 2000. Emanuel (2000) provides a qualitative explanation of the WISHE theory for hurricane development.

eppert and Wang, 2001. Keppert and Wang (2001) discuss the importance of vertical advection in generating supergradient winds in tropical cyclone boundary layers.

oskins, 1982. Hoskins (1982) discusses the semigeostrophic theory of frontogenesis and summarizes results from numerical models.

ouze. Houze’s *Cloud Dynamics* discusses convective storms at the graduate level.

lemp, 1987. Klemp (1987) describes the dynamics of tornadic thunderstorms.

Problems

9.1. Show by transforming from θ coordinates to height coordinates that the Ertel potential vorticity P is proportional to $F^2N_s^2 - S^4$. See (9.27).

9.2. Starting with the linearized Boussinesq equations for a basic state zonal flow that is a function of height, derive (9.35) and verify the form given for the Scorer parameter.

9.3. Show that for stationary flow over an isolated ridge in the broad ridge limit ($k_s \ll m_s$), the group velocity vector is directed upward so that energy cannot propagate upstream or downstream of the ridge.

9.4. An air parcel at 920 hPa with temperature 20°C is saturated (mixing ratio 16 g kg^{-1}). Compute θ_e for the parcel.

9.5. Suppose that the mass of air in an entraining cumulus updraft increases exponentially with height so that $m = m_0 e^{z/H}$, where $H = 8$ km and m_0 is the mass at a reference level. If the updraft speed is 3 m s^{-1} at 2-km height, what is its value at a height of 8 km, assuming that the updraft has zero net buoyancy?

9.6. Verify the approximate relationship between moist static energy and θ_e given by equation (2.71).

9.7. The azimuthal velocity component in some hurricanes is observed to have a radial dependence given by $v_\lambda = V_0(r_0/r)^2$ for distances from the center given by $r \geq r_0$. Letting $V_0 = 50\,\mathrm{m\,s^{-1}}$ and $r_0 = 50$ km, find the total geopotential difference between the far field $(r \rightarrow \infty)$ and $r = r_0$, assuming gradient wind balance and $f_0 = 5 \times 10^{-5}\,\mathrm{s^{-1}}$. At what distance from the center does the Coriolis force equal the centrifugal force?

9.8. Starting with equation (9.59), derive the angular momentum form of the gradient wind balance for an axisymmetric vortex given by equation (9.61).

MATLAB Exercises

M9.1. The MATLAB script **surface_front_1.m** demonstrates the concentration of an initial temperature gradient by an imposed deformation field. The temperature field is initially taken to have a linear decrease in *y*: $T(y) = 300 - 10^{-2}y$, where *y* is in kilometers. The horizontal wind is given by the streamfunction $\psi(x,y) = -15k^{-1}\sin(kx)\sin(my)$, where $k = 2\pi/3 \times 10^6$ and $m = \pi/3 \times 10^6$. Evaluate and contour-plot the deformation field for this streamfunction, where deformation is defined as $\partial u/\partial x - \partial v/\partial y$. Run the model and observe where the temperature gradient increases. Modify the code to plot the maximum temperature gradient as a function of time. For how long does the gradient increase approximately exponentially? What is the *e*-folding time for increase of the temperature gradient at the origin?

M9.2. The MATLAB script **profile_1.m** and the function **Tmoist.m** use pressure, temperature, and humidity data in the file **sounding.dat** to calculate and graph vertical profiles of temperature, dewpoint, and the temperature corresponding to pseudoadiabatic ascent from the lifting condensation level for a parcel lifted from the lowest level of the sounding. Modify this script to plot profiles of the potential temperature, the equivalent potential temperature, and the saturated potential temperature. *Hint:* Temperature at saturation can be calculated by inverting the expression for saturation vapor pressure, which is obtainable by integrating equation (D.4):

$$e_s = e_{s,tr}\exp\left[\frac{L_v}{R_v}\left(\frac{1}{T_{tr}} - \frac{1}{T}\right)\right]$$

where $e_{s,tr} = 6.11\,\mathrm{hPa}, T_{tr} = 273.16\,\mathrm{K}, \text{and } L_v = 2.5\times 10^6\,\mathrm{J\,kg^{-1}}$ are the saturation vapor pressure at the triple point temperature, the triple point temperature, and the latent heat of vaporization, respectively.

M9.3. For the thermodynamic sounding of Problem M9.2, compute the CAPE and the profile of vertical velocity for a parcel lifted from the lowest layer in the absence of entrainment. What is the maximum vertical velocity, and what is the distance that the parcel overshoots its level of neutral buoyancy?

M9.4. The MATLAB script **lee_wave_1.m** uses the Fourier series approach to solve for flow over an isolated ridge for the case of constant *N* and constant mean zonal flow. For the case of $\bar{u} = 50\,\mathrm{m\,s^{-1}}$, run the script for various ridge widths to determine the width is required for significant vertical propagation (i.e., at what ridge width do phase lines of vertical velocity begin to tilt with height)?

M9.5. The script **lee_wave_2.m** plots an approximate analytic solution for flow over a mountain in the wide ridge limit. The height profile is given by

$$h(x) = h_0 L^2 / \left(L^2 + x^2\right)$$

where L is the width scale of the ridge and h_0 is the ridge height. For a fixed ridge height of 2 km, vary the input zonal mean wind and comment on the dependence of the contoured solutions. (Note that the vertical scale is given in vertical wavelengths as in Figure 9.7.) Modify the script to vary the mountain height and determine what fraction of the vertical wavelength the mountain height must have in order for wave breaking to occur. (The onset of wave breaking occurs when the streamlines become vertical.)

M9.6. Consider stationary gravity waves forced by flow over a sinusoidal lower boundary for a case in which the static stability decreases rapidly with height at about the 6-km level. Thus, the buoyancy frequency is altitude dependent and the simple analytic solution (9.32) no longer applies. The MATLAB script named **linear_grav_wave_1.m** gives a highly accurate numerical solution for this situation.

(a) Describe the qualitative change in the wave behavior as the zonal wavelength is changed over the range of 10 to 100 km. Be sure to comment on momentum flux as well as on vertical velocity.

(b) Determine as accurately as you can what the minimum zonal wavelength is for vertical propagation to occur in the upper part of the domain (above the 6-km level).

(c) Determine how the amplitudes of the momentum flux and momentum flux convergence at z = 6 km change as the zonal wavelength is increased from 20 to 100 km. Do several different wavelengths so you can plot graphs of the dependence of the magnitudes of the momentum flux and zonal force on the wavelength of the sinusoidal topography. Use MATLAB to plot these two graphs.

---

[1] Some authors reserve this name for a version of the equations written in a transformed set of coordinates called geostrophic coordinates (e.g., Hoskins, 1975).

[2] A conservable variable is one that is conserved following the motion in the absence of sources and sinks (e.g., a chemical trace constituent).

[3] Wikipedia has a particularly good description of these relationships among thermodynamic variables, which result from constraints in the First Law of Thermodynamics and commutation of partial derivatives.

Chapter 10

The General Circulation

In its broadest sense, the *general circulation* of the atmosphere is usually considered to include the totality of motions that characterizes the global-scale atmospheric flow. Specifically, the study of the general circulation is concerned with the dynamics of climate—that is, with the temporally averaged structures of the fields of wind, temperature, humidity, precipitation, and other meteorological variables. The general circulation may thus be considered to consist of the flow averaged in time over a period sufficiently long to remove the random variations associated with individual weather systems but short enough to retain monthly and seasonal variations.

In the past, both observational and theoretical studies of the general circulation concentrated on the dynamics of the zonally averaged flow. The time-averaged circulation is, however, highly dependent on longitude due to longitudinally asymmetric forcing by orography and land–sea heating contrasts. The longitudinally dependent components of the general circulation may be separated into *quasi-stationary circulations*, which vary little in time; *monsoonal circulations*, which are seasonally reversing; and various subseasonal and interannual components, which together account for *low-frequency variability*. A complete understanding of the physical basis for the general circulation requires an explanation not only for the zonally averaged circulation but for the longitudinally and time-varying components as well.

Nevertheless, to introduce the study of the general circulation, it proves useful to isolate those processes that maintain the zonal-mean flow (i.e., the flow averaged around latitude circles). This approach follows naturally from the linear wave studies of previous chapters in which flow fields were split into zonal-mean and longitudinally dependent eddy components. In this chapter, however, we concentrate not on the development and motion of the eddies but on the influence of the eddies on the structure of the zonal-mean circulation. Focusing on the zonal mean allows us to isolate those features of the circulation that are not dependent on continentality and should thus be common to all thermally driven rotating fluid systems. In particular, we discuss the angular momentum and energy budgets of the zonally averaged flow. We also show that the mean meridional circulation (i.e., the circulation consisting of the zonal-mean vertical and meridional velocity components) satisfies a diagnostic equation analogous to the vertical motion equation of Section 6.5, but with the forcing determined by the distributions of diabatic heating and eddy heat and momentum fluxes.

Following our discussion of the zonal-mean circulation, we consider the longitudinally varying time-averaged circulation. In this chapter the primary emphasis is on extratropical aspects of the circulation; these can be discussed within the framework of quasi-geostrophic theory. The general circulation of the tropics is considered in Chapter 11.

10.1 The Nature of the Problem

Theoretical speculation on the nature of the general circulation has quite a long history. Perhaps the most important early work on the subject was that of an 18th-century Englishman, George Hadley. Hadley, in seeking a cause for the trade wind circulation, realized that this circulation must be a form of thermal convection driven by the difference in solar heating between equatorial and polar regions. He visualized the general circulation as consisting of a zonally symmetric overturning in which the heated equatorial air rises and flows poleward, where it cools, sinks, and flows equatorward again. At the same time, the Coriolis force deflects the poleward moving air at the upper levels to the east and the equatorward-moving air near the surface to the west. The latter is, of course, consistent with the observed surface winds in the trade wind zone, which are northeasterly in the Northern Hemisphere and southeasterly in the Southern Hemisphere. This type of circulation is now called a *Hadley circulation* or *Hadley cell.*

Although a circulation consisting of Hadley cells extending from equator to pole in each hemisphere is mathematically possible in the sense that such a circulation would not violate the laws of physics, the observed Hadley circulation is confined to the tropics. Evidence from a number of studies indicates that for conditions existing in Earth's atmosphere, a symmetric hemispheric-wide Hadley circulation would be baroclinically unstable. If such a circulation were to become established by some mechanism, it would quickly break down outside the tropics as baroclinic eddies developed and modified the zonal-mean circulation through their heat and momentum fluxes.

The observed general circulation thus cannot be understood purely in terms of zonally symmetric processes. Rather, it can be thought of qualitatively as developing through three-dimensional interactions among radiative and dynamical processes. In the mean, the net solar energy absorbed by the atmosphere and Earth must equal the infrared energy radiated back to space by the planet. The annually averaged solar heating is, however, strongly dependent on latitude, with a maximum at the equator and a minimum at the poles. The outgoing infrared radiation, however, is only weakly latitude dependent. Thus, there is a net radiation surplus in the equatorial region and a deficit in the polar region.

This differential heating warms the equatorial atmosphere relative to higher latitudes and creates a pole-to-equator temperature gradient. Therefore, it produces a growing store of zonal-mean available potential energy. At some point the westerly thermal wind (which must develop if the motion is to be balanced geostrophically in the presence of the pole-to-equator temperature gradient) becomes baroclinically unstable. As shown in Chapter 8, the resulting baroclinic waves transport heat poleward. These waves will intensify until their heat transport (together with the heat transported by planetary waves and ocean currents) is sufficient to balance the radiation deficit in the polar regions so that the pole-to-equator temperature gradient ceases to grow. At the same time, these perturbations convert potential energy into kinetic energy, thereby maintaining the kinetic energy of the atmosphere against the effects of frictional dissipation.

From a thermodynamic point of view, the atmosphere may be regarded as a "heat engine" that absorbs net heat at relatively warm temperatures in the tropics (primarily in the form of latent heat due to evaporation from the sea surface) and gives up heat at relatively cool temperatures in the extratropics. In this manner, net radiation generates available potential energy, which is in turn partially converted to kinetic energy, which does work to maintain the circulation against frictional dissipation. Only a small fraction of the solar energy input is actually converted to kinetic energy. Thus, from an engineer's viewpoint, the atmosphere is a rather inefficient heat engine. However, if due account is taken of the many constraints operating on atmospheric motions, it appears that the atmosphere may in fact generate kinetic energy about as efficiently as dynamically possible.

The preceding qualitative discussion suggests that the gross features of the general circulation outside the tropics can be understood on the basis of quasi-geostrophic theory, since, as we have seen, baroclinic instability is contained within the quasi-geostrophic framework. In view of this fact, and to keep the equations as simple as possible, our discussion of the zonally averaged and longitudinally varying components of the circulation outside the tropics will concentrate on those aspects that can be qualitatively represented by the quasi-geostrophic equations on the midlatitude β plane.

It should be recognized that a quasi-geostrophic model cannot provide a complete theory of the general circulation because in the quasi-geostrophic theory a number of assumptions are made concerning scales of motion that restrict in advance the possible types of solutions. Quantitative modeling of the general circulation requires complicated numerical models based on the primitive equations in spherical coordinates. The ultimate objective of these modeling efforts is to simulate the general circulation so faithfully that the climatological consequences of any change in the external parameters (e.g., the atmospheric concentration of carbon dioxide) can be predicted accurately. Present models can provide fairly accurate simulations of the current climate and plausible predictions of the response of the climate system to changes in external conditions. However, uncertainties in the representations of a number of physical processes, particularly clouds and precipitation, limit the confidence that can be placed in quantitative climate change predictions based on such models.

10.2 The Zonally Averaged Circulation

The observed global distribution of the longitudinally averaged zonal wind for the two solstice seasons was shown in Figure 6.1. Although there are important interhemispheric differences, the flow in both hemispheres is characterized by westerly jets with maximum zonal velocities at about 30° to 35° latitude near the tropopause. These westerlies vary in longitude, especially in the Northern Hemisphere (refer to Figure 6.2), but there is a significant zonally symmetric component, which we will refer to as the *mean zonal wind.*

Although departures of the time-averaged flow from zonal symmetry are important aspects of the general circulation, especially in the Northern Hemisphere, it is useful to first obtain an understanding of the dynamics of the zonally symmetric component before investigating the three-dimensional time-mean circulation. This section examines the dynamics of zonally symmetric motions using quasi-geostrophic theory and the log-pressure coordinate system introduced in Chapter 7, and shows that the meridional circulation associated with an axially symmetric vortex is dynamically analogous to the secondary divergent circulation in a baroclinic wave.

In the log-pressure coordinate system, the x and y components of the momentum equation, the hydrostatic approximation, the continuity equation, and the *thermodynamic* energy equation can be written as

$$Du/Dt - fv + \partial\Phi/\partial x = X \tag{10.1}$$

$$Dv/Dt + fu + \partial\Phi/\partial y = Y \tag{10.2}$$

$$\partial\Phi/\partial z = H^{-1}RT \tag{10.3}$$

$$\partial u/\partial x + \partial v/\partial y + \rho_0^{-1}\partial\left(\rho_0 w\right)/\partial z = 0 \tag{10.4}$$

$$DT/Dt + \left(\kappa T/H\right)w = J/c_p \tag{10.5}$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}$$

and *X* and *Y* designate the zonal and meridional components of drag due to small-scale eddies.

For convenience in this and the following chapters, we have dropped the asterisk notation used in Chapter 7 to distinguish the log-pressure coordinate from geometric height. Thus, *z* here designates the log-pressure variable defined in Section 7.4.1.

Analysis of the zonally averaged circulation involves study of the interaction of longitudinally varying disturbances (referred to as eddies and denoted by primed variables) with the longitudinally averaged flow (referred to as the mean flow and denoted by overbars). Thus, any variable *A* is expanded in the form $A = \overline{A} + A'$. This sort of average is a *Eulerian mean*, as it is evaluated at fixed latitude, height, and time. Eulerian mean equations are obtained by taking zonal averages of (10.1) through (10.5). Such averaging is facilitated by using (10.4) to expand the material derivative for any variable *A* in flux form as follows:

(10.6)

$$\begin{aligned}\rho_0\frac{DA}{Dt} &= \rho_0\left(\frac{\partial}{\partial t} + \mathbf{V}\cdot\boldsymbol{\nabla} + w\frac{\partial}{\partial z}\right)A + A\left[\boldsymbol{\nabla}\cdot(\rho_0\mathbf{V}) + \frac{\partial}{\partial z}(\rho_0 w)\right]\\ &= \frac{\partial}{\partial t}(\rho_0 A) + \frac{\partial}{\partial x}(\rho_0 Au) + \frac{\partial}{\partial y}(\rho_0 Av) + \frac{\partial}{\partial z}(\rho_0 Aw)\end{aligned}$$

where we recall that $\rho_0 = \rho_0(z)$.

Applying the zonal averaging operator then gives

(10.7)

$$\rho_0\overline{\frac{DA}{Dt}} = \frac{\partial}{\partial t}(\rho_0\overline{A}) + \frac{\partial}{\partial y}\left[\rho_0\left(\overline{A}\overline{v} + \overline{A'v'}\right)\right] + \frac{\partial}{\partial z}\left[\rho_0\left(\overline{A}\overline{w} + \overline{A'w'}\right)\right]$$

Here we have used the fact that $\partial\overline{(\;)}/\partial x = 0$, since quantities with the overbar are independent of *x*. We have also used the fact that for any variables *a* and *b*

$$\overline{ab} = \overline{(\overline{a} + a')(\overline{b} + b')} = \overline{\overline{a}\overline{b}} + \overline{\overline{a}b'} + \overline{a'\overline{b}} + \overline{a'b'} = \overline{a}\overline{b} + \overline{a'b'}$$

which follows from that fact that $\overline{a}$ and $\overline{b}$ are independent of *x* and $\overline{a'} = \overline{b'} = 0$ so that, for example, $\overline{\overline{a}b'} = \overline{a}\overline{b'} = 0$.

The terms involving $(\overline{v}, \overline{w})$ on the right in (10.7) can be rewritten in advective form with the aid of the zonal mean of (10.4):

(10.8)

$$\partial\overline{v}/\partial y + \rho_0^{-1}\partial(\rho_0\overline{w})/\partial z = 0$$

Applying the chain rule of differentiation to the mean terms on the right in (10.7) and substituting from (10.8), we can rewrite (10.7) as

(10.9)

$$\rho_0\overline{\frac{DA}{Dt}} = \rho_0\frac{\overline{D}}{Dt}(\overline{A}) + \frac{\partial}{\partial y}\left[\rho_0\left(\overline{A'v'}\right)\right] + \frac{\partial}{\partial z}\left[\rho_0\left(\overline{A'w'}\right)\right]$$

where

$$\frac{\overline{D}}{Dt} \equiv \frac{\partial}{\partial t} + \overline{v}\frac{\partial}{\partial y} + \overline{w}\frac{\partial}{\partial z} \tag{10.10}$$

is the rate of change following the mean meridional motion ($\overline{v}, \overline{w}$).

10.2.1 The Conventional Eulerian Mean

Applying the averaging scheme of (10.9) to (10.1) and (10.5), we obtain the zonal-mean zonal momentum and thermodynamic energy equations for quasi-geostrophic motions on the midlatitude β plane:

$$\partial\overline{u}/\partial t - f_0\overline{v} = -\partial\left(\overline{u'v'}\right)/\partial y + \overline{X} \tag{10.11}$$

$$\partial\overline{T}/\partial t + N^2HR^{-1}\overline{w} = -\partial\left(\overline{v'T'}\right)/\partial y + \overline{J}/c_p \tag{10.12}$$

where N is the buoyancy frequency defined by

$$N^2 \equiv \frac{R}{H}\left(\frac{\kappa T_0}{H} + \frac{dT_0}{dz}\right)$$

In (10.11) and (10.12), consistent with quasi-geostrophic scaling, we have neglected advection by the ageostrophic mean meridional circulation and vertical eddy flux divergences. It is easily confirmed that for quasi-geostrophic scales, these terms are small compared to the retained terms (see Problem 10.4). We have included the zonally averaged turbulent drag in (10.11) because stresses due to unresolved eddies (such as gravity waves) may be important not only in the boundary layer but also in the upper troposphere and in the middle atmosphere.

Similar scaling shows that the zonal average of the meridional momentum equation (10.2) can be approximated accurately by geostrophic balance:

$$f_0\overline{u} = -\partial\overline{\Phi}/\partial y$$

This can be combined with the hydrostatic approximation (10.3) to give the thermal wind relation

$$f_0\partial\overline{u}/\partial z + RH^{-1}\partial\overline{T}/\partial y = 0 \tag{10.13}$$

This relationship between zonal-mean wind and potential temperature distributions imposes a strong constraint on the ageostrophic mean meridional circulation ($\overline{v}, \overline{w}$). In the absence of a mean meridional circulation, the eddy momentum flux divergence in (10.11) and eddy heat flux divergence in (10.12) would tend separately to change the mean zonal wind and temperature fields, and thus would destroy thermal wind balance. The pressure gradient force that results from any small departure of the mean zonal wind from geostrophic balance will, however, drive a mean meridional circulation, which adjusts the mean zonal wind and temperature fields so that (10.13) remains valid. In many situations this compensation allows the mean zonal wind to remain unchanged even in the presence of large eddy heat and momentum fluxes. The mean meridional circulation thus plays exactly the same role in the zonal-mean circulation that the secondary divergent circulation plays in the synoptic-scale quasi-geostrophic system. In fact, for steady-state mean flow conditions, the $(\overline{v}, \overline{w})$ circulation must just balance the eddy forcing plus diabatic heating so that the balances in (10.11) and (10.12) are as follows:

- Coriolis force $(f_0\overline{v})$ ≈ divergence of eddy momentum fluxes
- Adiabatic cooling ≈ diabatic heating plus convergence of eddy heat fluxes

Analysis of observations shows that outside the tropics these balances appear to be approximately true above the boundary layer. Thus, changes in the zonal-mean flow arise from small imbalances between the forcing terms and the mean meridional circulation.

The Eulerian mean meridional circulation can be determined in terms of the forcing from an equation similar to the omega equation of Section 6.4. Before deriving this equation, it is useful to observe that the mean meridional mass circulation is nondivergent in the meridional plane. Thus, it can be represented in terms of a meridional mass transport streamfunction, which identically satisfies the continuity equation (10.8) by letting

$$\rho_0\overline{v} = -\frac{\partial\overline{\chi}}{\partial z}; \quad \rho_0\overline{w} = \frac{\partial\overline{\chi}}{\partial y} \tag{10.14}$$

The relationship of the sign of the streamfunction $\overline{\chi}$ to the sense of the mean meridional circulation is shown schematically in Figure 10.1.

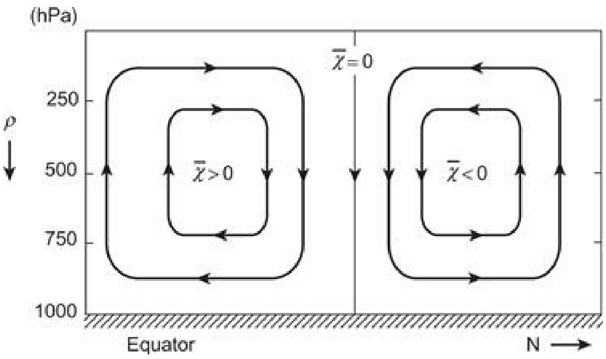


Figure 10.1 Relationship of the Eulerian mean meridional streamfunction to vertical and meridional motion.

The diagnostic equation for $\overline{\chi}$ is derived by first taking

$$f_0\frac{\partial}{\partial z}(10.11) + \frac{R}{H}\frac{\partial}{\partial y}(10.12)$$

and then using (10.13) to eliminate the time derivatives and (10.14) to express the mean meridional circulation in terms of $\overline{\chi}$. The resulting elliptic equation has the form

$$\begin{aligned}\frac{\partial^2\overline{\chi}}{\partial y^2} + \frac{f_0^2}{N^2}\rho_0\frac{\partial}{\partial z}\left(\frac{1}{\rho_0}\frac{\partial\overline{\chi}}{\partial z}\right) = \frac{\rho_0}{N^2}\Bigg[&\frac{\partial}{\partial y}\left(\frac{\kappa J}{H} - \frac{R}{H}\frac{\partial}{\partial y}\left(\overline{v'T'}\right)\right)\\ &- f_0\left(\frac{\partial^2}{\partial z\partial y}\left(\overline{u'v'}\right) - \frac{\partial\overline{X}}{\partial z}\right)\Bigg]\end{aligned} \tag{10.15}$$

Equation (10.15) can be used to qualitatively diagnose the mean meridional circulation. Because

$\overline{\chi}$ must vanish on the boundaries, it can be represented by a double Fourier series in *y* and *z*. Hence, the elliptic operator on the left side of (10.15) is approximately proportional to $-\overline{\chi}$, and (10.15) states qualitatively that

$$\overline{\chi} \propto -\frac{\partial}{\partial y}(\text{diabatic heating}) + \frac{\partial^2}{\partial y^2}(\text{large-scale eddy heat flux}) \\ + \frac{\partial^2}{\partial y \partial z}(\text{large-scale eddy momentum flux}) + \frac{\partial}{\partial z}(\text{zonal drag force})$$

Now, diabatic heating in the Northern Hemisphere decreases for increasing *y*. Thus, the first term on the right is positive and tends to force a mean meridional cell with $\overline{\chi} > 0$. This is referred to as a *thermally direct* cell, as warm air is rising and cool air is sinking. It is this process that primarily accounts for the Hadley circulation of the tropics as illustrated in Figure 10.2. For an idealized Hadley cell in the absence of eddy sources, the differential diabatic heating would be balanced only by adiabatic cooling near the equator and adiabatic warming at higher latitudes.

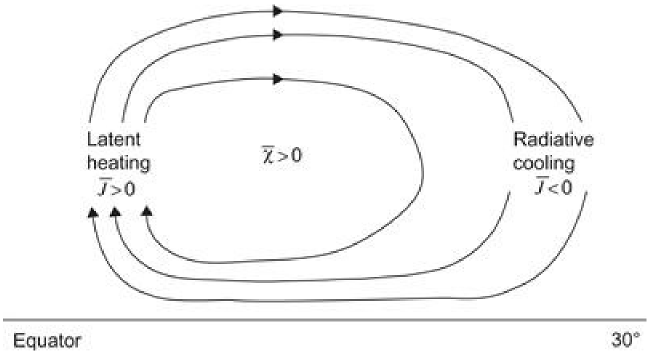


Figure 10.2 Schematic Eulerian mean meridional circulation showing the streamfunction for a thermally direct Hadley cell.

In the extratropical Northern Hemisphere, poleward eddy heat fluxes due to both transient synoptic-scale eddies and stationary planetary waves tend to transfer heat poleward, producing a maximum poleward heat flux $\overline{v'T'}$ in the lower troposphere at about 50° latitude as shown in Figure 10.3. Because $\overline{\chi}$ is proportional to the second derivative of $\overline{v'T'}$, which should be negative where $\overline{v'T'} > 0$, the term should tend to produce a mean meridional cell with $\overline{\chi} < 0$ centered in the lower troposphere at midlatitudes. Thus, the eddy heat flux tends to drive an indirect meridional circulation.

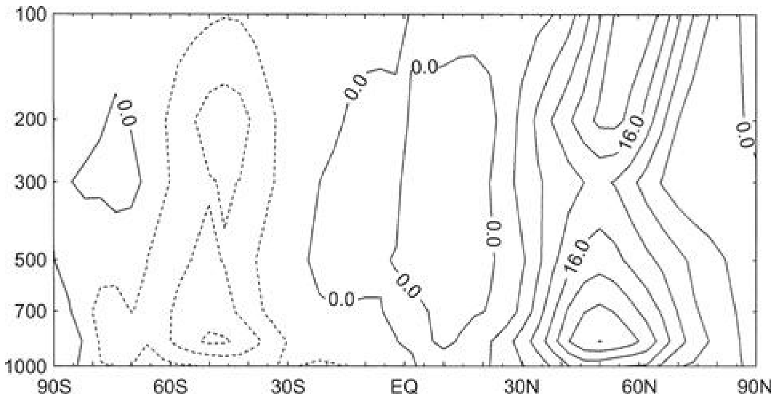


Figure 10.3 Observed northward eddy heat flux distribution (°Cm s^{-1}) for Northern Hemisphere winter.

Adapted from Schubert et al., 1990.

The existence of this indirect meridional circulation can be understood in terms of the need to maintain geostrophic and hydrostatic balance. North of the latitude where $\overline{v'T'}$ is a maximum, there is a convergence of eddy heat flux, while equatorward of that latitude, there is a divergence. Thus, eddy heat transport tends to reduce the pole-to-equator mean temperature gradient. If the mean zonal flow is to remain geostrophic, the thermal wind must then also decrease. In the absence of eddy momentum transport, this decrease in the thermal wind can only be produced by the Coriolis torque due to a mean meridional circulation with the sense of that in Figure 10.4. At the same time, it is not surprising to find that the vertical mean motions required by continuity oppose the temperature tendency associated with the eddy heat flux by producing adiabatic warming in the region of eddy heat flux divergence and adiabatic cooling in the region of eddy heat flux convergence.

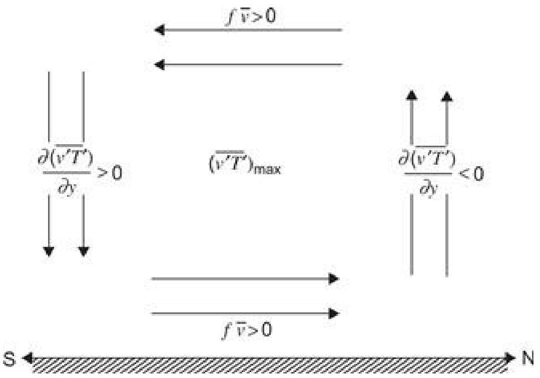

$f\overline{v}>0$
$\frac{\partial(\overline{v'T'})}{\partial y}>0$
$(\overline{v'T'})_{max}$
$\frac{\partial(\overline{v'T'})}{\partial y}<0$
$f\overline{v}>0$
S
N

Figure 10.4 Schematic Eulerian mean meridional circulation forced by poleward heat fluxes.

The next to last forcing term in (10.15) is proportional to the vertical gradient of the horizontal eddy momentum flux convergence. However, it can be shown (see Problem 10.5) that

$$-\frac{\partial^2}{\partial z\partial y}\left(\overline{u'v'}\right) = +\frac{\partial}{\partial z}\left(\overline{v'\zeta'}\right)$$

Thus, this term is proportional to the vertical derivative of the meridional vorticity flux. To interpret this eddy forcing physically, we suppose as shown in Figure 10.5 that the momentum flux convergence (or the vorticity flux) is positive and increasing with height. This will be true in the Northern Hemisphere troposphere poleward of the core of the jet stream, as $\overline{u'v'}$ tends to be poleward and to reach its maximum near the tropopause at about 30°(at the core of the mean jetstream) as shown in Figure 10.6. For this configuration of momentum flux, $\partial^2\overline{u'v'}/\partial y\partial z<0$ in the midlatitude troposphere, which again drives a mean meridional cell with $\overline{\chi}<0$. From (10.11) it is clear that the Coriolis force of this induced indirect meridional circulation is required to balance the acceleration due to the momentum flux convergence, which would otherwise increase the vertical shear of the mean zonal wind and destroy the thermal wind balance.

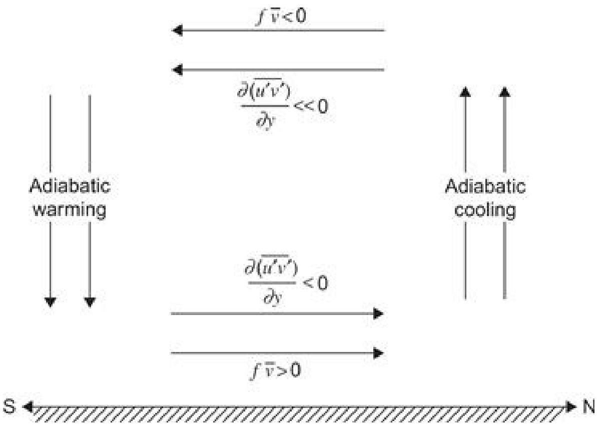

$f\bar{v}<0$
$\frac{\partial(\overline{u'v'})}{\partial y}<<0$
Adiabatic warming
Adiabatic cooling
$\frac{\partial(\overline{u'v'})}{\partial y}<0$
$f\bar{v}>0$
S
N

Figure 10.5 Schematic Eulerian mean meridional circulation forced by a vertical gradient in eddy momentum flux convergence.

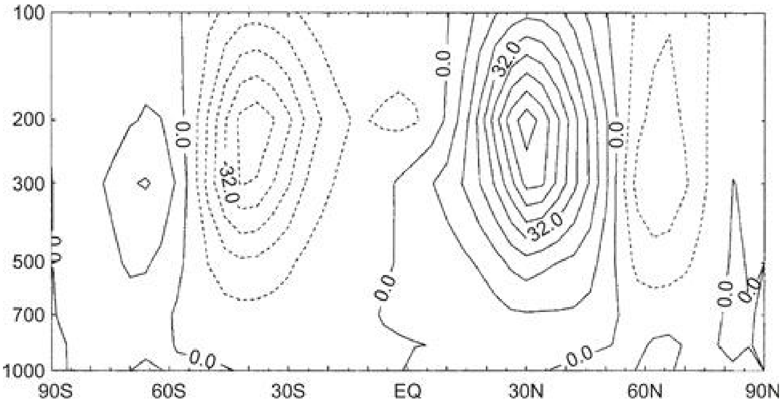


Figure 10.6 Observed northward eddy momentum flux distribution ($m^2\ s^{-2}$) for Northern Hemisphere winter.

Adapted from Schubert et al., 1990.

Thus, the combined eddy heat flux and the eddy momentum flux distributions tend to drive mean meridional cells in each hemisphere with rising motion poleward of 45° and sinking motion equatorward of 45°. This eddy forcing more than compensates the direct diabatic drive at midlatitudes and is responsible for the observed thermally indirect *Ferrel cell.*

The resultant observed climatology of the Eulerian mean meridional circulation is shown in Figure 10.7. It consists primarily of tropical Hadley cells driven by diabatic heating and eddy-driven midlatitude Ferrel cells. There are also minor thermally direct cells at polar latitudes. The meridional circulation in the winter is much stronger than that in the summer, especially in the Northern Hemisphere. This reflects the seasonal variation both in the diabatic and in the eddy flux forcing terms in (10.15).

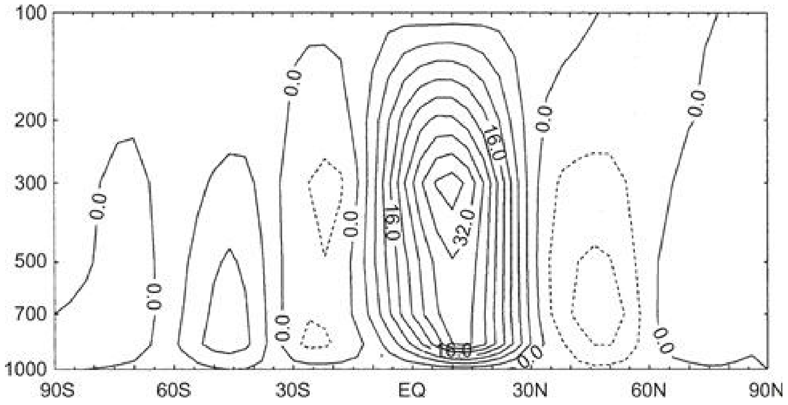


Figure 10.7 Streamfunction (units: 10^2 kg m^{-1} s^{-1}) for the observed Eulerian mean meridional circulation for Northern Hemisphere winter.

Based on the data of Schubert et al., 1990.

The Ferrel cell presents a conundrum, since the vertical circulations produced by individual baroclinic eddies are thermally direct, with warm air rising and cold air sinking. Yet, the Eulerian mean meridional circulation, which averages over these circulations, shows a thermally indirect circulation. This conundrum is resolved by examining the meridional circulation in isentropic coordinates (Figure 10.8), which reveals a thermally direct circulation in both the Hadley cells and in the extratropics. This shows that the Ferrel cell is an artifact of the Eulerian averaging process and that eddy fluxes of entropy are large compared to those of the mean circulation.

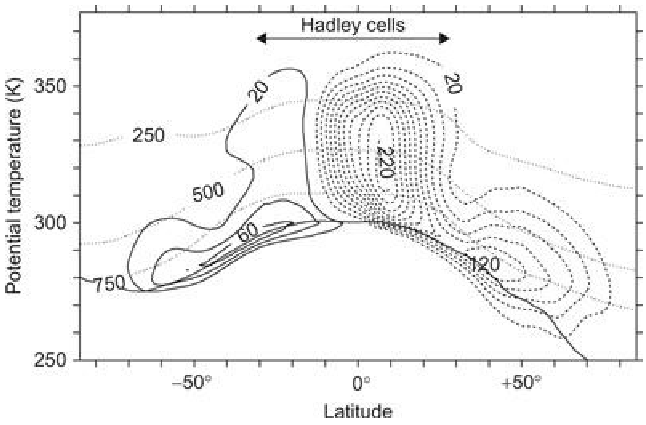


Figure 10.8 January time and zonal-mean isentropic mass flux streamfunction determined from ERA-40 reanalysis data 1980–2001). *Streamfunction contours* are shown every 20 × 10^9 kg s^{-1}, with implied clockwise circulation around negative values. *Dotted lines* show pressure surfaces and the *solid lower curve* is the median surface potential temperature.

Adapted from Schneider, 2006. Used with permission of Annual Reviews.

The zonal momentum balance in the upper troposphere in tropical and midlatitude cells is maintained by the balance between the Coriolis force caused by the mean meridional drift and the eddy momentum flux convergence. The heat balance is maintained by rising motion (adiabatic cooling) balancing the diabatic heating in the tropics and by the eddy heat flux convergence at high latitudes and subsidence (adiabatic warming) balancing the eddy heat flux divergence in the subtropics.

Because of the appearance of eddy flux terms in both mean momentum and thermodynamic energy equations, and the near cancellation of eddy and mean flow processes, it is rather inefficient to attempt to diagnose the net eddy forcing of the mean flow from the conventional Eulerian mean. It can be shown that similar eddy and mean flow compensation occurs in the Eulerian mean continuity equation for a long-lived tracer so that tracer transport calculations are also inefficient in this formulation.

10.2.2 The Transformed Eulerian Mean

An alternative approach to analysis of the zonal-mean circulation, which provides a clearer diagnosis of eddy forcing and also provides a more direct view of transport processes in the meridional plane, is the *transformed Eulerian mean* (TEM) formulation introduced by Andrews and McIntyre (1976). This transformation takes account of the fact that in (10.12) there tends to be a strong cancellation between the eddy heat flux convergence and adiabatic cooling, while the diabatic heating term is a small residual. Since in the mean an air parcel will rise to a higher equilibrium altitude only if its potential temperature is increased by diabatic heating, it is the *residual meridional circulation* associated with diabatic processes that is directly related to the mean meridional mass flow.

The TEM equations can be obtained from (10.11) and (10.12) by defining the residual circulation $(\overline{v}^*, \overline{w}^*)$ as follows:

(10.16a)

$$\overline{v}^* = \overline{v} - \rho_0^{-1} R H^{-1} \partial(\rho_0 \overline{v'T'} / N^2) / \partial z$$

(10.16b)

$$\overline{w}^* = \overline{w} + R H^{-1} \partial(\overline{v'T'} / N^2) / \partial y$$

The residual vertical velocity defined in this manner clearly represents that part of the mean vertical velocity with a contribution to adiabatic temperature change that is not canceled by the eddy heat flux divergence.

Substituting from (10.16a) into (10.11) and (10.12) to eliminate $(\overline{v}, \overline{w})$ yields the TEM equations

(10.17)

$$\partial \overline{u}/\partial t - f_0 \overline{v}^* = +\rho_0^{-1} \nabla \cdot \mathbf{F} + \overline{X} \equiv \overline{G}$$

(10.18)

$$\partial \overline{T}/\partial t + N^2 H R^{-1} \overline{w}^* = \overline{J}/c_p$$

(10.19)

$$\partial \overline{v}^*/\partial y + \rho_0^{-1} \partial \left(\rho_0 \overline{w}^*\right)/\partial z = 0$$

where $\mathbf{F} \equiv \mathbf{j} F_y + \mathbf{k} F_z$, the *Eliassen–Palm flux* (EP flux), is a vector in the meridional (y, z) plane, which for large-scale quasi-geostrophic eddies has the components

(10.20)

$$F_y = -\rho_0 \overline{u'v'}, \quad F_z = \rho_0 f_0 R \overline{v'T'} / (N^2 H)$$

and $\overline{G}$ in (10.17) designates the total zonal force due to both large- and small-scale eddies.

The TEM formulation clearly shows that the eddy heat and momentum fluxes do not act separately to drive changes in the zonal-mean circulation, but only in the combination given by the divergence of the EP flux. The fundamental role of the eddies is thus to exert a zonal force. This eddy forcing of the zonal-mean flow can be displayed conveniently by mapping the field of **F** and contouring the isolines of its divergence. When properly scaled by the basic state density, these contours give the zonal force per unit mass exerted by quasi-geostrophic eddies. The mean global EP flux divergence pattern for a Northern Hemisphere winter is shown in Figure 10.9. Note that in most of the extratropical troposphere, the EP flux is convergent so that the eddies exert a westward zonal force on the atmosphere. On the seasonal time scale the zonal force due to the EP flux divergence in (10.17) is nearly balanced by the eastward Coriolis force of the residual mean meridional circulation. Conditions for this balance are discussed in the next subsection.

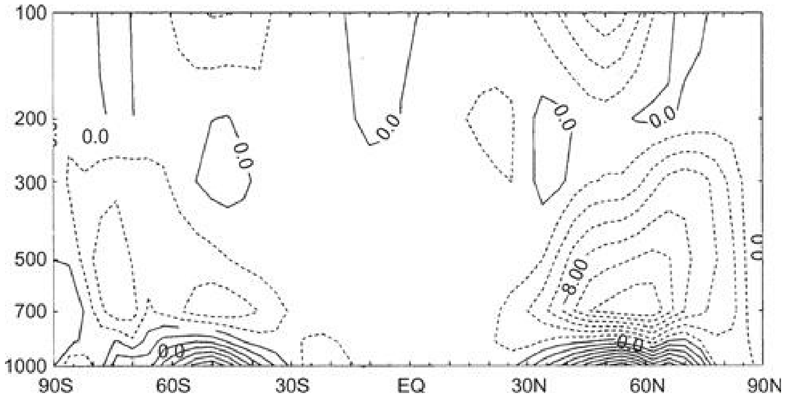


Figure 10.9 Eliassen–Palm flux divergence divided by the standard density ρ_0 for Northern Hemisphere winter. (Units: m s $^{-1}$ day $^{-1}$.)

Based on the data of Schubert et al., 1990.

The structure of the residual mean meridional circulation can be determined by defining a residual streamfunction:

$$\overline{\chi}^* \equiv \overline{\chi} + \rho_0 \frac{R}{H}\frac{\overline{v'T'}}{N^2}$$

It can then be shown by direct substitution into (10.14) and (10.15) that

$$\rho_0 \overline{v}^* = -\frac{\partial \overline{\chi}^*}{\partial z}; \quad \rho_0 \overline{w}^* = \frac{\partial \overline{\chi}^*}{\partial y}$$

and

$$\frac{\partial^2 \overline{\chi}^*}{\partial y^2} + \rho_0 \frac{f_0^2}{N^2}\frac{\partial}{\partial z}\left(\frac{1}{\rho_0}\frac{\partial \overline{\chi}^*}{\partial z}\right) = \frac{\rho_0}{N^2}\left[\frac{\partial}{\partial y}\left(\frac{\kappa \overline{J}}{H}\right) + f_0 \frac{\partial \overline{G}}{\partial z}\right] \tag{10.21}$$

The magnitudes of the diabatic and EP flux contributions to the source term on the right in (10.21) are generally larger in the winter hemisphere than in the summer hemisphere. In the Northern Hemisphere troposphere, the source terms are generally negative, whereas in the Southern Hemisphere, they are generally positive. This implies that $\overline{\chi}^*$ itself is positive in the Northern Hemisphere and negative in the Southern Hemisphere so that the residual meridional circulation consists of a single thermally direct overturning in each hemisphere, with the strongest cell in the winter hemisphere as shown in Figure 10.10.

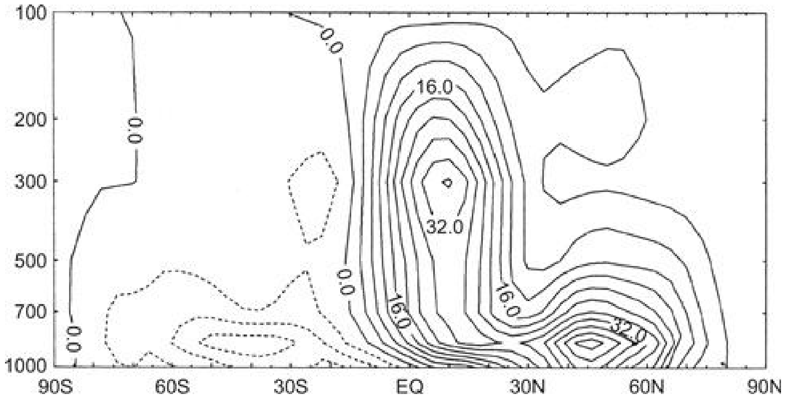


Figure 10.10 Residual mean meridional streamfunction (units: 10^2 kg m^{-1} s^{-1}) for Northern Hemisphere winter.

Based on the data of Schubert et al., 1990.

Unlike the conventional Eulerian mean, the residual mean vertical motion for time-averaged conditions is proportional to the rate of diabatic heating. It approximately represents the *diabatic circulation* in the meridional plane—that is, the circulation in which parcels that rise are diabatically heated and those that sink are diabatically cooled in order that their potential temperatures adjust to the local environment. The time-averaged residual mean meridional circulation thus approximates the mean motion of air parcels and thus, unlike the conventional Eulerian mean, provides an approximation to the mean advective transport of trace substances.

10.2.3 The Zonal-Mean Potential Vorticity Equation

Further insight into the nature of the extratropical zonal-mean circulation can be obtained by zonally averaging the quasi-geostrophic potential vorticity equation to obtain

$$\partial \overline{q} / \partial t = -\partial \left(\overline{q' v'} \right) / \partial y \tag{10.22}$$

where the zonal-mean potential vorticity is

$$\overline{q} = f_0 + \beta y + \frac{1}{f_0} \frac{\partial^2 \overline{\Phi}}{\partial y^2} + \frac{f_0}{\rho_0} \frac{\partial}{\partial z} \left(\frac{\rho_0}{N^2} \frac{\partial \overline{\Phi}}{\partial z} \right) \tag{10.23}$$

and the eddy potential vorticity is

$$q' = \frac{1}{f_0} \left(\frac{\partial^2 \Phi'}{\partial x^2} + \frac{\partial^2 \Phi'}{\partial y^2} \right) + \frac{f_0}{\rho_0} \frac{\partial}{\partial z} \left(\frac{\rho_0}{N^2} \frac{\partial \Phi'}{\partial z} \right) \tag{10.24}$$

The quantity $\overline{q'v'}$ on the right in (10.22) is the divergence of the meridional flux of potential vorticity. According to (10.22), for adiabatic quasi-geostrophic flow, the mean distribution of potential vorticity can be changed only if there is a nonzero flux of eddy potential vorticity. The zonal-mean potential vorticity, together with suitable boundary conditions on $\overline{\Phi}$, completely determines the distribution of zonal-mean geopotential and thus the zonal-mean geostrophic wind and temperature distributions. Thus, eddy-driven mean flow accelerations require nonzero potential vorticity fluxes.

It can be shown that the potential vorticity flux is related to the eddy momentum and heat fluxes. We first note that for quasi-geostrophic motions, the eddy horizontal velocities in the flux terms are geostrophic:

$$f_0 v' = \partial\Phi'/\partial x \quad \text{and} \quad f_0 u' = -\partial\Phi'/\partial y$$

Thus,

$$\overline{v'\frac{\partial^2\Phi'}{\partial x^2}} = \frac{1}{f_0}\overline{\frac{\partial\Phi'}{\partial x}\frac{\partial^2\Phi'}{\partial x^2}} = \frac{1}{2f_0}\overline{\frac{\partial}{\partial x}\left(\frac{\partial\Phi'}{\partial x}\right)^2} = 0$$

where we have used the fact that a perfect differential in *x* vanishes when averaged zonally. Thus,

$$\overline{q'v'} = \overline{\frac{v'}{f_0}\frac{\partial^2\Phi'}{\partial y^2}} + \frac{f_0}{\rho_0}\overline{v'\frac{\partial}{\partial z}\left(\frac{\rho_0}{N^2}\frac{\partial\Phi'}{\partial z}\right)}$$

We can use the chain rule of differentiation to rewrite the terms on the right in this expression as

$$\begin{aligned}\overline{\frac{v'}{f_0}\frac{\partial^2\Phi'}{\partial y^2}} &= \frac{1}{f_0^2}\left(\overline{\frac{\partial\Phi'}{\partial x}\frac{\partial^2\Phi'}{\partial y^2}}\right)\\ &= \frac{1}{f_0^2}\left[\frac{\partial}{\partial y}\left(\overline{\frac{\partial\Phi'}{\partial x}\frac{\partial\Phi'}{\partial y}}\right) - \frac{1}{2}\frac{\partial}{\partial x}\overline{\left(\frac{\partial\Phi'}{\partial y}\right)^2}\right] = -\frac{\partial}{\partial y}\left(\overline{u'v'}\right)\end{aligned}$$

and

$$\begin{aligned}\frac{f_0}{\rho_0}\overline{v'\frac{\partial}{\partial z}\left(\frac{\rho_0}{N^2}\frac{\partial\Phi'}{\partial z}\right)} &= \frac{1}{\rho_0}\left[\frac{\partial}{\partial z}\left(\frac{\rho_0}{N^2}\overline{\frac{\partial\Phi'}{\partial x}\frac{\partial\Phi'}{\partial z}}\right) - \frac{\rho_0}{2N^2}\frac{\partial}{\partial x}\overline{\left(\frac{\partial\Phi'}{\partial z}\right)^2}\right]\\ &= \frac{f_0}{\rho_0}\frac{\partial}{\partial z}\left(\frac{\rho_0}{N^2}\overline{v'\frac{\partial\Phi'}{\partial z}}\right)\end{aligned}$$

Thus,

$$\overline{q'v'} = -\frac{\partial\overline{u'v'}}{\partial y} + \frac{f_0}{\rho_0}\frac{\partial}{\partial z}\left(\frac{\rho_0}{N^2}\overline{v'\frac{\partial\Phi'}{\partial z}}\right) \tag{10.25}$$

so that it is not the momentum flux $\overline{u'v'}$ or heat flux $\overline{v'\partial\Phi'/\partial z}$ that drives net changes in the mean flow distribution, but rather the combination given by the potential vorticity flux. Under some circumstances the eddy momentum flux and eddy heat flux may individually be large, but the combination in (10.25) actually vanishes. This cancellation effect makes the traditional Eulerian mean formulation a poor framework for analysis of mean flow forcing by eddies.

Comparing (10.25) and (10.20), we see that the potential vorticity flux is proportional to the divergence of the EP flux vector:

$$\overline{q'v'} = \rho_0^{-1} \nabla \cdot \mathbf{F} \tag{10.26}$$

Thus, the contribution of large-scale motions to the zonal force in (10.17) equals the meridional flux of quasi-geostrophic potential vorticity. If the motion is adiabatic and the potential vorticity flux is nonzero, equation (10.22) shows that the mean flow distribution must change with time. For this reason, there cannot be complete compensation between Coriolis torque and zonal force terms in (10.17).

10.3 The Angular Momentum Budget

The previous section used the quasi-geostrophic version of the zonal-mean equations to show that large-scale eddies play an essential part in the maintenance of the zonal-mean circulation in the extratropics. In particular, we contrasted the mean flow forcing as represented by the conventional Eulerian mean and TEM formulations. This section expands our consideration of the momentum budget by considering the overall balance of *angular* momentum for the atmosphere and Earth combined. Thus, rather than simply considering the balance of momentum for a given latitude and height in the atmosphere, we must consider the transfer of angular momentum between Earth and the atmosphere, and the flow of angular momentum in the atmosphere.

It would be possible to utilize the complete spherical coordinate version of the TEM equations for this analysis, but we are concerned primarily with the angular momentum balance for a zonal ring of air extending from the surface to the top of the atmosphere. In that case it proves simpler to use the conventional Eulerian mean formulation. It also proves convenient to use a special vertical coordinate, called the *sigma coordinate*, in which Earth's surface is a coordinate surface.

Because the average rotation rate of Earth is itself observed to be very close to constant, the atmosphere must also on the average conserve its angular momentum. The atmosphere gains angular momentum from Earth in the tropics where the surface winds are easterly (i.e., where the angular momentum of the atmosphere is less than that of Earth) and gives up angular momentum to Earth in middle latitudes where the surface winds are westerly. Thus, there must be a net poleward transport of angular momentum within the atmosphere; otherwise, the torque due to surface friction would decelerate both easterlies and westerlies. Furthermore, the angular momentum given by Earth to the atmosphere in the belt of easterlies must just balance the angular momentum given up to Earth in the westerlies if the global angular momentum of the atmosphere is to remain constant.

In the equatorial regions the poleward angular momentum transport is divided between the advection of absolute angular momentum by the poleward flow in the axially symmetric Hadley circulation and transport by eddies. In midlatitudes, however, it is primarily the eddy motions that transport momentum poleward, and the angular momentum budget of the atmosphere must be qualitatively as shown in Figure 10.11.

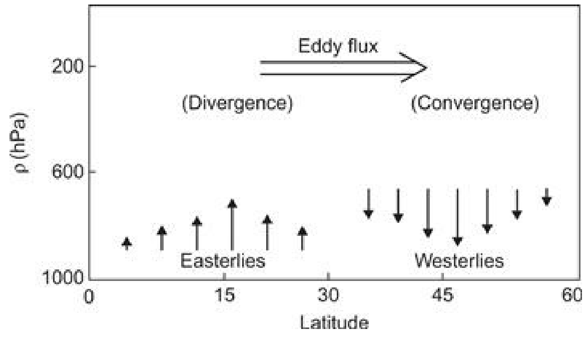


Figure 10.11 Schematic mean angular momentum budget for the atmosphere–Earth system.

As this figure suggests, there is a maximum poleward flux of angular momentum at about 30° latitude and a maximum horizontal flux convergence at about 45°. This maximum in the flux convergence is a reflection of the strong energy conversion in upper-level westerlies and is the mechanism whereby the atmosphere can maintain a positive zonal wind in the middle latitudes despite the momentum lost to the surface.

It is convenient to analyze the momentum budget in terms of absolute angular momentum. The absolute angular momentum per unit mass of atmosphere is

$$M = (\Omega a \cos\phi + u)\, a \cos\phi$$

where as before *a* is the radius of Earth. The crucial role of zonal eddy drag in maintaining the observed latitudinal profile of the mean zonal wind can be seen by evaluating the mean zonal velocity that would arise if a zonal ring of air at rest on the equator were to be displaced poleward, conserving *M*. In that case $u(\phi) = \Omega a \sin^2\phi / \cos\phi$, so that in an angular momentum–conserving Hadley circulation, $u \approx 130\,\mathrm{m\,s^{-1}}$ at 30° latitude, which is far greater than is observed.

Clearly, the absolute angular momentum must decrease as air parcels are advected poleward in the Hadley circulation. The absolute angular momentum of an individual air parcel can be changed only by torques caused by the zonal pressure gradient and eddy stresses. In isobaric coordinates, Newton’s second law in its angular momentum form is thus

$$\frac{DM}{Dt} = -a\cos\phi \left[\frac{\partial \Phi}{\partial x} + g\frac{\partial \tau_E^x}{\partial p}\right] \tag{10.27}$$

where τ_E^x is the zonal component of the vertical eddy stress, and it is assumed that horizontal eddy stresses are negligible compared to the vertical eddy stress.

10.3.1 Sigma Coordinates

In neither the isobaric nor the log-pressure coordinate system does the lower boundary exactly coincide with a coordinate surface. In analytical studies it is usual to assume that the lower boundary can be approximated as a constant pressure surface and to apply the approximate condition

$$\omega(p_s) \approx -\rho_0 g w(z_0)$$

as the lower boundary condition. Here we have assumed that the height of the ground z_0 is coincident with the pressure surface p_s–that is, where p_s is usually set equal to 1000 hPa. These assumptions are of course not strictly valid even when the ground is level. Pressure does change at the ground, but, more important, the height of the ground generally varies so that even if the pressure tendency were zero everywhere, the lower boundary condition should not be applied at a constant p_s. Rather, we should set $p_s = p_s(x, y)$. It is very inconvenient for mathematical analysis, however, to have a boundary condition that must be applied at a surface that is a function of the horizontal variables.

This problem can be overcome by defining a vertical coordinate that is proportional to pressure normalized by the surface pressure. The most common form of this coordinate is the *sigma coordinate*, defined as $\sigma \equiv p/p_s$, where $p_s(x, y, t)$ is the pressure at the surface. Thus, σ is a nondimensional-independent vertical coordinate that decreases upward from a value of $\sigma = 1$ at the ground to $\sigma = 0$ at the top of the atmosphere. In sigma coordinates the lower boundary condition will always apply exactly at $\sigma = 1$. Furthermore, the vertical σ velocity defined by

$$\dot{\sigma} \equiv D\sigma/Dt$$

will always be zero at the ground even in the presence of sloping terrain. Thus, the lower boundary condition in the σ system is merely

$$\dot{\sigma} = 0 \text{ at } \sigma = 1$$

To transform the dynamical equations from the isobaric system to the σ system, we first transform the pressure gradient force in a manner analogous to that shown in Section 1.4.3. Applying (1.36) with p replaced by Φ, s replaced by σ, and z replaced by p, we find that

$$(\partial\Phi/\partial x)_\sigma = (\partial\Phi/\partial x)_p + \sigma(\partial \ln p_s/\partial x)(\partial\Phi/\partial\sigma) \tag{10.28}$$

Because any other variable will transform in an analogous way, we can write the general transformation as

$$\boldsymbol{\nabla}_p(\ ) = \boldsymbol{\nabla}_\sigma(\ ) - \sigma\boldsymbol{\nabla}\ln p_s \partial(\ )/\partial\sigma \tag{10.29}$$

Applying the transformation (10.29) to the momentum equation (3.2), we get

$$\frac{D\mathbf{V}}{Dt} + f\mathbf{k}\times\mathbf{V} = -\boldsymbol{\nabla}\Phi + \sigma\boldsymbol{\nabla}\ln p_s \frac{\partial\Phi}{\partial\sigma} \tag{10.30}$$

where $\boldsymbol{\nabla}$ is now applied holding σ constant, and the total differential is

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{V}\cdot\boldsymbol{\nabla} + \dot{\sigma}\frac{\partial}{\partial\sigma} \tag{10.31}$$

The equation of continuity can be transformed to the σ system by first using (10.29) to express the divergence of the horizontal wind as

$$\boldsymbol{\nabla}_p\cdot\mathbf{V} = \boldsymbol{\nabla}_\sigma\cdot\mathbf{V} - \sigma(\boldsymbol{\nabla}\ln p_s)\cdot\partial\mathbf{V}/\partial\sigma \tag{10.32}$$

To transform the term $\partial\omega/\partial p$, we first note that since p_s does not depend on σ

$$\frac{\partial}{\partial p} = \frac{\partial}{\partial(\sigma p_s)} = \frac{1}{p_s}\frac{\partial}{\partial\sigma}$$

Thus, the continuity equation (3.5) can be written as

$$p_s(\boldsymbol{\nabla}_p \cdot \mathbf{V}) + \partial\omega/\partial\sigma = 0 \tag{10.33}$$

Now the sigma vertical velocity can be written as

$$\dot{\sigma} = \left(\frac{\partial\sigma}{\partial t} + \mathbf{V}\cdot\boldsymbol{\nabla}\sigma\right)_p + \omega\frac{\partial\sigma}{\partial p} = -\frac{\sigma}{p_s}\left(\frac{\partial p_s}{\partial t} + \mathbf{V}\cdot\boldsymbol{\nabla}p_s\right) + \frac{\omega}{p_s}$$

Differentiating the preceding with respect to σ, eliminating $\partial\omega/\partial\sigma$ with (10.33), and rearranging yields the transformed continuity equation

$$\frac{\partial p_s}{\partial t} + \boldsymbol{\nabla}\cdot(p_s\mathbf{V}) + p_s\frac{\partial\dot{\sigma}}{\partial\sigma} = 0 \tag{10.34}$$

With the aid of the equation of state and Poisson's equation–that is, (2.44)—the hydrostatic approximation can be written in the sigma system as

$$\frac{\partial\Phi}{\partial\sigma} = -\frac{RT}{\sigma} = -\frac{R\theta}{\sigma}(p/p_0)^{\kappa} \tag{10.35}$$

where $p_0 = 1000$ hPa.

Expanding the total derivative in (2.46), we may write the thermodynamic energy equation for sigma coordinates as

$$\frac{\partial\theta}{\partial t} + \mathbf{V}\cdot\boldsymbol{\nabla}\theta + \dot{\sigma}\frac{\partial\theta}{\partial\sigma} = \frac{J}{c_p}\frac{\theta}{T} \tag{10.36}$$

10.3.2 The Zonal-Mean Angular Momentum

We now transform the angular momentum equation (10.27) into sigma coordinates with the aid of (10.28) and (10.35) to yield

$$\left(\frac{\partial}{\partial t} + \mathbf{V}\cdot\boldsymbol{\nabla} + \dot{\sigma}\frac{\partial}{\partial\sigma}\right)M = -a\cos\phi\left(\frac{\partial\Phi}{\partial x} + \frac{RT}{p_s}\frac{\partial p_s}{\partial x} + \frac{g}{p_s}\frac{\partial\tau_E^x}{\partial\sigma}\right) \tag{10.37}$$

Multiplying the continuity equation (10.34) by M and adding the result to (10.37) multiplied by p_s, we obtain the flux form of the angular momentum equation[1]:

$$\begin{aligned}\frac{\partial(p_sM)}{\partial t} = &-\boldsymbol{\nabla}\cdot(p_sM\mathbf{V}) - \frac{\partial(p_sM\dot{\sigma})}{\partial\sigma}\\ &-a\cos\phi\left[p_s\frac{\partial\Phi}{\partial x} + RT\frac{\partial p_s}{\partial x}\right] - ga\cos\phi\frac{\partial\tau_E^x}{\partial\sigma}\end{aligned} \tag{10.38}$$

To obtain the zonal-mean angular momentum budget, we must average (10.38) in longitude. Using the spherical coordinate expansion for the horizontal divergence as given in Appendix C, we have

$$\nabla \cdot (p_s M \mathbf{V}) = \frac{1}{a\cos\phi}\left[\frac{\partial(p_s M u)}{\partial\lambda} + \frac{\partial(p_s M v\cos\phi)}{\partial\phi}\right] \tag{10.39}$$

We also observe that the bracketed term on the right in equation (10.38) can be rewritten as

$$\left[p_s\frac{\partial}{\partial x}(\Phi - RT) + \frac{\partial}{\partial x}(p_s RT)\right] \tag{10.40}$$

However, with the aid of the hydrostatic equation (10.35), we can write

$$(\Phi - RT) = \Phi + \sigma\partial\Phi/\partial\sigma = \partial\,(\sigma\Phi)/\partial\sigma$$

Thus, recalling that p_s does not depend on σ, we obtain

$$\left[p_s\frac{\partial\Phi}{\partial x} + RT\frac{\partial p_s}{\partial x}\right] = \left[\frac{\partial}{\partial\sigma}\left(p_s\sigma\frac{\partial\Phi}{\partial x}\right) + \frac{\partial}{\partial x}(p_s RT)\right] \tag{10.41}$$

Substituting from (10.39) and (10.41) into (10.38) and taking the zonal average gives

$$\begin{aligned}\frac{\partial\left(\overline{p_s M}\right)}{\partial t} = &-\frac{1}{\cos\phi}\frac{\partial}{\partial y}\left(\overline{p_s M v}\cos\phi\right)\\ &-\frac{\partial}{\partial\sigma}\left[\overline{p_s M\dot{\sigma}} + ga\cos\phi(\overline{\tau}^x_E) + (a\cos\phi)\overline{\sigma p_s\frac{\partial\Phi}{\partial x}}\right]\end{aligned} \tag{10.42}$$

The terms on the right in (10.42) represent the convergence of the horizontal flux of angular momentum and the convergence of the vertical flux of angular momentum, respectively.

Integrating (10.42) vertically from Earth's surface ($\sigma = 1$) to the top of the atmosphere ($\sigma = 0$), and recalling that $\dot{\sigma} = 0$ for $\sigma = 0, 1$, we have

$$\begin{aligned}\int_0^1 g^{-1}\frac{\partial}{\partial t}\overline{p_s M}\,d\sigma = &-(g\cos\phi)^{-1}\int_0^1\frac{\partial}{\partial y}\left(\overline{p_s M v}\cos\phi\right)d\sigma\\ &-a\cos\phi\left[\left(\overline{\tau}^x_E\right)_{\sigma=1} + \overline{p_s\partial h/\partial x}\right]\end{aligned} \tag{10.43}$$

where $h(x, y) = g^{-1}\Phi(x, y, 1)$ is the height of the lower boundary ($\sigma = 1$), and we have assumed that the eddy stress vanishes at $\sigma = 0$.

Equation (10.43) expresses the angular momentum budget for a zonal ring of air of unit meridional width, extending from the ground to the top of the atmosphere. In the long-term mean the three terms on the right, representing the convergence of the meridional flux of angular momentum, the torque due to small-scale turbulent fluxes at the surface, and the surface pressure torque, must balance. In the sigma coordinate system the surface pressure torque takes the particularly simple form $-\overline{p_s\partial h/\partial x}$. Thus, the pressure torque acts to transfer angular momentum from the atmosphere to the ground, provided that the surface pressure and the slope of the ground

$(\partial h/\partial x)$ are correlated positively. Observations indicate that this is generally the case in middle latitudes because there is a slight tendency for the surface pressure to be higher on the western sides of mountains than on the eastern sides (refer to Figure 4.11). In midlatitudes of the Northern Hemisphere the surface pressure torque provides nearly one-half of the total atmosphere–surface momentum exchange, but in the tropics and the Southern Hemisphere the exchange is dominated by turbulent eddy stresses.

The role of eddy motions in providing the meridional angular momentum transport necessary to balance the surface angular momentum sinks can be elucidated best if we divide the flow into zonal-mean and eddy components by letting

$$M = \overline{M} + M' = (\Omega a \cos\phi + \overline{u} + u')a\cos\phi$$
$$p_s v = \overline{(p_s v)} + (p_s v)'$$

where primes indicate deviations from the zonal-mean. Thus, the meridional flux becomes

$$\overline{(p_s M v)} = \left[\Omega a \cos\phi \overline{p_s v} + \overline{u}\overline{p_s v} + \overline{u'(p_s v)'}\right] a\cos\phi \tag{10.44}$$

The three meridional flux terms on the right in this expression are called the meridional Ω–momentum flux, the meridional drift, and the meridional eddy momentum flux, respectively.

The drift term is important in the tropics, but in midlatitudes it is small compared to the eddy flux and can be neglected in an approximate treatment. Furthermore, we can show that the meridional Ω–momentum flux does not contribute to the vertically integrated flux. Averaging the continuity equation (10.34) zonally and integrating vertically, we obtain

$$\overline{\frac{\partial p_s}{\partial t}} = -(\cos\phi)^{-1}\frac{\partial}{\partial y}\int_0^1 \overline{p_s v}\cos\phi\, d\sigma \tag{10.45}$$

Thus, for time-averaged flow—where the left side of (10.45) vanishes—there is no net mass flow across latitude circles. The vertically integrated meridional angular momentum flux is therefore given approximately by

$$\int_0^1 \overline{p_s M v}\, d\sigma \approx \int_0^1 \overline{u'(p_s v)'} a\cos\phi\, d\sigma \approx \int_0^1 a\cos\phi\, \overline{p}_s \overline{u'v'}\, d\sigma \tag{10.46}$$

where we have assumed that the fractional change in p_s is small compared to the change in v' so that $(p_s v)' \approx \overline{p}_s v'$. The angular momentum flux is thus proportional to the negative of the meridional component of the EP flux given in (10.20).

In the Northern Hemisphere, as shown in Figure 10.6, the eddy momentum flux in midlatitudes is positive and decreases in magnitude with latitude poleward of 30°. For quasi-geostrophic flow, positive eddy momentum flux requires that the eddies be asymmetric in the horizontal plane, with the trough and ridge axes tilting as indicated in Figure 10.12. When the troughs and ridges on the average have southwest to northeast phase tilt, the zonal flow will be larger than average ($u' > 0$) where the meridional flow is poleward ($v' > 0$) and less than average ($u' < 0$) where the flow is equatorward. Thus, $\overline{u'v'} > 0$, and the eddies will systematically transport positive zonal momentum poleward.

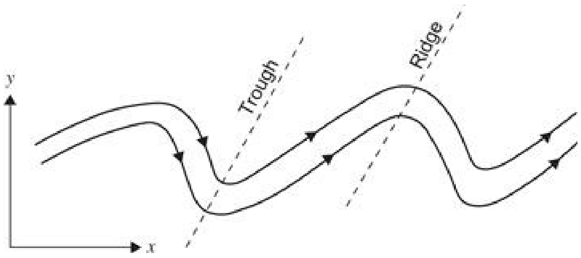


Figure 10.12 Schematic streamlines for a positive eddy momentum flux.

As shown in (10.42), the total vertical momentum flux consists of the flux due to large-scale motions $\overline{p_s M\dot{\sigma}} \approx \Omega a \cos\phi \overline{p_s\dot{\sigma}}$; the flux due to the pressure torque, $(a\cos\phi)\overline{\sigma p_s \partial\Phi / \partial x}$; and the flux due to small-scale turbulent stresses, $ga\cos\phi\overline{\tau_E^x}$. As mentioned previously, the last two are responsible for the transfer of momentum from Earth to the atmosphere in the tropics and from the atmosphere to Earth in midlatitudes. Outside the planetary boundary layer, however, the vertical momentum transport in the troposphere is due primarily to the vertical Ω–momentum flux, $\Omega a \cos\phi \overline{p_s\dot{\sigma}}$.

An estimate of the annual average of the zonal-mean total surface torque is shown in Figure 10.13. The total surface torque must be balanced by poleward fluxes of angular momentum in the atmosphere. Except for the belt within 10° of the equator, almost all of the poleward flux is accounted for by the eddy flux term given on the right side of (10.46). Thus, the momentum budget and the energy cycle both depend critically on the transports by the eddies.

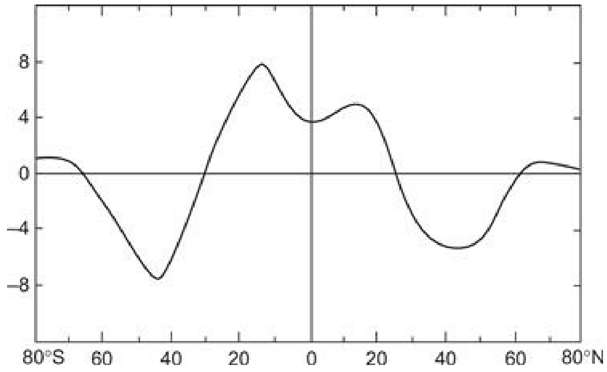


Figure 10.13 Latitudinal profile of the annual mean eastward torque (surface friction plus mountain torque) exerted on the atmosphere in units of 10^{18} m^2 kg s^{-2}.

Adapted from Oort and Peixoto, 1983.

10.4 The Lorenz Energy Cycle

The previous section discussed the interaction between zonally averaged flow and longitudinally varying eddy motions in terms of the angular momentum balance. It is also useful to examine the exchange of energy between the eddies and the mean flow. As in Section 10.2, we limit the analysis to quasi-geostrophic flow on a midlatitude β plane. Eulerian mean equations in log-pressure coordinates can then be written as

$$\partial\overline{u}/\partial t - f_0\overline{v} = -\partial\left(\overline{u'v'}\right)/\partial y + \overline{X} \tag{10.47}$$

$$f_0\overline{u} = -\partial\overline{\Phi}/\partial y \tag{10.48}$$

$$\frac{\partial}{\partial t}\left(\frac{\partial\overline{\Phi}}{\partial z}\right) + \overline{w}N^2 = \frac{\kappa}{H}\overline{J} - \frac{\partial}{\partial y}\left(\overline{v'\frac{\partial\Phi'}{\partial z}}\right) \tag{10.49}$$

$$\partial\overline{v}/\partial y + \rho_0^{-1}\partial(\rho_0\overline{w})/\partial z = 0 \tag{10.50}$$

Here we have used the hydrostatic approximation in (10.3) to express the temperature in (10.49) in terms of the differential geopotential thickness, and we have again neglected vertical eddy fluxes and advection by the mean meridional circulation in (10.47) and (10.49). We have, however, included a turbulent drag force, $\overline{X}$, in (10.47), as dissipation by unresolved turbulent eddies is an essential element in the energy balance.

To analyze the exchange of energy between mean flow and eddies, we require a similar set of dynamical equations for the eddy motion. For simplicity, assume that the eddies satisfy the following linearized set of equations[2]:

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)u' - \left(f_0 - \frac{\partial\overline{u}}{\partial y}\right)v' = -\frac{\partial\Phi'}{\partial x} + X' \tag{10.51}$$

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)v' + f_0u' = -\frac{\partial\Phi'}{\partial y} + Y' \tag{10.52}$$

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\frac{\partial\Phi'}{\partial z} + v'\frac{\partial}{\partial y}\left(\frac{\partial\overline{\Phi}}{\partial z}\right) + N^2w' = \frac{\kappa J'}{H} \tag{10.53}$$

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{1}{\rho_0}\frac{\partial(\rho_0 w')}{\partial z} = 0 \tag{10.54}$$

where *X′* and *Y′* are the zonally varying components of drag as a result of unresolved turbulent motions.

We now define a global average:

$$\langle\,\rangle \equiv A^{-1}\int_0^{\infty}\int_0^{D}\int_0^{L}(\,)dxdydz$$

where *L* is the distance around a latitude circle, *D* is the meridional extent of the midlatitude *β* plane, and *A* designates the total horizontal area of the *β* plane. Then for any quantity Ψ

$$\begin{aligned}&\langle\partial\Psi/\partial x\rangle = 0\\&\langle\partial\Psi/\partial y\rangle = 0, \quad \text{if } \Psi \text{ vanishes at } y = \pm D\\&\langle\partial\Psi/\partial z\rangle = 0, \quad \text{if } \Psi \text{ vanishes at } z = 0 \text{ and } z \to \infty\end{aligned}$$

An equation for the evolution of the mean flow kinetic energy can then be obtained by multiplying (10.47) by $\rho_0\overline{u}$ and (10.48) by $\rho_0\overline{v}$ and adding the results to get

$$\begin{aligned}\rho_0\frac{\partial}{\partial t}\left(\frac{\overline{u}^2}{2}\right) &= -\rho_0\overline{v}\frac{\partial\overline{\Phi}}{\partial y} - \rho_0\overline{u}\frac{\partial}{\partial y}(\overline{u'v'}) + \rho_0\overline{u}\overline{X}\\ &= -\frac{\partial}{\partial y}(\rho_0\overline{v}\overline{\Phi}) + \rho_0\overline{\Phi}\frac{\partial\overline{v}}{\partial y} - \frac{\partial}{\partial y}(\rho_0\overline{u}\overline{u'v'}) + \rho_0\overline{u'v'}\frac{\partial\overline{u}}{\partial y} + \rho_0\overline{u}\overline{X}\end{aligned}$$

Integrating over the entire volume, we get

$$\frac{d}{dt}\left\langle\frac{\rho_0\overline{u}^2}{2}\right\rangle = +\left\langle\rho_0\overline{\Phi}\frac{\partial\overline{v}}{\partial y}\right\rangle + \left\langle\rho_0\overline{u'v'}\frac{\partial\overline{u}}{\partial y}\right\rangle + \langle\rho_0\overline{u}\overline{X}\rangle \tag{10.55}$$

where we have assumed that $\overline{v} = 0$ and $\overline{u'v'} = 0$ for $y = \pm D$. Terms on the right side in (10.55) can be interpreted as the work done by the zonal-mean pressure force, the conversion of eddy kinetic energy to zonal-mean kinetic energy, and dissipation by the zonal-mean eddy stress. Alternatively, the first term on the right can be rewritten with the aid of the continuity equation to yield

$$\left\langle\rho_0\overline{\Phi}\frac{\partial\overline{v}}{\partial y}\right\rangle = -\left\langle\overline{\Phi}\frac{\partial\rho_0\overline{w}}{\partial z}\right\rangle = \left\langle\rho_0\overline{w}\frac{\partial\overline{\Phi}}{\partial z}\right\rangle = \frac{R}{H}\langle\rho_0\overline{w}\overline{T}\rangle$$

where we have assumed that $\rho_0\overline{w} = 0$ at $z = 0$, and $z \to \infty$. Thus, averaged over the whole domain the pressure work term is proportional to the correlation between the zonal-mean vertical mass flux $\rho_0\overline{w}$ and the zonal-mean temperature (or thickness). This term will be positive if on the average warm air is rising and cold air is sinking—that is, if there is a conversion from potential to kinetic energy.

Section 7.3.1 showed that in the quasi-geostrophic system the available potential energy is proportional to the square of the deviation of temperature from a standard atmosphere profile divided by the static stability. In terms of differential thickness the zonal-mean available potential energy is defined as

$$\overline{P} \equiv \frac{1}{2}\left\langle\frac{\rho_0}{N^2}\left(\frac{\partial\overline{\Phi}}{\partial z}\right)^2\right\rangle$$

Multiplying (10.49) through by $\rho_0(\partial\overline{\Phi}/\partial z)/N^2$ and averaging over space gives

$$\begin{aligned}\frac{d}{dt}\left\langle\frac{\rho_0}{2N^2}\left(\frac{\partial\overline{\Phi}}{\partial z}\right)^2\right\rangle = -\left\langle\rho_0\overline{w}\frac{\partial\overline{\Phi}}{\partial z}\right\rangle + \left\langle\frac{\rho_0\kappa\overline{J}}{N^2H}\left(\frac{\partial\overline{\Phi}}{\partial z}\right)\right\rangle\\ -\left\langle\frac{\rho_0}{N^2}\frac{\partial\overline{\Phi}}{\partial z}\frac{\partial}{\partial y}\left(\overline{v'\frac{\partial\Phi'}{\partial z}}\right)\right\rangle\end{aligned} \tag{10.56}$$

The first term on the right is just equal and opposite to the first term on the right in (10.55), which confirms that this term represents a conversion between zonal-mean kinetic and potential energies. The second term involves the correlation between temperature and diabatic heating; it expresses the generation of zonal-mean potential energy by diabatic processes. The final term, which involves the meridional eddy heat flux, expresses the conversion between zonal-mean and eddy potential energy.

That the second term on the right in (10.55) and the final term in (10.56) represent conversion between zonal-mean and eddy energies can be confirmed by performing analogous operations on the eddy equations (10.51), (10.52), and (10.53) to obtain equations for the eddy kinetic and available potential energies:

$$\begin{aligned}\frac{d}{dt}\left\langle\rho_0\frac{\overline{u'^2}+\overline{v'^2}}{2}\right\rangle=&+\left\langle\rho_0\overline{\Phi'\left(\frac{\partial u'}{\partial x}+\frac{\partial v'}{\partial y}\right)}\right\rangle\\&-\left\langle\rho_0\overline{u'v'}\frac{\partial\overline{u}}{\partial y}\right\rangle+\left\langle\rho_0\left(\overline{u'X'}+\overline{v'Y'}\right)\right\rangle\end{aligned}\tag{10.57}$$

$$\begin{aligned}\frac{d}{dt}\left\langle\frac{\rho_0}{2N^2}\overline{\left(\frac{\partial\Phi'}{\partial z}\right)^2}\right\rangle=&-\left\langle\rho_0\overline{w'\frac{\partial\Phi'}{\partial z}}\right\rangle+\left\langle\frac{\rho_0\kappa\overline{J'\partial\Phi'/\partial z}}{N^2H}\right\rangle\\&-\left\langle\frac{\rho_0}{N^2}\left(\frac{\partial^2\overline{\Phi}}{\partial z\partial y}\right)\left(\overline{v'\frac{\partial\Phi'}{\partial z}}\right)\right\rangle\end{aligned}\tag{10.58}$$

If we set $w' = 0$ at $z = 0$, the first term on the right in (10.57) can be rewritten using the continuity equation (10.54) as

$$\left\langle\rho_0\overline{\Phi'\left(\frac{\partial u'}{\partial x}+\frac{\partial v'}{\partial y}\right)}\right\rangle=-\left\langle\overline{\Phi'\frac{\partial(\rho_0w')}{\partial z}}\right\rangle=\left\langle\rho_0\overline{w'\left(\frac{\partial\Phi'}{\partial z}\right)}\right\rangle$$

which is equal to minus the first term on the right in (10.58). Thus, this term expresses the conversion between eddy kinetic and eddy potential energy for the Eulerian mean formulation. Similarly, the last term in (10.58) is equal to minus the last term in (10.56) and thus represents conversion between eddy and zonal-mean available potential energy.

The Lorenz energy cycle can be expressed compactly by defining zonal-mean and eddy kinetic and available potential energies:

$$\begin{aligned}&\overline{K}\equiv\left\langle\rho_0\frac{\overline{u}^2}{2}\right\rangle,\quad K'\equiv\left\langle\rho_0\frac{\overline{u'^2}+\overline{v'^2}}{2}\right\rangle\\&\overline{P}\equiv\frac{1}{2}\left\langle\frac{\rho_0}{N^2}\left(\frac{\partial\overline{\Phi}}{\partial z}\right)^2\right\rangle,\quad P'\equiv\frac{1}{2}\left\langle\frac{\rho_0}{N^2}\overline{\left(\frac{\partial\Phi'}{\partial z}\right)^2}\right\rangle\end{aligned}$$

the energy transformations

$$[\overline{P}\cdot\overline{K}]\equiv\left\langle\rho_0\overline{w}\frac{\partial\overline{\Phi}}{\partial z}\right\rangle,\quad[P'\cdot K']\equiv\left\langle\rho_0\overline{w'\frac{\partial\Phi'}{\partial z}}\right\rangle$$

$$[K'\cdot\overline{K}]\equiv\left\langle\rho_0\overline{u'v'}\frac{\partial\overline{u}}{\partial y}\right\rangle,\quad[P'\cdot\overline{P}]\equiv\left\langle\frac{\rho_0}{N^2}\overline{v'\frac{\partial\Phi'}{\partial z}}\frac{\partial^2\overline{\Phi}}{\partial y\partial z}\right\rangle$$

and the sources and sinks

$$\overline{R}\equiv\left\langle\frac{\rho_0}{N^2}\frac{\kappa\overline{J}}{H}\frac{\partial\overline{\Phi}}{\partial z}\right\rangle,\quad R'\equiv\left\langle\frac{\rho_0}{N^2}\overline{\frac{\kappa J'}{H}\frac{\partial\Phi'}{\partial z}}\right\rangle$$

$$\overline{\varepsilon}\equiv\left\langle\rho_0\overline{u}\overline{X}\right\rangle,\quad\varepsilon'\equiv\left\langle\rho_0\left(\overline{u'X'}+\overline{v'Y'}\right)\right\rangle$$

Equations (10.55) through (10.58) can then be expressed in the simple form

$$d\overline{K}/dt = [\overline{P}\cdot\overline{K}] + [K'\cdot\overline{K}] + \overline{\varepsilon} \tag{10.59}$$

$$d\overline{P}/dt = -[\overline{P}\cdot\overline{K}] + [P'\cdot\overline{P}] + \overline{R} \tag{10.60}$$

$$dK'/dt = [P'\cdot K'] - [K'\cdot\overline{K}] + \varepsilon' \tag{10.61}$$

$$dP'/dt = -[P'\cdot K'] - [P'\cdot\overline{P}] + R' \tag{10.62}$$

Here $[A\cdot B]$ designates conversion from energy form *A* to form *B*.

Adding (10.59) through (10.62), we obtain an equation for the rate of change of total energy (kinetic plus available potential):

$$d(\overline{K} + K' + \overline{P} + P')/dt = \overline{R} + R' + \overline{\varepsilon} + \varepsilon' \tag{10.63}$$

For adiabatic inviscid flow, the right side vanishes and the total energy $\overline{K} + K' + \overline{P} + P'$ is conserved. In this system the zonal-mean kinetic energy does not include a contribution from the mean meridional flow because the zonally averaged meridional momentum equation was replaced by the geostrophic approximation. (Likewise, use of the hydrostatic approximation means that neither the mean nor the eddy vertical motion is included in the total kinetic energy.) Thus, the quantities that are included in the total energy depend on the particular model used. For any model the definitions of energy must be consistent with the approximations employed.

In the long-term mean, the left side of (10.63) must vanish. Thus, the production of available potential energy by zonal-mean and eddy diabatic processes must balance the mean plus eddy kinetic energy dissipation:

$$\overline{R} + R' = -\overline{\varepsilon} - \varepsilon' \tag{10.64}$$

Because solar radiative heating is a maximum in the tropics, where the temperatures are high, it is clear that $\overline{R}$, the generation of zonal-mean potential energy by the zonal-mean heating, will be positive. For a *dry* atmosphere in which eddy diabatic processes are limited to radiation and diffusion *R′*, the diabatic production of eddy available potential energy should be negative because the thermal radiation emitted to space from the atmosphere increases with increasing temperature and thus tends to reduce horizontal temperature contrasts in the atmosphere. For Earth's atmosphere, however, the presence of clouds and precipitation greatly alters the distribution of *R′*. Present estimates (Figure 10.14) suggest that in the Northern Hemisphere *R′* is positive and nearly half as large as $\overline{R}$. Thus, diabatic heating generates both zonal-mean and eddy available potential energy.

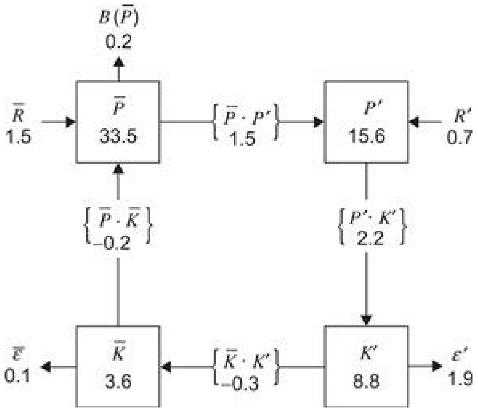


Figure 10.14 Observed mean energy cycle for the Northern Hemisphere. Numbers in *squares* are energy amounts in units of 10^5 J m^{-2}. Numbers *next to arrows* are energy transformation rates in units of W m^{-2}. $B(\overline{P})$ represents a net energy flux into the Southern Hemisphere. Other symbols are defined in the text.

Adapted from Oort and Peixoto, 1974. Copyright © American Geophysical Union. Used with permission.

Equations (10.59) through (10.62) together provide a complete description of the quasi-geostrophic energy cycle from the conventional Eulerian mean point of view. The content of these equations is summarized by means of the four-box diagram in Figure 10.14. In this diagram the squares represent reservoirs of energy, and the arrows indicate sources, sinks, and conversions. The observed direction of the conversion terms in the troposphere for the Northern Hemisphere annual mean is indicated by arrows. It should be emphasized that the direction of the various conversions cannot be theoretically deduced by reference to the energy equations alone. It also should be emphasized that the conversion terms given here are a result of the particular type of zonal average model used. The analogous energy equations for the TEM equations have rather different conversions. Thus, the energy transformations given in the present analysis should not be regarded as fundamental properties of the atmosphere but rather as properties of the Eulerian mean system.

Nevertheless, because the conventional Eulerian mean model is generally used as a basis for the study of baroclinic waves, the four-box energy diagram presented here does provide a useful framework for considering the role of weather disturbances in maintenance of the general circulation. The observed energy cycle as summarized in Figure 10.14 suggests the following qualitative picture:

1. The zonal-mean diabatic heating generates mean zonal available potential energy through a net heating of the tropics and cooling of the polar regions.
2. Baroclinic eddies transport warm air poleward and cold air equatorward, and transform the mean available potential energy to eddy available potential energy.
3. At the same time, eddy available potential energy is transformed into eddy kinetic energy by the vertical motions in the eddies.
4. The zonal kinetic energy is maintained primarily by the conversions from eddy kinetic energy due to the correlation $\overline{u'v'}$. This is discussed further in the next section.
5. The energy is dissipated by surface and internal friction in the eddies and mean flow.

In summary, the observed atmospheric energy cycle as given by the Eulerian mean formulation is consistent with the notion that baroclinically unstable eddies are the primary disturbances responsible for the energy exchange in midlatitudes. It is through the eddy motions that the kinetic energy lost through turbulent stresses is replaced, and it is the eddies that are primarily responsible for the poleward heat transport to balance the radiation deficit in the polar regions. In addition to transient baroclinic eddies, forced stationary orographic waves and free Rossby waves may also contribute substantially to the poleward heat flux. The direct conversion of mean available potential energy to mean kinetic energy by symmetric overturning is, however, small and negative in the middle latitudes but positive in the tropics, where it plays an important role in the maintenance of the mean Hadley circulation.

10.5 Longitudinally Dependent Time-Averaged Flow

So far in this chapter we have concentrated on the zonally averaged component of the general circulation. For a planet with a longitudinally uniform surface, the flow averaged over a season should be completely characterized by the zonally averaged component of the circulation, since for such a hypothetical planet the statistics of zonally asymmetric transient eddies (i.e., weather disturbances) should be independent of longitude. On Earth, however, large-scale topography and continent–ocean heating contrasts provide strong forcing for longitudinally asymmetric planetary-scale time mean motions. Such motions, usually referred to as *stationary waves*, are especially strong in the Northern Hemisphere during the winter season.

Observations indicate that tropospheric stationary waves generally tend to have an equivalent barotropic structure; that is, wave amplitude generally increases with height, but phase lines tend to be vertical. Although nonlinear processes may be significant in the formation and maintenance of stationary waves, the climatological stationary wave pattern can to a first approximation be described in terms of forced barotropic Rossby waves. When superposed on zonal-mean circulation, such waves produce local regions of enhanced and diminished time mean westerly winds, which strongly influence the development and propagation of transient weather disturbances. They thus represent essential features of the climatological flow.

10.5.1 Stationary Rossby Waves

The most significant of the time mean zonally asymmetric circulation features is the pattern of stationary planetary waves excited in the Northern Hemisphere by the flow over the Himalayas and the Rockies. It was shown in Section 5.7.2 that the quasi-stationary wave pattern along 45° latitude could be accounted for to a first approximation as the forced wave response when mean westerlies impinge on the large-scale topography. More detailed analysis suggests that zonally asymmetric heat sources also contribute to the forcing of the climatological stationary wave pattern. Some controversy remains, however, concerning the relative importance of heating and orography in forcing the observed stationary wave pattern. The two processes are difficult to separate because the pattern of diabatic heating is influenced by orography.

The discussion of topographic Rossby waves in Section 5.7.2 used a β-plane channel model in which it was assumed that wave propagation was parallel to latitude circles. In reality, however, large-scale topographic features and heat sources are confined in latitude as well as longitude, and the stationary waves excited by such forcing may propagate energy meridionally as well as zonally. For a quantitatively accurate analysis of the barotropic Rossby wave response to a local source, it is necessary to use the barotropic vorticity equation in spherical coordinates and to include the latitudinal dependence of the mean zonal wind. The mathematical analysis for such a situation is beyond the scope of this book. It is possible, however, to obtain a qualitative notion of the nature of the wave propagation for this case by generalizing the β-plane analysis of Section 5.7. Thus, rather than assuming that propagation is limited to a channel of specified width, we assume that the β plane extends to plus and minus infinity in the meridional direction and that Rossby waves can propagate both northward and southward without reflection from artificial walls.

The free barotropic Rossby wave solution then has the form of (5.109) and satisfies the dispersion relation of (5.110), where l is the meridional wave number, which is now allowed to vary. From (5.112) it is clear that for a specified zonal wave number, k, the free solution is stationary for l given by

$$l^2 = \beta/\overline{u} - k^2 \tag{10.65}$$

Thus, for example, westerly flow over an isolated mountain that primarily excites a response at a given k will produce stationary waves with both positive and negative l satisfying (10.65). As remarked in Section 5.7.1, although Rossby wave phase propagation relative to the mean wind is always westward, this is not true of the group velocity. From (5.110) we readily find that the x and y components of group velocity are

$$c_{gx} = \frac{\partial \nu}{\partial k} = \overline{u} + \beta \frac{\left(k^2 - l^2\right)}{\left(k^2 + l^2\right)^2} \tag{10.66}$$

$$c_{gy} = \frac{\partial \nu}{\partial l} = \frac{2\beta k l}{\left(k^2 + l^2\right)^2} \tag{10.67}$$

For stationary waves, these may be expressed alternatively with the aid of equation (10.65) as

$$c_{gx} = \frac{2\overline{u}k^2}{\left(k^2 + l^2\right)}, \quad c_{gy} = \frac{2\overline{u}kl}{\left(k^2 + l^2\right)} \tag{10.68}$$

The group velocity vector for stationary Rossby waves is perpendicular to the wave crests. It always has an eastward zonal component and a northward or southward meridional component depending on whether l is positive or negative. The magnitude is given by

$$|\mathbf{c_g}| = 2\overline{u}\cos\alpha \tag{10.69}$$

(see Problem 10.9). Here, as shown in Figure 10.15 for the case of positive l, α is the angle between lines of constant phase and the y axis.

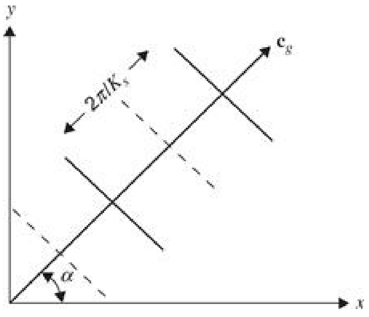


Figure 10.15 Stationary plane Rossby wave in a westerly flow. Ridges (*solid lines*) and troughs (*dashed lines*) are oriented at

an angle α to the y axis; the group velocity relative to the ground, c_g, is oriented at an angle α to the x axis. The wavelength is 2 π/K_s.

Adapted from Hoskins, 1983.

Because energy propagates at the group velocity, (10.68) indicates that the stationary wave response to a localized topographic feature should consist of two wave trains, one with $l > 0$ extending eastward and northward and the other with $l < 0$ extending eastward and southward. An example computed using spherical geometry is given in Figure 10.16. Although the positions of individual troughs and ridges remain fixed for stationary waves, the wave trains in this example do not decay in time, as the effects of dissipation are counteracted by energy propagation from the source at the Rossby wave group velocity.

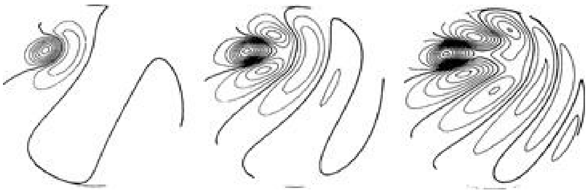


Figure 10.16 Vorticity pattern generated on a sphere when a constant angular velocity westerly flow impinges on a circular forcing centered at 30° N and 45° W of the central point. Left to right, the response at 2, 4, and 6 days after the forcing is switched on. Five contour intervals correspond to the maximum vorticity response that would occur in 1 day if there were no wave propagation. *Heavy lines* correspond to zero contours. The pattern is drawn on a projection in which the sphere is viewed from infinity.

After Hoskins, 1983.

For the climatological stationary wave distribution in the atmosphere, the excitation comes from a number of sources, both topographic and thermal, distributed around the globe. Thus, it is not easy to trace out distinct paths of wave propagation. Nevertheless, detailed calculations using spherical geometry suggest that two-dimensional barotropic Rossby wave propagation provides a reasonable first approximation for the observed departure of the extratropical time mean flow from zonal symmetry.

Rossby waves excited by isolated orographic features also play a significant role in the momentum budget. Letting the amplitude coefficient Ψ be real in (5.109), the meridional momentum flux can be expressed as

$$\overline{u'v'} = -\overline{(\partial\psi'/\partial x)\,(\partial\psi'/\partial y)} = -\Psi^2 kl/2$$

From (10.68) it is then verified readily that if $\overline{u} > 0$,

$$\begin{aligned} c_{gy} &> 0 \quad \text{implies} \quad \overline{u'v'} < 0 \\ c_{gy} &< 0 \quad \text{implies} \quad \overline{u'v'} > 0 \end{aligned}$$

Thus, westerly momentum converges into the wave source region (where the energy flux is divergent). This eddy momentum flux convergence is necessary to balance the momentum lost to the surface through the pressure torque mechanism discussed in Section 10.3.

10.5.2 Jet Stream and Storm Tracks

When the longitudinally asymmetric geopotential anomalies associated with stationary waves are added to the zonal-mean geopotential distribution, the resulting time mean field includes local regions of enhanced meridional geopotential gradient that are manifested in the wind field of the Northern Hemisphere by the Asian and North American jetstreams. The existence of these two jets can be inferred from the January mean 500-hPa geopotential height field shown in Figure 6.3. Note the strong meridional gradients in height associated with the troughs centered just east of the Asian and North American continents (the same features can be seen in annual mean charts, although with somewhat reduced intensity). The zonal flow associated with these semipermanent troughs is illustrated in Figure 6.2. In addition to the two intense jet cores in the western Pacific and western Atlantic, there is a third weaker jet centered over North Africa and the Middle East. Figure 6.2 shows dramatically the large deviations from zonal symmetry in the jet stream structure. In midlatitudes the zonal wind speed varies by nearly a factor of three between the core of the Asian jet and the area of low wind speed in western North America.

Although, as was argued earlier, the climatological stationary wave distribution on which the Asian and North American jets are superposed is apparently forced primarily by orography, the structure of the jets also appears to be influenced by continent–ocean heating contrasts. Thus, the strong vertical shear in Asian and North American jets reflects a thermal wind balance consistent with the very strong meridional temperature gradients that occur in winter near the eastern edges of the Asian and North American continents due to the contrast between warm water to the southeast and cold land to the northwest. A satisfactory description of the jet streams must account, however, not only for their thermal structure but for the westerly acceleration that air parcels must experience as they enter the jet and the deceleration as they leave the jet core.

To understand the quasi-geostrophic momentum budget in the jet streams and its relationship to the observed distribution of weather, we consider the zonal component of the momentum equation, which (if we neglect the β effect) may be written in the form of

$$\frac{D_g u_g}{Dt} = f_0(v - v_g) \equiv f_0 v_a \tag{10.70}$$

where v_a is the meridional component of the ageostrophic wind. This equation indicates that the westerly acceleration $(Du_g / Dt) > 0$) that air parcels experience as they enter the jet can only be provided by a poleward ageostrophic wind component $(v_a > 0)$; conversely, the easterly acceleration that air parcels experience as they leave the jet requires an equatorward ageostrophic motion. This meridional flow, together with its accompanying vertical circulation, is illustrated in Figure 10.17. Note that this secondary circulation is thermally direct upstream of the jet core. A magnitude of $v_a \sim 2 - 3\,\mathrm{m\,s^{-1}}$ is required to account for the observed zonal wind acceleration. This is an order of magnitude stronger than the zonal-mean indirect cell (Ferrel cell) that prevails in midlatitudes. Downstream of the jet core, however, the secondary circulation is thermally indirect but much stronger than the zonally averaged Ferrel cell. It is interesting to note that the vertical motion pattern on the poleward (cyclonic shear) side of the jet is similar to that associated with deep transient baroclinic eddies in the sense that subsidence occurs to the west of the stationary trough associated with the jet, and ascent occurs east of the trough (e.g., Figure 6.18).

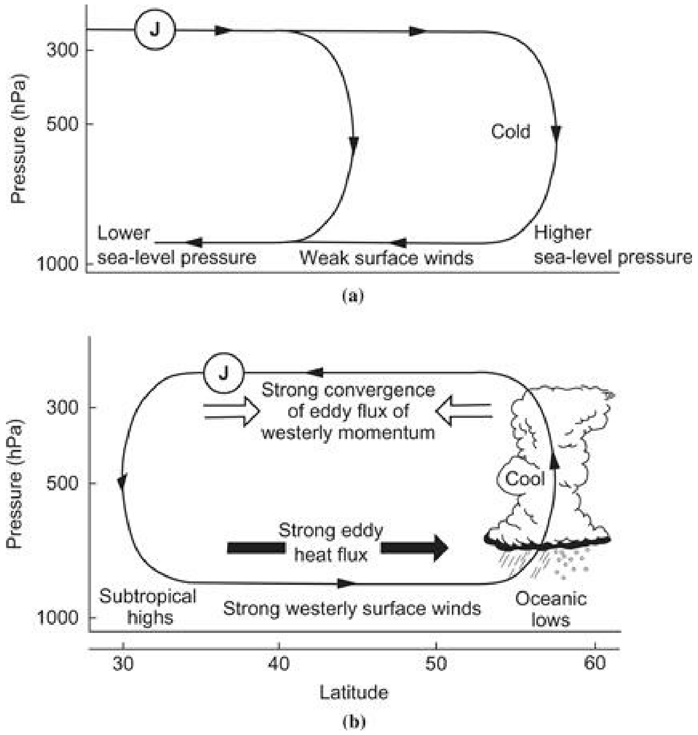

J
300
500
1000
Pressure (hPa)
Cold
Lower
sea-level pressure
Weak surface winds
Higher
sea-level pressure
(a)
J
Strong convergence
of eddy flux of
westerly momentum
Cool
Strong eddy
heat flux
Subtropical
highs
Strong westerly surface winds
Oceanic
lows
30
40
50
60
Latitude
(b)

Figure 10.17 Meridional cross-sections showing the relationship between the time mean secondary meridional circulation (*continuous thin lines with arrows*) and the jet stream ("J") at locations (a) upstream and (b) downstream from the jet stream cores.

After Blackmon et al., 1977. Copyright © American Meteorological Society. Reprinted with permission.

Because the growth rate of baroclinically unstable synoptic-scale disturbances is proportional to the strength of the basic state thermal wind, it is not surprising that the Pacific and Atlantic jet streams are important source regions for storm development. Typically, transient baroclinic waves develop in the jet entrance region, grow as they are advected downstream, and decay in the jet exit region. The role of these transient eddies in maintenance of the jet stream structure is rather complex. Transient eddy heat fluxes, which are strong and poleward in the storm tracks, appear to act to weaken the climatological jets. The transient eddy vorticity flux in the upper troposphere, however, appears to act to maintain the jet structure. In both cases the secondary ageostrophic circulation associated with the jet tends to partly balance the influence of the transient eddy fluxes in order to maintain the mean thermal and momentum balances.

10.6 Low-Frequency Variability

An understanding of the general circulation requires consideration not only of the zonal-mean and stationary wave components and their variations with the annual cycle, but also of irregular variability on time scales longer than that of individual transient eddies. The term *low-frequency variability* is generally used to describe such components of the general circulation. The observed spectrum of low-frequency variability ranges from weather anomalies lasting only 7 to 10 days to interannual variability on the scale of several years (see Section 11.1.6).

One possible cause of atmospheric low-frequency variability is forcing due to anomalies in sea surface temperature (SST), which themselves arise from coherent air–sea interaction. Because of the large thermal inertia of the oceanic surface mixed layer, such anomalies tend to have time scales much longer than those associated with subseasonal variations in the atmosphere; they are thought to be of greatest significance on the seasonal and interannual time scales.

It is believed, however, that substantial variability on subseasonal time scales can arise in midlatitudes in the absence of anomalous SST forcing as a result of internal nonlinear atmospheric dynamics, although SST anomalies may tend to favor the occurrence of certain types of variations. One example of internally generated low-frequency variability is the forcing of large-scale anomalies by potential vorticity fluxes of high-frequency transient waves. This process appears to be important in the maintenance of high-amplitude quasi-stationary wave disturbances called *blocking* patterns. Some types of blocking may also be related to special nonlinear wave patterns called *solitary waves*, in which damping by Rossby wave dispersion is balanced by intensification due to nonlinear advection. Although most internal mechanisms involve nonlinearity, there is some evidence that the longitudinally dependent time mean flow may be unstable to linear barotropic normal modes that are stationary in space but oscillate in time at low frequencies. Such modes, which are global in scale, may be responsible for some observed teleconnection patterns.

10.6.1 Climate Regimes

It has long been observed that extratropical circulation appears to alternate between a so-called high-index state, corresponding to a circulation with strong zonal flow and weak waves, and a "low-index" state, with weak zonal flow and high-amplitude waves. This behavior suggests that more than one climate regime exists consistent with a given external forcing and that the observed climate may switch back and forth between regimes in a chaotic fashion. Whether the high-index and low-index states actually correspond to distinct quasi-stable atmospheric climate regimes is a matter of controversy.

The concept of climate regimes can also be demonstrated in a highly simplified model of the atmosphere developed by Charney and DeVore (1979). They examined the equilibrium mean states that can result when a damped topographic Rossby wave interacts with the zonal-mean flow. Their model is an extension of the topographic Rossby wave analysis given in Section 5.7.2. In this model the wave disturbance is governed by (5.118), which is the linearized form of the barotropic vorticity equation (5.113) with weak damping added. The zonal-mean flow is governed by the barotropic momentum equation:

$$\frac{\partial \overline{u}}{\partial t} = -D(\overline{u}) - \kappa(\overline{u} - U_e) \tag{10.71}$$

where the first term on the right designates forcing by interaction between the waves and the mean flow, and the second term represents a linear relaxation toward an externally determined basic state flow, U_e.

The zonal mean equation (10.71) can be obtained from (5.113) by dividing the flow into zonal-mean and eddy parts and taking the zonal average to get

$$\frac{\partial}{\partial t}\left(-\frac{\partial \overline{u}}{\partial y}\right) = -\frac{\partial}{\partial y}\left(\overline{v_g' \zeta_g'}\right) - \frac{f_0}{H}\frac{\partial}{\partial y}\left(\overline{v_g' h_T}\right)$$

which, after integrating in y and adding the external forcing term, yields (10.71) with

$$D(\overline{u}) = -\overline{v_g' \zeta_g'} - (f_0/H)\,\overline{v_g' h_T} \tag{10.72}$$

As shown in Problem 10.5, the eddy vorticity flux, that is, the first term on the right in (10.72), is proportional to the divergence of the eddy momentum flux. The second term, which is sometimes referred to as the *form drag*, is the equivalent in the barotropic model of the surface pressure torque term in the angular momentum balance equation (10.43).

If h_T and the eddy geostrophic streamfunction are assumed to consist of single harmonic wave components in x and y, as given by (5.115) and (5.116), respectively, the vorticity flux vanishes, and with the aid of (5.119) the form drag can be expressed as

$$D(\overline{u}) = -\left(\frac{f_0}{H}\right)\overline{v_g' h_T} = \left(\frac{rK^2 f_0^2}{2\overline{u}H^2}\right)\frac{h_0^2 \cos^2 ly}{\left[(K^2 - K_s^2)^2 + \varepsilon^2\right]} \tag{10.73}$$

where r is the spin-down rate due to boundary layer dissipation, ε is as defined in equation (5.119), and K_s is the resonant stationary Rossby wave number defined in (5.112).

It is clear from (10.73) that the form drag will have a strong maximum when $\overline{u} = \beta/K^2$, as shown schematically in Figure 10.18. The last term on the right in (10.71), however, will decrease linearly as $\overline{u}$ increases. Thus, for suitable parameters there will be three values of $\overline{u}$ (labeled A, B, and C in Figure 10.18) for which the wave drag just balances the external forcing so that steady-state solutions may exist. By perturbing the solution about the points A, B, and C, it is shown easily (Problem 10.12) that point B corresponds to an unstable equilibrium, whereas the equilibria at points A and C are stable. Solution A corresponds to a low-index equilibrium, with high-amplitude waves analogous to a blocking regime. Solution C corresponds to a high-index equilibrium with strong zonal flow and weak waves. Thus, for this very simple model there are two possible “climates” associated with the same forcing.

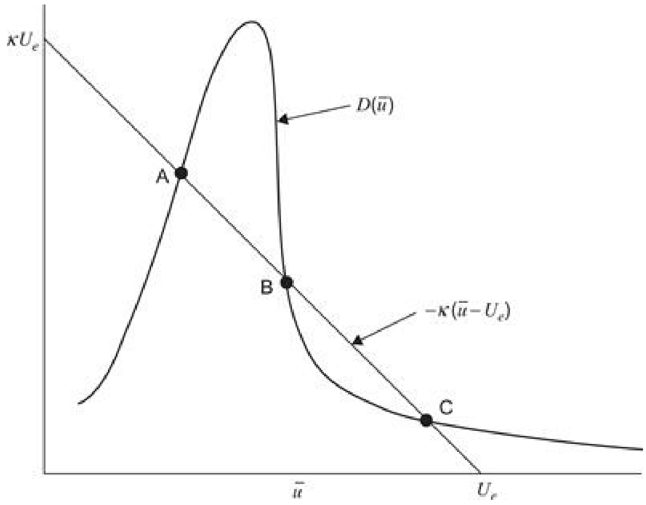


Figure 10.18 Schematic graphical solution for steady-state solutions of the Charney–DeVore model.

Adapted from Held, 1983.

The Charney–DeVore model is a highly oversimplified model of the atmosphere. Models that contain more degrees of freedom do not generally have multiple steady solutions. Rather, there is a tendency for the (unsteady) solutions to cluster about certain climate regimes and to shift between regimes in an unpredictable fashion. Such behavior is characteristic of a wide range of nonlinear dynamical systems and is popularly known as *chaos* (see Section 13.7).

10.6.2 Annular Modes

If a second meridional mode is added to the Charney–DeVore model (see Exercise M10.2), oscillatory solutions can be obtained that bear a qualitative resemblance to the observed leading modes of variability in the extratropical atmospheric circulation in both hemispheres. These observed modes, referred to as *annular modes*, are characterized by geopotential anomalies of opposite signs in the polar and midlatitude regions and by corresponding mean zonal wind anomalies with opposite signs equatorward and poleward of about 45° latitude (Figure 10.19). The annular modes exist year-round in the troposphere, but are strongest in the winter when they extend well into the stratosphere, especially in the Northern Hemisphere. The zonally symmetric mean flow anomalies associated with the annular modes are apparently maintained by anomalous eddy momentum fluxes, which are themselves influenced by the zonally symmetric flow anomalies.

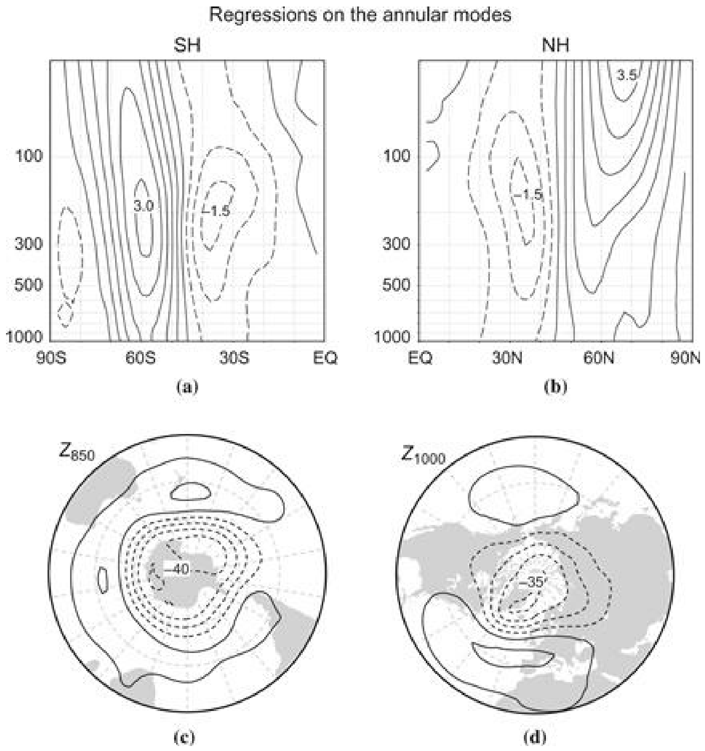


Figure 10.19 (a,b) Latitude–height cross-sections showing typical amplitudes of annular mode anomalies for mean-zonal wind (ordinate labels are pressure in hPa, contour interval 0.5m s^{-1}). (c,d) Lower tropospheric geopotential height (contour interval 10 m). The Southern Hemisphere (SH) is on the *left* and the Northern Hemisphere (NH) is on the *right*. Phase shown corresponds to the high-index state (strong polar vortices). The low-index state would have opposite signed anomalies.

After Thompson and Wallace, 2000. Copyright © American Meteorological Society. Reprinted with permission.

There is some evidence for downward propagation of the wintertime annular modes, suggesting that circulation changes in the stratosphere may precede annular mode changes in the troposphere. Dynamical linkages between the stratosphere and the troposphere are discussed further in Chapter 12.

10.6.3 Sea Surface Temperature Anomalies

Sea surface temperature anomalies influence the atmosphere by altering the flux of latent and sensible heat from the ocean, and thus providing anomalous heating patterns. The efficacy of such anomalies in exciting global-scale responses depends on their ability to generate Rossby waves. A thermal anomaly can generate a Rossby wave response only by perturbing the vorticity field. This requires that the thermal anomaly produce an anomalous vertical motion field, which in turn produces anomalous vortex tube stretching.

For low-frequency disturbances the thermodynamic energy equation (10.5) may be approximated as

$$\mathbf{V}\cdot\nabla T + wN^2HR^{-1} \approx J/c_p \tag{10.74}$$

Thus, diabatic heating can be balanced by horizontal temperature advection or by adiabatic cooling due to vertical motion. The ability of diabatic heating produced by a sea surface temperature anomaly to generate Rossby waves depends on which of these processes dominates. In the extratropics, SST anomalies primarily generate low-level heating, which is balanced mainly by horizontal temperature advection. In the tropics, positive SST anomalies are associated with enhanced convection, and the resulting diabatic heating is balanced by adiabatic cooling. Tropical anomalies have their greatest effect in the Western Pacific, where the average sea surface temperature is very high so that even a small positive anomaly can generate large increases in evaporation due to the exponential increase of saturation vapor pressure with temperature. By continuity of mass, the upward motion in cumulonimbus convection requires convergence at low levels and divergence in the upper troposphere.

The low-level convergence acts to sustain the convection by moistening and destabilizing the environment, whereas the *upper-level divergence* generates a vorticity anomaly. If the mean flow is westerly in the region of upper-level divergence, the forced vorticity anomaly will form a stationary Rossby wave train. The observed upper tropospheric height anomalies during the Northern Hemisphere winter associated with such an anomaly are shown schematically in Figure 10.20. The pattern strongly suggests a train of stationary Rossby waves that emanates from the equatorial source region and follows a great circle path, as predicted by barotropic Rossby wave theory (see Section 10.5.1). In this manner, tropical SST anomalies may generate low-frequency variability in the extratropics. It is possible that the effects of SST anomalies and of internal variability are not completely independent. In particular, it is more likely that the atmosphere will tend to preferentially reside in those regimes whose anomalous flow is correlated with the pattern in Figure 10.20 than would be the case in the absence of SST anomalies.

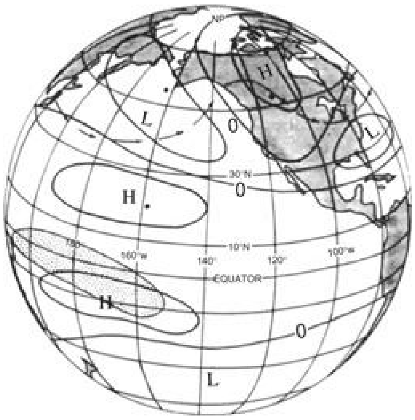


Figure 10.20 Pattern of middle and upper tropospheric height anomalies for Northern Hemisphere winter during an ENSO

event in the tropical Pacific. The region of enhanced tropical precipitation is shown by *shading*. *Arrows* depict a 200-hPa streamline for the anomaly conditions. “H” and “L” designate anomaly highs and lows, respectively. The anomaly pattern propagates along a great circle path with an eastward component of group velocity, as predicted by stationary Rossby wave theory.

After Horel and Wallace, 1981. Copyright © American Meteorological Society. Reprinted with permission.

10.7 Numerical Simulation of the General Circulation

The only practical manner in which to quantitatively simulate the present climate, or to predict possible climate modifications resulting from intentional or unintentional human intervention, is by numerical simulation with the aid of computers. Atmospheric general circulation models (AGCMs) are similar to large-scale numerical weather prediction models (see Chapter 13) in that they attempt to explicitly simulate synoptic-scale weather disturbances. However, whereas weather prediction is an *initial value* problem, which requires that the evolution of the flow be computed from a specified initial state, general circulation modeling is usually performed as a *boundary value* problem in which the average circulation is computed for specified external forcing conditions.

In many AGCMs the sea surface temperature is treated as a specified forcing. In reality, of course, there are strong interactions between the atmosphere and the ocean: The winds drive currents, which influence the sea surface temperature distribution, which in turn influences the atmosphere. Such interaction may be simulated using *coupled climate models*, also known as Earth system models (ESMs), which include flux couplers that link the atmosphere with other constituent components of the climate system, including the ocean, land surface, biosphere, and cryosphere. Unlike the dynamical cores of AGCMs, which involve numerical solution of the equations of dynamic meteorology and may be compared against known analytical nonlinear solutions on the sphere, there are no known solutions of the full coupled system. Therefore, for ESMs, validation typically comprises a comparison of simulated statistics with observations during the instrumental record.

Atmospheric model intercomparison projects (AMIPs) provide a standardized experimental design that is meant to facilitate comparing the performance of different models by fixing many aspects of the simulation. Experiments are configured with prescribed sea surface temperature and sea–ice boundary conditions, along with the concentration and distribution of radiatively important gases such as ozone and carbon dioxide. Similarly, the coupled model intercomparison project (CMIP) provides a standardized protocol for comparing ESMs. The collected datasets for AMIPs and CMIPs provide a large sample of information that has proven useful for climate studies and model comparisons; data from these experiments are freely available on the Web.

Interest in the development of AGCMs and ESMs derives not only from the pursuit of basic research but also from societal interest in predicting future climate. Periodic reports from the Intergovernmental Panel on Climate Change (IPCC) provide a comprehensive summary of the results from these model projections, and also a thorough review of the latest results in climate research. Forecasts described in these reports have been consistent in projecting, relative to preindustrial times, about 2.0 to 4.5°C warming of the global-mean surface temperature due to a presumed doubling of carbon dioxide. Projections also show a small, but nontrivial, probability that the warming may be much larger. An introduction to climate sensitivity analysis in Section 10.8 will serve to provide a quantitative basis for understanding the source of this disturbingly wide range of uncertainty.

Because of its enormous complexity and many important applications, general circulation modeling has become a complicated activity that cannot possibly be covered adequately in a short space. Here we can only give a summary of the primary physical processes represented and present an example of an application in climate modeling. A brief discussion of the technical aspects of the formulation of numerical prediction models is given in Chapter 13.

10.7.1 Dynamical Formulation

Most general circulation models are based on the primitive equations in the σ-coordinate form introduced in Section 10.3.1. As was pointed out in that section, σ coordinates make it possible to retain the dynamical advantages of pressure coordinates but simplify the specification of boundary conditions at the surface.

The minimum prediction equations for a σ-coordinate GCM are the horizontal momentum equation (10.30), the mass continuity equation (10.34), the thermodynamic energy equation (10.36), and a moisture continuity equation, which can be expressed as

(10.75)

$$\frac{D}{Dt}(q_v) = P_v$$

where q_v is the water vapor mixing ratio and P_v is the sum of all sources and sinks. In addition, we require the hydrostatic equation (10.35) to provide a diagnostic relationship between the geopotential and temperature fields. Finally, a relationship is needed to determine the evolution of the surface pressure $p_s(x, y, t)$. This is given by integrating (10.34) vertically and using the boundary conditions $\dot{\sigma} = 0$ at $\sigma = 0$ and 1 to obtain

(10.76)

$$\frac{\partial p_s}{\partial t} = -\int_0^1 \nabla \cdot (p_s \mathbf{V}) d\sigma$$

Vertical variations are generally represented by dividing the atmosphere into a number of levels and utilizing a finite difference grid. AGCMs typically have prediction levels at 1- to 3-km intervals extending from the surface to about a 30-km altitude. Some models, however, have many more levels extending nearly to the mesopause. Horizontal resolution of global models varies widely, from an effective grid size of several hundred kilometers to less than 100 km.

10.7.2 Physical Processes and Parameterizations

The various types of surface and atmospheric processes represented in a typical AGCM and the interactions among these processes are shown schematically in Figure 10.21. The most important classes of physical processes are (1) radiative, (2) cloud and precipitation, and (3) turbulent mixing and exchange.

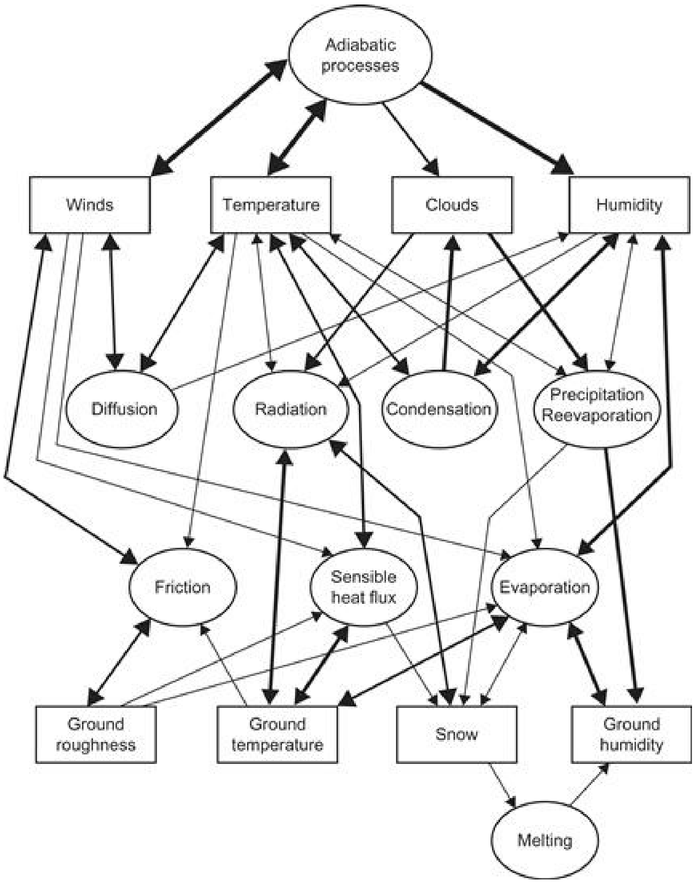

Adiabatic processes
Winds
Temperature
Clouds
Humidity
Diffusion
Radiation
Condensation
Precipitation Reevaporation
Friction
Sensible heat flux
Evaporation
Ground roughness
Ground temperature
Snow
Ground humidity
Melting

Figure 10.21 Schematic of processes commonly included in AGCMs and their interactions. The *thickness* of each arrow gives a rough indication of the importance of the interaction that a particular arrow represents.

After Simmons and Bengtsson, 1984. Copyright © Cambridge University Press. Used with permission.

As pointed out in Section 10.1, the fundamental process that drives the circulation of the atmosphere is the differential radiative heating that results in a net gain in heat at low latitudes relative to high latitudes. The general circulation of the atmosphere and the oceans provides the meridional and vertical heat transfer required for balance.

Half of the solar radiation absorbed at the surface is used to evaporate water and thus to moisten the atmosphere. Solar heating is realized in the atmosphere primarily in the form of latent heat release associated with convective clouds. The global distribution of evaporation clearly depends on the sea surface temperature, which is itself dependent on the general circulation of the oceans as well as on interactions with the atmosphere. That is why, for detailed understanding of the influence of large-scale air–sea interactions on climate, it is important to run coupled atmosphere–ocean GCMs in which the sea surface temperature is predicted. At present, however, most AGCMs use externally specified monthly or seasonal mean sea surface temperatures. Over land, however, the surface temperature adjusts very quickly to changes in the fluxes of solar and infrared radiation and is determined from a surface energy–balance equation.

Atmospheric radiative heating by solar radiation and heating and cooling by long-wave thermal radiation are computed using radiative transfer models of varying sophistication. Zonally averaged distributions of radiatively important constituents such as carbon dioxide, ozone, and even cloudiness are often employed, but the more complete models do utilize model-predicted zonally and time-varying cloudiness in their radiation codes.

Boundary layer fluxes of momentum, heat, and moisture are parameterized in most AGCMs by the use of bulk aerodynamic formulas (see Section 8.3.1). Typically the fluxes are specified to be proportional to the magnitude of the horizontal velocity at the lowest atmospheric level times the difference between the field variable at the boundary and its value at the lowest atmospheric level. In some models the boundary layer is explicitly resolved by locating several prediction levels within the lowest 2 km and utilizing the model-predicted boundary layer static stability in the parameterization of turbulent fluxes.

The hydrological cycle is usually represented by a combination of parameterization and explicit prediction. The water vapor mixing ratio is generally one of the explicitly predicted fields. The distributions of layer clouds and large-scale precipitation are then determined from the predicted distribution of humidity by requiring that when the predicted humidity exceeds 100%, enough vapor is condensed to reduce the mixing ratio to saturation or less. Parameterizations in terms of the mean state thermal and humidity structure must be used to represent the distributions of convective clouds and precipitation.

10.8 Climate Sensitivity, Feedbacks, and Uncertainty

As discussed previously, ESMs predict warming of about 3°C under a scenario of doubling carbon dioxide concentration over preindustrial values, but the probability distribution of forecasts is skewed, with a long tale extending toward much larger warming. To understand the source of this uncertainty, we briefly review climate sensitivity analysis, which is typically defined as the change in the global-mean temperature, from one equilibrium state to another, due to an imposed forcing that changes the heat budget of the Earth system. Assuming that the change is small, we may estimate it by the leading-order Taylor series approximation

$$T(R_0+\delta R) = T(R_0) + \left.\frac{\partial T}{\partial R}\right|_{R_0} \delta R \tag{10.77}$$

Here, the global-mean surface temperature, T, in equilibrium, is assumed to depend only on the parameter R, which may be taken as the net radiative heating of the surface, and R_0 is the value for the unperturbed climate. The linear response to perturbing the radiative heating by δR is then given by

$$\delta T = \left.\frac{\partial T}{\partial R}\right|_{R_0} \delta R = \lambda\,\delta R \tag{10.78}$$

where λ is the climate sensitivity parameter, which determines the change in temperature for a given amount of forcing. A common experiment with climate models used to determine λ involves perturbing the carbon dioxide field by instantaneously doubling the concentration and then integrating the model to a new equilibrium state (when the global-mean surface temperature reaches a new time-averaged value). Results show an initial, relatively fast, response due to strong greenhouse forcing, with about half of the warming occurring in about 50 years, followed by a slower asymptotic approach to the new equilibrium value, which may take over 1000 years.

The evolution to a new equilibrium state involves an interaction between various aspects of the Earth system, including feedback processes, which will be discussed next. In the absence of feedbacks, the experiments described previously produce a warming of about $\delta T_0 = 1.2$°C due to δR_0 of roughly 4 W m^{-2}, yielding a "control" climate sensitivity of 0.3 $K(W\ m^{-2})^{-1}$. Feedbacks may increase or decrease this control value by an amount proportional to the "output," δT:

$$\delta T = \delta T_0 + f \delta T \tag{10.79}$$

where "feedback factor" f controls the amplification. Feedbacks are due to physical processes that depend on temperature, such as clouds, which affect the radiative forcing. Therefore, we take $R(\alpha_i(T))$, where α_i represents the physical process, so that application of the chain rule in (10.78) gives

$$\delta T \;=\; \lambda \delta R_0 + \lambda \sum_i \frac{\partial R}{\partial \alpha_i}\frac{\partial \alpha_i}{\partial T}\delta T \tag{10.80}$$

This result is the same as (10.79) provided that the feedback factor is defined by

$$f \;=\; \lambda \sum_i \frac{\partial R}{\partial \alpha_i}\frac{\partial \alpha_i}{\partial T} \tag{10.81}$$

Solving (10.79) for δT gives

$$\delta T \;=\; \frac{\delta T_0}{1 - f} \tag{10.82}$$

If the feedback parameter is positive, the feedbacks amplify the control response to a given forcing; $f \geq 1$ represents unphysical solutions. Note that since $1/(1-f)$ is a nonlinear function,[3] small changes in f may lead to large changes in δT. This fact is graphically depicted in Figure 10.22 and reveals the essence of the uncertainty in climate forecasts. Climate models approximate the natural system, with many sources of uncertainty, which implies uncertainty in the feedback factor (assumed Gaussian in Figure 10.22). This feedback uncertainty maps into a non-Gaussian distribution in δT with a "long tail" of large warming (8–12° C) relative to the peak of the distribution (3° C). Reducing the long tail requires narrowing the distribution for the feedback factor. Since the total feedback factor is a sum of many feedback processes (see 10.80), this presents a daunting task in complex Earth System models, but may be more tractable in simpler models that are properly calibrated.

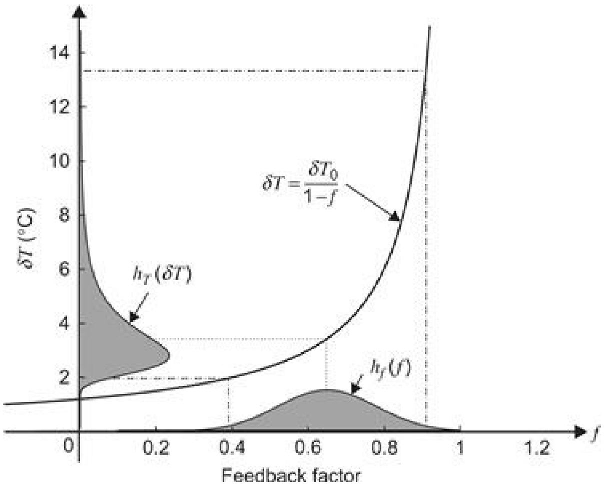


Figure 10.22 How (10.82) (*solid curve*) maps uncertainty in the feedback factor, f (abscissa) onto uncertainty in the temperature equilibrium change, δT (ordinate). *Gray shading* gives the probability density function, with a Gaussian distribution assumed for the feedback factor for purpose of illustration. The control equilibrium temperature change is taken to be 3°C. *Dotted* (*dashed*) *lines* show the mean (95% confidence interval) for the feedback factor.

Adapted with permission from Roe and Baker, 2007.

Of the many feedbacks in ESMs, the following are currently thought to be the most important:

1. The **water vapor feedback** results from the fact that, through the Clausius-Clapeyron equation, the saturation vapor pressure of water vapor increase with temperature. Water vapor is the most important greenhouse gas in Earth's atmosphere, so unless the relative humidity changes, a positive feedback results when the temperature increases because of, for example, CO_2 forcing, since the amount of water vapor in the atmosphere increases, producing more warming. Conversely, forcing that produces cooling reduces the saturation vapor pressure, leading to more cooling.
2. The **ice–albedo feedback** is due to the fact that ice strongly reflects solar radiation so that decreasing the amount of ice reduces the planetary albedo, thereby increasing the temperature and producing more ice loss. Conversely, increasing the amount of ice increases the planetary albedo, thereby decreasing the temperature, which further increases the amount of ice.
3. **Cloud feedback** comes in a wide variety, but one particularly simple one involves the areal coverage of low clouds. Low clouds emit radiation near the surface temperature, so they have little effect on outgoing long-wave radiation, but since they are highly reflective, they can reduce the net incoming solar radiation. Therefore, if warming results in substantially less low cloud, a positive feedback occurs since the planetary albedo decreases, producing more warming. Conversely, if warming results in more low cloud, a negative feedback results.
4. The **lapse-rate feedback** derives from the fact that, especially in the tropics, the lapse rate tends to follow a saturated adiabat. Since saturated adiabats become less steep with increasing temperature (more latent heat is available to offset adiabatic cooling for lifted parcels), a negative feedback results because smaller lapse rates result in less warming. Conversely, decreasing the temperature increases the lapse rate, resulting in more cooling.

Suggested References

ames. James, *Introduction to Circulating Atmospheres*, provides an excellent introduction to observational and theoretical aspects of the global general circulation.

orenz. Lorenz, *The Nature and Theory of the General Circulation of the Atmosphere*, although somewhat out of date, contains an excellent survey of the subject, including observational and theoretical aspects.

andall. Randall (ed.), *General Circulation Model Development*, contains articles describing all aspects of general circulation modeling.

chneider, 2006. Schneider (2006) provides a modern review of the general circulation.

ashington and Parkinson. Washington and Parkinson, *An Introduction to Three-Dimensional Climate Modeling*, provide an excellent text, which covers the physical basis and computational aspects of general circulation modeling.

Problems

10.1. Starting with the isobaric version of the thermodynamic energy equation (2.42), derive the log-pressure version (10.5).

10.2. Show that in the σ-coordinate system a mass element $\rho_0 dxdydz$ takes the form $-g^{-1}p_s dxdyd\sigma$.

10.3. Compute the mean zonal wind $\overline{u}$ at the 200-hPa level at 30° N under the assumption that $\overline{u}=0$ at the equator and that the absolute angular momentum is independent of latitude. What is the implication of this result for the role of eddy motions?

10.4. Show by scale analysis that advection by the mean meridional circulation can be neglected in the zonally averaged equations (10.11) and (10.12) for quasi-geostrophic motions.

10.5. Show that for quasi-geostrophic eddies the next to last term in square brackets on the right side in (10.15) is proportional to the vertical derivative of the eddy meridional relative vorticity flux.

10.6. From equations (10.16) through (10.19), derive the governing equation for the residual streamfunction (10.21).

10.7. Using the observed data given in Figure 10.13, compute the time required for each possible energy transformation or loss to restore or deplete the observed energy stores. (A watt equals 1 Js $^{-1}$.)

10.8. Compute the surface torque per unit horizontal area exerted on the atmosphere by topography for the following distribution of surface pressure and surface height:

$$p_s = p_0 + \hat{p}\sin kx, \quad h = \hat{h}\sin\left(kx-\gamma\right)$$

where p_0 = 1000 hPa, $\hat{p}=10$ hPa, $\hat{h}=2.5\times10^3$ m, γ = $\pi/6$ rad, and k = $1/(a\cos\phi)$, where ϕ = $\pi/4$ radiansis the latitude, and a is Earth's radius. Express the answer in kg s^{-2}.

10.9. Starting from (10.66) and (10.67), show that the group velocity relative to the ground for stationary Rossby waves is perpendicular to the wave crests and has a magnitude given by (10.69).

10.10. Consider a thermally stratified liquid contained in a rotating annulus of inner radius 0.8 m, outer radius 1.0 m, and depth 0.1 m. The temperature at the bottom boundary is held constant at T_0. The fluid is assumed to satisfy the equation of state (10.76) with $\rho_0=10^3$ kg m^{-3} and $\varepsilon = 2\times10^{-4}\,\mathrm{K}^{-1}$. If the temperature increases linearly with height along the outer radial boundary at a rate of 1° C cm^{-1} and is constant with height along the inner radial boundary, determine the geostrophic velocity at the upper boundary for a rotation rate of Ω = 1 rad s^{-1}. (Assume that the temperature depends linearly on radius at each level.)

10.11. Show by considering $\partial \overline{u}/\partial t$ for small perturbations about the equilibrium points in Figure 10.18 that point B is an unstable equilibrium point while points A and C are stable.

MATLAB Exercises

M10.1. The MATLAB script **topo_wave_1.m** uses finite differencing to solve the linearized vorticity equation (5.118) on a midlatitude β plane for the case of a constant mean zonal wind and forcing by a circular mountain with mountain height given by $h_T(x,y) = h_0 L^2 \left(L^2 + x^2 + y^2\right)^{-1}$, where h_0 and L are constants characterizing the mountain height and horizontal scale, respectively. Run the script for mean zonal winds of 5, 10, 15, and 20 ms^{-1}. In each case estimate the horizontal wavelength and group velocity of the northward and southward Rossby wave trains that form in the lee of the mountain. Compare your results with the corresponding expressions given in Section 10.5.1.

M10.2. The two-meridional-mode version of the Charney and Devore (1979) model described in Section 10.6.1 is given in the MATLAB script **C_D_model.m**. Solutions for this model may be steady state, periodic in time, or irregularly varying, depending on the forcing of the first mode zonal-mean streamfunction (`zf(1)` in the script) and the initial amplitude of this mode (`zinit(1)` in the script). Run this script for forcing values of `zf(1) = 0.1,0.2,0.3,0.4,0.5`. In each case do two runs with initial conditions `zinit(1) = zf(1)` and `zinit(1) = 0.1`, respectively. For each of the 10 cases, note whether the solution is steady, periodic, or irregular. Compute the time average of the zonal mean zonal wind for nondimensional time interval $2000 < t < 3000$ for each of these cases. Are your results consistent with Figure 10.18? Note for each case whether the streamfunction tends to be in phase or out of phase with the topography and whether the results are qualitatively consistent with the solution for topographic Rossby waves given in Section 5.7.2.

M10.3. The MATLAB script **baroclinic_model_1.m** provides a simple illustration of the effect of baroclinic eddies on the mean flow. The script extends the two-level baroclinic instability model discussed in Section 7.2 by calculating the evolution of the mean zonal flow components U_m and U_T caused by meridional vorticity and heat fluxes associated with unstable baroclinic waves. The calculation is done for a midlatitude β plane with weak Ekman layer damping. The eddies are governed by the linearized model of Section 7.2 (with fixed zonal wavelength of 6000 km), but the zonal-mean flow is affected by eddy–mean flow interactions and evolves as a result of eddy heat and momentum fluxes. In this model the mean thermal wind is relaxed toward a baroclinically unstable radiative equilibrium state (designated by `U0rad` in the code), which has a sinusoidal distribution in the meridional direction, and the mean wind components U_m and U_T are assumed to vanish at the northern and southern boundaries. Run the model for four cases with `U0rad = 10, 20, 30, and 40 ms`$^{-1}$, respectively. In each case run the model long enough so that the eddy kinetic energy equilibrates. Show in a table the maximum values for U_m and U_T for each of these cases. Explain these results in terms of the baroclinic instability theory of Section 7.2.

---

[1] It may be shown (Problem 10.2) that in sigma coordinates, the mass element $\rho_0 dxdydz$ takes the form $-g^{-1} p_s dxdyd\sigma$. Thus, p_s in sigma space plays a role similar to density in physical space.

[2] A similar analysis can be carried out for the nonlinear case.

[3] In the limit of the small feedback factor, $1/(1-f) \approx 1 + f$. This is linear feedback, which is consistent with (10.79) when the feedback is proportional to the control value, $\delta T = \delta T_0 + f\delta T_0$, and with (10.80) when the processes, α_i, do not depend on

temperature.

Chapter 11

Tropical Dynamics

Throughout the previous chapters of this book, we have emphasized circulation systems of the extratropical regions (i.e., the regions poleward of about 30° latitude). This emphasis should not be regarded as an indication of a lack of interesting motion systems in the tropics, but is a result, rather, of the relative complexity of the dynamics of tropical circulations. There is no simple theoretical framework analogous to quasi-geostrophic theory that can be used to provide an overall understanding of large-scale tropical motions.

Outside the tropics, the primary energy source for synoptic-scale disturbances is the zonal available potential energy associated with the latitudinal temperature gradient. Observations indicate that latent heat release and radiative heating are usually secondary contributors to the energetics of extratropical synoptic-scale systems. In the tropics, however, the storage of available potential energy is small due to the very small temperature gradients in the tropical atmosphere. Latent heat release appears to be the primary energy source, at least for those disturbances that originate within the equatorial zone. Most latent heat release in the tropics occurs in association with convective cloud systems, although much of the actual precipitation falls from mesoscale regions of strati-form clouds within such systems; the cloud systems are themselves generally embedded in large-scale circulations.

The diabatic heating associated with tropical precipitation not only drives a local response in the atmospheric circulation, but through excitation of equatorial waves may also induce a remote response. Thus, there is a strong interaction among cumulus convection, mesoscale, and large-scale circulations, which is of primary importance for understanding tropical motion systems. Furthermore, the distribution of diabatic heating in the tropics is strongly influenced by sea surface temperature (SST) variations, and these in turn are strongly influenced by the motion of the atmosphere.

An understanding of tropical circulations thus requires consideration of equatorial wave dynamics, interactions of cumulus convection and mesoscale circulations with large-scale motions, and air–sea interactions. Detailed treatment of all these topics is beyond the scope of an introductory text. Nevertheless, because the tropics play a fundamental role in the general circulation of the atmosphere, and coupling between tropical and middle latitudes is an important consideration in extratropical extended-range forecasting, some discussion of the tropics is required even in a text with an extratropical emphasis.

Of course, it is not always possible to distinguish clearly between tropical and extratropical systems. In subtropical regions ($\sim 30^\circ$ latitude), circulation systems characteristic of both tropical and extratropical regions may be observed depending on the season and geographical location. To keep the discussion in this chapter as simple as possible, we focus primarily on the zone well equatorward of 30° latitude where the influence of middle-latitude systems should be minimum.

11.1 The Observed Structure of Large-Scale Tropical Circulations

Because of the nature of their energy sources, as well as the smallness of the Coriolis parameter, large-scale equatorial motion systems have several distinctive characteristic structural features that are quite different from those of midlatitude systems. Many of these can be understood in terms of the equatorial wave modes discussed in Section 11.3. Before discussing equatorial wave theory, however, it is useful to review some of the major observed circulation features of the tropical atmosphere.

11.1.1 The Intertropical Convergence Zone

Traditionally, the tropical general circulation was thought to consist of a thermally direct Hadley circulation in which air in the lower troposphere in both hemispheres moved equatorward toward the *intertropical convergence zone* (ITCZ), where by continuity considerations it was forced to rise uniformly and move poleward, thus transporting heat away from the equator in the upper troposphere in both hemispheres. This simple model of large-scale overturning is not, however, consistent with the observed vertical profile of equivalent potential temperature (θ_e). As indicated in Figure 11.1, the mean tropical atmosphere is conditionally stable above about 600 hPa. Thus, a large-scale upward mass flow, were it to exist, would be up the gradient of θ_e in the upper troposphere and would actually *cool* the upper troposphere in the region of the ITCZ. Such a circulation could not generate potential energy and would not, therefore, satisfy the heat balance in the equatorial zone.

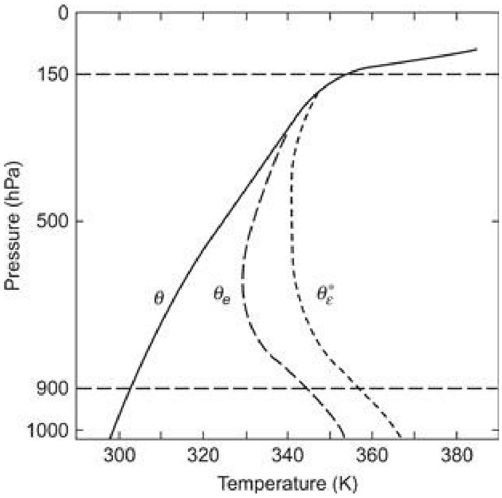


Figure 11.1 Typical sounding in the tropical atmosphere showing the vertical profiles of potential temperature θ, equivalent potential temperature θ_e, and equivalent potential temperature θ_e^* of a hypothetically saturated atmosphere with the same temperature at each level. This figure should be compared with Figure 2.8, which shows similar profiles for a midlatitude squall line sounding.

After Ooyama, 1969. Copyright © American Meteorological Society. Reprinted with permission.

It appears that the only way in which heat can effectively be brought from the surface to the upper troposphere in the ITCZ is through pseudoadiabatic ascent in the cores of large cumulonimbus clouds (often referred to as “hot towers”). For such motions, cloud parcels approximately conserve θ_e. They can, therefore, arrive in the upper troposphere with moderate temperature excesses. Thus, the heat balance of the equatorial zone can be accounted for, at least qualitatively, provided that the vertical motion in the ITCZ is confined primarily to the updrafts of individual convective cells. Riehl and Malkus (1958) estimated that only 1500–5000 individual hot towers would need to exist simultaneously around the globe to account for the required vertical heat transport in the ITCZ.

This view of the ITCZ as a narrow zonal band of vigorous cumulus convection has been confirmed beyond a doubt by observations, particularly satellite cloud photos. An example is given in Figure 11.2, which shows infrared cloud brightness temperatures in the tropics for the period August 14 to December 17, 1983. Low brightness temperatures signal the presence of deep anvil clouds characteristic of convective storms. The ITCZ appears as a line of deep convective cloud extending across the Atlantic and Pacific oceans between about 5° and 10°N.

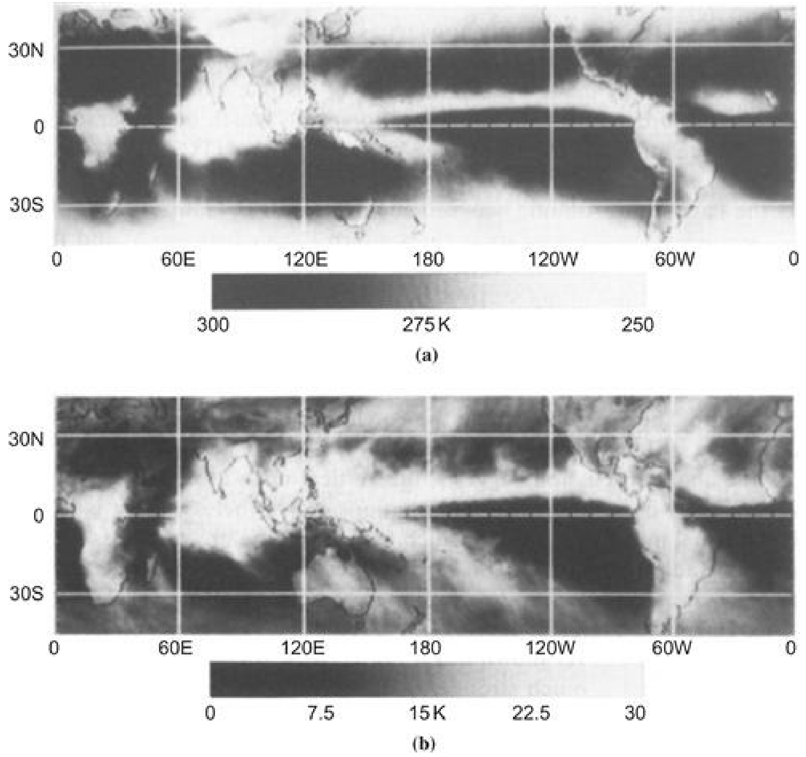


Figure 11.2 (a) Time-mean IR brightness temperature and (b) three-hour standard deviations about the time-mean IR brightness temperature for August 14 to December 17, 1983. Low values of brightness temperature indicate the presence of high cold cirrus anvil clouds.

After Salby et al., 1991. Copyright © American Meteorological Society. Reprinted with permission.

Observations indicate that within the ITCZ, precipitation greatly exceeds the moisture supplied by evaporation from the ocean surface below. Thus, much of the vapor necessary to maintain the convection in the ITCZ must be supplied by the converging trade wind flow in the lower troposphere. In this manner the large-scale flow provides the latent heat supply required to sustain the convection. The convective heating in turn produces large-scale midtropospheric temperature perturbations and (through hydrostatic adjustment) surface and upper-level pressure perturbations, which maintain the low-level inflow.

The preceding description of the ITCZ is actually oversimplified. In reality, the ITCZ over the oceans rarely appears as a long unbroken band of heavy convective cloudiness, and it almost never is found right at the equator. Rather, it is usually made up of a number of distinct cloud clusters, with scales of the order of a few hundred kilometers, which are separated by regions of relatively clear skies. The strength of the ITCZ is also quite variable in both space and time. It is particularly persistent and well defined over the Pacific and Atlantic between about 5° and 10°N latitude (as in Figure 11.2) and occasionally appears in the western Pacific between 5° and 10°S.

Figure 11.2 shows that not only is the mean deep convection associated with the ITCZ found in the 5° to 10°N latitude belt, but during the period shown in Figure 11.2, the standard deviation of deep convection is also a maximum there. This is consistent with the idea that the ITCZ is the locus of transient cloud clusters, rather than simply a region of steady-state precipitation and mean uplift. The dry zones along the equator in the oceanic regions are a particularly striking feature in Figure 11.2.

As the preceding discussion suggests, the vertical mass flux associated with the ITCZ has important regional variations. Nevertheless, there is a significant zonal mean component, which constitutes the upward mass flux of the mean Hadley circulation. This Hadley circulation consists of overturning thermally direct cells in the low latitudes of both hemispheres, as shown in Figure 10.7. The center of the Hadley circulation is located at the mean latitude of the ITCZ. As Figure 10.7 shows, the winter hemisphere branch of the Hadley cell is much stronger than the summer hemisphere branch. Observations indicate that two Hadley cells, symmetric about the equator, are rarely observed even in the equinoctial seasons. Rather, the northern cell dominates from November to March, the southern cell dominates from May to September, and rapid transitions occur in April and October (see Oort, 1983).

11.1.2 Equatorial Wave Disturbances

The variance observed in the cloudiness associated with the ITCZ, as illustrated in Figure 11.2b, is generally caused by transient precipitation zones associated with weak equatorial wave disturbances that propagate westward along the ITCZ. That such westward-propagating disturbances exist and are responsible for a large part of the cloudiness in the ITCZ can be seen easily by viewing time-longitude sections constructed from daily satellite pictures cut into thin zonal strips. An example is shown in Figure 11.3. The well-defined bands of cloudiness that slope from right to left down the page define the locations of the cloud clusters as functions of longitude and time. Clearly much of the cloudiness in the 5 to 10°N latitude zone of the Pacific is associated with westward-moving disturbances. The slope of the cloud lines in Figure 11.3 implies a westward propagation speed of about 8 to 10ms^{-1}. The longitudinal separation of the cloud bands is about 3000 to 4000 km, corresponding to a period range of about 4 to 5 days for this type of disturbance.

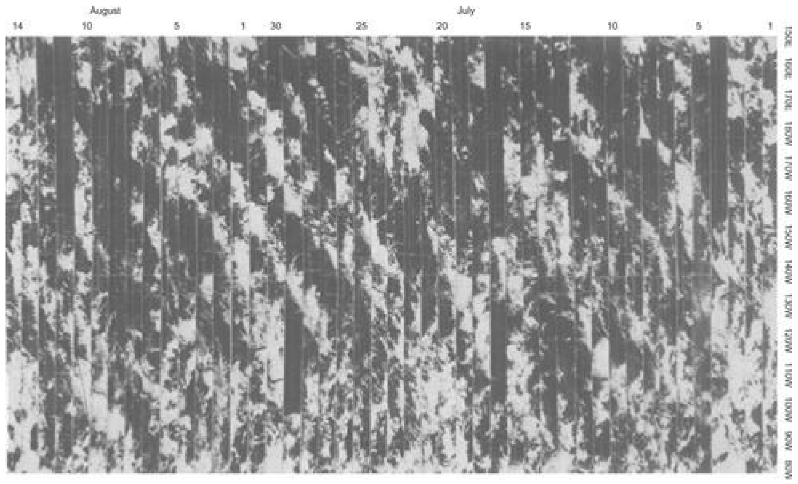


Figure 11.3 Time-longitude sections of satellite photographs for the period July 1 to August 14, 1967, in the 5 to 10°N latitude band of the Pacific. The westward progression of the cloud clusters is indicated by the bands of cloudiness sloping down the page from right to left.

After Chang, 1970. Copyright © American Meteorological Society. Reprinted with permission.

Diagnostic studies suggest that these westward-propagating wave disturbances are generally driven by the release of latent heat in their accompanying convective precipitation areas. The vertical structure of a typical equatorial wave disturbance is shown in schematic form in Figure 11.4. Vertical motion in such disturbances is proportional to the diabatic heating rate; thus, the maximum vertical velocities occur in the convective zone. By mass continuity there must be convergence at low levels in the convective zone and divergence in the upper levels. Hence, provided that the absolute vorticity has the sign of f, the divergence term in the vorticity equation will induce cyclonic vorticity tendencies in the lower troposphere and anticyclonic vorticity tendencies in the upper troposphere. The process of adjustment between mass and velocity fields will then generate a low-level trough and an upper-level ridge.[1] Thus, the thickness (or layer mean temperature) in the convective zone must be greater than in the surrounding environment.

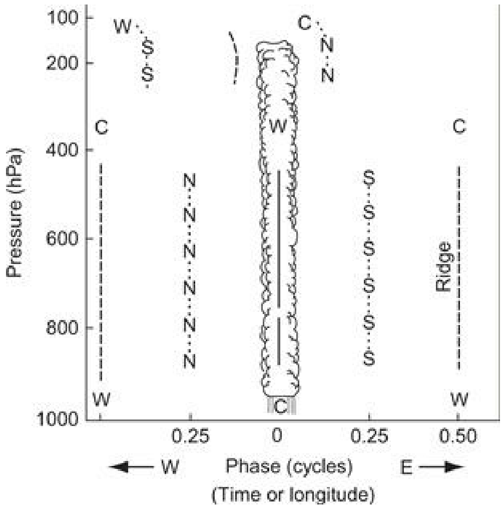


Figure 11.4 Schematic model for equatorial wave disturbances showing trough axis (*solid line*), ridge axis (*dashed lines*), and axes of northerly and southerly wind components (*dotted lines*). Regions of warm and cold air are designated by W and C, respectively. The axes shown are for northerly (N) and southerly (S) wind components.

After Wallace, 1971. Copyright © American Geophysical Union. Used with permission.

In the convectively active areas of equatorial waves there is net upward motion, and midtropospheric temperatures are warmer than average (although generally by less than 1°C). The correlations between temperature and vertical motion, and between temperature and diabatic heating, are thus both positive, and the potential energy generated by the diabatic heating is immediately converted to kinetic energy—that is, the $P' \cdot K'$ conversion balances R' in (10.62). There is, in this approximation, no storage in the form of available potential energy. The energy cycle of these disturbances, therefore, differs remarkably from that of midlatitude baroclinic systems in which the available potential energy greatly exceeds the kinetic energy.

For latent heat release by cumulonimbus clouds to be an effective energy source for large-scale disturbances, there must be an interaction between the convective scale and the large scale, as mentioned in Section 9.7.2. In such interaction large-scale convergence at low levels moistens and destabilizes the environment so that small-scale thermals can easily reach the level of free convection and produce deep cumulus convection. The cumulus cells, in turn, act cooperatively to provide the large-scale heat source that drives the secondary circulation responsible for the low-level convergence.

A typical vertical profile of divergence in the precipitation zone of a synoptic-scale equatorial wave disturbance in the western Pacific is shown in Figure 11.5. Convergence is not limited to low-level frictional inflow in the planetary boundary layer, but extends up to nearly 400 hPa, which is the height where the hot towers achieve their maximum buoyancy. The deep convergence implies that there must be substantial entrainment of midtropospheric air into the convective cells. Because the midtropospheric air is relatively dry, this entrainment will require considerable evaporation of liquid water to bring the mixture of cloud and environment air to saturation. It will thus reduce the buoyancy of the cloud air and may in fact produce negatively buoyant convective downdrafts if there is sufficient evaporative cooling. However, in the large cumulonimbus clouds present in equatorial waves, the central core updrafts are protected from entrainment by the surrounding cloud air so that they can penetrate nearly to the tropopause without dilution by environmental air. These undiluted cores constitute the "hot towers" referred to in Section 11.1.1. Because the hot towers are responsible for much of the vertical heat and mass transport above the boundary layer in the ITCZ, and the wave disturbances contain most of the active convective precipitation areas along the ITCZ, it is obvious that equatorial waves play an essential role in the general circulation of the atmosphere.

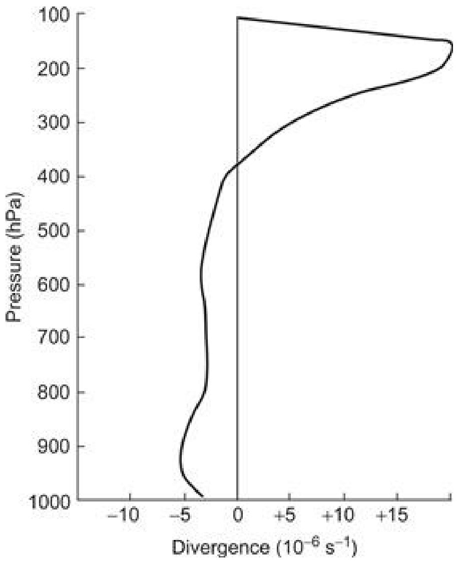


Figure 11.5 Vertical profile of 4°-square area average divergence based on composites of many equatorial disturbances.

Adapted with permission from *Williams, 1971.*

11.1.3 African Wave Disturbances

The considerations of the previous subsection are valid for ITCZ disturbances over most regions of the tropical oceans. In the region of the North African continent, however, local effects due to surface conditions create a unique situation that requires separate discussion. During the Northern Hemisphere summer, intense surface heating over the Sahara generates a strong positive meridional temperature gradient in the lower troposphere between the equator and about 25°N.

The resulting easterly thermal wind is responsible for the existence of a strong easterly jet core near 650 hPa centered near 16°N as shown in Figure 11.6. Synoptic-scale disturbances are observed to form and propagate westward in the cyclonic shear zone to the south of this jet core.

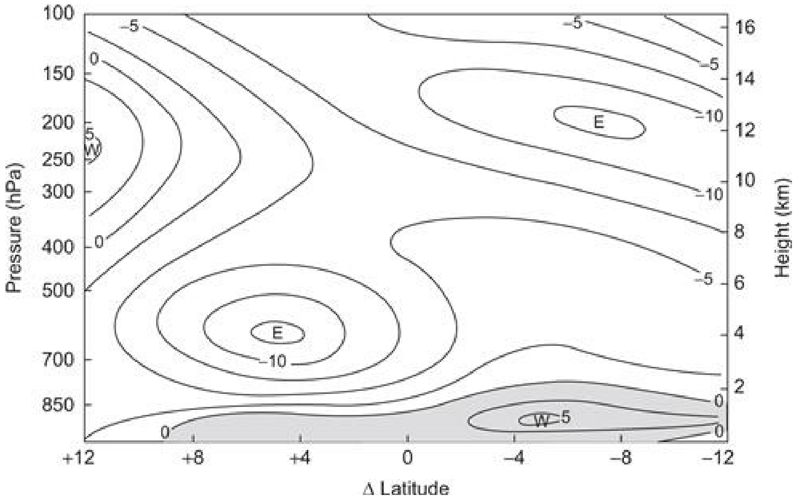


Figure 11.6 Mean-zonal wind distribution in the North African region (30°W to 10°E longitude) for the period August 23 to September 19, 1974. Latitude is shown relative to latitude of maximum disturbance amplitude at 700 hPa (about 12°N). The contour interval is $2.5 \mathrm{ms}^{-1}$.

After Reed et al., 1977. Copyright © American Meteorological Society. Reprinted with permission.

Occasionally, such disturbances are progenitors of tropical storms and hurricanes in the western Atlantic. The average wavelength of observed African wave disturbances is about 2500 km, and the westward propagation speed is about $8 \mathrm{ms}^{-1}$, implying a period of about 3.5 days. The disturbances have horizontal velocity perturbations that reach maximum amplitude at the 650-hPa level, as indicated in Figure 11.7. Although there is considerable organized convection associated with these waves, they do not appear to be driven primarily by latent heat release, but depend, rather, on barotropic and baroclinic conversions of energy from the easterly jet.

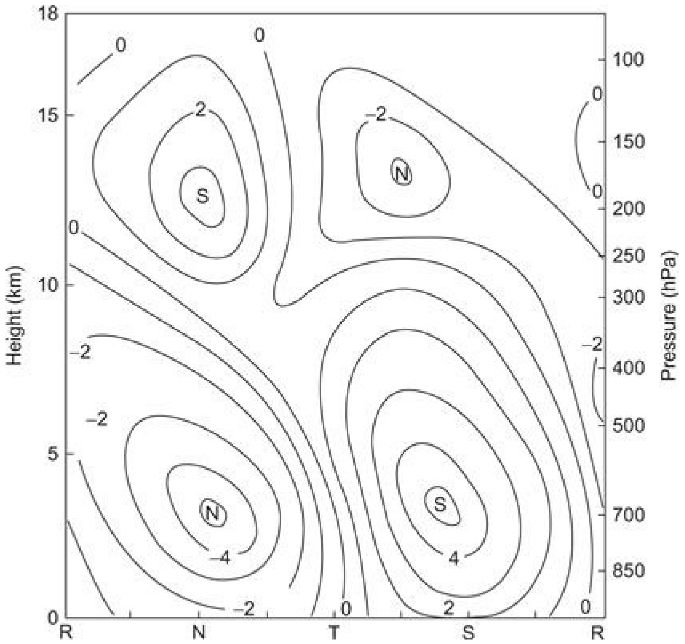


Figure 11.7 Vertical cross-section along the reference latitude of Figure 11.6 showing perturbation meridional velocities in ms^{-1}. R, N, T, and S refer to ridge, northwind, trough, and southwind sectors of the wave, respectively.

After Reed et al., 1977. Copyright © American Meteorological Society. Reprinted with permission.

In Figure 11.8 the absolute vorticity profile for the African easterly jet shown in Figure 11.6 is plotted. The shaded region indicates the area in which the vorticity gradient is negative. Thus, it is clear that the African jet satisfies the necessary condition for barotropic instability discussed in Section 7.4.2.[2] Baroclinic instability due to the strong easterly shear in the lower troposphere also appears to play a role in these disturbances. Thus, both barotropic and baroclinic conversions from the mean flow energy appear to be important for the generation of African wave disturbances.

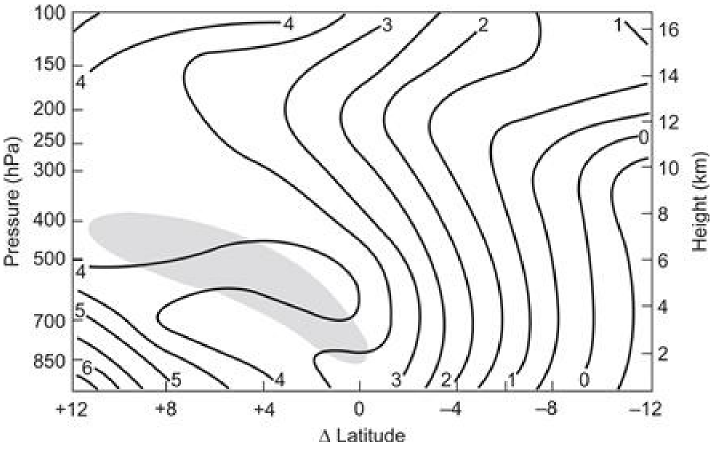


Figure 11.8 Absolute vorticity (units $10^{-5}\mathrm{s}^{-1}$) corresponding to the mean wind field of Figure 11.6. Shading shows region where $\beta - \partial^2\bar{u}/\partial y^2$ is negative.

After Reed et al., 1977. Copyright © American Meteorological Society. Reprinted with permission.

Because such disturbances continue to exist in the absence of strong mean wind shears after they have propagated westward into the Atlantic, it is unlikely that either baroclinic or barotropic instability continues to be the primary energy source for their maintenance. Rather, diabatic heating through precipitating convective systems appears to be the main energy source for such waves over the ocean.

11.1.4 Tropical Monsoons

The term *monsoon* is commonly used in a rather general sense to designate any seasonally reversing circulation system. The basic drive for a monsoon circulation is provided by the contrast in the thermal properties of the land and sea surfaces. Because the thin layer of soil that responds to seasonal changes in surface temperature has a small heat capacity compared to the heat capacity of the upper layer of the ocean that responds on a similar timescale, the absorption of solar radiation raises the surface temperature over land much more rapidly than over the ocean. The warming of the land relative to the ocean leads to enhanced cumulus convection and thus to latent heat release, which produces warm temperatures throughout the troposphere.

Much of the tropics is influenced by monsoons. The most extensive monsoon circulation by far is the Asian monsoon. This monsoon completely dominates the climate of the Indian subcontinent, producing warm wet summers and cool dry winters. An idealized model of the structure of the

Asian summer monsoon is indicated in Figure 11.9. As indicated in Figure 11.9, the 1000 to 200-hPa thickness is larger over the land than over the ocean. As a result there is a pressure gradient force at the upper levels directed from the land to the ocean.

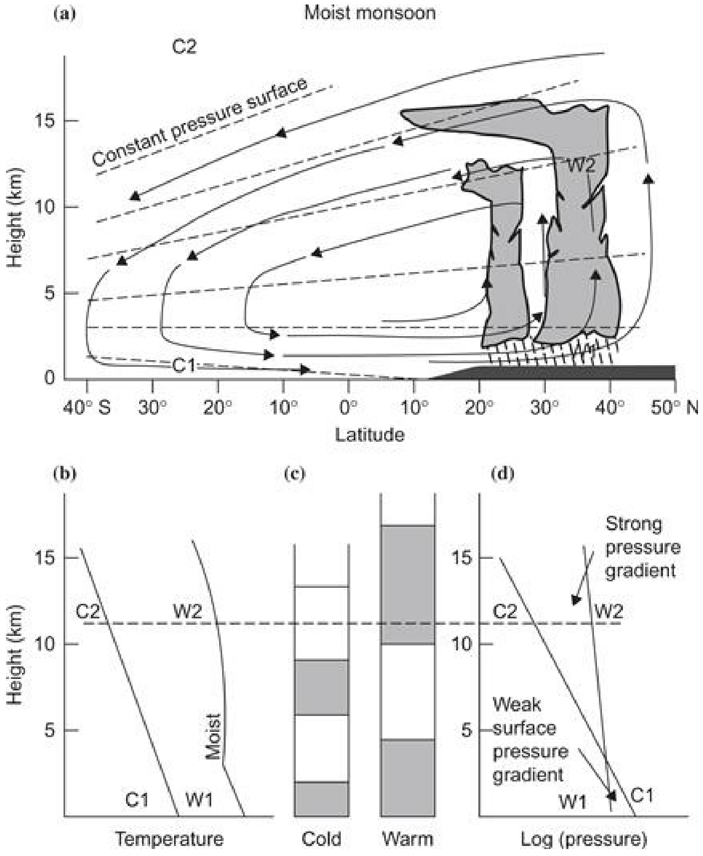

(a)
Moist monsoon
C2
Constant pressure surface
W2
C1
Height (km)
0
5
10
15
40° S
30°
20°
10°
0°
10°
20°
30°
40°
50° N
Latitude
(b)
(c)
(d)
C2
W2
C1
W1
Moist
Temperature
Cold
Warm
Strong pressure gradient
Weak surface pressure gradient
Log (pressure)

Figure 11.9 Schematic representation of the moist phase of the Asian monsoon. (a) *Arrows* show the meridional circulation, *dashed lines* the isobars. (b) Temperature profiles for columns C1 to C2 and W1 to W2, respectively. (c) Schematic of the mass distribution in the cold and warm sectors. (d) Horizontal pressure difference as a function of height.

After Webster and Fasullo, 2003.

The divergent wind that develops in response to this pressure gradient (shown by the arrows in Figure 11.9) causes a net mass transport out of the air column above the continent and thereby generates a surface low over the continent (often called a *thermal low*). A convergent wind then develops at low levels. This low-level flow produces a convergence of moisture, which by increasing the equivalent potential temperature in the boundary layer makes the environment more favorable for development of the cumulus convection, which is the primary energy source for the monsoon circulation.

The low-level convergence and upper-level divergence over the continent constitute a secondary circulation that concentrates cyclonic vorticity at the lower levels and anticyclonic vorticity at the upper levels. Thus, the vorticity adjusts toward geostrophic balance. From Figure 11.9 it is clear that a positive correlation exists between the vertical motion and the temperature field. Therefore, the monsoon circulation converts eddy potential energy to eddy kinetic energy, just as midlatitude baroclinic eddies do.

Unlike the case of baroclinic eddies, however, the primary energy cycle of the monsoons does not involve the zonal-mean potential or kinetic energy. Rather, eddy potential energy is generated directly by diabatic heating (latent and radiative heating); the eddy potential energy is converted to eddy kinetic energy by a thermally direct secondary circulation; and the eddy kinetic energy is frictionally dissipated. (A portion of the eddy kinetic energy may be converted to zonal kinetic energy.) In a dry atmosphere, monsoon circulations would still exist; however, because the diabatic heating would then be confined to a shallow layer near the surface, they would be much weaker than the observed monsoons. The presence of cumulus convection and its concomitant latent heat release greatly amplifies the eddy potential energy generation and makes the summer monsoons among the most important features of the global circulation.

In the winter season the thermal contrast between the land and the sea reverses so that the circulation is just opposite to that shown in Figure 11.9. As a result the continents are cool and dry and the precipitation is found over the relatively warm oceans.

11.1.5 The Walker Circulation

The pattern of diabatic heating in the equatorial regions exhibits strong departures from zonal symmetry. These are caused by longitudinal variations in sea surface temperature due mainly to the effects of wind-driven ocean currents. Such SST variations produce zonally asymmetric atmospheric circulations, which in some regions dominate over the Hadley circulation. Of particular significance is the east–west overturning along the equator, which is shown schematically in Figure 11.10. Several overturning cells are indicated, which are associated with diabatic heating over equatorial Africa, Central and South America, and the Maritime Continent (i.e., the Indonesian region). The dominant cell in both zonal scale and amplitude, however, is that in the equatorial Pacific. This cell is referred to as *Walker circulation* after G. T. Walker, who first documented the surface pressure pattern associated with it.

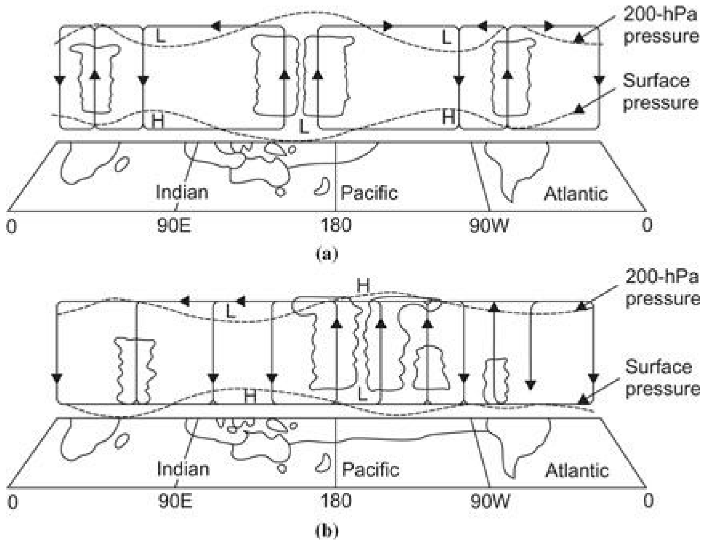


Figure 11.10 Schematic diagrams of the Walker circulations along the equator for (a) normal conditions and (b) El Niño conditions.

From Webster, 1983, and Webster and Chang, 1988. Copyright © American Meteorological Society. Reprinted with permission.

As suggested by Figure 11.10, this pressure pattern consists of low surface pressure in the western Pacific and high surface pressure in the eastern Pacific. The resulting westward-directed pressure gradient force drives mean surface easterlies in the equatorial Pacific, which are much stronger than the zonal-mean surface easterlies, and by horizontal vapor transport provides a moisture source for convection in the western Pacific, in addition to that provided by the high evaporation rates caused by the high sea surface temperatures in that region.

The wind stress due to the time-mean equatorial easterly surface winds over the Pacific has a strong influence on the heat balance of the ocean surface layer. It advects warm surface waters into the western Pacific and produces poleward drifts in the oceanic Ekman layer which by continuity drive an equatorial upwelling. This upwelling accounts for the cold tongue of water along the equator, which in turn is a major reason for the equatorial dry zone exhibited earlier in Figure 11.2.

11.1.6 El Niño and the Southern Oscillation

The east–west pressure gradient associated with the Walker circulation undergoes an irregular interannual variation. This global-scale "see-saw" in pressure, and its associated changes in patterns of wind, temperature, and precipitation, was named the *southern oscillation* by Walker. This oscillation can be clearly seen by comparing time series of surface pressure anomalies (i.e., departures from the long-term mean) for locations on the western and eastern sides of the equatorial Pacific. As shown in the upper portion of Figure 11.11, surface pressure anomalies at Darwin, Australia, and Tahiti are negatively correlated and have strong variations in the period range of 2 to 5 years. During periods of large pressure difference between these stations, the Walker circulation is unusually strong and has the same general structure as its time-mean pattern. In periods of weak pressure difference the Walker circulation weakens and the region of maximum rainfall shifts eastward (see Figure 11.10b).

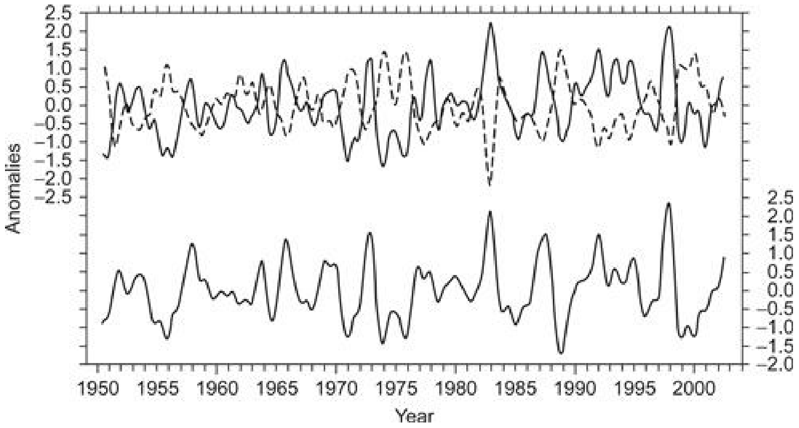


Figure 11.11 Time series showing SST anomalies (°C) in the eastern Pacific (*lower curve*) and anomalies in sea-level pressures (hPa) at Darwin (*upper solid curve*) and Tahiti (*upper dashed curve*). Data are smoothed to eliminate fluctuations with periods less than a year.

Figure courtesy of Dr. Todd Mitchell, University of Washington.

The weakening of trade winds during periods of small Darwin–Tahiti pressure difference reduces wind-driven oceanic upwelling and deepens the oceanic thermocline in the eastern Pacific. This in turn, as shown in the lower portion of Figure 11.11, leads to an increase in the sea surface temperature that is referred to as *El Niño* (Spanish for "the child"). This term was originally applied to a warming of the coastal waters of Peru and Ecuador that occurs annually near Christmas time (hence, El Niño refers to the Christ child), but is now used in a more general sense for the large-scale oceanic anomalies associated with the weak wind phase of the Southern Oscillation. In the opposite phase, the strengthening of trade winds associated with large Darwin–Tahiti pressure differences enhances oceanic upwelling and deepens the oceanic thermocline. This causes a cooling of the sea surface temperatures in the equatorial Pacific that is now referred to as *La Niña*.

The ocean anomalies associated with weakening in the trade winds may begin near the coast, but over the course of several months spread westward along the equator to produce large-scale positive SST anomalies over much of the equatorial Pacific. These SST anomalies in turn cause a further weakening of the trade winds. This entire complex of atmospheric and oceanic variations is now referred to as ENSO (for El Niño—Southern Oscillation). It is a dramatic example of interannual climate variability associated with atmosphere–ocean coupling.

The leading theoretical model for ENSO is the "delayed oscillator" model. In this model the sea surface temperature anomaly, T, in the eastern Pacific satisfies an equation of the form

$$\frac{dT}{dt} = bT(t) - cT(t - \tau)$$

where b and c are positive constants and τ is a time delay determined by the adjustment time for the equatorial ocean. The first term on the right represents a positive feedback associated with changes in the Darwin–Tahiti pressure difference. This term represents the atmosphere–ocean coupling through which an initial weakening of the wind causes an increase of the SST, which induces a further weakening of the wind and thus a further increase in the SST. The second term provides a negative feedback due to the adjustment in the thermocline depth (and thus the ocean temperature) caused by propagating equatorial waves (see Section 11.4) in the ocean that are excited by the SST changes. The time delay in the negative feedback term is determined by the time that it takes for wave energy excited by air–sea interaction in the eastern Pacific to propagate to the western boundary of the ocean, undergo reflection, and propagate back to the region of origin. For realistic parameters the delayed oscillator model leads to ENSO oscillations in the period range of 3 to 4 years. This highly simplified model qualitatively accounts for the average characteristics of an ENSO cycle, but cannot, however, account for the observed irregularity of ENSO.

In addition to its profound effects in the equatorial region, ENSO is associated with a wide range of interannual climate anomalies in the extratropics. Thus, the development of models that show skill in predicting ENSO several months in advance is of considerable practical importance.

11.1.7 Equatorial Intraseasonal Oscillation

In addition to the interannual variability associated with El Niño, the equatorial circulation has an important intraseasonal oscillation, which occurs on a timescale of 30 to 60 days and is often referred to as the Madden–Julian oscillation (MJO) in honor of the meteorologists who first described it. The structure of the equatorial intraseasonal oscillation is shown schematically in Figure 11.12, which shows the time development of the oscillation in the form of longitude–height sections along the equator, with time increasing at an interval of about 10 days for each panel from top to bottom. The circulations in Figure 11.12 are intended to represent anomalies from the time-mean equatorial circulation.

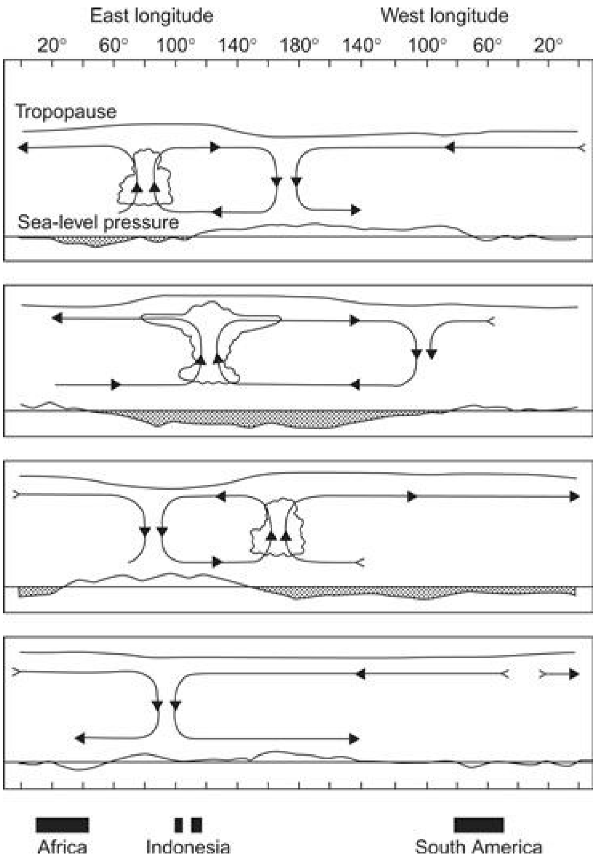


Figure 11.12 Longitude-height section of the anomaly pattern associated with the tropical intraseasonal oscillation (MJO). Reading downward, the panels represent a time sequence with intervals of about 10 days. *Streamlines* show the west–east circulation, the *wavy top line* represents the tropopause height, and the *bottom line* represents surface pressure (with *shading* showing below-normal surface pressure).

After Madden, 2003; adapted from Madden and Julian, 1972.

The oscillation originates with the development of a surface low-pressure anomaly over the Indian Ocean, accompanied by enhanced boundary layer moisture convergence, increased convection, warming of the troposphere, and raising of the tropopause height. The anomaly pattern gradually moves eastward at about $5 \mathrm{m s}^{-1}$ and reaches maximum intensity over the western Pacific. As the anomaly moves over the cooler waters of the central Pacific, the anomalous convection gradually weakens, although a circulation disturbance continues eastward and can sometimes be traced completely around the globe. The observed intraseasonal oscillation is known to be associated with equatorial Rossby and Kelvin waves (see Section 11.4). However, a completely satisfactory theory for the oscillation has not yet been developed.

11.2 Scale Analysis of Large-Scale Tropical Motions

Despite the uncertainties involved in the interaction between convective and synoptic scales, some information on the character of synoptic-scale motions in the tropics can be obtained through the methods of scale analysis. The scaling arguments can be carried out most conveniently if the governing equations are written in the log-pressure coordinate system introduced in Section 7.4.1:

$$\left(\frac{\partial}{\partial t} + \mathbf{V}\cdot\boldsymbol{\nabla} + w^*\frac{\partial}{\partial z^*}\right)\mathbf{V} + f\mathbf{k}\times\mathbf{V} = -\boldsymbol{\nabla}\Phi \tag{11.1}$$

$$\partial\Phi/\partial z^* = RT/H \tag{11.2}$$

$$\partial u/\partial x + \partial v/\partial y + \partial w^*/\partial z^* - w^*/H = 0 \tag{11.3}$$

$$\left(\frac{\partial}{\partial t} + \mathbf{V}\cdot\boldsymbol{\nabla}\right)T + w^* N^2 H/R = J/c_p \tag{11.4}$$

We wish to compare the magnitudes of the various terms in (11.1) through (11.4) for synoptic-scale motions in the tropics. We first note that an upper limit on the vertical velocity scale W is imposed by the continuity equation (11.3). Thus, following the discussion of Section 4.5,

$$\partial u/\partial x + \partial v/\partial y \leq U/L$$

However, for motions with vertical scales comparable to the density scale height H,

$$\partial w^*/\partial z^* - w^*/H \sim W/H$$

so that the vertical velocity scale must satisfy the constraint $W \leq HU/L$ if the horizontal divergence and the vertical stretching terms are to balance in the continuity equation. We next define characteristic scales for the various field variables as follows:

$H \sim 10^4$ m	vertical length scale
$L \sim 10^6$ m	horizontal length scale
$U \sim 10\,\mathrm{m\,s^{-1}}$	horizontal velocity scale
$W \leq HU/L$	vertical velocity scale
$\delta\Phi$	geopotential fluctuation scale
$L/U \sim 10^5$ s	timescale for advection

The magnitudes chosen for the horizontal length and velocity scales are typical for observed values in synoptic-scale systems both in the tropics and at midlatitudes. We now wish to show how the corresponding characteristic scales for vertical velocity and geopotential fluctuations are limited by the dynamic constraints imposed by conservation of mass, momentum, and thermodynamic energy.

We can estimate the magnitude of the geopotential fluctuation $\delta\Phi$ by scaling the terms in the momentum equation (11.1). For this purpose it is convenient to compare the magnitude of the horizontal inertial acceleration,

$$(\mathbf{V}\cdot\boldsymbol{\nabla})\mathbf{V} \sim U^2/L$$

with each of the other terms in (11.1) as follows:

$$|\partial\mathbf{V}/\partial t| \,/\, |(\mathbf{V}\cdot\boldsymbol{\nabla})\mathbf{V}| \sim 1 \tag{11.5}$$

$$|w^* \partial \mathbf{V}/\partial z^*| / |(\mathbf{V}\cdot\boldsymbol{\nabla})\mathbf{V}| \sim WL/UH \leq 1 \tag{11.6}$$

$$|f\mathbf{k}\times\mathbf{V}| / |(\mathbf{V}\cdot\boldsymbol{\nabla})\mathbf{V}| \sim fL/U = \mathrm{Ro}^{-1} \leq 1 \tag{11.7}$$

$$|\boldsymbol{\nabla}\Phi| / |(\mathbf{V}\cdot\boldsymbol{\nabla})\mathbf{V}| \sim \delta\Phi/U^2 \tag{11.8}$$

We have shown previously that in middle latitudes where $f \sim 10^{-4}\,\mathrm{s^{-1}}$, the Rossby number Ro is small so that to first approximation the Coriolis force and pressure gradient force terms balance. In that case, $\delta\Phi$ ~*fUL*. In the equatorial region, however, $f \leq 10^{-5}\,\mathrm{s^{-1}}$ and the Rossby number is of order unity or greater. Therefore, it is not appropriate to assume that the Coriolis force balances the pressure gradient force. In fact, (11.5) through (11.8) show that if the pressure gradient force is to be balanced in (11.1), the geopotential perturbation must scale as $\delta + \Phi \sim U^2 \sim 100\,\mathrm{m^2\,s^{-2}}$, and geopotential perturbations associated with equatorial synoptic-scale disturbances will be an order of magnitude smaller than those for midlatitude systems of similar scale.

This constraint on the amplitude of geopotential fluctuations in the tropics has profound consequences for the structure of synoptic-scale tropical motion systems. These consequences can be understood easily by applying scaling arguments to the thermodynamic energy equation. It is first necessary to obtain an estimate of the temperature fluctuations. The hydrostatic approximation (11.2) implies that for systems whose vertical scale is comparable to the scale height,

$$T = (H/R)\partial\Phi/\partial z^* \sim (\delta\Phi/R) \sim U^2/R \sim 0.3\,\mathrm{K} \tag{11.9}$$

Therefore, deep tropical systems are characterized by practically negligible synoptic-scale temperature fluctuations. Referring to the thermodynamic energy equation, we find that for such systems

$$\left(\frac{\partial}{\partial t} + \mathbf{V}\cdot\boldsymbol{\nabla}\right)T \sim 0.3\,\mathrm{K\,d^{-1}}$$

In the absence of precipitation, the diabatic heating is caused primarily by the emission of long wave radiation, which tends to cool the troposphere at a rate of $J/c_p \sim 1\,\mathrm{K\,d^{-1}}$. Because the actual temperature fluctuations are small, this radiative cooling must be approximately balanced by adiabatic warming due to subsidence. Thus, to a first approximation, (11.4) becomes a diagnostic relationship for w^*:

$$w^*(N^2H/R) = J/c_p \tag{11.10}$$

For the tropical troposphere, $N^2H/R \sim 3\,\mathrm{K\,km^{-1}}$, so the vertical motion scale must satisfy $W \sim 0.3\,\mathrm{cm\,s^{-1}}$ and $WL/UH \sim 0.03$ in (11.6). Therefore, in the absence of precipitation, the vertical motion is constrained to be even smaller than in extratropical synoptic systems of a similar scale. Not only can the vertical advection term be neglected in (11.1), but from the continuity equation (11.3), the divergence of the horizontal wind is $\sim 3\times 10^{-7}\,\mathrm{s^{-1}}$. Thus, the flow is nearly nondivergent.

The quasi-nondivergent nature of the flow in the absence of convective disturbances in the tropics makes it possible to simplify the governing equations for that situation. A theorem of

Helmholtz[3] states that any velocity field can be divided into a *nondivergent* part $\mathbf{V}_\psi$ and an *irrotational* part $\mathbf{V}_e$ such that

$$\mathbf{V} = \mathbf{V}_\psi + \mathbf{V}_e$$

where $\nabla \cdot \mathbf{V}_\psi = 0$ and $\nabla \times \mathbf{V}_e = 0$.

For a two-dimensional velocity field the nondivergent part can be expressed in terms of a *streamfunction* ψ defined by letting

$$\mathbf{V}_\psi = \mathbf{k} \times \nabla \psi \tag{11.11}$$

or in Cartesian components:

$$u_\psi = -\partial \psi / \partial y, \quad v_\psi = \partial \psi / \partial x$$

from which it is verified easily that $\nabla \cdot \mathbf{V}_\psi = 0$ and $\zeta = \mathbf{k} \cdot \nabla \times \mathbf{V}_\psi = \nabla^2 \psi$. Because the isolines of ψ correspond to streamlines for the nondivergent velocity and the distance separating the isolines of ψ is inversely proportional to the magnitude of the nondivergent velocity, the spatial distribution of $\mathbf{V}_\psi$ can be pictured easily by plotting lines of constant ψ on a synoptic chart.

We now approximate **V** by its nondivergent part $\mathbf{V}_\psi$ in (11.1) and neglect the small vertical advection term to obtain an approximate momentum equation valid for synoptic-scale motions in the tropics outside regions of precipitation:

$$\frac{\partial \mathbf{V}_\psi}{\partial t} + (\mathbf{V}_\psi \cdot \nabla)\mathbf{V}_\psi + f\mathbf{k} \times \mathbf{V}_\psi = -\nabla \Phi \tag{11.12}$$

Using the vector identity

$$(\mathbf{V} \cdot \nabla)\mathbf{V} = \nabla \left(\frac{\mathbf{V} \cdot \mathbf{V}}{2} \right) + \mathbf{k} \times \mathbf{V}\zeta$$

we can rewrite (11.12) as

$$\frac{\partial \mathbf{V}_\psi}{\partial t} = -\nabla \left(\Phi + \frac{\mathbf{V}_\psi \cdot \mathbf{V}_\psi}{2} \right) - \mathbf{k} \times \mathbf{V}_\psi (\zeta + f) \tag{11.13}$$

We next take $\mathbf{k} \cdot \nabla \times$ (11.13) to obtain the vorticity equation valid for nondivergent flow:

$$\left(\frac{\partial}{\partial t} + \mathbf{V}_\psi \cdot \nabla \right) (\zeta + f) = 0 \tag{11.14}$$

This equation shows that in the absence of condensation heating, synoptic-scale circulations in the tropics in which the vertical scale is comparable to the scale height of the atmosphere must be barotropic; absolute vorticity is conserved following the nondivergent horizontal wind. Such disturbances cannot convert potential energy to kinetic energy. They must be driven by barotropic conversion of mean-flow kinetic energy or by lateral coupling either to midlatitude systems or to precipitating tropical disturbances.

Because both the nondivergent velocity and the vorticity can be expressed in terms of the streamfunction, (11.14) requires only the field of ψ at any level in order to make a prediction. The pressure distribution is neither required nor predicted. Rather, it must be determined diagnostically. The relationship of the pressure and streamfunction fields can be obtained by taking $\nabla \cdot$ (11.13). This yields a diagnostic relationship between the geopotential and streamfunction fields, which is usually referred to as the *nonlinear balance equation*:

$$\nabla^2\left[\Phi + \frac{1}{2}(\nabla\psi)^2\right] = \nabla \cdot \left[\left(f + \nabla^2\psi\right)\nabla\psi\right] \tag{11.15}$$

For the special case of stationary circularly symmetric flow, (11.15) is equivalent to the gradient wind approximation. Unlike the gradient wind, however, the balance in (11.15) does not require information on trajectory curvature and thus can be solved for Φ from knowledge of the instantaneous distribution of ψ on an isobaric surface. Alternatively, if the Φ distribution is known, (11.15) can be solved for ψ. In this case the equation is quadratic so that there are generally two possible solutions, which correspond to the normal and anomalous gradient wind cases.

Such a balance condition is valid only when the previous scaling arguments apply. These, however, have been based on the assumptions that the depth scale is comparable to the scale height of the atmosphere and that the horizontal scale is of order 1000 km. There is a special class of planetary-scale motions for which the divergence term in the vorticity equation is important even outside of regions of active precipitation (see Section 11.4). For such motions the pressure field cannot be diagnosed from a balance relationship. Rather, the pressure distribution must be predicted from the primitive equation form of the dynamics equations.

For precipitating synoptic-scale systems in the tropics, the preceding scaling considerations require extensive modification. Precipitation rates in such systems are typically of the order 2 cm d^{-1}. This implies condensation of $m_w = 20$ kg water for an atmospheric column of 1 m^2 cross-section. Because the latent heat of condensation is $L_c \approx 2.5 \times 10^6\ \mathrm{J\,kg^{-1}}$, this precipitation rate implies an addition of heat energy to the atmospheric column of

$$m_w L_c \sim 5 \times 10^7\ \mathrm{J\,m^{-2}\,d^{-1}}$$

If this heat is distributed uniformly over the entire atmospheric column of mass $p_0/g \approx 10^4\ \mathrm{kg\,m^{-2}}$, then the average heating rate per unit mass of air is

$$J/c_p \approx \left[L_c m_w / c_p\left(p_0/g\right)\right] \sim 5\ \mathrm{K\,d^{-1}}$$

In reality the condensation heating due to deep convective clouds is not distributed evenly over the entire vertical column, but is a maximum between 300 and 400 hPa, where the heating rate can be as high as 10 K d^{-1}. In this case the approximate thermodynamic energy equation (11.10) implies that the vertical motion on the synoptic scale in precipitating systems must have a magnitude of order $W \sim 3\mathrm{cm\ s^{-1}}$ in order that the adiabatic cooling can balance the condensation heating in the 300- to 400-hPa layer. Therefore, the average vertical motion in precipitating disturbances in the tropics is an order of magnitude larger than the vertical motion outside the disturbances. As a result the flow in these disturbances has a relatively large divergent component so that the barotropic vorticity equation (11.14) is no longer a reasonable approximation, and the full primitive equations must be used to analyze the flow.

11.3 Condensation Heating

The manner in which the atmosphere is heated by condensation of water vapor depends crucially on the nature of the condensation process. In particular, it is necessary to differentiate between latent heat release through large-scale vertical motion (i.e., synoptic-scale forced uplift) and the latent heat release due to deep cumulus convection. The former process, which is generally associated with midlatitude synoptic systems, can be incorporated easily into the thermodynamic energy equation in terms of the synoptic-scale field variables. The large-scale heating field resulting from the cooperative action of many cumulonimbus cells, however, requires representation of this type of latent heating in terms of the synoptic-scale field variables, which is much more difficult.

Before considering the problem of condensation heating by cumulus convection, it is worth indicating briefly how the condensation heating by large-scale forced uplift can be included in a prediction model. The approximate thermodynamic energy equation for a pseudoadiabatic process given in Section 2.9.1 states that

$$\frac{D\ln\theta}{Dt} \approx -\left(\frac{L_c}{c_pT}\right)\frac{Dq_s}{Dt} \tag{11.16}$$

It may be shown that an *equivalent static stability*, Γ_e, may be defined so that (11.16) can be written as

$$\left(\frac{\partial}{\partial t}+\mathbf{V}\cdot\nabla\right)\theta + w\Gamma_e \approx 0 \tag{11.17}$$

where

$$\Gamma_e \approx \begin{cases} (\Gamma_s/\Gamma_d)\,\partial\theta_e/\partial z & \text{for } q \geq q_s \text{ and } w > 0 \\ \partial\theta/\partial z & \text{for } q < q_s \text{ or } w < 0 \end{cases}$$

Here, Γ_s and Γ_d are the lapse rates for pseudoadiabatic and dry ascent, respectively. Thus, in the case of condensation due to large-scale forced ascent ($\Gamma_e > 0$), the thermodynamic energy equation has essentially the same form as for adiabatic motions except that the static stability is replaced by the equivalent static stability. As a consequence, the local temperature changes induced by the forced ascent will be smaller than for the case of forced dry ascent with the same lapse rate.

If, however, $\Gamma_e < 0$, the atmosphere is conditionally unstable and condensation will occur primarily through cumulus convection. In that case, the vertical velocity must be that of the individual cumulus updrafts, not the synoptic-scale w. Thus, a simple formulation of the thermodynamic energy equation in terms of only the synoptic-scale variables is not possible. We can still, however, simplify the thermodynamic energy equation to some extent. We recall from Section 11.2—see Eq. (11.10)—that due to the smallness of temperature fluctuations in the tropics, the adiabatic cooling and diabatic heating terms must approximately balance. Thus, (11.16) becomes approximately

$$w\frac{\partial \ln\theta}{\partial z} \approx -\frac{L_c}{c_pT}\frac{Dq_s}{Dt} \tag{11.18}$$

The synoptic-scale vertical velocity w that appears in (11.18) is the average of very large vertical motions in the active convection cells and small vertical motions in the environment. Thus, if we let w' be the vertical velocity in the convective cells and the $\overline{w}$ be vertical velocity in the environment, we have

$$w = aw' + (1-a)\overline{w} \tag{11.19}$$

where a is the fractional area occupied by the convection. Approximating the change in q_s following the motion by $w\frac{\partial q_s}{\partial z}$, we can then write (11.18) in the form

$$w\frac{\partial \ln\theta}{\partial z} \approx -\frac{L_c}{c_pT}aw'\frac{\partial q_s}{\partial z} \tag{11.20}$$

The problem now is to express the condensation heating term on the right in (11.20) in terms of synoptic-scale field variables.

This problem of *parameterizing* the cumulus convective heating is one of the most challenging areas in tropical meteorology. A simple approach that has been used successfully in some theoretical studies[4] is based on the fact that, because the storage of water in the clouds is rather small, the total vertically integrated heating rate due to condensation must be approximately proportional to the net precipitation rate:

$$-\int_{z_c}^{z_T} (\rho a w' \partial q_s/\partial z)\, dz = P \tag{11.21}$$

where z_c and z_T represent the cloud base and cloud top heights, respectively, and P is the precipitation rate (kg m^{-2} s^{-1}).

Because relatively little moisture goes into changing the atmospheric vapor mixing ratio, the net precipitation rate must approximately equal the moisture convergence into an atmospheric column plus surface evaporation:

$$P = -\int_0^{z_m} \boldsymbol{\nabla}\cdot(\rho q \mathbf{V})dz + E \tag{11.22}$$

where E is the evaporation rate (kg m^{-2} s^{-1}) and z_m is the top of the moist layer ($z_m \approx 2$ km over much of the equatorial oceans). Substituting into (11.22) from the approximate continuity equation for q,

$$\boldsymbol{\nabla}\cdot(\rho q\mathbf{V}) + \partial(\rho q w)/\partial z \approx 0 \tag{11.23}$$

we obtain

$$P = (\rho w q)_{z_m} + E \tag{11.24}$$

Using (11.24) we can relate the vertically averaged heating rate to the synoptic-scale variables $w(z_m)$ and $q(z_m)$.

We still, however, need to determine distribution of the heating in the vertical. The most common approach is to use an empirically determined vertical distribution based on observations. In that case, equation (11.16) can be written as

$$\left(\frac{\partial}{\partial t}+\mathbf{V}\cdot\boldsymbol{\nabla}\right)\ln\theta+w\frac{\partial\ln\theta}{\partial z}=\frac{L_c}{\rho c_p T}\eta\left(z\right)\left[\left(\rho w q\right)_{z_m}+E\right] \tag{11.25}$$

where $\eta(z)=0$ for $z<z_c$ and $z>z_T$ and $\eta(z)$ for $z_c\leq z\leq z_T$ is a weighting function that must satisfy

$$\int_{z_c}^{z_T}\eta(z)dz=1$$

Recalling that the diabatic heating must be approximately balanced by adiabatic cooling as indicated in (11.20), we see from (11.25) that $\eta(z)$ will have a vertical structure similar to that of the large-scale vertical mass flux, ρw. Observations indicate that for many tropical synoptic-scale disturbances, $\eta(z)$ reaches its maximum at about the 400-hPa level, consistent with the divergence pattern shown earlier in Figure 11.5.

The preceding formulation is designed to model average tropical conditions. In reality, the vertical distribution of diabatic heating is determined by the local distribution of cloud heights. Thus, the cloud height distribution is apparently a key parameter in cumulus parameterization. A cumulus parameterization scheme in which this distribution is determined in terms of the large-scale variables was developed by Arakawa and Schubert (1974). A number of other schemes have been suggested in the past decade. Discussion of these is beyond the scope of this text.

11.4 Equatorial Wave Theory

Equatorial waves are an important class of eastward- and westward-propagating disturbances in the atmosphere and in the ocean that are trapped about the equator (i.e., they decay away from the equatorial region). Diabatic heating by organized tropical convection can excite atmospheric equatorial waves, whereas wind stresses can excite oceanic equatorial waves. Atmospheric equatorial wave propagation can cause the effects of convective storms to be communicated over large longitudinal distances, thus producing remote responses to localized heat sources. Furthermore, by influencing the pattern of low-level moisture convergence, atmospheric equatorial waves can partly control the spatial and the temporal distribution of convective heating. Oceanic equatorial wave propagation, however, can cause local wind stress anomalies to remotely influence the thermocline depth and the sea surface temperature, as was discussed in Section 11.1.6.

11.4.1 Equatorial Rossby and Rossby–Gravity Modes

A complete development of equatorial wave theory would be rather complicated. In order to introduce equatorial waves in the simplest possible context, we here use a shallow water model, analogous to that introduced in Section 4.5, and concentrate on the *horizontal* structure. Vertical propagation in a stratified atmosphere is discussed in Chapter 12. For simplicity, we consider the linearized momentum and continuity equations for a fluid system of mean depth h_e in a motionless basic state. Because we are interested only in the tropics, we utilize Cartesian geometry on an *equatorial β plane* Section 5.7. In this approximation, terms proportional to cos ϕ are replaced by unity, and terms proportional to sin ϕ are replaced by y/a, where y is the distance from the equator and a is the radius of Earth. The Coriolis parameter in this approximation is given by

$$f \approx \beta y \tag{11.26}$$

where $\beta \equiv 2\Omega/a$, and Ω is the angular velocity of Earth. The resulting linearized shallow water equations (Section 4.5) for perturbations on a motionless basic state of mean depth h_e may be written as

$$\partial u'/\partial t - \beta y v' = -\partial \Phi' \partial x \tag{11.27}$$

$$\partial v'/\partial t + \beta y u' = -\partial \Phi' \partial y \tag{11.28}$$

$$\partial \Phi'/\partial t + g h_e \left(\partial u'/\partial x + \partial v'/\partial y\right) = 0 \tag{11.29}$$

where $\Phi' = gh'$ is the geopotential disturbance, and primed variables designate perturbation fields.

The x and t dependence may be separated by specifying solutions in the form of zonally propagating waves:

$$\begin{pmatrix} u' \\ v' \\ \Phi' \end{pmatrix} = \begin{bmatrix} \hat{u}(y) \\ \hat{v}(y) \\ \hat{\Phi}(y) \end{bmatrix} \exp[i(kx - \nu t)] \tag{11.30}$$

Substitution of (11.30) into (11.27), (11.28), and (11.29) then yields a set of ordinary differential equations in y for the meridional structure functions $\hat{u}, \hat{v}, \hat{\Phi}$:

$$-i\nu\hat{u} - \beta y \hat{v} = -ik\hat{\Phi} \tag{11.31}$$

$$-i\nu\hat{v} + \beta y \hat{u} = -\partial\hat{\Phi}/\partial y \tag{11.32}$$

$$-i\nu\hat{\Phi} + g h_e \left(ik\hat{u} + \partial\hat{v}/\partial y\right) = 0 \tag{11.33}$$

If (11.31) is solved for $\hat{u}$ and the result substituted into (11.32) and (11.33), we obtain

$$\left(\beta^2 y^2 - \nu^2\right)\hat{v} = ik\beta y\hat{\Phi} + i\nu\partial\hat{\Phi}/\partial y \tag{11.34}$$

$$\left(\nu^2 - g h_e k^2\right)\hat{\Phi} + i\nu g h_e \left(\frac{\partial\hat{v}}{\partial y} - \frac{k}{\nu}\beta y \hat{v}\right) = 0 \tag{11.35}$$

Finally, (11.35) can be substituted into (11.34) to eliminate $\hat{\Phi}$, yielding a second-order differential equation in the single unknown, $\hat{v}$:

$$\frac{\partial^2 \hat{v}}{\partial y^2} + \left[\left(\frac{\nu^2}{gh_e} - k^2 - \frac{k}{\nu}\beta\right) - \frac{\beta^2 y^2}{gh_e}\right]\hat{v} = 0 \tag{11.36}$$

Because (11.36) is homogeneous, we expect that nontrivial solutions satisfying the requirement of decay at large $|y|$ will exist only for certain values of ν, corresponding to frequencies of the normal mode disturbances.

Before discussing this equation in detail, it is worth considering the asymptotic limits that occur when either $h_e \to \infty$ or $\beta = 0$. In the former case, which is equivalent to assuming that the motion is nondivergent, (11.36) reduces to

$$\frac{\partial^2 \hat{v}}{\partial y^2} + \left[-k^2 - \frac{k}{\nu}\beta\right]\hat{v} = 0$$

Solutions exist of the form $\hat{v} \sim \exp(ily)$, provided that ν satisfies the Rossby wave dispersion relationship, $\nu = -\beta k/(k^2 + l^2)$. This illustrates that for nondivergent barotropic flow, equatorial dynamics is in no way special. Earth's rotation enters only in the form of the β effect; it is not dependent on f. However, if $\beta = 0$, all influence of rotation is eliminated and (11.36) reduces to the shallow water gravity wave model, which has nontrivial solutions for

$$\nu = \pm\left[gh_e\left(k^2 + l^2\right)\right]^{1/2}$$

Returning to (11.36), we seek solutions for the meridional distribution of $\hat{v}$, subject to the boundary condition that the disturbance fields vanish for $|y| \to \infty$. This boundary condition is necessary because the approximation $f \approx \beta y$ is not valid for latitudes much beyond $\pm 30°$, so that solutions must be trapped equatorially if they are to be good approximations to the exact solutions on the sphere. Equation (11.36) differs from the classic equation for a harmonic oscillator in y because the coefficient in square brackets is not a constant but is a function of y. For sufficiently small y, this coefficient is positive and solutions oscillate in y, whereas for large y, solutions either grow or decay in y. Only the decaying solutions, however, can satisfy the boundary conditions.

It turns out[5] that solutions to (11.36), which satisfy the condition of decay far from the equator, exist only when the constant part of the coefficient in square brackets satisfies the relationship

$$\frac{\sqrt{gh_e}}{\beta}\left(-\frac{k}{\nu}\beta - k^2 + \frac{\nu^2}{gh_e}\right) = 2n + 1; \quad n = 0,1,2,\ldots \tag{11.37}$$

which is a cubic dispersion equation determining the frequencies of permitted equatorially trapped free oscillations for zonal wave number k and meridional mode number n. These solutions can be expressed most conveniently if y is replaced by the nondimensional meridional coordinate

$$\xi \equiv \left(\beta\big/\sqrt{gh_e}\right)^{1/2} y$$

Then the solution has the form

$$\hat{v}(\xi) = v_0 \mathrm{H}_n(\xi)\exp\left(-\xi^2/2\right) \tag{11.38}$$

where v_0 is a constant with velocity units, and $H_n(\xi)$ designates the nth *Hermite polynomial.* The first few of these polynomials have the values

$$H_0 = 1, \quad H_1(\xi) = 2\xi, \quad H_2(\xi) = 4\xi^2 - 2$$

Thus, the index n corresponds to the number of nodes in the meridional velocity profile in the domain $|y| < \infty$.

In general the three solutions of (11.37) can be interpreted as eastward- and westward-moving equatorially trapped gravity waves, and westward-moving equatorial Rossby waves. The case $n = 0$ (for which the meridional velocity perturbation has a Gaussian distribution centered at the equator) must be treated separately. In this case the dispersion relationship (11.37) factors as

$$\left(\frac{\nu}{\sqrt{gh_e}} - \frac{\beta}{\nu} - k\right)\left(\frac{\nu}{\sqrt{gh_e}} + k\right) = 0 \tag{11.39}$$

The root $\nu/k = -\sqrt{gh_e}$, corresponding to a westward-propagating gravity wave, is not permitted, as the second term in parentheses in (11.39) was implicitly assumed not to vanish when (11.34) and (11.35) were combined to eliminate $\hat{\Phi}$. The roots given by the first term in parentheses in (11.39) are

$$\nu = k\sqrt{gh_e}\left[\frac{1}{2} \pm \frac{1}{2}\left(1 + \frac{4\beta}{k^2\sqrt{gh_e}}\right)^{1/2}\right] \tag{11.40}$$

The positive root corresponds to an eastward-propagating equatorial inertia–gravity wave, whereas the negative root corresponds to a westward-propagating wave, which resembles an inertia–gravity wave for long zonal scale ($k \rightarrow 0$) and resembles a Rossby wave for zonal scales characteristic of synoptic-scale disturbances. This mode is generally referred to as a *Rossby–gravity wave*. The horizontal structure of the westward-propagating $n = 0$ solution is shown in Figure 11.13, whereas the relationship between frequency and zonal wave number for this and several other equatorial wave modes is shown in Figure 11.14.

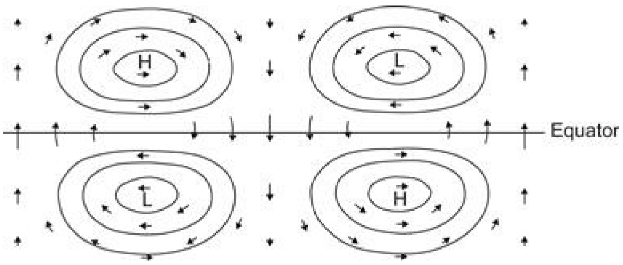


Figure 11.13 Plan view of horizontal velocity and height perturbations associated with an equatorial Rossby–gravity wave.

Adapted from Matsuno, 1966. Used with permission of the Japan Society for the Promotion of Science.

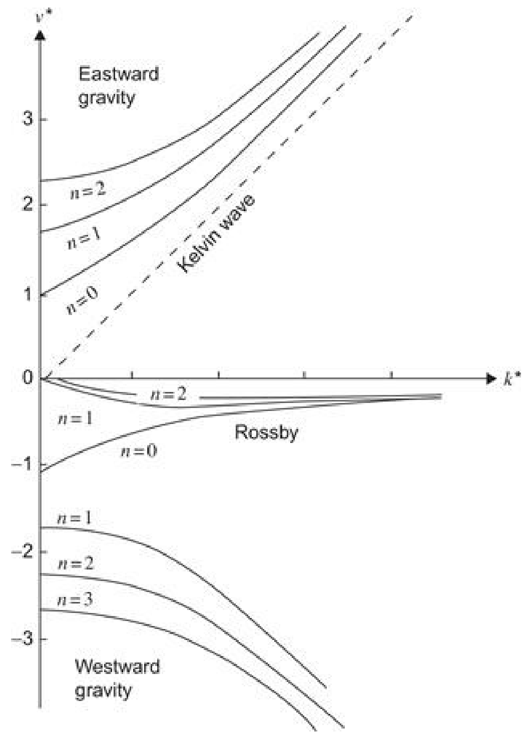


Figure 11.14 Dispersion diagram for free equatorial waves. Frequency and zonal wave numbers have been nondimensionalized by defining $\nu^* \equiv \nu/\left(\beta\sqrt{gh_e}\right)^{1/2}$ $k^* \equiv k\left(\sqrt{gh_e}/\beta\right)^{1/2}$. *Curves* show dependence of frequency on zonal wave number for eastward and westward gravity modes and for Rossby and Kelvin modes. (k^* axis tic marks at unit interval with 0 on left.)

11.4.2 Equatorial Kelvin Waves

In addition to the modes discussed in the previous section, there is another equatorial wave that is of great practical importance. For this mode, which is called the equatorial *Kelvin wave*, the meridional velocity perturbation vanishes and (11.31), (11.32), and (11.33) are reduced to the simpler set

$$-i\nu\hat{u} = -ik\hat{\Phi} \tag{11.41}$$

$$\beta y\hat{u} = -\partial\hat{\Phi}/\partial y \tag{11.42}$$

$$-i\nu\hat{\Phi} + gh_e\left(ik\hat{u}\right) = 0 \tag{11.43}$$

Combining (11.41) and (11.43) to eliminate $\hat{\Phi}$, we see that the Kelvin wave dispersion equation is identical to that for ordinary shallow water gravity waves:

$$c^2 \equiv (\nu/k)^2 = gh_e \tag{11.44}$$

According to (11.44), the phase speed *c* can be either positive or negative. However, if (11.41) and (11.42) are combined to eliminate $\hat{\Phi}$, we obtain a first-order equation for determining the meridional structure:

$$\beta y\hat{u} = -c\partial\hat{u}/\partial y \tag{11.45}$$

which may be integrated immediately to yield

$$\hat{u} = u_o \exp\left(-\beta y^2/2c\right) \tag{11.46}$$

where u_0 is the amplitude of the perturbation zonal velocity at the equator. Equation (11.46) shows that if solutions decaying away from the equator are to exist, the phase speed must be positive ($c > 0$). Thus, Kelvin waves are eastward propagating and have zonal velocity and geopotential perturbations that vary in latitude as Gaussian functions centered on the equator. The *e*-folding decay width is given by

$$Y_K = |2c/\beta|^{1/2}$$

which for a phase speed $c = 30 \text{ m s}^{-1} Y_K \approx 1600$ km.

The perturbation wind and geopotential structure for the Kelvin wave are shown in plan view in Figure 11.15. In the zonal direction the force balance is exactly that of an eastward-propagating shallow water gravity wave. A vertical section along the equator would thus be the same as that shown in Figure 5.10. The meridional force balance for the Kelvin mode is an exact geostrophic balance between the zonal velocity and the meridional pressure gradient. It is the change in sign of the Coriolis parameter at the equator that permits this special type of equatorial mode to exist.

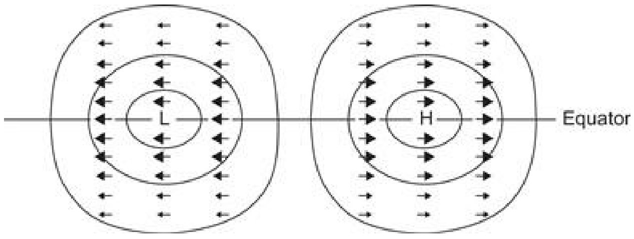


Figure 11.15 Plan view of horizontal velocity and height perturbations that are associated with an equatorial Kelvin wave.

Adapted from Matsuno, 1966.

11.5 Steady Forced Equatorial Motions

Not all zonally asymmetric circulations in the tropics can be explained on the basis of inviscid equatorial wave theory. For quasi-steady circulations the zonal pressure gradient force must be balanced by turbulent drag rather than by inertia. The Walker circulation may be regarded as a quasi-steady equa\-torially trapped circulation that is generated by diabatic heating. The simplest models of such circulations specify the diabatic heating and use the equations of equatorial wave theory to compute the atmospheric response. These models, however, neglect the fact that the distribution of diabatic heating is highly dependent on the mean wind and equivalent potential temperature distributions in the boundary layer. These in turn depend on the surface pressure and the moisture distribution, which are themselves dependent on the motion field. Thus, in a consistent model the diabatic heating cannot be regarded as an externally specified quantity, but must be obtained as part of the solution by, for example, using the cumulus parameterization scheme illustrated in Section 11.3.

As indicated in (11.25), this scheme requires information on the vertical distribution of convective heating in order to solve for the temperature perturbation throughout the troposphere. There is, however, evidence that the essential features of stationary equatorial circulations can be partly explained on the basis of a model that involves only the boundary layer. This is perhaps not surprising since the maintenance of a convective system depends on evaporation and moisture convergence in the boundary layer. Over the tropical oceans the boundary layer can be approximated as a mixed layer of about 2-km depth. This is capped by an inversion across which there is a density discontinuity $\delta\rho$ (see Figure 8.2). The virtual temperature in the mixed layer is strongly correlated with the sea surface temperature. If we assume that the pressure field is uniform at the top of the mixed layer, the surface pressure will be determined by hydrostatic mass adjustment within the mixed layer. The resulting pressure perturbation in the layer depends on the density discontinuity at the top of the layer and the deviation, h, of the layer depth from its mean depth H_b. In addition, there will be a contribution from the perturbation virtual potential temperature, θ_v, within the mixed layer. Thus, the perturbation geopotential in the mixed layer can be expressed as

$$\Phi = g\left(\delta\rho/\rho_0\right)h - \Gamma\theta_v \tag{11.47}$$

where $\Gamma \equiv (gH_b/\theta_0)$ is a constant and ρ_0 and θ_0 are constant mixed layer reference values of density and potential temperature, respectively.

According to (11.47), positive sea surface temperature anomalies and negative boundary layer height anomalies will produce low surface pressures and vice versa. If the boundary layer depth does not vary too much, the surface pressure gradient will thus tend to be proportional to the sea surface temperature gradient. The dynamics of steady circulations in such a mixed layer can be approximated by a set of linear equations analogous to the equatorial wave equations (11.27), (11.28), and (11.29), but with the time derivative terms replaced by linear damping terms.

Thus, in the momentum equations the surface eddy stress is taken to be proportional to the mean velocity in the mixed layer. In the continuity equation the perturbation in the mixed layer height is proportional to the mass convergence in the layer, with a coefficient that is smaller in the presence of convection than in its absence because of ventilation of the boundary layer by convection. The x and y components of the vertically averaged momentum equation for steady motions in the mixed layer can then be written as

$$\alpha u - \beta y v + \partial\Phi/\partial x = 0 \tag{11.48}$$

$$\alpha v + \beta y u + \partial\Phi/\partial y = 0 \tag{11.49}$$

whereas the continuity equation can be expressed as

$$\alpha h + H_b\left(1-\varepsilon\right)\left(\partial u/\partial x + \partial v/\partial y\right) = 0 \tag{11.50}$$

where ε is a coefficient that is zero in the absence of convection and of order 3/4 when convection occurs. Substituting from (11.47) into (11.50) yields

$$\alpha\Phi + c_b^2(1-\varepsilon)(\partial u/\partial x + \partial v/\partial y) = -\alpha\Gamma\theta_v \tag{11.51}$$

where $c_b^2 \equiv g(\delta\rho/\rho_0)H_b$ is the square of the phase speed for gravity waves propagating along the inversion at the top of the mixed layer. See Battisti et al. (1999) for a more complete discussion.

Equations (11.48), (11.49), and (11.51) are a closed set for prediction of the boundary layer variables u, v, Φ for a specified boundary layer perturbation virtual temperature θ_v. However, because the parameter ε depends on the presence or absence of convection, the system can only be solved by iteration through using (11.22) to test for the presence of convection. This model can be used to compute the steady surface circulation. A sample calculation of the anomalous circulation for temperature anomalies corresponding to a typical ENSO event is shown in Figure 11.16. Note that the region of convergence is narrower than the region of warm SST anomalies due to the convective feedback in the boundary layer model.

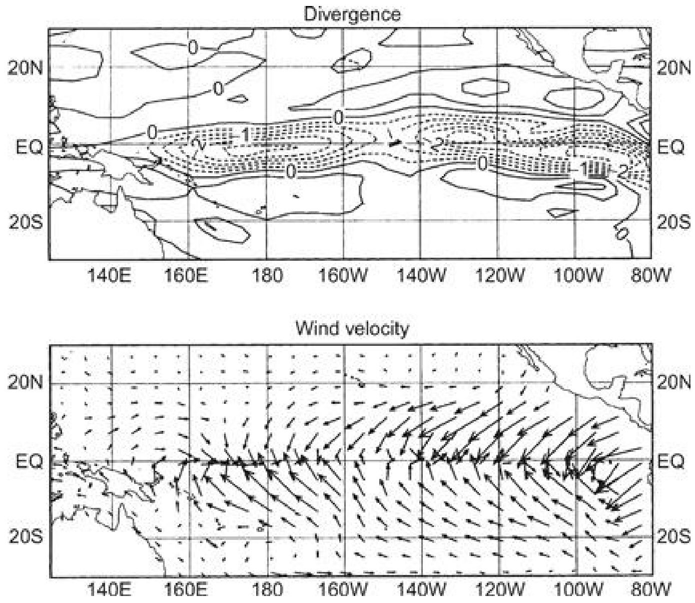


Figure 11.16 Steady surface circulation forced by sea surface temperature anomalies characteristic of El Niño in the equatorial Pacific.

Courtesy of D. Battisti.

Suggested References

hilander. Philander, *El Niño, La Niña, and the Southern Oscillation*, is a well-written text that covers the dynamics of both atmospheric and oceanographic aspects of ENSO at an advanced level.

renberth, 1991. Trenberth, 1991 is an excellent review article on ENSO.

Vallace, 1971. Wallace, 1971 is a detailed review of the structure of synoptic-scale tropospheric wave disturbances in the equatorial Pacific.

Vebster and Fasullo, 2003. Webster and Fasullo, 2003 present an excellent review of the structure and dynamics of tropical monsoons.

Problems

11.1 Suppose that the relative vorticity at the top of an Ekman layer at 15°N is $\zeta = 2 \times 10^{-5}\,\mathrm{s}^{-1}$. Let the eddy viscosity coefficient be $K_m = 10\,\mathrm{m^2\,s^{-1}}$ and the water vapor mixing ratio at the top of the Ekman layer be 12gkg $^{-1}$. Use the method of Section 11.3 to estimate the precipitation rate due to moisture convergence in the Ekman layer.

11.2 As mentioned in Section 11.1.3, barotropic instability is a possible energy source for some equatorial disturbances. Consider the following profile for an easterly jet near the equator:

$$\bar{u}\left(y\right) = -u_0 \sin^2\left[l\left(y - y_0\right)\right]$$

where u_0, y_0, and l are constants and y is the distance from the equator. Determine the necessary conditions for this profile to be barotropically unstable.

11.3 Show that the nonlinear terms in the balance equation (11.15)

$$G\left(x, y\right) \equiv -\nabla^2\left(\frac{1}{2}\boldsymbol{\nabla}\psi\cdot\boldsymbol{\nabla}\psi\right) + \boldsymbol{\nabla}\cdot\left(\boldsymbol{\nabla}\psi\nabla^2\psi\right)$$

may be written in Cartesian coordinates as

$$G\left(x, y\right) = 2\left[\left(\partial^2\psi/\partial x^2\right)\left(\partial^2\psi/\partial y^2\right) - \left(\partial^2\psi/\partial x\partial y\right)^2\right]$$

11.4 ith the aid of the results of Problem 11.3, show that if f is assumed to be constant the balance equation (11.15) is equivalent to the gradient wind equation equation (3.15) for a circularly symmetric regular low with geopotential perturbation given by

$$\Phi = \Phi_0\left(x^2 + y^2\right)\Big/L^2$$

where Φ_0 is a constant geopotential and L a constant length scale. *Hint:* Assume that $\psi(x, y)$ has the same functional dependence on (x, y) as *does* Φ.

11.5 Starting from the perturbation equations (11.27), (11.28) and (11.29), show that the sum of kinetic plus available potential energy is conserved for equatorial waves. Hence, show that for the Kelvin wave there is an equipartition of energy between kinetic and available potential energy.

11.6 Solve for the meridional dependence of the zonal wind and geopotential perturbations for a Rossby–gravity mode in terms of the meridional velocity distribution (11.38).

11.7 Use the linearized model (11.48) and (11.49) to compute the meridional distribution of divergence in the mixed layer for a situation in which the geostrophic wind is given by $u_g = u_o \exp\left(-\beta y^2/2c\right),\ v_g = 0$, where u_0 and c are constants.

11.8 Show that the frequency of the $n = 1$ equatorial Rossby mode is given approximately by $\nu = -k\beta\left(k^2 + 3\beta/\sqrt{gh_e}\right)^{-1}$ and use this result to solve for the $\hat{u}\left(y\right)$ and $\hat{\Phi}(y)$ fields in terms of $\hat{v}(y)$. *Hint:* Use the fact that the Rossby wave phase speed is much less than $\sqrt{gh_e}$.

MATLAB Exercises

M11.1 The MATLAB script **profile_2.m** and the function **Tmoist.m** use December through March seasonal mean pressure and temperature data for Samoa (14°S, 171°W) in the file

tropical_temp.dat, together with an assumed constant relative humidity of 80%, to calculate and graph vertical profiles of temperature, dewpoint, and the temperature corresponding to pseudoadiabatic ascent from the lifting condensation level for a parcel lifted from the lowest level of the sounding. Modify this script to plot profiles of the potential temperature, the equivalent potential temperature, and the saturation equivalent potential temperature. *Hint:* See Problem M9.2.

M11.2 For the thermodynamic sounding of Exercise M11.1, compute the CAPE and the profile of vertical velocity for a parcel lifted from the lowest layer in the absence of entrainment. What is the maximum vertical velocity and what is the distance that the parcel overshoots its level of neutral buoyancy? What can you conclude about convection associated with this mean tropical sounding?

M11.3 Consider a simple nondivergent flow given by the streamfunction $\psi = A \sin kx \sin ly$. The MATLAB script **nonlinear_balance.m** solves the nonlinear balance equation (11.15) for the corresponding geopotential Φ, assuming that the Coriolis parameter is a constant corresponding to its value at 30°N. Run this script for several values of the amplitude, *A*, in the range $0.4 \times 10^7 \text{ to } 4.0 \times 10^7 \,\mathrm{m^2\,s^{-1}}$. Note how the geopotential field depends on *A* For the case $A = 4.0 \times 10^7 \mathrm{m^2 s^{-1}}$ use the **gradient** function in MATLAB to compute the geostrophic wind and thus find the ageostrophic wind. Plot these as quiver plots and explain why the ageostrophic wind has the structure seen in your plot.

M11.4 A simple model for the horizontal flow in the mixed layer on an equatorial β plane forced by an equatorial wave mode is given by the following equations:

$$\frac{\partial u}{\partial t} = -\alpha u + \beta y v - \frac{\partial \Phi}{\partial x}$$

$$\frac{\partial v}{\partial t} = -\alpha v - \beta y u - \frac{\partial \Phi}{\partial y}$$

where $\Phi(x, y, t)$ is given by (11.30) for the appropriate mode. The MATLAB script **equatorial_mixed_layer.m** solves these equations numerically for the horizontal velocity components and divergence field in the mixed layer forced by a specified Rossby–gravity wave ($n = 0$ mode) geopotential perturbation corresponding to a westward-propagating disturbance of zonal wavelength 4000 km and $\sqrt{gh_e} = 18 \,\mathrm{m\,s^{-1}}$. Run the code for a long enough time so that the solution becomes periodic in time. Rerun the model for zonal wavelength 10,000 km and $\sqrt{gh_e} = 36 \,\mathrm{m\,s^{-1}}$. In each case compare the frequency of the oscillation with the Coriolis parameter at the latitude of maximum convergence forced by the wave. What can you conclude from these results?

M11.5 Modify the MATLAB script of M11.4 to compute the mixed layer velocity and divergence pattern for the $n = 1$ Rossby mode for zonal wavelength 4000 km and $\sqrt{gh_e} = 18 \,\mathrm{m\,s^{-1}}$ using the formulas for frequency and geopotential derived in Problem 11.8.

M11.6 The MATLAB script **forced_equatorial_mode2.m** shows the time development of the velocity and height perturbations produced by a transient localized mass source for shallow water waves on an equatorial β plane. The model is based on equations (11.48), (11.49), and (11.51), but with time tendency terms included ($\varepsilon = 0$ is set here, but you can experiment with different values if you wish). The source is located at the equator at $x = 0$. It rises smoothly in amplitude for the first 2.5

days and then decreases to 0 at day 5. Modify the script to contour the divergence and the vorticity. Run the script for a 10-day period and interpret the results in terms of equatorial wave theory by noting the structure and speeds of energy propagation for the disturbances to the east and west of the source.

---

[1] The terms *trough* and *ridge* as used by tropical meteorologists designate pressure minima and maxima, respectively, just as in midlatitudes. In the easterlies of the Northern Hemisphere tropics, however, the zonal-mean pressure increases with latitude so that the pattern of isobars depicting a tropical trough will resemble the pattern associated with a ridge in middle latitudes (i.e., there is a poleward deflection of the isobars).

[2] It should be noted here that the profile shown in Figure 11.6 is not a zonal mean. Rather, it is a time mean for a limited longitudinal domain. Provided that the longitudinal scale of variation of this time-mean zonal flow is large compared to the scale of the disturbance, the time-mean flow may be regarded as a locally valid basic state for linear stability calculations.

[3] See, for example, Bourne and Kendall (1968, p. 190).

[4] See, for example, Stevens and Lindzen (1978).

[5] See Matsuno (1966).

Chapter 12

Middle Atmosphere Dynamics

The *middle atmosphere* is generally regarded as the region extending from the tropopause (about 10–16km altitude, depending on latitude) to about 100km. The bulk of the middle atmosphere consists of two main layers, the *stratosphere* and the *mesosphere*, which are distinguished on the basis of temperature stratification (Figure 12.1). The stratosphere, which has large static stability associated with an overall increase of temperature with height, extends from the tropopause to the *stratopause* at about 50km. The mesosphere, which has a lapse rate similar to that in the troposphere, extends from the stratopause to the *mesopause* at about 80 to 90km.

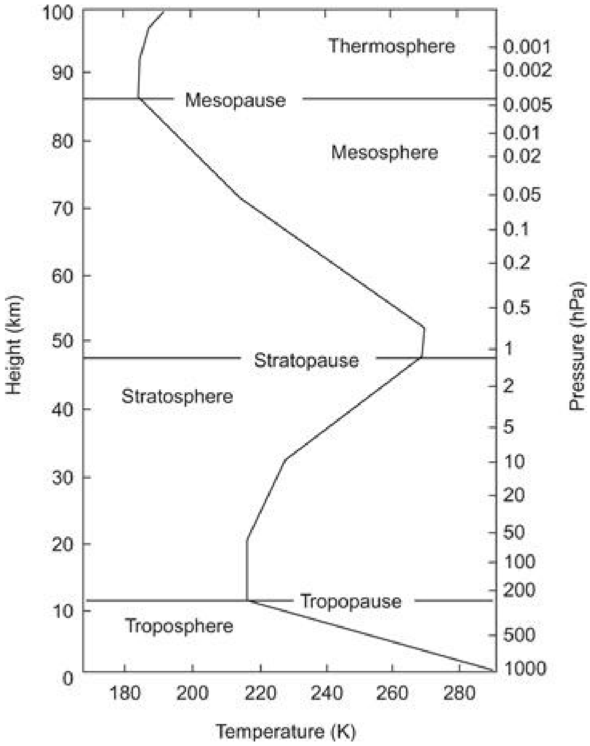


Figure 12.1 Midlatitude mean temperature profile.

Based on the U.S. Standard Atmosphere, 1976.

Previous chapters of this text have focused almost exclusively on the dynamics of the troposphere. The troposphere accounts for about 85% of the total mass of the atmosphere and virtually all atmospheric water. There can be little doubt that processes occurring in the troposphere are primarily responsible for weather disturbances and climate variability. Nevertheless the middle atmosphere cannot be neglected. The troposphere and the middle atmosphere are linked through radiative and dynamical processes that must be represented in global forecast and climate models. They are also linked through the exchange of trace substances that are important in the photochemistry of the ozone layer. This chapter focuses primarily on dynamical processes in the lower part of the middle atmosphere and their links to the troposphere.

12.1 Structure and Circulation of the Middle Atmosphere

Zonal-mean temperature cross-sections for January and July in the lower and middle atmosphere are shown in the upper panels of Figures 12.2 and 12.3, respectively. Because very little solar radiation is absorbed in the troposphere, the thermal structure of the troposphere is maintained by an approximate balance among infrared radiative cooling, vertical transport of sensible and latent heat away from the surface by small-scale eddies, and large-scale heat transport by synoptic-scale eddies. The net result is a mean temperature structure in which the surface temperature has its maximum in the equatorial region and decreases toward both the winter and the summer poles. There is also a rapid decrease in altitude, with a lapse rate of about 6° C km^{-1}.

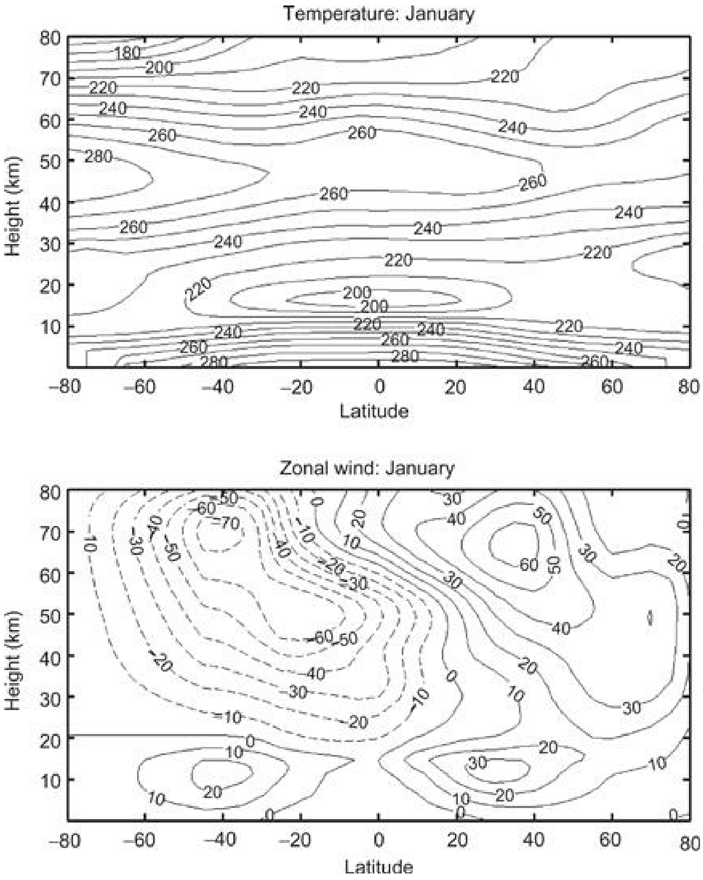


Figure 12.2 Observed monthly and zonally averaged temperature (K) and zonal wind m s^{-1}) for January.

Based on Fleming et al., 1990.

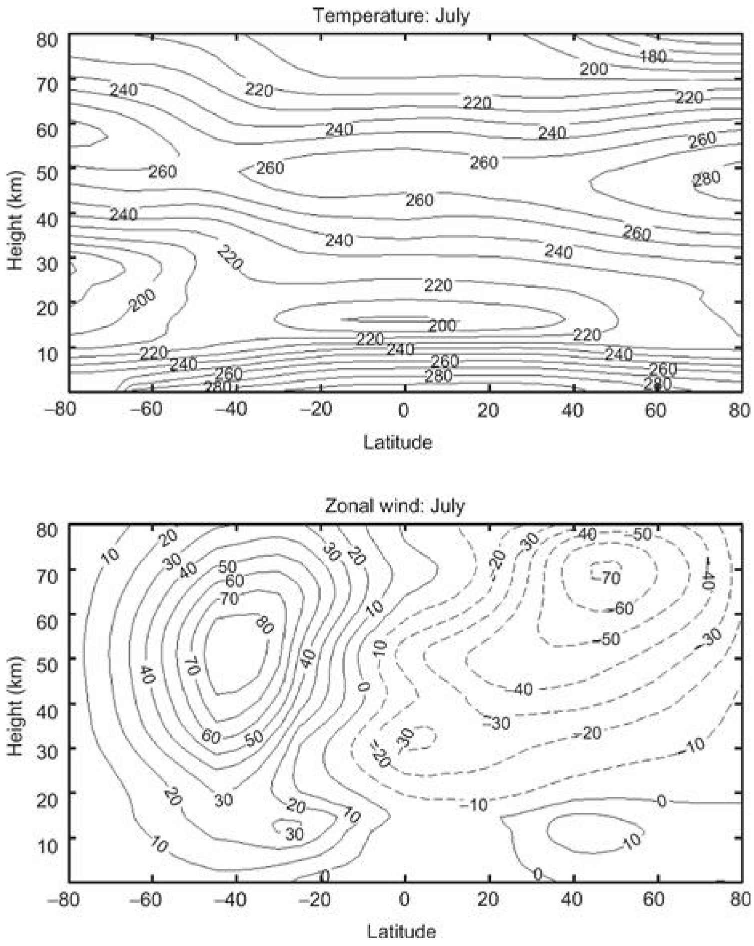


Figure 12.3 Observed monthly and zonally averaged temperature (K) and zonal wind (m s^{-1}) for July.

Based on Fleming et al., 1990.

In the stratosphere, however, infrared radiative cooling is in the mean balanced primarily by radiative heating due to the absorption of solar ultraviolet radiation by ozone. As a result of the solar heating in the ozone layer, the mean temperature in the stratosphere increases with height to a maximum at the stratopause near 50km. Above the stratopause, temperature decreases with height due to the reduced solar heating of ozone.

The meridional temperature structure in the middle atmosphere is also quite different from that in the troposphere. In the lower stratosphere, where the temperature is strongly influenced by upper tropospheric processes, there is a temperature minimum at the equator and maxima at the summer pole and in midlatitudes of the winter hemisphere. Above about 30 hPa the temperature decreases uniformly from summer pole to winter pole, in qualitative accord with radiative equilibrium conditions.

Mean zonal wind climatologies for the middle atmosphere are usually derived from the satellite-observed temperature field. This is done using the geostrophic winds on an isobaric surface in the lower stratosphere (obtained from conventional meteorological analyses) as a lower boundary condition and integrating the thermal wind equation vertically. January and July mean zonal wind cross-sections are shown in the lower panels of Figures 12.2 and 12.3, respectively. The main features are an easterly jet in the summer hemisphere and a westerly jet in the winter hemisphere, with maxima in the wind speeds occurring near the 60-km level. Of particular significance are the high-latitude westerly jets in the winter hemispheres. These polar night jets provide wave guides for the vertical propagation of quasi-stationary planetary waves. In the Northern Hemisphere the EP flux convergence (see Section 10.2.2) due to such waves occasionally leads to rapid deceleration of the mean zonal flow and an accompanying *sudden stratospheric warming* in the polar region, as discussed in Section 12.4.

The zonal mean flow in the equatorial middle atmosphere is strongly influenced by vertically propagating inertia–gravity waves and by equatorial wave modes, especially the Kelvin and Rossby–gravity modes. These waves interact with the mean flow to produce a long-period oscillation called *quasi-biennial oscillation*. This oscillation produces large year-to-year variability in the mean zonal wind in the equatorial middle atmosphere, which is not shown in the long-term means of Figure 12.3.

12.2 The Zonal-Mean Circulation of the Middle Atmosphere

As discussed in Section 10.1, the general circulation of the atmosphere considered as a whole can be regarded to a first approximation as the atmospheric response to the diabatic heating caused by absorption of solar radiation at the surface. Thus, it is reasonable to say that the atmosphere is driven by differential diabatic heating. For an open subregion of the atmosphere, such as the middle atmosphere, it is not correct, however, to assume that the circulation is driven by diabatic heating. It is, rather, necessary to consider the transfer of momentum and energy between that subregion and the rest of the atmosphere.

In the absence of eddy motions, the zonal-mean temperature of the middle atmosphere would relax to a radiatively determined state in which, except for a small lag due to thermal inertia, the temperature would correspond to an annually varying radiative equilibrium following the annual cycle in solar heating. The circulation in such a situation would consist only of a zonal-mean zonal flow in thermal wind balance with the meridional temperature gradient. Neglecting small effects of the annual cycle, there would be no meridional or vertical circulation and no stratosphere–troposphere exchange for such a hypothetical state.

A cross-section of the radiatively determined temperature during a Northern Hemisphere winter is shown in Figure 12.4. It should be compared to the observed temperature profile for the same season that was shown in Figure 12.2. Although the rather uniform increase of temperature from winter pole to summer pole in the region of 30- to 60-km altitude is qualitatively consistent with the radiatively determined distribution, the actual temperature difference between the two poles is much smaller than in the radiatively determined state. Above 60km even the sign of the gradient is opposite to that in the radiative solution; the observed temperatures near the summer polar mesopause are much colder than those near the winter polar mesopause.

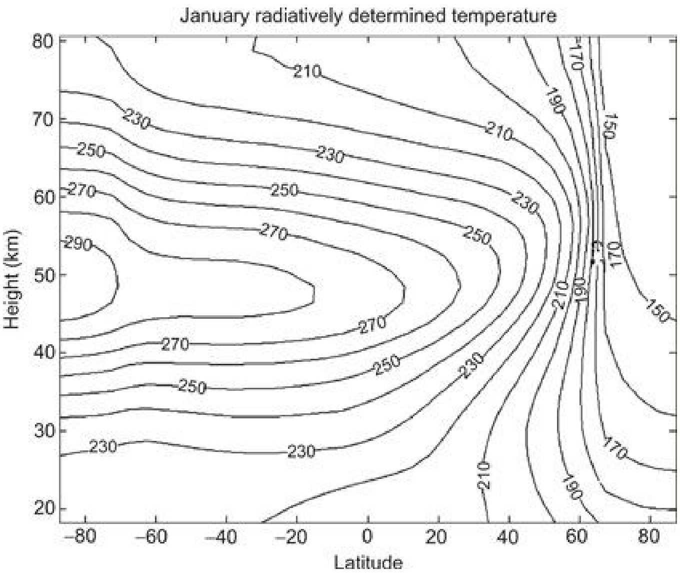


Figure 12.4 Radiatively determined middle atmosphere temperature distribution (K) for northern winter soltice from a radiative model that is time-marched through an annual cycle. Realistic tropospheric temperatures and cloudiness are used to determine the upward radiative flux at the tropopause.

Based on Shine, 1987.

Departures from this radiatively determined state must be maintained by eddy transports. Thus, rather than *causing* the mean circulation, the radiative heating and cooling patterns observed in the middle atmosphere are a *result* of the eddies driving the flow away from a state of radiative balance. This eddy-driven circulation has meridional and vertical wind components that induce substantial local departures from radiative equilibrium, especially in the winter stratosphere and in the mesosphere in both winter and summer.

12.2.1 Lagrangian Motion of Air Parcels

Viewed in the conventional Eulerian mean framework (Section 10.2.1), the time-averaged zonal-mean temperature distribution in the middle atmosphere is determined by the net balance among net radiative heating or cooling, eddy heat and adiabatic heating or cooling by the mean meridional circulation $(\overline{v}, \overline{w})$. That this framework does not provide a useful model for transport in the meridional plane can be seen easily by considering steady-state adiabatic motion in the presence of large-scale waves. According to (10.12), in such a situation waves that have a positive poleward heat flux must drive a nonzero $\overline{w}$, with upwelling at high latitudes and downwelling at low latitudes corresponding to the regions of heat flux convergence and divergence, respectively. However, if the flow is adiabatic, there can be no motion across the isentropes, and thus in steady state there can be no net vertical transport even though $\overline{w}$ is nonzero. Thus, $\overline{w}$ clearly does not provide an approximation to the vertical motion of air parcels (i.e., the vertical transport) in such a situation.

But how can vertical transport vanish when the Eulerian mean vertical motion is finite? The resolution of this “nontransport” paradox can be illustrated by considering the kinematics of adiabatic flow in the presence of a stationary large-scale wave superposed on a background westerly flow as shown in Figure 12.5a. Parcels moving along the streamlines labeled S_1, S_2, and S_3 will oscillate about their mean latitudes, moving equatorward and upward, then poleward and downward following the up and down displacements of the isentropic surfaces associated with the wave. However, over a full wave period there can be no net vertical motion following the parcels, as the parcels must remain on the same isentropic surface. Hence there can be no vertical transport. If, however, an average vertical velocity is computed by taking the mean along a constant latitude circle, then as clearly shown in Figure 12.5a, on the latitude circle ϕ_3 poleward of the maximum in wave amplitude, the Eulerian mean is upward, $\overline{w}(\phi_3) > 0$, as regions of upward motion dominate over regions of downward motion at that latitude.

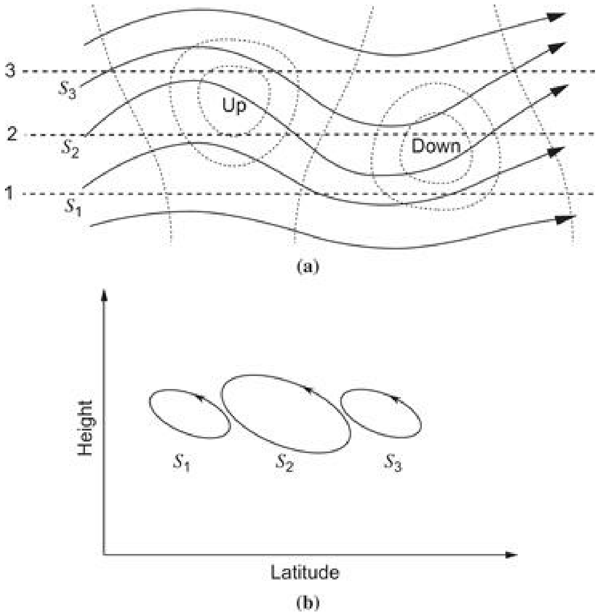


Figure 12.5 Parcel motions for an adiabatic planetary wave in a westerly zonal flow. (a) *Solid lines* labeled S_1, S_2, and S_3 are parcel trajectories, *heavy dashed lines* are latitude circles, and *light dashed lines* are contours of the vertical velocity field. (b) Projection of parcel oscillations on the meridional plane.

Conversely, on the latitude circle ϕ_1 equatorward of the maximum in wave amplitude, the Eulerian mean is downward $\overline{w}(\phi_1) < 0$. The conventional Eulerian mean circulation thus suggests misleadingly that a trace constituent would be transported upward poleward of the latitude of maximum wave amplitude and downward equatorward of that latitude. In reality, air parcels in this idealized example of adiabatic wave motion are not undergoing net vertical displacements, but are simply oscillating back and forth along the same isentropic surface with trajectories whose meridional projections are closed ellipses as shown in Figure 12.5b. In such a situation there is no net vertical parcel motion, and thus no net vertical transport.

Only if there is net diabatic heating or cooling can there be mean transport across the isentropes. In the winter stratosphere, where stationary planetary wave amplitudes tend to be large and there is diabatic heating at low latitudes and diabatic cooling at high latitudes, the meridional projection of actual parcel motions is a combination of the elliptical orbits of Figure 12.5b and mean vertical drift as indicated in Figure 12.6, which clearly shows that a vertically stratified long-lived trace constituent is transported upward at low latitudes and downward at high latitudes because of vertical transport of fluid parcels associated with diabatic heating and cooling. Thus, in order to represent net transport effects efficiently, a zonal averaging process should yield an average vertical circulation that mimics the net cross-isentropic parcel motion.

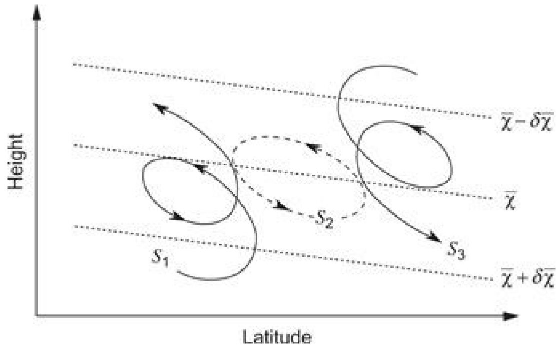

Height
$\overline{\chi}-\delta\overline{\chi}$
$\overline{\chi}$
$\overline{\chi}+\delta\overline{\chi}$
S_1
S_2
S_3
Latitude

Figure 12.6 Projections of parcel motions on the meridional plane for planetary waves in westerlies with diabatic heating at low latitudes and diabatic cooling at high latitudes. *Thin dashed lines* show tilting of isosurfaces of tracer mixing ratio due to transport by the diabatic circulation.

12.2.2 The Transformed Eulerian Mean

In many circumstances the transformed Eulerian mean (TEM) equations introduced in Section 10.2.2 provide a useful model for the study of global-scale middle atmospheric transport. Recall that in this formulation the zonal-mean momentum, mass continuity, thermodynamic energy, and thermal wind equations have the form[1]

$$\partial\overline{u}/\partial t - f_0\overline{v}^* = \rho_0^{-1}\nabla\cdot\mathbf{F} + \overline{X} \equiv \overline{G} \tag{12.1}$$

$$\partial\overline{T}/\partial t + N^2HR^{-1}\overline{w}^* = -\alpha_r\left[\overline{T} - \overline{T}_r\,(y,z,t)\right] \tag{12.2}$$

$$\partial\overline{v}^*/\partial y + \rho_0^{-1}\partial(\rho_0\overline{w}^*)/\partial z = 0 \tag{12.3}$$

$$f_0\partial\overline{u}/\partial z + RH^{-1}\partial\overline{T}/\partial y = 0 \tag{12.4}$$

Here the residual circulation $(\overline{v}^*, \overline{w}^*)$ is as defined in (10.16a,b), **F** designates the EP flux due to large-scale eddies, $\overline{X}$ is the zonal force due to small-scale eddies (e.g., gravity wave drag), and $\overline{G}$ designates the total *zonal force*. In (12.2) the diabatic heating is approximated in terms of a Newtonian relaxation proportional to the departure of the zonal mean temperature $\overline{T}(y,z,t)$ from its radiative equilibrium value $\overline{T}_r(y,z,t)$, where α_r is the Newtonian cooling rate.

To understand how eddies can lead to departures of the zonal-mean temperature distribution in the middle atmosphere from its radiatively determined state, we use the TEM system of equations to consider an idealized model of extratropical forcing. The dependence of the circulation on forcing frequency can be examined in an idealized model in which $\overline{T}_r$ is taken to be a function of height alone, and the forcing has a simple harmonic dependence in time with frequency σ. Then $\overline{w}^*$ and $\overline{G}$ are of the form

$$\begin{bmatrix} \overline{w}^* \\ \overline{G} \end{bmatrix} = Re\begin{bmatrix} \hat{w}\,(\phi,z) \\ \hat{G}\,(\phi,z) \end{bmatrix} e^{i\sigma t} \tag{12.5}$$

where $\hat{w}$ and $\hat{G}$ are complex latitude- and height-dependent amplitudes. Combining (12.1) and (12.3) to eliminate $\overline{v}^*$ and combining the resulting equation with (12.2) and (12.4) to eliminate $\overline{T}$ yields a partial differential equation for $\hat{w}$:

(12.6)

$$\frac{\partial}{\partial z}\left(\frac{1}{\rho_0}\frac{\partial(\rho_0 \hat{w})}{\partial z}\right)+\left(\frac{i\sigma}{i\sigma+\alpha_r}\right)\left(\frac{N^2}{f_0^2}\right)\frac{\partial^2 \hat{w}}{\partial y^2}=\frac{1}{f_0}\frac{\partial}{\partial y}\left(\frac{\partial \hat{G}}{\partial z}\right)$$

The structure of $\hat{w}$ in the meridional plane depends on the latitudinal and vertical distribution of the forcing $\hat{G}$. The differential equation (12.6) is elliptic in form. Such a form implies that a forcing applied locally in any region will produce a nonlocal response in the form of a large-scale circulation that acts to maintain the zonal flow in thermal wind balance with the meridional temperature gradient.

As suggested by the coefficients of (12.6), the character of the nonlocal response depends on the ratio of the frequency of the forcing to the Newtonian relaxation rate. We consider three cases:

1. *High-frequency variations* $\sigma \gg \alpha_r$: The motions are approximately adiabatic (as is true in sudden stratospheric warmings)—that is,

   $$\frac{i\sigma}{i\sigma+\alpha_r}\rightarrow 1$$

   Away from the forcing region, the right side of (12.6) vanishes so that the two terms on the left must balance each other. For this balance to hold, a simple scaling argument indicates that in midlatitudes the vertical scale δz of the response must be related to the horizontal scale as $\delta z \sim (f_0/N)\,\delta y \sim 10^{-2}\delta y$.
2. *Low-frequency variations (e.g., the annual cycle)*: The Newtonian relaxation coefficient has a magnitude of order $\alpha_r \approx 1/(20\ \mathrm{days})$ in the lower stratosphere, increasing to $\alpha_r \approx 1/(5\ \mathrm{days})$ near the stratopause. Thus $\alpha_r > \sigma$ and the coefficient multiplying the second term on the left in (12.6) is reduced from its value in case (1) so that the vertical penetration scale is increased.
3. *Steady-state regime* $\sigma/\alpha_r \rightarrow 0$: In this limit $\partial\overline{u}/\partial t = 0$, and rather than using (12.6), it is simpler to note that (12.1) in this case reduces to a simple balance between the Coriolis force and the zonal drag force:

   (12.7)

   $$-f_0\overline{v}^* = \overline{G}$$

   Combining this equation with the continuity equation (12.3) and requiring that

   $$\rho_0\overline{w}^* \rightarrow 0 \quad \text{as} \quad z \rightarrow \infty$$

   it follows that

   (12.8)

   $$\rho_o\overline{w}^* = -\frac{\partial}{\partial y}\left[\frac{1}{f_0}\int_z^\infty \rho_0\overline{G}dz'\right]$$

   According to (12.8), $\overline{w}^*$ is zero in the regions *above* a localized forcing region, whereas it is constant for regions directly below it; hence the term *downward control* is sometimes used to refer to this steady-state situation.

   Substitution of (12.8) into the first law of thermodynamics (12.2) and neglecting the time-dependent term yield an expression that explicitly shows the dependence on the zonal force distribution of the departure of the time and zonally averaged temperature from radiative equilibrium:

$$(\overline{T}-\overline{T}_r)=\frac{N^2H}{\alpha_r\rho_0R}\frac{\partial}{\partial y}\left[\frac{1}{f_0}\int_z^{\infty}\rho_0\overline{G}dz'\right] \tag{12.9}$$

Thus, in steady state the departure of temperature from radiative equilibrium at a given level depends on the meridional gradient of the zonal force distribution in the column above that level.

For mathematical simplicity, the variation of the Coriolis parameter with latitude and other effects of spherical geometry have been neglected in deriving the equations of this section. It is straightforward to extend this model to spherical coordinates. Figure 12.7 shows streamlines of the meridional mass circulation in response to isolated extratropical forcing for three cases corresponding to the differing frequency regimes discussed earlier, but with spherical geometry retained.

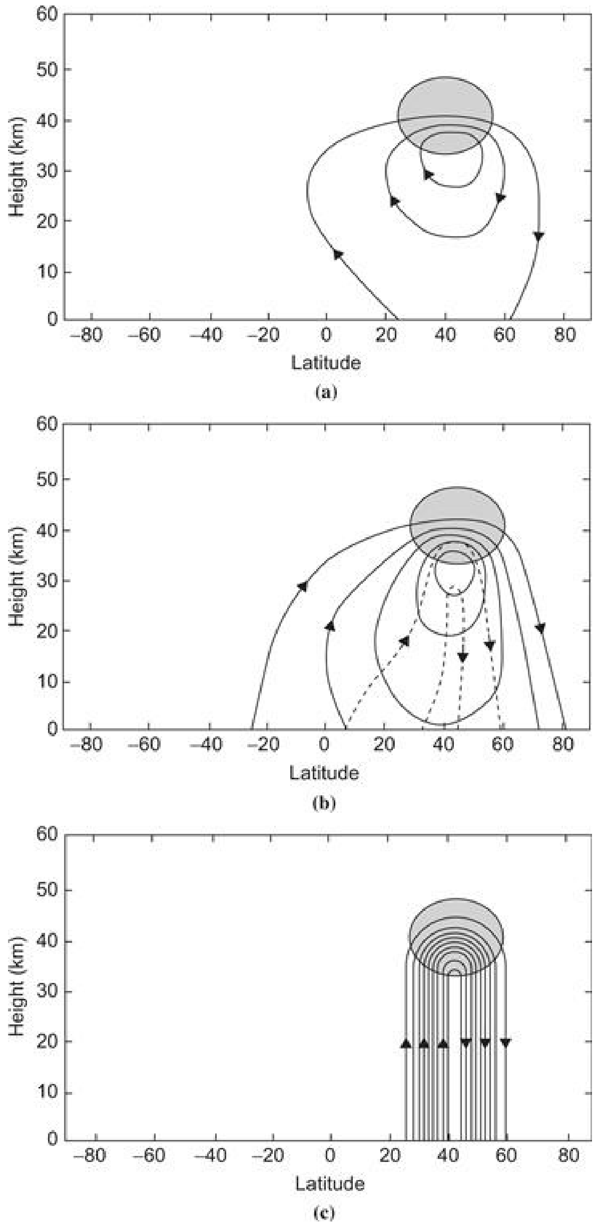


Figure 12.7 Response of the zonally symmetric circulation to a westward force applied in the *shaded* region. Contours are *streamlines* of the TEM meridional circulation, with the same contour interval used in each panel. (a) Adiabatic response for high-frequency forcing. (b) Response for annual frequency forcing and 20-day radiative damping timescale; *solid and dashed contours* show the response that is in phase and 90° out of phase, respectively, with the forcing. (c) Response to steady-state forcing.

After Holton et al., 1995. Used with permission of American Geophysical Union.

A comparison of Figures 12.2 and 12.4 shows that the largest departures from radiative equilibrium occur in the summer and winter mesosphere and in the polar winter stratosphere. According to (12.9), these are the locations and seasons when zonal forcing must be strongest. The zonal force in the mesosphere is believed to be caused primarily by vertically propagating internal gravity waves. These transfer momentum from the troposphere into the mesosphere, where wave breaking produces strong zonal forcing. The zonal force in the winter stratosphere is due primarily to stationary planetary Rossby waves. These, as discussed in Section 12.3, can propagate vertically provided that the mean zonal wind is westerly and less than a critical value that depends strongly on the wavelength of the waves. Thus, in the extratropical stratosphere we expect a strong annual cycle in $\delta\overline{T}$, with large values (i.e., strong departure from radiative equilibrium) in winter and small values in summer.

This is indeed observed (see Figures 12.2 and 12.4). Furthermore, because eddy forcing maintains the observed temperature above its radiative equilibrium in the extratropical stratosphere, there is radiative cooling, and from (12.3) the residual vertical motion must be downward. By mass continuity it is then required that the residual vertical motion be upward in the tropics, implying that the temperature must be below radiative equilibrium in that region. Note that it is the dynamical driving by extratropical eddies, rather than local forcing, that is responsible for the upward residual motion and net radiative heating in the tropical stratosphere.

12.2.3 Zonal-Mean Transport

Viewed from the TEM perspective, the overall meridional circulation in the winter stratosphere is qualitatively as pictured in Figure 12.8. The residual circulation transfers mass and trace chemicals upward across the tropopause in the tropics and downward in the extratropics. This vertical circulation is closed in the lower stratosphere by a poleward meridional drift balanced by EP flux convergence. That this schematic picture gives a qualitatively correct view can be ascertained by examining the zonal-mean mixing ratio distribution of any long-lived vertically stratified trace species. As an example, the distribution of N_2O is shown in Figure 12.9. N_2O is produced at the surface and is mixed uniformly in the troposphere, but then it decays with height in the stratosphere due to photochemical dissociation. Thus, as shown in Figure 12.9, the mixing ratio decreases upward in the stratosphere. Note, however, that surfaces of constant mixing ratio are displaced upward in the tropics and downward at higher latitudes, suggesting that the mean meridional mass transport is upward in the tropics and downward in the extratropics as suggested in Figure 12.9.

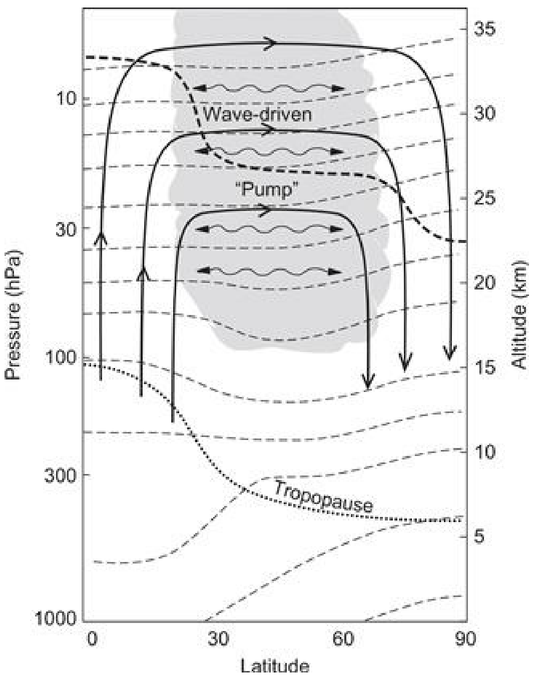


Figure 12.8 Schematic cross-section of the wave-driven circulation in the middle atmosphere and its role in transport. *Thin dashed lines* denote potential temperature surfaces. *Dotted line* is the tropopause. *Solid lines* are contours of the TEM meridional circulation driven by the wave-induced forcing (*shaded region*). *Wavy double-headed arrows* denote meridional transport and mixing by eddy motions. *Heavy dashed line* shows an isopleth of mixing ratio for a long-lived tracer.

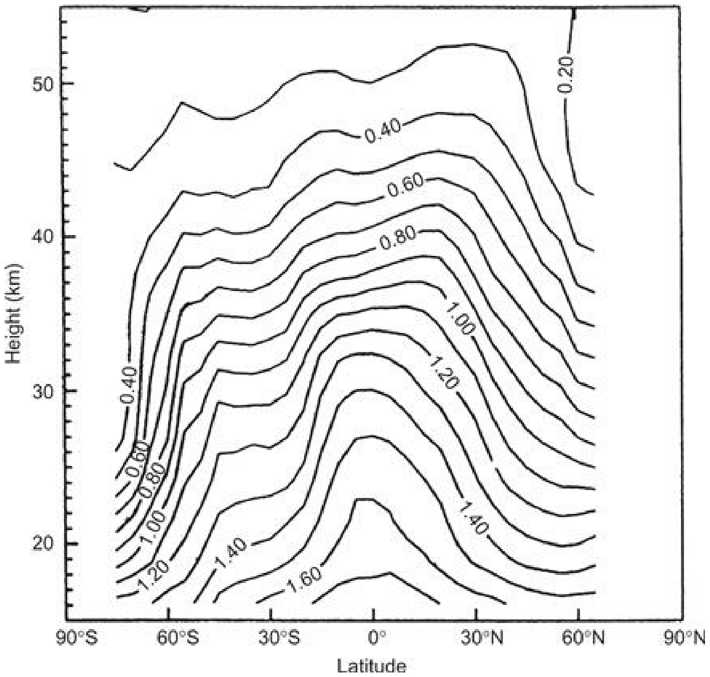


Figure 12.9 October zonal mean cross-section of methane (ppmv) from observations by the HALOE on the UARS. Note the strong vertical stratification due to photochemical destruction in the stratosphere. The upward-bulging mixing ratio isopleths in the equatorial region are evidence of upward mass flow in the equatorial region, whereas the downward-sloping isopleths in high latitudes are evidence of subsidence in the polar regions. The region of flattened isopleths in the midlatitudes of the Southern Hemisphere is evidence for quasi-adiabatic wave transport due to wintertime planetary wave activity.

After *Norton, 2003*.

In addition to the slow meridional drift by the residual meridional velocity shown in Figure 12.8, tracers in the winter stratosphere are also subject to rapid quasi-isentropic transport and mixing by breaking planetary waves. Such eddy transport, which is highly variable in space and time, must be included for quantitatively accurate modeling of transport within the stratosphere.

In the mesosphere the residual circulation is dominated by a single circulation cell with upward motion in the summer polar region, meridional drift from the summer to the winter hemisphere, and subsidence in the winter polar region. This circulation, like the residual circulation in the stratosphere, is eddy driven. However, in the mesosphere it appears that the dominant eddies are vertically propagating internal gravity waves, which have shorter scales in space and time than the planetary waves that dominate the eddy activity in the stratosphere.

12.3 Vertically Propagating Planetary Waves

In Section 12.1 we pointed out that the predominant eddy motions in the stratosphere are vertically propagating quasi-stationary planetary waves (Rossby waves) and that these waves are confined to the winter hemisphere. To understand the absence of synoptic-scale motions and the confinement of stationary planetary waves to the winter hemisphere, it is necessary to examine the conditions under which planetary waves can propagate vertically.

12.3.1 Linear Rossby Waves

To analyze planetary wave propagation in the stratosphere, it is convenient to write the equations of motion in the log-pressure coordinate system introduced in Section 7.4.1. For analysis of extratropical planetary wave motions in the middle atmosphere, we may refer the motions to the midlatitude β plane and use the quasi-geostrophic potential vorticity equation, which in log-pressure coordinates can be written as

(12.10)

$$\left(\frac{\partial}{\partial t}+\mathbf{V}_g\cdot\boldsymbol{\nabla}\right)q=0$$

where

$$q\equiv\nabla^2\psi+f+\frac{f_0^2}{\rho_0N^2}\frac{\partial}{\partial z}\left(\rho_0\frac{\partial\psi}{\partial z}\right)$$

Here $\psi=\Phi/f_0$ is the geostrophic streamfunction and f_0 is a constant midlatitude reference value of the Coriolis parameter. We now assume that the motion consists of a small-amplitude disturbance superposed on a constant zonal-mean flow. Thus, letting $\psi=-\overline{u}y+\psi', q=\overline{q}+q'$ and linearizing (12.10), we find that the perturbation q' field must satisfy

(12.11)

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)q'+\beta\frac{\partial\psi'}{\partial x}=0$$

where

$$q' \equiv \nabla^2 \psi' + \frac{f_0^2}{\rho_0 N^2} \frac{\partial}{\partial z} \left(\rho_0 \frac{\partial \psi'}{\partial z} \right) \tag{12.12}$$

Equation (12.11) has solutions in the form of harmonic waves with zonal and meridional wave numbers k and l and zonal phase speed c_x:

$$\psi' (x, y, z, t) = \Psi (z)\, e^{i(kx+ly-kc_x t)+z/2H} \tag{12.13}$$

Here the factor $e^{z/2H}$ (which is proportional to $\rho_0^{-1/2}$) is included to simplify the equation for the vertical dependence. Substituting (12.13) into (12.11) yields

$$d^2\Psi/dz^2 + m^2\Psi = 0 \tag{12.14}$$

where

$$m^2 \equiv \frac{N^2}{f_0^2} \left[\frac{\beta}{(\overline{u} - c_x)} - \left(k^2 + l^2\right) \right] - \frac{1}{4H^2} \tag{12.15}$$

Referring back to Section 7.4, we recall that $m^2 > 0$ is required for vertical propagation, and in that case, m is the vertical wave number; that is, solutions of (12.14) have the form $\Psi(z) = Ae^{imz}$, where A is a constant amplitude and the sign of m is determined by requiring that the vertical component of group velocity be positive. For stationary waves ($c_x = 0$), we see from (12.15) that vertically propagating modes can exist only for mean zonal flows that satisfy the condition

$$0 < \overline{u} < \beta \left[\left(k^2 + l^2\right) + f_0^2/\left(4N^2H^2\right) \right]^{-1} \equiv U_c \tag{12.16}$$

where U_c is called the *Rossby critical velocity*. Thus, vertical propagation of stationary waves can occur only in the presence of westerly winds weaker than a critical value that depends on the horizontal scale of the waves. In the summer hemisphere the stratospheric mean zonal winds are easterly so that stationary planetary waves are all trapped vertically.

In the real atmosphere the mean zonal wind is not constant, but depends on latitude and height. However, both observational and theoretical studies suggest that (12.16) still provides a qualitative guide for estimating vertical propagation of planetary waves, even though the actual critical velocity may be larger than indicated by the β-plane theory.

12.3.2 Rossby Wavebreaking

The term *wavebreaking* for Rossby waves simply refers to a rapid, irreversible deformation of material contours. Since by (12.10) potential vorticity is approximately conserved in Rossby waves, isolines of potential vorticity on isentropic surfaces approximate material contours, and wavebreaking can best be illustrated by considering the field of potential vorticity. Wavebreaking may occur when the disturbance fields reach amplitudes for which nonlinear effects can no longer be neglected in the dynamical equations. For example, if the flow is divided into mean and disturbance parts, and nonlinear terms are included, the equation for conservation of quasi-geostrophic potential vorticity (12.10) becomes

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)q'+v'\frac{\partial\overline{q}}{\partial y}=-u'\frac{\partial q'}{\partial x}-v'\frac{\partial q'}{\partial y} \tag{12.17}$$

For steady waves propagating relative to the ground at zonal-phase speed c_x, the variation of phase in time and space is given by $\phi=k(x-c_xt)$, where k is the zonal wave number, and it is readily verified that

$$\frac{\partial}{\partial t}=-c_x\frac{\partial}{\partial x}$$

so that in the linearized version of (12.17) there is a balance between advection of the disturbance potential vorticity q' by the Doppler-shifted mean wind and the advection of mean potential vorticity by the disturbance meridional wind:

$$(\overline{u}-c_x)\frac{\partial q'}{\partial x}=-v'\frac{\partial\overline{q}}{\partial y} \tag{12.18}$$

The validity of the linear approximation can thus be assessed by comparing the sizes of the two terms on the right side of (12.17) with either term in (12.18). Linearity holds provided that

$$|\overline{u}-c_x|\gg|u'| \tag{12.19a}$$

and

$$\partial\overline{q}/\partial y\gg|\partial q'/\partial y| \tag{12.19b}$$

Basically these criteria require that the slope of the material contours in the x, y plane must be small.

As indicated in (12.13), in an atmosphere with constant mean zonal wind, vertically propagating linear Rossby waves have amplitudes increasing exponentially in height. Thus, at some altitude the disturbance amplitude will become sufficiently large so that wavebreaking must occur. In the real atmosphere, however, the mean-zonal flow varies in both latitude and height, and this variation is crucial for understanding the distribution and mean-flow forcing provided by Rossby wavebreaking. The simplest example of Rossby wavebreaking occurs in the presence of a *critical surface* along which the Doppler-shifted phase speed of the wave vanishes $(\overline{u}-c_x=0)$. In that case (12.19a) cannot be satisfied even for small-amplitude waves.

For understanding wave behavior near critical lines it is helpful to generalize the Rossby wave analysis of Section 12.3.1 to a situation such as that shown in the seasonal mean cross-sections of Figure 12.2 in which the zonal wind depends on latitude and height, $\overline{u}=\overline{u}(y,z)$. The x and t dependence in (12.11) can then be separated by seeking solutions of the form

$$\psi'=e^{z/2H}\mathrm{Re}\left[\Psi(y,z)\,e^{ik(x-c_xt)}\right] \tag{12.20}$$

We obtain

$$\frac{\partial^2\Psi}{\partial y^2}+\frac{f_0^2}{N^2}\frac{\partial^2\Psi}{\partial z^2}+n_k^2\Psi=0 \tag{12.21}$$

where the small vertical variation of N^2 is neglected, and

$$n_k^2(y,z)=(\overline{u}-c_x)^{-1}\partial\overline{q}/\partial y-k^2-f_0^2/(4H^2N^2) \tag{12.22}$$

Equation (12.21) has a form similar to that of the equation governing the two-dimensional propagation of light waves in a medium with variable refractive index, n_k. The propagation of linear Rossby wave EP flux in that case can be shown to be along rays that behave somewhat like light rays. Thus, wave activity will tend to propagate along rays that bend toward regions of large positive n_k^2 and avoid regions of negative n_k^2. For stationary Rossby waves ($c_x = 0$) of low zonal wave number, n_k^2 is positive in a region with westerly winds that are not too strong, and increases to infinity along a critical surface where the mean flow vanishes. Thus, the index of refraction for wave activity is positive in the winter hemisphere, but increases rapidly toward the equatorial zero wind line. As a result, Rossby wave activity tends to propagate upward and equatorward, and wavebreaking occurs in the vicinity of the equatorial critical line.

12.4 Sudden Stratospheric Warmings

In the lower stratosphere the temperature is a minimum at the equator and has maxima at the summer pole and at about 45° latitude in the winter hemisphere (see Figure 12.2). From thermal wind considerations the rapid decrease of temperature poleward of 45° in winter requires a zonal vortex with strong westerly shear with height.

In the Northern Hemisphere, every other year or so, this normal winter pattern of a cold polar stratosphere with a westerly vortex is interrupted in a spectacular manner in midwinter. Within the space of a few days the polar vortex becomes highly distorted and breaks down (see Figure 12.10) with an accompanying *large-scale* warming of the polar stratosphere, which can quickly reverse the meridional temperature gradient and (through thermal wind balance) create a circumpolar easterly current. Warmings of 40K in a few days have occurred at the 50-hPa level.

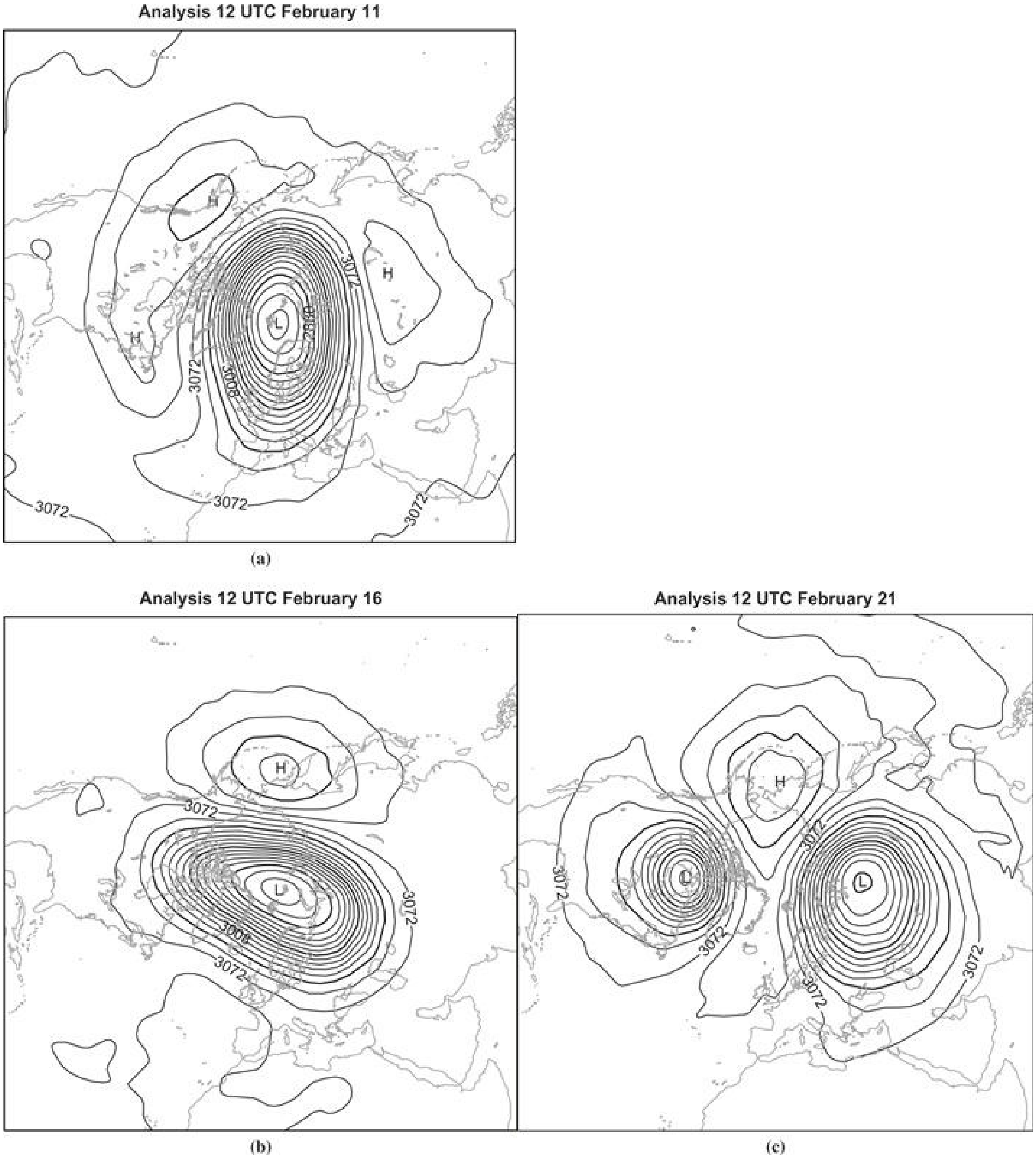

Analysis 12 UTC February 11
H
H
H
L
3072
2848
3008
3072
3072
3072
(a)
Analysis 12 UTC February 16
H
L
3072
3008
3072
3072
(b)
Analysis 12 UTC February 21
H
L
L
3072
3072
3072
3072
3072
(c)

Figure 12.10 10-hPa geopotential height analyses for days in 1979: February 11 (a), February 16 (b), and February 21 (c) at 12 UTC, showing breakdown of the polar vortex associated with a wave number 2's sudden stratospheric warming. Contour interval: 16 dam.

Analysis from ERA-40 reanalysis courtesy of the European Centre for Medium-Range Weather Forecasts.

Numerous observational studies of sudden warmings confirm that enhanced propagation of planetary waves from the troposphere, primarily zonal wave numbers 1 and 2, is essential for the development of the warmings. Since major warmings are observed primarily in the Northern Hemisphere, it is logical to conclude that enhanced wave propagation into the stratosphere is due to topographically forced waves, which are much stronger in the Northern Hemisphere than in the Southern Hemisphere. Even in the Northern Hemisphere, however, it is apparently only in certain winters that conditions are right to produce sudden warmings.

It is generally accepted that sudden warmings are an example of transient mean-flow forcing due to planetary wave driving. Section 10.2.3 showed that in order for planetary waves to decelerate the zonal-mean circulation, there must be a nonzero equatorward eddy potential vorticity flux (i.e., a net EP flux convergence). We further showed that for steady nondissipative waves, the divergence of the EP flux vanishes. For normal stationary planetary waves in the stratospheric polar night jet, this constraint should be at least approximately satisfied, as radiative and frictional dissipation are rather small. Thus, the strong interaction between the waves and the mean flow that occurs during the course of a sudden warming must be associated with wave transience (i.e., change of amplitude with respect to time) and wave dissipation. Most of the dramatic mean-flow deceleration that occurs during a sudden warming is caused by amplification of quasi-stationary planetary waves in the troposphere followed by propagation into the stratosphere.

This process can be understood by considering the interaction between zonal mean and eddy potential vorticities. We assume that on the short time scales characteristic of sudden warmings, diabatic processes can be neglected. We also assume that the eddy motion is approximately governed by the linearized eddy potential vorticity equation and that the zonal mean is changed only by the eddy flux convergence. The eddies and mean flow are thus related by the quasi-linear system

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)q'+v'\frac{\partial\overline{q}}{\partial y}=-S' \tag{12.23}$$

where again

$$q'\equiv\nabla^2\psi'+\frac{f_0^2}{\rho_0}\frac{\partial}{\partial z}\left(\frac{\rho_0}{N^2}\frac{\partial\psi'}{\partial z}\right) \tag{12.24}$$

and an eddy damping term, $-S'$, has been added to represent the effects of both mechanical and thermal dissipation. (For constant and equal Rayleigh friction and Newtonian cooling, $S'=\alpha q'$, where α is the dissipation rate coefficient.) Multiplying (12.23) by q' and averaging zonally after noting that

$$\overline{u}\,\overline{q'\frac{\partial q'}{\partial x}}=\frac{\overline{u}}{2}\frac{\partial\overline{q'^2}}{\partial x}=0$$

yields with the aid of (10.26)

$$\frac{\partial A}{\partial t}+\boldsymbol{\nabla}\cdot\mathbf{F}=D \tag{12.25}$$

where *wave activity*, *A*, and *dissipation*, *D*, are defined, respectively, as

$$A\equiv\frac{\rho_0\overline{(q')^2}}{2\partial\overline{q}/\partial y},\quad\text{and}\quad \mathrm{D}\equiv-\frac{\rho_0\overline{S'q'}}{\partial\overline{q}/\partial y} \tag{12.26}$$

For linear planetary waves satisfying the quasi-geostrophic equations, it can be shown that

$$\mathbf{F}=(0,c_{gy},c_{gz})\,A \tag{12.27}$$

so that the EP flux is given by the wave activity times the group velocity in the meridional (y,z) plane. It is the flux of wave activity—not the flux of wave energy—that is fundamental for wave–mean-flow interaction.

Equation (12.25) shows that if dissipation and transience vanish, the EP flux divergence must vanish. Substitution into (12.1) then gives the *nonacceleration theorem*, which states that there is no wave-driven mean-flow acceleration for steady $(\partial A/\partial t = 0)$ and conservative ($D = 0$) waves. The EP flux divergence, which must be nonzero for wave-induced forcing, is dependent on wave transience and dissipation.

In sudden stratospheric warmings the planetary wave amplitudes increase rapidly in time. During the Northern Hemisphere winter, quasi-stationary planetary waves of zonal wave numbers 1 and 2 are produced in the troposphere by orographic forcing. These waves propagate vertically into the stratosphere, implying a local increase in the wave activity. Thus, $\partial A/\partial t > 0$, implying from (12.25) that, for quasi-conservative flow, the flux of quasi-geostrophic potential vorticity is negative and that the Eliassen–Palm flux field is convergent:

$$\rho_0 \overline{v'q'} = \boldsymbol{\nabla} \cdot \mathbf{F} < 0 \tag{12.28}$$

Usually $D < 0$ also, so dissipation and wave growth both produce a convergent EP flux (i.e., an equatorward potential vorticity flux).

According to the transformed-Eulerian-mean zonal-momentum equation (12.1), EP flux convergence leads to a deceleration of the zonal-mean zonal wind, partially offset by the Coriolis torque $f\overline{v}^*$. As a result, the polar night jet weakens, allowing even more waves to propagate into the stratosphere. At some point, the mean zonal wind changes sign so that a critical layer is formed. Stationary linear waves can no longer propagate beyond the level where this occurs, and strong EP flux convergence and even faster easterly acceleration then occurs below the critical layer.

Figure 12.11a illustrates the deceleration of the zonal-mean wind due to the Eliassen–Palm flux convergence caused by upward propagation of transient planetary wave activity. Note that the deceleration is spread out over a broader range in height than the eddy forcing. This reflects the elliptical nature of the equation for mean zonal-wind acceleration that can be derived from the transformed Eulerian mean equations. Figure 12.11b shows the TEM residual meridional circulation and the pattern of temperature perturbation associated with the zonal wind deceleration. The thermal wind relation implies that a deceleration of the polar night jet leads to a warming in the polar stratosphere and a cooling in the equatorial stratosphere, with a compensating warming in the polar mesosphere and cooling in the equatorial mesosphere. Eventually, as more of the flow becomes easterly, waves can no longer propagate vertically. The wave-induced residual circulation then decreases, and radiative cooling processes are able to slowly reestablish the normal cold polar temperatures. Thermal wind balance then implies that the normal westerly polar vortex is also reestablished.

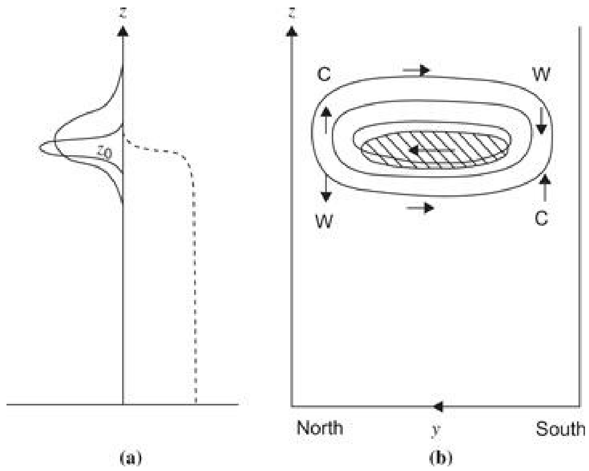


Figure 12.11 Schematic of transient wave, mean–flow interactions occurring during a stratospheric warming. (a) Height profiles of EP flux (*dashed line*), EP flux divergence (*heavy line*), and mean zonal flow acceleration (*thin line*); z_0 is the height reached by the leading edge of the wave packet at the time pictured. (b) Latitude–height cross-section showing region where the EP flux is convergent (*hatching*), contours of induced zonal acceleration (*lines*), and induced residual circulation (arrows). Regions of warming (W) and cooling (C) are also shown.

From Andrews et al., 1987.

In some cases the wave amplification may be large enough to produce a polar warming but insufficient to lead to reversal of the mean zonal wind in the polar region. Such “minor warmings” occur every winter and are generally followed by a quick return to normal winter circulation. A “major warming” in which the mean zonal flow reverses at least as low as the 30-hPa level in the polar region seems to occur only about once every couple of years. If the major warming occurs sufficiently late in the winter, the westerly polar vortex may not be restored at all before the normal seasonal circulation reversal.

12.5 Waves in the Equatorial Stratosphere

Section 11.4 discussed equatorially trapped waves in the context of shallow water theory. Under some conditions, however, equatorial waves (both gravity and Rossby types) may propagate vertically, and the shallow water model must be replaced by a continuously stratified atmosphere examine the vertical structure. It turns out that vertically propagating equatorial waves share a number of physical properties with ordinary gravity modes. Section 5.6 discussed vertically propagating gravity waves in the presence of rotation for a simple situation in which the Coriolis parameter was assumed to be constant and the waves were assumed to be sinusoidal in both x and y.

We found that such inertia–gravity waves can propagate vertically only when the wave frequency satisfies the inequality $f < \nu < N$. Thus, at the middle latitudes, waves with periods that are in the range of several days are generally vertically trapped—that is, they are not able to propagate significantly into the stratosphere. As the equator is approached, however, the decreasing Coriolis frequency should allow vertical propagation to occur for lower-frequency waves. Therefore, in the equatorial region there is a possibility for the existence of long-period, vertically propagating internal gravity waves.

As in Section 11.4, we consider linearized perturbations on an equatorial β plane. The linearized equations of motion, continuity equation, and first law of thermodynamics can then be expressed in log-pressure coordinates as

$$\partial u'/\partial t - \beta y v' = -\partial \Phi'/\partial x \tag{12.29}$$

$$\partial v'/\partial t + \beta y u' = -\partial \Phi'/\partial y \tag{12.30}$$

$$\partial u'/\partial x + \partial v'/\partial y + \rho_0^{-1}\partial\left(\rho_0 w'\right)/\partial z = 0 \tag{12.31}$$

$$\partial^2 \Phi'/\partial t \partial z + w' N^2 = 0 \tag{12.32}$$

We again assume that the perturbations are zonally propagating waves, but we now assume that they also propagate vertically with vertical wave number *m*. Due to the basic state density stratification, there will also be an amplitude growth in height proportional to $\rho_0^{-1/2}$. Thus, the *x*,*y*,*z*, and *t* dependencies can be separated as

$$\begin{pmatrix} u' \\ v' \\ w' \\ \Phi' \end{pmatrix} = e^{z/2H} \begin{bmatrix} \hat{u}(y) \\ \hat{v}(y) \\ \hat{w}(y) \\ \hat{\Phi}(y) \end{bmatrix} \exp\left[i\left(kx + mz - \nu t\right)\right] \tag{12.33}$$

Substituting from (12.33) into (12.29) through (12.32) yields a set of ordinary differential equations for the meridional structure:

$$-i\nu\hat{u} - \beta y \hat{v} = -ik\hat{\Phi} \tag{12.34}$$

$$-i\nu\hat{v} + \beta y \hat{u} = -\partial\hat{\Phi}/\partial y \tag{12.35}$$

$$\left(ik\hat{u} + \partial\hat{v}/\partial y\right) + i\left(m + i/2H\right)\hat{w} = 0 \tag{12.36}$$

$$\nu\left(m - i/2H\right)\hat{\Phi} + \hat{w}N^2 = 0 \tag{12.37}$$

12.5.1 Vertically Propagating Kelvin Waves

For Kelvin waves the aforementioned perturbation equations can be simplified considerably. Setting $\hat{v} = 0$ and eliminating $\hat{w}$ between (12.36) and (12.37), we obtain

$$-i\nu\hat{u} = -ik\hat{\Phi} \tag{12.38}$$

$$\beta y \hat{u} = -\partial\hat{\Phi}/\partial y \tag{12.39}$$

$$-\nu\left(m^2 + 1/4H^2\right)\hat{\Phi} + \hat{u}kN^2 = 0 \tag{12.40}$$

-

Equation (12.38) can be used to eliminate Φ in (12.39) and (12.40). This yields two independent equations that the field of $\hat{u}$ must satisfy. The first of these determines the meridional distribution of $\hat{u}$ and is identical to (11.47). The second is simply the dispersion equation

$$c^2\left(m^2+1/4H^2\right)-N^2=0 \tag{12.41}$$

where, as in Section 11.4, $c^2=(\nu^2/k^2)$.

If we assume that $m^2 \gg 1/4H^2$, as is true for most observed stratospheric Kelvin waves, (12.41) reduces to the dispersion relationship for internal gravity waves (5.66) in the hydrostatic limit $(|k| \ll |m|)$. For waves in the stratosphere that are forced by disturbances in the troposphere, the energy propagation (i.e., the group velocity) must have an upward component. Therefore, according to the arguments of Section 5.4, the phase velocity must have a downward component. We showed in Section 11.4 that Kelvin waves must propagate eastward $(c>0)$ if they are to be trapped equatorially. However, eastward phase propagation requires $m<0$ for downward phase propagation. Thus, the vertically propagating Kelvin wave has phase lines that tilt eastward with height as shown in Figure 12.12.

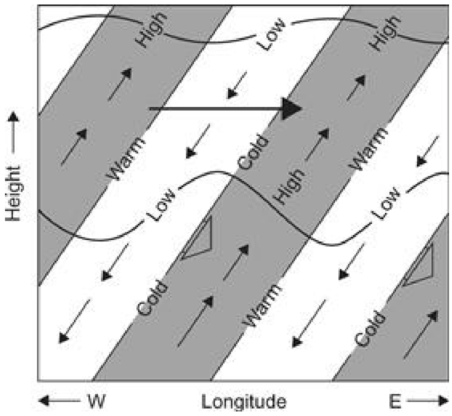


Figure 12.12 Longitude–height section along the equator showing pressure, temperature, and wind perturbations for a thermally damped Kelvin wave. *Heavy wavy lines* indicate material lines; *short black arrows* show phase propagation. Areas of high pressure are *shaded*. Length of the *small thin arrows* is proportional to the wave amplitude, which decreases with height due to damping. The *large shaded arrow* indicates the net mean flow acceleration due to the wave stress divergence.

12.5.2 Vertically Propagating Rossby–Gravity Waves

For all other equatorial modes, (12.34) through (12.37) can be combined in a manner exactly analogous to that described for the shallow water equations in Section 11.4.1. The resulting meridional structure equation is identical to (11.38) if we again assume that $m^2 \gg 1/4H^2$ and set

$$gh_e=N^2/m^2$$

For the $n=0$ mode the dispersion relation (11.41) then implies that

$$|m|=N\nu^{-2}\left(\beta+\nu k\right) \tag{12.42}$$

When $\beta=0$ we again recover the dispersion relationship for hydrostatic internal gravity waves. The role of the β effect in (12.42) is to break the symmetry between eastward $(\nu>0)$ and westward $(\nu<0)$ propagating waves. Eastward propagating modes have shorter vertical wavelengths than westward propagating modes. Vertically propagating $n=0$ modes can exist only for $c=\nu/k>-\beta/k^2$. Because $k=s/a$, where s is the number of wavelengths around a latitude circle, this condition implies that for $\nu < 0$ solutions exist only for frequencies satisfying the inequality

$$|\nu| < 2\Omega/s \tag{12.43}$$

For frequencies that do not satisfy (12.43), the wave amplitude will not decay away from the equator and it is not possible to satisfy boundary conditions at the pole.

After some algebraic manipulation, the meridional structure of the horizontal velocity and geopotential perturbations for the $n = 0$ mode can be expressed as

$$\begin{pmatrix} \hat{u} \\ \hat{v} \\ \hat{\Phi} \end{pmatrix} = v_0 \begin{pmatrix} i\,|m|\,N^{-1}\nu y \\ 1 \\ i\nu y \end{pmatrix} \exp\left(-\frac{\beta\,|m|\,y^2}{2N}\right) \tag{12.44}$$

The westward-propagating $n = 0$ mode is generally referred to as the Rossby–gravity mode.[2] For upward energy propagation this mode must have downward phase propagation ($m < 0$) just like an ordinary westward-propagating internal gravity wave. The resulting wave structure in the x,z plane at a latitude north of the equator is shown in Figure 12.13. Of particular interest is the fact that poleward-moving air has positive temperature perturbations and vice versa so that the eddy heat flux contribution to the vertical EP flux is positive.

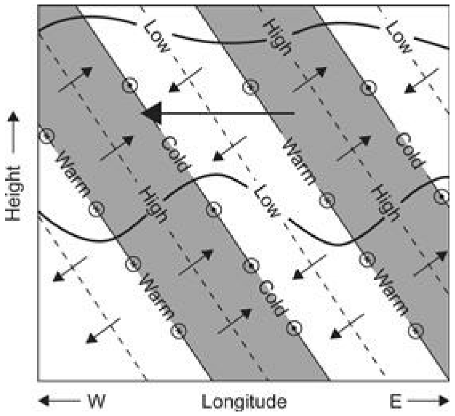


Figure 12.13 Longitude–height section along a latitude circle north of the equator showing pressure, temperature, and wind perturbations for a thermally damped Rossby–gravity wave. Areas of high pressure are *shaded*. *Small arrows* indicate zonal and vertical wind perturbations with length proportional to the wave amplitude. Meridional wind perturbations are shown by *arrows pointed into* the page (northward) and *out of the page* (southward). The *large shaded arrow* indicates the net mean flow acceleration due to the wave stress divergence.

12.5.3 Observed Equatorial Waves

Both Kelvin and Rossby–gravity modes have been identified in observational data from the equatorial stratosphere. The observed stratospheric Kelvin waves are primarily of zonal wave number $s = 1$ and have periods in the range of 12 to 20 days. An example of zonal wind oscillations caused by the passage of Kelvin waves at a station near the equator is shown in the form of a time–height section in Figure 12.14. During the observational period shown in Figure 12.14, the westerly phase of the quasi-biennial oscillation (see Section 12.6) is descending so that at each level there is a general increase in the mean zonal wind with increasing time. However, superposed on this secular trend is a large fluctuating component with a period between speed maxima of about 12 days and a vertical wavelength (computed from the tilt of the oscillations with height) of about 10 to 12km.

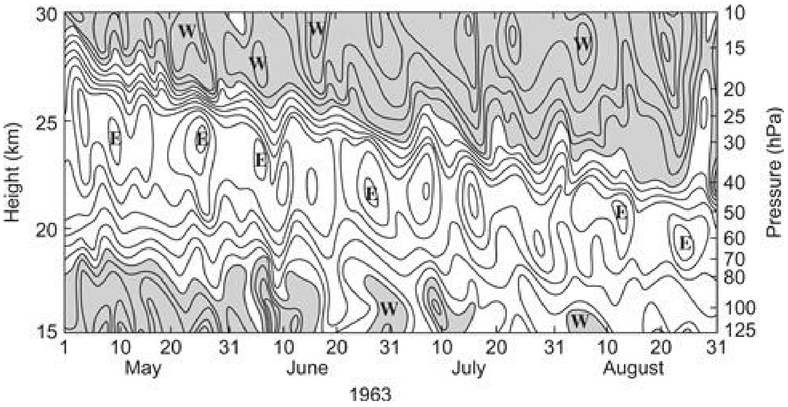


Figure 12.14 Time–height section of zonal wind at Canton Island (3° S). Isotachs at intervals of 5ms^{-1}. Westerlies are *shaded.*

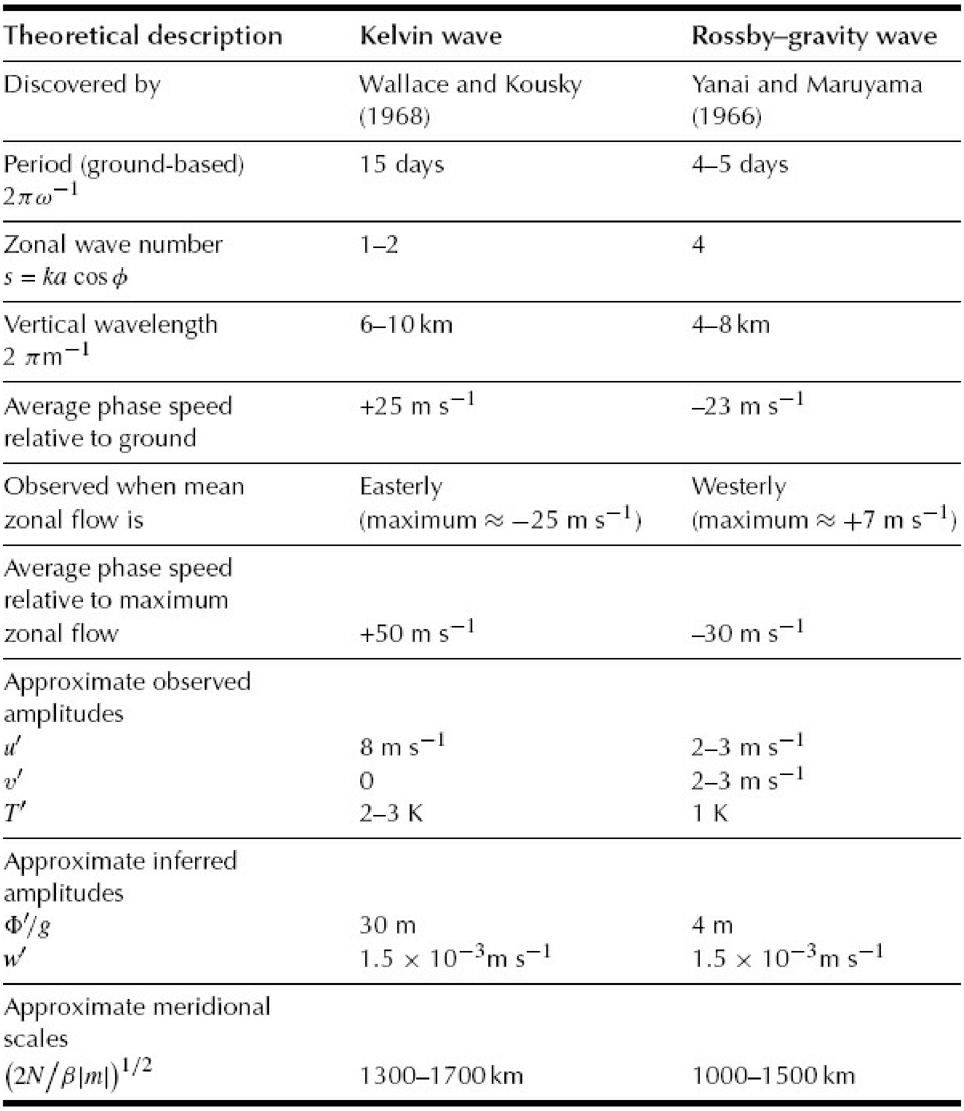
Courtesy of J. M. Wallace and V. E. Kousky.

Observations of the temperature field for the same period reveal that the temperature oscillation leads the zonal wind oscillation by 1/4 cycle (i.e., maximum temperature occurs prior to maximum westerlies), which is just the phase relationship required for upward-propagating Kelvin waves (see Figure 12.12). Additional observations from other stations indicate that these oscillations do propagate eastward at the theoretically predicted speed. Therefore, there can be little doubt that the observed oscillations are Kelvin waves.

The existence of the Rossby–gravity mode has been confirmed in observational data from the stratosphere in the equatorial Pacific. This mode is identified most easily in the meridional wind component, as v' is a maximum at the equator for the Rossby–gravity mode. The observed Rossby–gravity waves have $s = 4$, vertical wavelengths in the range of 6 to 12km, and a period range of 4 to 5 days. Kelvin and Rossby–gravity waves each have significant amplitude only within about 20° latitude of the equator.

A more complete comparison of observed and theoretical properties of the Kelvin and Rossby–gravity modes is presented in Table 12.1. In comparing theory and observation, it must be recalled that it is the frequency relative to the mean flow, not relative to the ground, that is dynamically relevant.

Table 12.1 Characteristics of the Dominant Observed Planetary-Scale Waves in the Equatorial Lower Stratosphere



Theoretical description	Kelvin wave	Rossby–gravity wave
Discovered by	Wallace and Kousky (1968)	Yanai and Maruyama (1966)
Period (ground-based) $2\pi\omega^{-1}$	15 days	4–5 days
Zonal wave number $s = ka\cos\phi$	1–2	4
Vertical wavelength $2\,\pi m^{-1}$	6–10 km	4–8 km
Average phase speed relative to ground	$+25$ m s^{-1}	-23 m s^{-1}
Observed when mean zonal flow is	Easterly (maximum ≈ -25 m s^{-1})	Westerly (maximum $\approx +7$ m s^{-1})
Average phase speed relative to maximum zonal flow	$+50$ m s^{-1}	-30 m s^{-1}
Approximate observed amplitudes		
u'	8 m s^{-1}	2–3 m s^{-1}
v'	0	2–3 m s^{-1}
T'	2–3 K	1 K
Approximate inferred amplitudes		
Φ'/g	30 m	4 m
w'	1.5×10^{-3} m s^{-1}	1.5×10^{-3} m s^{-1}
Approximate meridional scales $(2N/\beta\lvert m\rvert)^{1/2}$	1300–1700 km	1000–1500 km

From Andrews et al., 1987.

It appears that Kelvin and Rossby–gravity waves are excited by oscillations in the large-scale convective heating pattern in the equatorial troposphere. Although these waves do not contain much energy compared to typical tropospheric weather disturbances, they are the predominant disturbances of the equatorial stratosphere and, through their vertical energy and momentum transport, play a crucial role in the general circulation of the stratosphere. In addition to the stratospheric modes considered here, there are higher-speed Kelvin and Rossby–gravity modes, which are important in the upper stratosphere and mesosphere. There is also a broad spectrum of equatorial gravity waves, which appears to be important for the momentum balance of the equatorial middle atmosphere.

12.6 The Quasi-Biennial Oscillation

The search for periodic oscillations in the atmosphere has a long history. Aside, however, from the externally forced diurnal and annual components and their harmonics, no compelling evidence exists for truly periodic atmospheric oscillations. The phenomenon that perhaps comes closest to exhibiting periodic behavior not associated with a periodic forcing function is the quasi-biennial oscillation (QBO) in the mean zonal winds of the equatorial stratosphere. This oscillation has the following observed features.

1. Zonally symmetric easterly and westerly wind regimes alternate regularly with periods varying from about 24 to 30 months.
2. Successive regimes first appear above 30km but propagate downward at a rate of 1km mo^{-1}.
3. Downward propagation occurs without loss of amplitude between 30 and 23km, but there is rapid attenuation below 23km.
4. Oscillation is symmetric about the equator with a maximum amplitude of about 20 m s^{-1} and an approximately Gaussian distribution in latitude with a half-width of about 12°.

This oscillation is best depicted by means of a time–height section of the zonal wind speed at the equator as shown in Figure 12.15. It is apparent from this figure that the vertical shear of the wind is quite strong at the level where one regime is replacing the other. Because the QBO is zonally symmetric and causes only very small mean meridional and vertical motions, the QBO mean zonal wind and temperature fields satisfy the thermal wind balance equation. For the equatorial β plane this has the form—compare with (10.13):

$$\beta y \partial \overline{u}/\partial z = -RH^{-1}\partial \overline{T}/\partial y$$

For equatorial symmetry $\partial \overline{T}/\partial y = 0$ at $y = 0$, and by L'Hopital's rule thermal wind balance at the equator has the form

$$\partial \overline{u}/\partial z = -R(H\beta)^{-1}\partial^2 \overline{T}/\partial y^2 \tag{12.45}$$

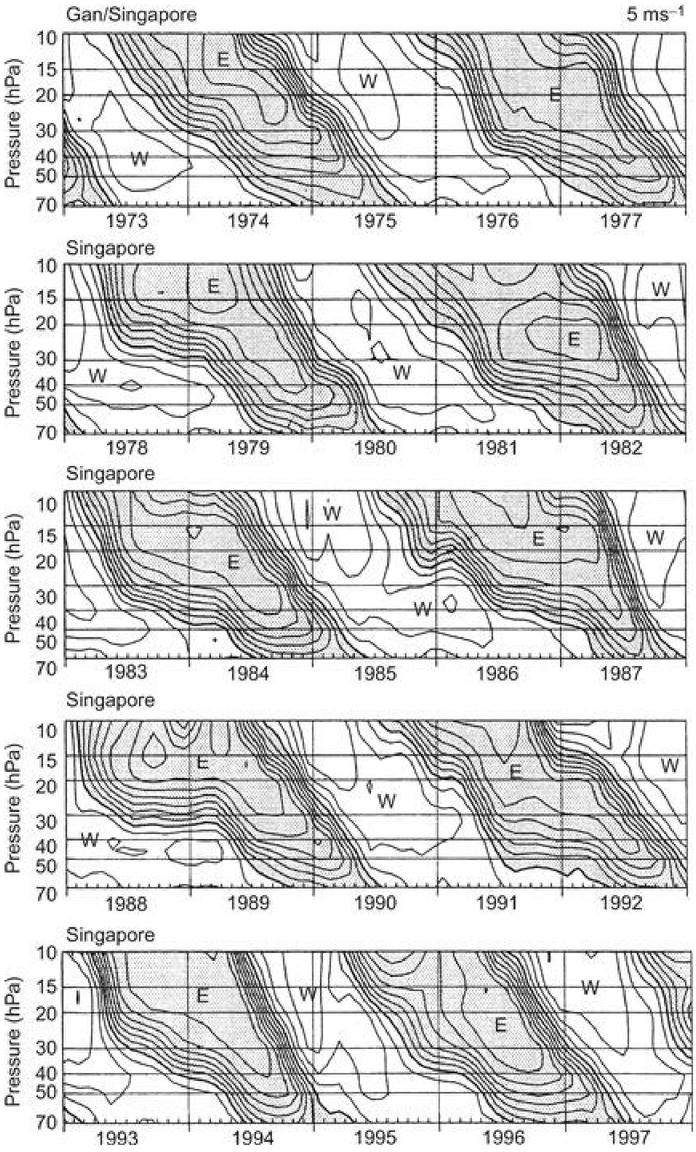


Figure 12.15 Time–height section of departure of monthly mean zonal winds (m s^{-1}) for each month from the long-term average for that month at equatorial stations. Note the alternating downward-propagating westerly (W) and easterly (E) regimes.

After Dunkerton, 2003. Data provided by B. Naujokat.

Equation (12.45) can be used to estimate the magnitude of the QBO temperature perturbation at the equator. The observed magnitude of vertical shear of the mean zonal wind at the equator is ~ 5 m s^{-1} km^{-1}, and the meridional scale is $\sim$ 1200 km, from which (12.45) shows that the temperature perturbation has an amplitude $\sim$ 3 K at the equator. Because the second derivative of temperature has the opposite sign to that of the temperature at the equator, the westerly and easterly shear zones have warm and cold equatorial temperature anomalies, respectively.

The main factors that a theoretical model of the QBO must explain are the approximate biennial periodicity, the downward propagation without loss of amplitude, and the occurrence of zonally symmetric westerlies at the equator. Because a zonal ring of air in westerly motion at the equator has an angular momentum per unit mass greater than that of Earth, no plausible zonally symmetric advection process could account for the westerly phase of the oscillation. Therefore, there must be a vertical transfer of momentum by eddies to produce the westerly accelerations in the downward-propagating shear zone of the QBO.

Observational and theoretical studies have confirmed that vertically propagating equatorial Kelvin and Rossby–gravity waves provide a significant fraction of the zonal momentum sources necessary to drive the QBO. From Figure 12.12 it is clear that Kelvin waves with upward energy propagation transfer westerly momentum upward (i.e., u' and w' are positively correlated so that $u'w' > 0$). Thus, the Kelvin waves can provide a source of westerly momentum for the QBO.

Vertical momentum transfer by the Rossby–gravity mode requires special consideration. Examination of Figure 12.13 shows that $\overline{u'w'} > 0$ also for the Rossby–gravity mode. The total effect of this wave on the mean flow cannot be ascertained from the vertical momentum flux alone, but rather the complete vertical EP flux must be considered. This mode has a strong poleward heat flux ($\overline{v'T'} > 0$), which provides an upward-directed EP flux contribution. This dominates over the vertical momentum flux, and the net result is that the Rossby–gravity mode transfers easterly momentum upward and can provide a momentum source for the easterly phase of the QBO. Observed Kelvin and Rossby–gravity wave momentum fluxes are not, however, sufficient to account for the observed zonal accelerations of the QBO. Additional wave sources, such as gravity waves generated by convective storms, must also contribute to the forcing of the QBO.

It was pointed out in Section 12.4 that quasi-geostrophic wave modes do not produce any net mean flow acceleration unless the waves are transient or are damped mechanically or thermally. Similar considerations apply to gravity waves and to the equatorial Kelvin and Rossby–gravity modes. Stratospheric waves are subject to thermal damping by infrared radiation, and to both thermal and damping by small-scale turbulent motions. Such damping is strongly dependent on the Doppler-shifted frequency of the waves.

As the Doppler-shifted frequency decreases, the vertical component of group velocity also decreases, and there is a longer time available for the wave energy to be damped as it propagates through a given vertical distance. Thus, eastward-propagating gravity waves and Kelvin waves tend to be damped preferentially in westerly shear zones, where their Doppler-shifted frequencies decrease with height. The momentum flux convergence associated with this damping provides a westerly acceleration of the mean flow, and thus causes the westerly shear zone to descend. Similarly, the westward-propagating gravity waves and Rossby–gravity waves are damped in easterly shear zones, thereby causing an easterly acceleration and descent of the easterly shear zone. We conclude that the QBO is indeed excited primarily by vertically propagating waves through wave transience and damping, which causes westerly accelerations in westerly shear zones and easterly accelerations in easterly shear zones.

This process of wave, mean-flow interaction can be elucidated by considering the heavy wavy lines shown earlier in Figures 12.12 and 12.13. These lines indicate the vertical displacement of horizontal surfaces of fluid parcels (material surfaces) by the velocity field associated with the waves. (For sufficiently weak thermal damping they are approximately the same as isentropic surfaces.) The wavy lines show that the maximum upward displacement occurs 1/4 cycle after the maximum upward perturbation velocity. In the Kelvin wave case (Figure 12.12), positive pressure perturbations coincide with negative material surface slopes. Thus, the fluid below a wavy material line exerts an eastward-directed pressure force on the fluid above. Because the wave amplitude decreases with height, this force is larger for the lower of the two material lines in Figure 12.12. There will thus be a net westerly acceleration of the block of fluid contained between the two wavy material lines shown in Figure 12.12.

For the Rossby–gravity wave, however, positive pressure perturbations coincide with positive slopes of the material lines. There is thus a westward-directed force exerted by the fluid below the lines on the fluid above as shown in Figure 12.13. In this case the result is a net easterly acceleration of the fluid contained between the two wavy material lines shown in Figure 12.13. Thus, by considering the stresses acting across a material surface corrugated by the waves, it is possible to deduce the mean-flow acceleration caused by the waves without explicit reference to the EP fluxes of the waves.

How such a mechanism can cause a mean zonal flow oscillation when equal amounts of easterly and westerly momentum are transferred upward across the tropopause by the waves can be seen qualitatively by considering Figure 12.16. If initially, as shown in Figure 12.16a, the mean zonal wind is weak and westerly, the eastward-propagating waves will be preferentially damped at lower altitude and produce a westerly acceleration, which will move downward in time as the mean westerlies intensify and the wave-driven acceleration is concentrated at ever lower altitudes. The westward-propagating waves, however, initially penetrate to higher altitudes, where they produce easterlies, which also move downward in time. Eventually the westerlies are damped out as they approach the tropopause, and the stratosphere is dominated by easterlies so that the eastward-propagating waves can penetrate to high altitude and produce a new westerly phase. In this manner the mean zonal wind is forced to oscillate back and forth between westerlies and easterlies with a period that depends primarily on the vertical momentum transport and other properties of the waves, not on an oscillating external forcing.

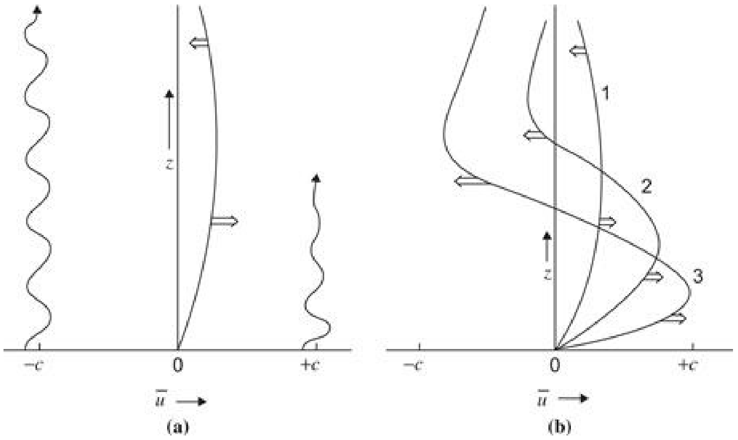


Figure 12.16 Schematic representation of wave-driven accelerations that lead to the zonal wind QBO. Eastward- and westward-propagating gravity waves of phase speeds $+c$ and $-c$, respectively, propagate upward and are dissipated at rates dependent on the Doppler-shifted frequency. (a) Initial weak westerly current selectively damps the eastward-propagating wave and leads to westerly acceleration at lower levels and easterly acceleration at higher levels. (b) Descending westerly shear zones block penetration of eastward-propagating waves, whereas westward-propagating waves produce descending easterlies aloft. *Broad arrows* show locations and direction of maxima in mean wind acceleration. *Wavy lines* indicate relative penetration of waves.

After Plumb, 1982.

12.7 Trace Constituent Transport

The study of global transport involves the motion of atmospheric tracers, which are defined as chemical or dynamical quantities that label fluid parcels. Chemical tracers consist of minor atmospheric species that have significant spatial variability in the atmosphere. Dynamical tracers (potential temperature and potential vorticity) are properties of the flow field that are conserved following the motion under certain conditions. These also can be useful for interpretation of transport.

12.7.1 Dynamical Tracers

Potential temperature, defined in (2.44), can be thought of as a label for the vertical position of an air parcel. Because the atmosphere is stably stratified, potential temperature increases monotonically with height (slowly in the troposphere and rapidly in the stratosphere as shown in Figure 12.17) and thus can be used as an independent vertical coordinate: the isentropic coordinates introduced in Section 4.6.1. A parcel moving adiabatically remains on a surface of constant potential temperature and can be "tagged" by its value of potential temperature. Thus the motion of such a parcel is two-dimensional when viewed in isentropic coordinates. Potential temperature surfaces are quasi-horizontal, but they do move up and down in physical space with adiabatic changes in temperature. Thus, it is useful to reference trace constituent data to potential temperature, rather than pressure or altitude, as reversible variations at local altitudes or pressures caused by adiabatic vertical displacements associated with transient motions (e.g., gravity waves) are then accounted for.

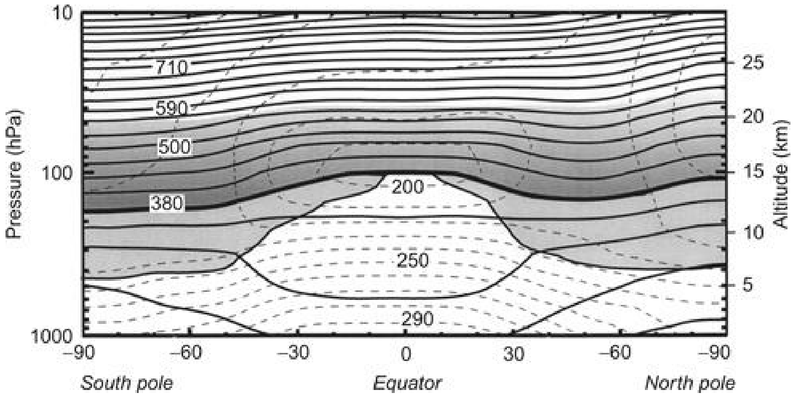


Figure 12.17 Latitude–height section showing January mean potential temperature surfaces (*solid contours*) and temperature (*dashed contours*). Heavy line shows the 380-K θ surface, above which all θ surfaces lie completely in the middle atmosphere. The *light shaded* region below the 380 K surface is the lowermost stratosphere, where Θ surfaces span the tropopause (shown by the *line* bounding the lower side of the shaded region).

Courtesy of C. Appenzeller.

The other commonly used dynamical tracer, potential vorticity (PV), is conserved for adiabatic, frictionless flow. As defined in equation (4.25), PV normally has a positive gradient in the meridional direction on an isentropic surface; it is negative in the Southern Hemisphere, zero at (or near) the equator, and positive in the Northern Hemisphere. PV has strong gradients in both height and latitude. Because a parcel moving in adiabatic frictionless flow conserves PV as well as potential temperature, its motion must lie parallel to isocontours of PV on isentropic surfaces. Rapid meridional transport implies production of strong PV anomalies as high polar values of PV are advected equatorward or low equatorial values are advected poleward. As discussed earlier, the meridional gradient of the background PV field resists meridional displacements through the production of Rossby waves. Thus, regions of strong PV gradients on isentropic surfaces can act as semipermeable "barriers" to transport. This PV barrier effect is one of the reasons that eddy diffusion is often not a good model for meridional transport.

Because PV depends only on the distribution of horizontal winds and temperatures on isentropic surfaces, its distribution can be determined from conventional meteorological observations. Thus, the evolution of PV on an isentropic surface can be used as a surrogate for the study of trace constituent transport on isentropic surfaces in circumstances in which there are inadequate observations of relevant chemical tracers. In addition, for dynamical studies, PV has the important property, not shared by other tracers, that it not only is advected by the flow, but actually determines the flow field. Thus, the distribution of PV on isentropic surfaces can be “inverted” to yield the wind and temperature fields. Changes in the PV distribution are then said to “induce” changes in the wind and temperature fields. In the quasi-geostrophic case the induced changes are required to maintain geostrophic and hydrostatic balance. More generally, such changes preserve a higher-order balance among the dynamical fields.

12.7.2 Chemical Tracers

Much can be learned about the nature of large-scale transport by considering the climatological distribution of quasi-conservative chemical tracers in the atmosphere. The distribution of such trace substances is dependent on the competition between dynamical and chemical processes. This competition can be approximately measured by comparing the characteristic dynamical and chemical timescales. *Chemical timescale* is the average time for replacement or removal of a tracer due to chemical sources and sinks. *Dynamical timescale* is the average time for advective and diffusive processes to transport the tracer from equator to pole or across a scale height in the vertical. The role of transport in determining tracer climatology depends on the nature and distribution of tracer sources and sinks, and on the relative magnitudes of the timescales for dynamical and chemical processes.

If the chemical timescale is much shorter than the dynamical timescale, the tracer will be in photochemical equilibrium and transport does not directly influence its distribution. Transport can, however, play an important indirect role by partly determining the concentrations of other species that participate in the photochemical production or loss of the tracer in question.

If the chemical timescale is much longer than the dynamical timescale, the tracer will be passively advected by the flow field. In the absence of localized sources and sinks, such a tracer would eventually become well mixed due to the dispersive effects of transport. It is for this reason that species such as nitrous oxide (N_2O) tend to have uniform concentrations in the troposphere and are thus not useful for tropospheric transport studies.

When the chemical and dynamical timescales are comparable, the observed species concentration depends on the net effects of chemical sources and sinks and of transport. In many cases of interest, the ratio of chemical and dynamical timescales changes drastically between the troposphere and the stratosphere, and with altitude in the stratosphere. Tracers that have long lifetimes in the troposphere and the lower stratosphere are referred to as *long-lived tracers*, as the bulk of the tracer mass is contained in the troposphere and lower stratosphere, and the lifetime is determined mainly by the (slow) flux of the tracer to altitudes where the chemical timescales become comparable to or shorter than dynamical timescales.

12.7.3 Transport in the Stratosphere

Transport processes are conveniently divided between those that involve mean motions of the atmosphere, or *advection*, and those that may be characterized as turbulent, or diffusive in nature. In the case of point sources, such as volcanic eruptions, the distinction is quite clear; advection moves the center of mass of the plume along the direction of the average wind, whereas turbulent diffusion disperses the plume in the plane orthogonal to the average wind. On a global scale, however, the distinction between advective and diffusive processes is not always clear. Because the atmosphere is characterized by spatially and temporally varying motions with a wide range of scales, there is no obvious physical separation between “mean” and “turbulent” motions.

In practice, those transport processes that are explicitly resolved by the particular observational network or transport model being utilized are often regarded as the advective motions, whereas the remaining unresolved motions are assumed to be diffusive. The effects of the unresolved motions must then be parameterized in terms of the mean motions in some fashion. Usually this involves assuming that tracer fluxes by the unresolved motions are proportional to the gradient of the resolved tracer distribution. However, this approach is not always physically justified. A major problem in the modeling of global transport is accurate representation of the contribution of the unresolved eddy motions to the total transport.

As discussed in Section 12.2, the global-scale residual meridional circulation in the middle atmosphere is driven by wave-induced zonal forces associated with Rossby waves and gravity waves. Not surprisingly, the residual circulation plays an essential role in the meridional and vertical transport of trace chemical constituents in the middle atmosphere. Additionally, the waves responsible for the zonal force that drives the residual circulation are also responsible for the quasi-isentropic stirring and mixing that is associated with wavebreaking. Thus, understanding of transport involves both eddy and mean-flow transport effects.

In dynamical studies it is usual to characterize a chemical constituent by the volume mixing ratio (or mole fraction), defined as $\chi \equiv n_T/n_A$, where n_T and n_A designate the number densities (molecules m^{-3}) for the trace constituent and air, respectively. The mixing ratio is conserved following the motion in the absence of sources and sinks and thus satisfies the simple tracer continuity equation

$$\frac{D\chi}{Dt} = S \tag{12.46}$$

where S designates the sum of all chemical sources and sinks.

As in the case of the dynamical variables discussed in Section 12.2, it is useful to define a longitudinally averaged mixing ratio $\overline{\chi}$ and a disturbance or eddy ratio χ' such that $\chi = \overline{\chi} + \chi'$. Again, it proves useful to use the residual mean meridional circulation $(\overline{v}^*, \overline{w}^*)$ defined in (10.16a,b). The zonal mean tracer continuity equation in the TEM framework can then be written as

$$\frac{\partial \overline{\chi}}{\partial t} + \overline{v}^* \frac{\partial \overline{\chi}}{\partial y} + \overline{w}^* \frac{\partial \overline{\chi}}{\partial z} = \overline{S} + \frac{1}{\rho_0} \boldsymbol{\nabla} \cdot \mathbf{M} \tag{12.47}$$

Here, **M** represents the diffusive effects of the eddies, plus advective effects not represented by the residual meridional circulation. In models the term involving **M** is often represented by meridional and vertical eddy diffusion, with empirically determined diffusion coefficients.

To appreciate the role of the wave-induced global circulation in determining the distribution of long-lived tracers in the middle atmosphere, it is useful to consider a hypothetical atmosphere in which there are no wave motions and thus no wave-induced zonal force. In that case, as argued in Section 12.2, the middle atmosphere would relax to radiative equilibrium, the residual circulation would vanish, and the distribution of the tracer would be determined at each altitude by a balance between slow upward diffusion and photochemical destruction. Thus, tracer mixing ratio surfaces would, in an annual mean, tend to be close to horizontal. This is to be contrasted to observed distributions, which are characterized by mixing ratio surfaces that bow upward in the tropics and slope downward toward both poles (e.g., Figure 12.9).

As discussed in Section 12.2, the wave-induced global-scale circulation consists of upward and poleward motion across the isentropes in low latitudes, accompanied by diabatic heating, and downward motion across the isentropes at high latitudes, accompanied by diabatic cooling. Actual parcel trajectories, of course, do not follow the zonally averaged motion, but are influenced by the three-dimensional wave motion. Nevertheless the diabatic circulation defined by the mean diabatic heating and cooling closely approximates the global transport circulation. For seasonal and longer timescales the TEM residual circulation generally provides a good approximation to the diabatic circulation and is generally simpler to compute from standard meteorological analyses. For shorter-period phenomena in which the temperature tendency is large, the residual circulation is no longer a good approximation to the diabatic circulation.

The previous conceptual model of global transport is clearly supported by long-lived tracer observations as shown earlier in Figure 12.9. In middle latitudes there are regions in which tracer mixing ratio isopleths are nearly horizontal, reflecting the horizontal homogenizing role of meridional dispersion by planetary wavebreaking in the surf zone. The same processes also tend to homogenize the PV distribution on isentropes in the so-called surf zone in midlatitudes where Rossby wavebreaking tends to occur. The surf zone is bounded at both low and high latitudes by regions of strong meridional gradients of long-lived tracers and of PV. The existence of such gradients is evidence that there is only weak mixing into and out of the tropics and into and out of the polar winter vortex. Thus, these locations are sometimes referred to as “transport barriers." The strong PV gradients, strong winds, and strong wind shears that occur along the transport barriers at the subtropical and polar edges of the surf zone all act to suppress wavebreaking, and thus to minimize mixing and sustain the strong gradients at those locations.

Suggested References

Andrews et al. Andrews, Holton, and Leovy, *Middle Atmosphere Dynamics* present a graduate-level treatment of the dynamics of the stratosphere and mesosphere.

Brasseur and Solomon. Brasseur and Solomon’s *Aeronomy of the Middle Atmosphere* has an excellent discussion of the chemistry of the stratosphere at an advanced level.

Problems

12.1. Suppose that temperature increases linearly with height in the layer between 20 and 50 km at a rate of 2 $\mathrm{K\,km^{-1}}$. If the temperature is 200 K at 20 km, find the value of the scale height H for which the log-pressure height z coincides with actual height at 50 km. (Assume that z coincides with the actual height at 20 km and let $g = 9.81\ \mathrm{m\,s^{-2}}$ be a constant.)

12.2. Find the Rossby critical velocities for zonal wave numbers 1, 2, and 3 (i.e., for 1, 2, and 3 wavelengths around a latitude circle). Let the motion be referred to a β plane centered at 45° N, scale height $H = 7$ km, buoyancy frequency $N = 2 \times 10^{-2}\ \mathrm{s^{-1}}$, and infinite meridional scale ($l = 0$).

12.3. Suppose that a stationary linear Rossby wave is forced by flow over sinusoidal topography with height $h(x) = h_0 \cos(kx)$, where h_0 is a constant and k is the zonal wave number. Show that the lower boundary condition on the streamfunction ψ can be expressed in this case as

$$(\partial\psi/\partial z) = -hN^2/f_0$$

Using this boundary condition and an appropriate upper boundary condition, solve for $\psi(x,z)$ in the case $|m| \gg (1/2H)$ using the equations of Section 12.3.1. How does the position of the trough relative to the mountain ridge depend on the sign of m^2 for the limit $|m| \gg (1/2H)$.

12.4. Consider a very simple model of a steady-state mean meridional circulation for zonally symmetric flow in a midlatitude channel bounded by walls at $y=0,\ \pi/l$ and $z=0,\ \pi/m$. We assume that the zonal mean zonal flow $\overline{u}$ is in thermal wind balance and that the eddy momentum and heat fluxes vanish. For simplicity, we let $\rho_0 = 1$ (Boussinesq approximation) and let the zonal force due to small-scale motions be represented by a linear drag: $\overline{X} = -\gamma\overline{u}$. We assume that the diabatic heating has the form $J/c_p = (H/R)J_0 \cos ly \sin mz$, and we let N and f be constants. Equations (12.1) through (12.4) then yield the following:

$$-f_0\overline{v}^* = -\gamma\overline{u}$$

$$+N^2HR^{-1}\overline{w}^* = +\overline{J}/c_p$$

$$\overline{v}^* = -\frac{\partial\overline{\chi}^*}{\partial z};\quad \overline{w}^* = \frac{\partial\overline{\chi}^*}{\partial y}$$

$$f_0\partial\overline{u}/\partial z + RH^{-1}\partial\overline{T}/\partial y = 0$$

Assuming that there is no flow through the walls, solve for the residual circulation defined by $\overline{\chi}^*$, $\overline{v}^*$, and $\overline{w}^*$.

12.5. For the situation of Problem 12.4, solve for the steady-state zonal wind and temperature fields $\overline{u}$ and $\overline{T}$.

12.6. Find the geopotential and vertical velocity fluctuations for a Kelvin wave of zonal wave number 1, phase speed 40 m s^{-1}, and zonal velocity perturbation amplitude 5 m s^{-1}. Let $N^2 = 4 \times 10^{-4}$ s^{-2}.

12.7. For the situation of Problem 12.6 compute the vertical momentum flux $M \equiv \rho_0\overline{u'w'}$. Show that M is constant with height.

12.8. Determine the form for the vertical velocity perturbation for the Rossby–gravity wave corresponding to the u', v', and Φ' perturbations given in (12.44).

12.9. For a Rossby–gravity wave of zonal wave number 4 and phase speed -20 m s^{-1}, determine the latitude at which the vertical momentum flux $M \equiv \rho_0\overline{u'w'}$ is a maximum.

12.10. Suppose that the mean zonal wind shear in the descending westerlies of the equatorial QBO can be represented analytically on the equatorial β plane in the form $\partial u/\partial z = \Lambda \exp\left(-y^2/L^2\right)$ where L = 1200 km. Determine the approximate meridional dependence of the corresponding temperature anomaly for $|y| \ll L$.

12.11. Estimate the TEM residual vertical velocity in the westerly shear zone of the equatorial QBO assuming that radiative cooling can be approximated by Newtonian cooling with a 20-day relaxation time, that the vertical shear is 20 m s^{-1} per 5km, and that the meridional half-width is 12° latitude.

MATLAB Exercises

M12.1. The MATLAB script **topo_Rossby_wave.m** plots solutions for various fields for a stationary linear Rossby wave forced by flow over an isolated ridge. A β–plane channel model is used following the discussion in Section 12.3.1. Run the script for mean wind values of 5, 10, 15, 20, and 25 m s^{-1}. Describe how the geopotential and potential temperature perturbations change with changing mean wind speed. Explain these results based on the arguments given in Section 12.3.1.

M12.2. For the situation of Problem 12.4 let $J_0 = 10^{-6}\ s^{-3}$, $N = 10^{-2}\ s^{-1}$, $f = 10^{-4}\ s^{-1}$, $l = 10^{-6}$ m^{-1}, $m = \pi/H$, where $H = 10^4$ m, and $\gamma\ 10^{-5}\ s^{-1}$. Make contour plots of the fields $\overline{u}, \overline{v}^*, \overline{w}^*, \overline{T}$ and discuss the relationship of these fields to the imposed diabatic heating.

M12.3. The MATLAB script **sudden_warming_model.m** provides an analog to the sudden stratospheric warmings on an extratropical β plane centered at 60° N. The amplitude of a single planetary wave of zonal wave number $s = 1$ or $s = 2$ is specified at the bottom boundary (taken to be the 16-km level), and a time integration is carried out to determine the evolution of the zonal mean wind in the stratosphere forced by the EP flux of the specified wave. For the cases $s = 1$ and 2, run the model for geopotential heights between 100 and 400 m at a 50-m interval. For each case note whether the flow tends toward steady state or repeated sudden warmings. Modify the code to plot the time–height evolution of the EP flux divergence for the case $s = 2$ and forcing of 200 m.

M12.4. The MATLAB script **qbo_model.m** is a simplified one-dimensional model of the equatorial QBO originally introduced by Plumb (1977). In this model the mean zonal flow is forced by two waves of equal amplitude at the lower boundary and equal and opposite phase speeds. The initial mean wind profile has a weak westerly shear. If the model is run for sufficiently weak forcing, the mean wind approaches a steady state, but for forcing greater than a critical amplitude, a downward-propagating oscillatory solution occurs, with period dependent on the amplitude of forcing. Run the script for a range of forcing between 0.04 and 0.40 and determine the approximate minimum forcing for a large-amplitude oscillatory motion to occur, and the dependence of period on the forcing amplitude. Modify the script to compute the time mean of the zonal wind over the course of several oscillations. Try a case in which the eastward forcing is 0.15 and the westward forcing is 0.075. How does the time mean wind change in this case?

---

[1] As in Chapter 10, we express the log-pressure coordinate simply as z rather than z^*.

[2] Some authors use this term to describe both eastward and westward $n = 0$ waves.

Chapter 13

Numerical Modeling and Prediction

Dynamical meteorology provides the theoretical basis and methodology for modern weather forecasting. Stated simply, the objective of dynamical forecasting is to predict the future state of the atmospheric circulation from knowledge of its present state by use of numerical approximations to the dynamical equations. Fulfilling this objective requires observations of the initial state of the field variables, a closed set of prediction equations relating the field variables, and a method of integrating the equations in time to obtain the future distribution of the field variables.

Numerical prediction is a highly specialized field, which is continually evolving. Operational forecast centers utilize complex prediction models that require the largest available supercomputers to obtain a solution. It is difficult to provide more than a superficial introduction to such models in an introductory text. Fortunately, however, many aspects of numerical prediction can be illustrated using a simple model, such as the barotropic vorticity equation. In fact, this equation was the basis of the earliest operational numerical prediction models.

13.1 Historical Background

The British scientist L. F. Richardson made the first attempt to predict the weather numerically. His book, *Weather Prediction by Numerical Process*, published in 1922, is the classic treatise in this field. In his work Richardson showed how the differential equations governing atmospheric motions could be written approximately as a set of algebraic difference equations for values of the tendencies of various field variables at a finite number of points in space. Given the observed values of the field variables at these grid points, the tendencies could be calculated numerically by solving the algebraic difference equations.

By extrapolating the computed tendencies ahead a small increment in time, an estimate of the fields at a short time in the future could be obtained. The new values of the field variables could then be used to recompute the tendencies, which could in turn be used to extrapolate further ahead in time, and so on. Even for short-range forecasting over a small area of Earth, this procedure requires an enormous number of arithmetic operations. Richardson did not foresee the development of high-speed digital computers. He estimated that a workforce of 64,000 people would be required just to keep up with the weather on a global basis!

Despite the tedious labor involved, Richardson worked out one example forecast for surface pressure tendencies at two grid points. Unfortunately, the results were very poor. Predicted pressure changes were an order of magnitude larger than those observed. At the time this failure was thought to be due primarily to the poor initial data available, especially the absence of upper-air soundings. However, it is now known that there were other, even more serious problems with Richardson's scheme.

After Richardson's failure to obtain a reasonable forecast, numerical prediction was not again attempted for many years. Finally, after World War II interest in numerical prediction revived partly because of the vast expansion of the meteorological observation network, which provided much improved initial data. Even more important was the development of digital computers, which made the enormous volume of arithmetic operations required in a numerical forecast feasible. At the same time it was realized that Richardson's scheme was not the simplest possible scheme for numerical prediction. His equations not only governed the slow-moving meteorologically important motions, but also included high-speed sound and gravity waves as solutions. Such waves are in nature usually very weak in amplitude. However, for reasons that will be explained later, if Richardson had carried his numerical calculation beyond the initial time step, these oscillations would have amplified spuriously, thereby introducing so much "noise" in the solution that the meteorologically relevant disturbances would have been obscured.

The American meteorologist J. G. Charney showed in 1948 how the dynamical equations could be simplified by systematic introduction of the geostrophic and hydrostatic approximations so that sound and gravity oscillations were filtered out. The equations that resulted from Charney's filtering approximations were essentially those of the quasi-geostrophic model. Thus, Charney's approach utilized the conservative properties of potential vorticity. A special case of this model, the *equivalent barotropic model*, was used in 1950 to make the first successful numerical forecast.

This model provided forecasts of the geopotential near 500 hPa. Thus, it did not forecast "weather" in the usual sense. It could, however, be used by forecasters as an aid in predicting the local weather associated with large-scale circulations. Later multilevel versions of the quasi-geostrophic model provided explicit predictions of surface pressure and temperature distributions, but the accuracy of such predictions was limited because of the approximations inherent in the quasi-geostrophic model.

With the development of vastly more powerful computers and more sophisticated modeling techniques, numerical forecasting has now returned to models that are quite similar to Richardson's formulation and are far more accurate than quasi-geostrophic models. Nevertheless, it is still worth considering the simplest filtered model, the barotropic vorticity equation, to illustrate some of the technical aspects of numerical prediction in a simple context.

13.2 Numerical Approximation of the Equations of Motion

The equations of motion are an example of a general class of systems known as *initial value problems*. A system of differential equations is referred to as an initial value problem when the solution depends not only on boundary conditions but also on the values of the unknown fields or their derivatives at some initial time. Clearly, weather forecasting is a primary example of a *nonlinear* initial value problem. Due to its nonlinearity, even the simplest forecast equation, the barotropic vorticity equation, is rather complicated to analyze. Fortunately, general aspects of the numerical solution of initial value problems can be illustrated using linearized prototype equations that are much simpler than the barotropic vorticity equation.

13.2.1 Finite Differences

The equations of motion involve terms that are quadratic in the dependent variables (the advection terms). Such equations generally cannot be solved analytically. Rather, they must be approximated by some suitable discretization and solved numerically. The simplest form of discretization is the *finite difference* method.

To introduce the concept of finite differencing, we consider a field variable $\Psi(x)$, which is a solution to some differential equation in the interval $0 \leq x \leq L$. If this interval is divided into J subintervals of length δx, then $\psi(x)$ can be approximated by a set of $J + 1$ values as $\Psi_j = \psi(j\delta x)$, which are the values of the field at the $J + 1$ grid points given by $x = j\delta x, j = 0, 1, 2, \ldots, J,$ where

$\delta x = L/J$. Provided that δx is sufficiently small compared to the scale on which ψ varies, the $J + 1$ grid point values should provide good approximations to $\psi(x)$ and its derivatives.

The limits of accuracy of a finite difference representation of a continuous variable can be assessed by noting that the field can also be approximated by a finite Fourier series expansion:

(13.1)

$$\psi(x) = \frac{a_0}{2} + \sum_{m=1}^{J/2}\left[a_m \cos\frac{2\pi m x}{L} + b_m \sin\frac{2\pi m x}{L}\right]$$

The available $J + 1$ values of Ψ_j are just sufficient to determine the $J + 1$ coefficients in (13.1). That is, it is possible to determine a_0 plus a_m and b_m for wave numbers $m = 1, 2, 3, \ldots, J/2$. The shortest wavelength component in (13.1) has wavelength $L/m = 2L/J = 2\delta x$. Thus, the shortest wave that can be resolved by finite differencing has a wavelength twice that of the grid increment. Accurate representation of derivatives is only possible, however, for wavelengths much greater than $2\delta x$.

We now consider how the grid point variable Ψ_j can be used to construct a finite difference approximation to a differential equation. That is, we wish to represent derivatives such as $d\psi/dx \text{ and } d^2\psi/dx^2$ in terms of the finite difference fields. We first consider the Taylor series expansions about the point x_0:

(13.2)

$$\psi(x_0 + \delta x) = \psi(x_0) + \psi'(x_0)\delta x + \psi''(x_0)\frac{(\delta x)^2}{2} + \psi'''(x_0)\frac{(\delta x)^3}{6} + \mathrm{O}\left[(\delta x)^4\right]$$

(13.3)

$$\psi(x_0 - \delta x) = \psi(x_0) - \psi'(x_0)\delta x + \psi''(x_0)\frac{(\delta x)^2}{2} - \psi'''(x_0)\frac{(\delta x)^3}{6} + \mathrm{O}\left[(\delta x)^4\right]$$

where the primes indicate differentiation with respect to x, and $\mathrm{O}[(\delta x)^4]$ means that terms with order of magnitude $(\delta x)^4$ or less are neglected.

Subtracting (13.3) from (13.2) and solving for $\psi'(x)$ give a finite difference expression for the first derivative of the form

(13.4)

$$\psi'(x_0) = [\psi(x_0 + \delta x) - \psi(x_0 - \delta x)]/(2\delta x) + \mathrm{O}\left[(\delta x)^2\right]$$

while adding the same two expressions and solving for $\psi''(x)$ give a finite difference expression for the second derivative of the form

(13.5)

$$\psi''(x_0) = [\psi(x_0 + \delta x) - 2\psi(x_0) + \psi(x_0 - \delta x)]/(\delta x)^2 + \mathrm{O}\left[(\delta x)^2\right]$$

Because the difference approximations in (13.4) and (13.5) involve points at equal distances on either side of x_0, they are called *centered differences*. These approximations neglect terms of order $(\delta x)^2$. We thus say that the truncation error is order $(\delta x)^2$. Higher accuracy can be obtained by decreasing the grid interval, but at the cost of increasing the density of grid points. Alternatively, it is possible to obtain higher-order accuracy without decreasing the grid spacing by writing formulas analogous to (13.2) and (13.3) for the interval $2\delta x$ and using these together with (13.2) and (13.3) to eliminate error terms less than order $(\delta x)^4$. This approach, however, has the disadvantage of producing more complicated expressions and can be difficult to implement near boundary points.

13.2.2 Centered Differences: Explicit Time Differencing

As a prototype model, consider the linear one-dimensional advection equation

$$\partial q/\partial t + c\partial q/\partial x = 0 \tag{13.6}$$

with c a specified speed and $q(x, 0)$ a known initial condition. This equation can be approximated to second-order accuracy in x and t by the centered difference equation

$$\left[q(x, t+\delta t) - q(x, t-\delta t)\right]/(2\delta t) = -c\left[q\left(x+\delta x, t\right) - q\left(x-\delta x, t\right)\right]/(2\delta x) \tag{13.7}$$

The original differential equation (13.6) is thus replaced by a set of algebraic equations (13.7), which can be solved to determine solutions for a finite set of points that define a grid mesh in x and t (see Figure 13.1). For notational convenience it is useful to identify points on the grid mesh by indices m, s. These are defined by letting $x = m\delta x$, where $m = 0, 1, 2, 3, \ldots, M$, and $t = s\delta t$; $s = 0, 1, 2, 3, \ldots, S$. Writing $\hat{q}_{m,s} \equiv q(m\delta x, s\delta t)$, the difference equation (13.7) can then be expressed as

$$\hat{q}_{m,s+1} - \hat{q}_{m,s-1} = -\sigma\left(\hat{q}_{m+1,s} - \hat{q}_{m-1,s}\right) \tag{13.8}$$

where $\sigma \equiv c\delta t/\delta x$ is the *Courant number*. This form of time differencing is referred to as the *leapfrog* method, as the tendency at time step s is given by the difference in values computed for time steps $s + 1$ and $s - 1$ (i.e., by leaping across point s).

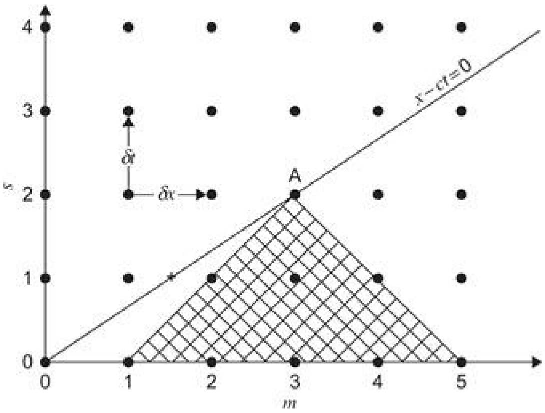


Figure 13.1 Grid in $x - t$ space showing the domain of dependence of the explicit finite-difference solution at $m = 3$ and $s = 2$ for the one-dimensional linear advection equation. *Solid circles* show grid points. The *sloping line* is a characteristic curve along which $q(x, t) = q(0, 0)$, and the “+” shows an interpolated point for the semi-Lagrangian differencing scheme. In this example the leapfrog scheme is unstable because the finite difference solution at point A does not depend on $q(0, 0)$.

Leapfrog differencing cannot be used for the initial time $t = 0$ ($s = 0$), as $\hat{q}_{m,-1}$ is not known. For the first time step, an alternative method is required such as the forward difference approximation

$$\hat{q}_{m,1} - \hat{q}_{m,0} \equiv -(\sigma/2)(\hat{q}_{m+1,0} - \hat{q}_{m-1,0}) \tag{13.9}$$

The centered difference scheme for the advection equation (13.8) is an example of an *explicit* time differencing scheme. In an explicit difference scheme, the value of the predicted field at a given grid point for time step $s + 1$ depends only on the known values of the field at previous time steps. (In the case of the leapfrog method the fields at time step s and $s - 1$ are used.) The difference equation can then be solved simply by marching through the grid and obtaining the solution for each point in turn. The explicit leapfrog scheme is thus simple to solve. However, as shown in the next section, it has disadvantages in that it introduces a spurious "computational" mode and has stringent requirements on the maximum value permitted for the Courant number. There are a number of alternative explicit schemes that do not introduce a computational mode (e.g., see Problem 13.3), but these still require that the Courant number be sufficiently small.

13.2.3 Computational Stability

Experience shows that solutions to finite difference approximations such as (13.8) will not always resemble solutions to the original differential equations even when the finite difference increments in space and time are very small. It turns out that the character of the solutions depends critically on the *computational stability* of the difference equations. If the difference equations are not stable, numerical solutions will exhibit exponential growth even when, as in the case of the linear advection equation, the original differential equation system has solutions whose amplitudes remain constant in time.

In the example (13.8), stability considerations place stringent limitations on the value of the parameter σ, as we show later. If the initial condition is specified as

$$q(x,0) = \mathrm{Re}\left[\exp(ikx)\right] = \cos(kx)$$

The analytic solution of (13.6) that satisfies this initial condition is

$$q(x,t) = \mathrm{Re}\left\{\exp\left[ik(x-ct)\right]\right\} = \cos(kx-ct) \tag{13.10}$$

We now compare (13.10) with the solution of the finite difference system (13.8) and (13.9). In finite difference form the initial condition is

$$\hat{q}_{m,0} = \exp(ikm\delta x) = \exp(ipm) \tag{13.11}$$

where $p \equiv k\delta x$. Noting that the analytic solution (13.10) is separable in x and t, we consider solutions of (13.8) and (13.9) of the form

$$\hat{q}_{m,s} = B^s \exp(ipm) \tag{13.12}$$

where B is a complex constant. Substituting into (13.8) and dividing through by the common factor B^{s-1}, we obtain a quadratic equation in B:

$$B^2 + (2i\sin\theta_p)B - 1 = 0 \tag{13.13}$$

where sin $\theta_p \equiv \sigma \sin p$. Equation (13.13) has two roots, which may be expressed in the form

$$B_1 = \exp(-i\theta_p), \quad B_2 = -\exp(+i\theta_p)$$

The general solution of the finite difference equation is thus

$$\hat{q}_{m,s} = [CB_1^s + DB_2^s]\exp(ipm) = Ce^{i(pm-\theta_p s)} + D(-1)^s e^{i(pm+\theta_p s)} \tag{13.14}$$

where C and D are constants to be determined by the initial conditions (13.11) and the first time step (13.9). The former gives $C + D = 1$, whereas the latter gives

$$Ce^{-i\theta_p} - De^{+i\theta_p} = 1 - i\sin\theta_p \tag{13.15}$$

Thus,

$$C = \frac{(1+\cos\theta_p)}{2\cos\theta_p}, \quad D = -\frac{(1-\cos\theta_p)}{2\cos\theta_p} \tag{13.16}$$

From inspection of (13.14) it is clear that the solution will remain finite for $s \to \infty$ provided that θ_p is real. If θ_p is imaginary, then one term in (13.14) will grow exponentially and the solution will be unbounded for $s \to \infty$. This sort of behavior is referred to as *computational instability*. Now because

$$\theta_p = \sin^{-1}(\sigma \sin p) \tag{13.17}$$

θ_p will be real only for $|\sigma \sin p| \leq 1$, which can only be valid for all waves (i.e., for all p) if $\sigma \leq 1$. Thus, computational stability of the difference equation (13.8) requires that

$$\sigma = c\delta t/\delta x \leq 1 \tag{13.18}$$

which is referred to as the Courant–Friedrichs–Levy (CFL) stability criterion.

The CFL criterion for this example states that for a given space increment δx, the time step δt must be chosen so that the dependent field will be advected a distance less than one grid length per time step. The restriction on σ given by (13.18) can be understood physically by considering the character of the solution in the x, t plane as shown in Figure 13.1. In the situation shown in this figure, $\sigma = 1.5$. Examination of the centered difference system (13.8) shows that the numerical solution at point A in Figure 13.1 depends only on grid points within the cross-hatch region. However, because A lies on the *characteristic* line $x - ct = 0$, the true solution at point A depends only on the initial condition at $x = 0$ (i.e., a parcel that is initially at the origin will be advected to the point $3\delta x$ in time $2\delta t$). The point $x = 0$ is outside the *domain of influence* of the numerical solution. Hence, the numerical solution cannot possibly faithfully reproduce the solution to the original differential equation, since, as shown in Figure 13.1, the value at point A in the numerical solution has no dependence on the conditions at $x = 0$. Only when the CFL criterion is satisfied will the domain of influence of the numerical solution include the characteristic lines of the analytic solution.

Although the CFL condition (13.18) guarantees stability of the centered difference approximation to the one-dimensional advection equation, in general the CFL criterion is only a necessary and not sufficient condition for computational stability. Other ways of finite differencing the one-dimensional advection equation may lead to more stringent limits on σ than given in (13.18).

The existence of computational instability is one of the prime motivations for using filtered equations. In the quasi-geostrophic system, no gravity or sound waves occur. Thus, the speed c in (13.18) is just the maximum wind speed. Typically, $c < 100\,\mathrm{m\,s^{-1}}$ so that for a grid interval of 200km, a time increment of over 30 min is permissible. However, in the nonhydrostatic equations commonly used in cloud-resolving models, the solution would have characteristics corresponding to acoustic modes, and to ensure that the domain of influence included such characteristics, c would need to be set equal to the speed of sound, which is the fastest wave described in that set of equations. In that case, $c \approx 300\,\mathrm{ms^{-1}}$, and for a 1-km vertical grid interval, a time step of only a few seconds would be permitted.

13.2.4 Implicit Time Differencing

The spurious computational mode introduced by the leapfrog time differencing scheme does not occur in a number of other explicit schemes, such as the Euler backward scheme discussed in Problem 13.3. Such schemes do, however, retain the time step limitation imposed by the CFL condition. This restriction and the computational mode are eliminated by utilization of an alternative finite differencing scheme called the *trapezoidal implicit scheme*. For the linear advection equation (13.6), this scheme can be written in the form

$$\frac{(\hat{q}_{m,s+1} - \hat{q}_{m,s})}{\delta t} = -\frac{c}{2}\left[\frac{(\hat{q}_{m+1,s+1} - \hat{q}_{m-1,s+1})}{2\delta x} + \frac{(\hat{q}_{m+1,s} - \hat{q}_{m-1,s})}{2\delta x}\right] \tag{13.19}$$

Substituting the trial solution (13.12) into (13.19) yields

$$B^{s+1} = \left[\frac{1 - i\,(\sigma/2)\sin p}{1 + i\,(\sigma/2)\sin p}\right] B^s \tag{13.20}$$

where as before $\sigma = c\delta t/\delta x$ and $p = k\delta x$. Defining

$$\tan\theta_p \equiv (\sigma/2)\sin p \tag{13.21}$$

and eliminating the common term B^s in (13.20), it can be shown that

$$B = \left(\frac{1 - i\tan\theta_p}{1 + i\tan\theta_p}\right) = \exp\left(-2i\theta_p\right) \tag{13.22}$$

so that the solution may be expressed simply as

$$\hat{q}_{m,s} = A\exp\left[ik(m\delta x - 2\theta_p s/k)\right] \tag{13.23}$$

Equation (13.19) involves only two time levels. Hence, unlike (13.14), the solution yields only a single mode, which has phase speed $c' = 2\theta_p/(k\delta t)$. According to (13.21), θ_p remains real for all values of δt. (This should be contrasted to the situation for the explicit scheme given by (13.17).) Thus, the implicit scheme is absolutely stable. The truncation errors, however, can become large if θ_p is not kept sufficiently small (see Problem 13.9). A disadvantage of the implicit scheme is that the integration cannot proceed by marching through the grid, as in the explicit case. In (13.19) there are terms involving the $s + 1$ time level on both sides of the equal sign, and these involve the values at a total of three grid points. Thus, the system (13.19) must be solved simultaneously for all points on the grid. If the grid is large, this may involve inverting a very large matrix, and so is computationally intensive. Furthermore, it is usually not feasible to utilize the implicit approach for nonlinear terms, such as the advection terms in the momentum equations. Semi-implicit schemes in which the linear terms are treated implicitly and the nonlinear terms explicitly are discussed briefly in Section 13.5.

13.2.5 The Semi-Lagrangian Integration Method

The differencing schemes discussed previously are Eulerian schemes in which the time integration is carried out by computing the tendencies of the predicted fields at a set of grid points fixed in space. Although it would be possible in theory to carry out predictions in a Lagrangian framework by following a set of marked fluid parcels, in practice this is not a viable alternative, since shear and stretching deformations tend to concentrate marked parcels in a few regions. Thus, it is difficult to maintain uniform resolution over the forecast region. It is possible to take advantage of the conservative properties of Lagrangian schemes, while maintaining uniform resolution, by employing a semi-Lagrangian technique. This approach permits relatively long time steps while retaining numerical stability and high accuracy.

The semi-Lagrangian method can be illustrated in a very simple fashion with the one-dimensional advection equation (13.6). According to this equation, the field q is conserved following the zonal flow at speed c. Thus, for any grid point, $x_m = m\delta x$, and time $t_s = s\delta t$:

$$q(x_m, t_s + \delta t) = q(\tilde{x}_m^s, t_s) \tag{13.24}$$

Here, $\tilde{x}_m^s$ is the location at time t_s for the air parcel that is located at point x_m at time $t_s + \delta t$. This position in general does not lie on a grid point (see the "plus" marked on Figure 13.1), so evaluation of the right-hand side of (13.24) requires interpolating from the grid point values at time t. For $c > 0$ the position $\tilde{x}_m^s$ lies between the grid points x_{m-p} and x_{m-p-1}, where p is the integer part of the expression $c\delta t/\delta x$ (a measure of the number of grid points traversed in a time step). If linear interpolation is used,

$$q(\tilde{x}_m^s, t_s) = \alpha q(x_{m-p-1}, t_s) + (1 - \alpha) q(x_{m-p}, t_s)$$

where $\alpha = (x_{m-p} - \tilde{x}_m^s)/\delta x$. Thus, in Figure 13.1, $p = 1$, and to predict q at point A, data are interpolated between the points $m = 1$ and $m = 2$ to the point shown by the "+".

In an actual prediction model the velocity field is predicted rather than known, as in this simple example. Thus, for a two-dimensional field

$$q(x, y, t + \delta t) = q(x - u\delta t, y - v\delta t, t) \tag{13.25}$$

where the velocity components at time *t* can be used to estimate the fields at $t+\delta t$; once these are obtained they can be used to provide more accurate approximations to the velocities on the right in (13.25). The right side in (13.25) is again estimated by interpolation, which now must be carried out in two dimensions.

As shown in Figure 13.1, the semi-Lagrangian scheme guarantees that the domain of influence in the numerical solution corresponds to that of the physical problem. Thus, the scheme is computational stability for time steps much longer than possible with an explicit Eulerian scheme. The semi-Lagrangian scheme also preserves the values of conservative properties quite accurately and is particularly useful for accurately advecting trace constituents such as water vapor.

13.2.6 Truncation Error

To be useful it is necessary not only that a numerical solution be stable but that it also provide an accurate approximation to the true solution. The difference between the numerical solution to a finite difference equation and the solution to the corresponding differential equation is called the *discretization error*. If this error approaches zero with δt and δx, the solution is called *convergent*. The difference between a differential equation and the finite difference analog to it is referred to as a *truncation error* because it arises from truncating the Taylor series approximation to the derivatives. If this error approaches zero as δt and δx go to zero, the scheme is called *consistent*. According to the Lax equivalence theorem,[1] if the finite difference formulation satisfies the consistency condition, then stability is the necessary and sufficient condition for convergence. Thus, if a finite difference approximation is consistent and stable, one can be certain that the discretization error will decrease as the difference intervals are decreased, even if it is not possible to determine the error exactly.

Because numerical solutions are as a rule sought only when analytic solutions are unavailable, it is usually not possible to determine the accuracy of a solution directly. For the linear advection equation with constant advection speed considered in Section 13.2.3, it is possible, however, to compare the solutions of the finite difference equation (13.8) and the original differential equation (13.6). We can then use this example to investigate the accuracy of the difference method introduced earlier.

From the preceding discussion we can already conclude that the magnitude of the truncation error in the present case will be of order $\delta x^2 \text{ and } \delta t^2$. It is possible to obtain more precise information on accuracy from examination of the solution (13.14). Note that for $\theta_p \to 0, C \to 1 \text{ and } D \to 0$. The part of the solution proportional to *C* is the *physical mode*. The part proportional to *D* is called the *computational mode* because it has no counterpart in the analytic solution to the original differential equation. It arises because centered time differencing has turned a differential equation that is first order in time into a second-order finite difference equation. The accuracy of the finite difference solution depends not only on the smallness of *D* and the closeness of *C* to unity but on the match between the phase speed of the physical mode and the phase speed in the analytic solution. The phase of the physical mode is given in (13.14) by

$$pm - \theta_p s = (p/\delta x)\,(m\delta x - \theta_p s\delta x/p) = k(x - c'\,t)$$

where $c' = \theta_p \delta x/(p\delta t)$ is the phase speed of the physical mode. Its ratio to the true phase speed is

$$c'/c = \theta_p \delta x/(pc\delta t) = \sin^{-1}(\sigma \sin p)/(\sigma p)$$

so that $c'/c \to 1$ as $\sigma p \to 0$. The dependence of c'/c and $|D|/|C|$ on wavelength is shown in Table 13.1 for the particular case where $\sigma = 0.75$.

Table 13.1 Centered Finite Difference Accuracy

$L/\delta x$	p	θ_p	c'/c	$\lvert D\rvert/\lvert C\rvert$
2	π	π	–	∞
4	$\pi/2$	0.848	0.720	0.204
8	$\pi/4$	0.559	0.949	0.082
16	$\pi/8$	0.291	0.988	0.021
32	$\pi/16$	0.147	0.997	0.005

Note: Results apply to the advection equation as a function of resolution for $\sigma = 0.75$.

It is clear from Table 13.1 that phase speed and amplitude errors both increase as the wavelength decreases. Short waves move slower than long waves in the finite difference solution, even though in the original equation all waves move at speed *c*. This dependence of phase speed on wavelength in the difference solution is called *numerical dispersion*. It is a serious problem for numerical modeling of any advected field that has sharp gradients (and thus large-amplitude short-wave components).

Short waves also suffer from having significant amplitude in the computational mode. This mode, which has no counterpart in the solution to the original differential equation, propagates opposite to the direction of the physical mode and changes sign from one time step to the next. This behavior makes it easy to recognize when the computational mode has significant amplitude.

13.3 The Barotropic Vorticity Equation in Finite Differences

The simplest example of a dynamical forecast model is the barotropic vorticity equation (11.14), which for a Cartesian β plane can be written in the form

$$\frac{\partial \zeta}{\partial t} = -F(x, y, t) \tag{13.26}$$

where

$$F(x, y, t) = \mathbf{V}_\psi \cdot \boldsymbol{\nabla}(\zeta + f) = \frac{\partial}{\partial x}(u_\psi \zeta) + \frac{\partial}{\partial y}(v_\psi \zeta) + \beta v_\psi \tag{13.27}$$

and $u_\psi = -\partial\psi/\partial y, v_\psi = \partial\psi/\partial x,$ and $\zeta = \nabla^2\psi$. We have here used the fact that the horizontal velocity is nondivergent ($\partial u_\psi/\partial x + \partial v_\psi/\partial y = 0$) to write the advection term in flux form. The advection of absolute vorticity *F*(*x*,*y*,*t*) may be calculated provided that we know the field of $\psi(x, y, t)$. Equation (13.26) can then be integrated forward in time to yield a prediction for ζ. It is then necessary to solve the Poisson equation $\zeta = \boldsymbol{\nabla}^2\psi$ to predict the streamfunction.

A straightforward solution method is the leapfrog scheme discussed in Section 13.2.2. This requires writing (13.27) in finite difference form. Suppose that the horizontal *x*, *y* space is divided into a grid of $(M+1)\times(N+1)$ points separated by distance increments δx and δy. Then we can write the coordinate position of a given grid point as $x_m = m\delta x, y_n = n\delta y,$ where $m = 0, 1, 2, \ldots, M$ and $n = 0, 1, 2, \ldots, N$. Thus, any point on the grid is uniquely identified by the indices (m, n). A portion of such a grid space is shown in Figure 13.2.

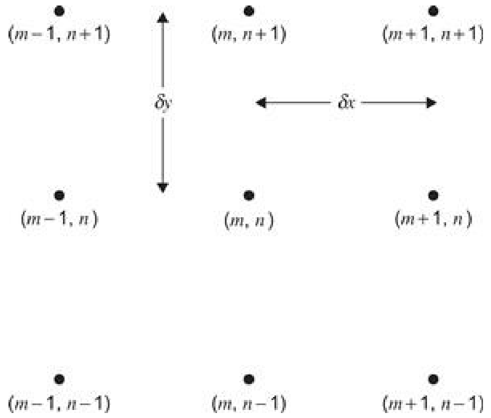


Figure 13.2 Portion of a two-dimensional (*x*, *y*) grid mesh for solution of the barotropic vorticity equation.

Centered difference formulas of the type (13.4) can then be used to approximate derivatives in the expression *F*(*x*,*y*,*t*). For example, if we assume that $\delta x = \delta y \equiv d$,

(13.28)

$$\begin{aligned} u_\psi &\approx u_{m,n} = -(\psi_{m,n+1} - \psi_{m,n-1})/2d \\ v_\psi &\approx v_{m,n} = +(\psi_{m+1,n} - \psi_{m-1,n})/2d \end{aligned}$$

Similarly, with the aid of (13.5) we find that the horizontal Laplacian can be approximated as

(13.29)

$$\nabla^2\psi \approx (\psi_{m+1,n}+\psi_{m-1,n}+\psi_{m,n+1}+\psi_{m,n-1}-4\psi_{m,n})/d^2 = \zeta_{m,n}$$

The finite difference form of the Laplacian is proportional to the difference between the value of the function at the central point and the average value at the four surrounding grid points. If there are $(M-1)\times(N-1)$ interior grid points, then (13.29) yields a set of $(M-1)\times(N-1)$ simultaneous equations, which together with suitable boundary conditions determine $\psi_{m,n}$ for a given array $\zeta_{m,n}$. This set can be solved by standard methods of matrix inversion.

Before expressing the advection term *F*(*x*,*y*,*t*) in finite difference form, it is worth noting that if ψ is taken to be constant on the northern and southern boundaries of the *β*-plane channel, it is easily shown by integration over the area of the channel that the average value of *F* is zero. This implies that the mean vorticity is conserved for the channel. It is also possible, with a little more algebra, to show that the mean kinetic energy and the mean square vorticity (called the *enstrophy*) are conserved.

For accuracy of long-term integrations, it is desirable that any finite difference approximation to *F* satisfy the same conservation constraints as the original differential form; otherwise, the finite difference solution would not be conservative. The mean vorticity, for example, might then drift systematically in time purely because of the nature of the finite difference solution. Finite difference schemes that simultaneously conserve vorticity, kinetic energy, and enstrophy have been designed. They are, however, rather complex. For our purposes it is sufficient to note that by writing the advection in the flux form (13.27) and using centered space differences we can conserve both mean vorticity and mean kinetic energy:

$$\begin{aligned} F_{m,n} = \frac{1}{2d} & [(u_{m+1,n}\zeta_{m+1,n} - u_{m-1,n}\zeta_{m-1,n}) \\ & + (v_{m,n+1}\zeta_{m,n+1} - v_{m,n-1}\zeta_{m,n-1})] + \beta v_{m,n} \end{aligned} \tag{13.30}$$

It is verified readily that if ψ satisfies periodic boundary conditions, there is a cancellation of terms when (13.30) is summed over the domain. Thus,

$$\sum_{m=1}^{M}\sum_{n=1}^{N} F_{m,n} = 0 \tag{13.31}$$

Therefore, mean vorticity is conserved (except for errors introduced by time differencing) when (13.30) is used as the finite difference form of the advection term. This form also conserves mean kinetic energy (see Problem 13.2). Enstrophy is not conserved in this difference formulation, and it is conventional to add a small diffusion term in order to control any numerically generated increase in enstrophy.

The procedure for preparing a numerical forecast with the barotropic vorticity equation can now be summarized as follows:

1. The observed geopotential field at the initial time is used to compute the initial streamfunction $\psi_{m,n}(t=0)$ at all grid points.
2. $F_{m,n}$ is evaluated at all grid points.
3. $\zeta_{m,n}(t+\delta t)$ is determined using centered differencing except at the first time step when a forward difference must be used.
4. The simultaneous set (13.29) is solved for $\psi_{m,n}(t+\delta t)$.
5. The predicted array of $\psi_{m,n}$ is used as data, and steps 2 through 4 are repeated until the desired forecast period is reached. For example, a 24-h forecast with 30-min time increments would require 48 cycles.

13.4 The Spectral Method

In the finite difference method, the dependent variables are specified on a set of grid points in space and time, and derivatives are approximated using finite differences. An alternative approach, referred to as the spectral method, involves representing the spatial variations of the dependent variables in terms of finite series of orthogonal functions called *basis functions*. For the Cartesian geometry of a midlatitude *β*-plane channel, the appropriate set of basis functions is a double Fourier series in *x* and *y*. For spherical Earth, however, the appropriate basis functions are the spherical harmonics.

A finite difference approximation is *local* in the sense that the finite difference variable $\Psi_{m,n}$ represents the value of $\psi(x,y)$ at a particular point in space, and the finite difference equations determine the evolution of the $\Psi_{m,n}$ for all grid points. The spectral approach, however, is based on *global* functions—that is, the individual components of the appropriate series of basis functions. In the case of Cartesian geometry, for example, these components determine the amplitudes and phases of the sinusoidal waves that when summed determine the spatial distribution of the dependent variable. The solution proceeds by determining the evolution of a finite number of

Fourier coefficients. Because the distribution in wave number space of the Fourier coefficients for a given function is referred to as its *spectrum*, it is appropriate to call this approach the spectral method.

At low resolution the spectral method is generally more accurate than the grid point method, partly because, for linear advection, the numerical dispersion discussed in Section 13.2.4 can be severe in a grid point model but does not occur in a properly formulated spectral model. For the range of resolutions used commonly in forecast models the two approaches are comparable in accuracy, and each has its advocates.

13.4.1 The Barotropic Vorticity Equation in Spherical Coordinates

The spectral method is particularly advantageous for solution of the vorticity equation. When the proper set of basis functions is chosen, it is trivial to solve the Poisson equation for the streamfunction. This property of the spectral method not only saves computer time but eliminates the truncation error associated with finite differencing the Laplacian operator.

In practice the spectral method is applied most frequently to global models. This requires use of spherical harmonics, which are more complicated than Fourier series. To keep the discussion as simple as possible, it is again useful to consider the barotropic vorticity equation as a prototype forecast model in order to illustrate the spectral method on the sphere.

The barotropic vorticity equation in spherical coordinates may be expressed as

$$\frac{D}{Dt}(\zeta + 2\Omega \sin\phi) = 0 \tag{13.32}$$

where, as usual, $\zeta = \nabla^2\psi$, with ψ a streamfunction, and

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \frac{u}{a\cos\phi}\frac{\partial}{\partial\lambda} + \frac{v}{a}\frac{\partial}{\partial\phi} \tag{13.33}$$

It turns out to be convenient to use $\mu \equiv \sin\phi$ as the latitudinal coordinate, in which case the continuity equation can be written

$$\frac{1}{a}\frac{\partial}{\partial\lambda}\left(\frac{u}{\cos\phi}\right) + \frac{1}{a}\frac{\partial}{\partial\mu}(v\cos\phi) = 0 \tag{13.34}$$

so that the streamfunction is related to the zonal and meridional velocities according to

$$\frac{u}{\cos\phi} = -\frac{1}{a}\frac{\partial\psi}{\partial\mu}; \quad v\cos\phi = \frac{1}{a}\frac{\partial\psi}{\partial\lambda} \tag{13.35}$$

The vorticity equation can then be expressed as

$$\frac{\partial\nabla^2\psi}{\partial t} = \frac{1}{a^2}\left[\frac{\partial\psi}{\partial\mu}\frac{\partial\nabla^2\psi}{\partial\lambda} - \frac{\partial\psi}{\partial\lambda}\frac{\partial\nabla^2\psi}{\partial\mu}\right] - \frac{2\Omega}{a^2}\frac{\partial\psi}{\partial\lambda} \tag{13.36}$$

where

$$\nabla^2 \psi = \frac{1}{a^2} \left\{ \frac{\partial}{\partial \mu} \left[(1-\mu^2) \frac{\partial \psi}{\partial \mu} \right] + \frac{1}{1-\mu^2} \frac{\partial^2 \psi}{\partial \lambda^2} \right\} \tag{13.37}$$

The appropriate orthogonal basis functions are the *spherical harmonics*, which are defined as

$$Y_\gamma (\mu, \lambda) \equiv P_\gamma (\mu) e^{im\lambda} \tag{13.38}$$

where $\gamma \equiv (n, m)$ is a vector containing the integer indices for the spherical harmonics. These are given by $m = 0, \pm 1, \pm 2, \pm 3, \ldots, n = 1, 2, 3, \ldots,$ where it is required that $|m| \le n$. Here, P_γ designates an associated Legendre function of the first kind of degree *n*. From (13.38) it is clear that *m* designates the zonal wave number. It can be shown[2] that $n - |m|$ designates the number of nodes of P_γ in the interval $-1 < \mu < 1$ (i.e., between the poles), and thus measures the meridional scale of the spherical harmonic. The structures of a few spherical harmonics are shown in Figure 13.3.

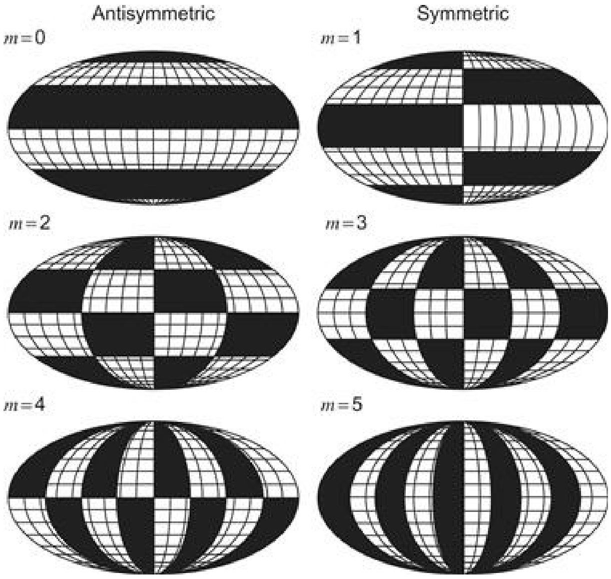


Figure 13.3 Patterns of positive and negative regions for the spherical harmonic functions with $n = 5$ and $m =$ 0, 1, 2, 3, 4, 5.

After Washington and Parkinson, 1986; adapted from Baer, 1972. Copyright © American Meteorological Society. Reprinted with permission.

An important property of the spherical harmonics is that they satisfy the relationship

$$\nabla^2 Y_\gamma = -\frac{n(n+1)}{a^2} Y_\gamma \tag{13.39}$$

so that the Laplacian of a spherical harmonic is proportional to the function itself, which implies that the vorticity associated with a particular spherical harmonic component is simply proportional to the streamfunction for the same component.

In the spectral method on the sphere, the streamfunction is expanded in a finite series of spherical harmonics by letting

$$\psi(\lambda,\mu,t)=\sum_{\gamma}\psi_{\gamma}(t)Y_{\gamma}(\mu,\lambda) \tag{13.40}$$

where ψ_γ is the complex amplitude for the Y_γ spherical harmonic and the summation is over both *n* and *m*. The individual spherical harmonic coefficients ψ_γ are related to the streamfunction $\psi(\lambda,\mu)$ through the inverse transform

$$\psi_{\gamma}(t)=\frac{1}{4\pi}\int_{S}Y_{\gamma}^{*}\psi(\lambda,\mu,t)dS \tag{13.41}$$

where $dS = d\mu d\lambda$ and Y_γ^* designates the complex conjugate of Y_γ.

13.4.2 Rossby–Haurwitz Waves

Before considering numerical solution of the barotropic vorticity equation, it is worth noting that an exact analytic solution of the nonlinear equation can be obtained in the special case where the streamfunction is equal to a single spherical harmonic mode. Thus, we let

$$\psi(\lambda,\mu,t)=\psi_{\gamma}(t)\,e^{im\lambda}P_{\gamma}(\mu) \tag{13.42}$$

Substituting from (13.42) into (13.36) and applying (13.39), we find that the nonlinear advection term is identically zero so that the amplitude coefficient satisfies the ordinary linear equation

$$-n\,(n+1)\,d\psi_{\gamma}/dt=-2\Omega im\psi_{\gamma} \tag{13.43}$$

which has the solution $\psi_\gamma(t)=\psi_\gamma(0)\exp(i\nu_\gamma t)$, where

$$\nu_{\gamma}=2\Omega m/\left[n\,(n+1)\right] \tag{13.44}$$

is the dispersion relationship for *Rossby–Haurwitz waves*, which is the name given to planetary waves on a sphere. (This should be compared to (5.110) with $\bar{u}=0$, which is the equivalent expression for a midlatitude *β*-plane.) Because the horizontal scale of a spherical harmonic mode is proportional to n^{-1}, (13.44) shows that a single mode propagates westward on a sphere at a speed that is approximately proportional to the square of the horizontal scale. This solution also suggests why for some problems the spectral method is superior to the finite difference method at coarse resolution. A model containing even a single Fourier component can represent a realistic meteorological field (the Rossby wave), while many grid points are required for an equivalent representation in finite differences.

13.4.3 The Spectral Transform Method

When many spherical harmonic modes are present, the solution of (13.36) by a purely spectral method requires evaluation of the nonlinear interactions among various modes due to the advection term. It turns out that the number of interaction terms increases as the square of the number of modes retained in the series (13.40) so that this approach becomes computationally inefficient for models with the sort of spatial resolution required for prediction of synoptic-scale weather disturbances. The spectral transform method overcomes this problem by transforming between spherical harmonic wave number space and a latitude–longitude grid at every time step and carrying out the multiplications in the advection term in grid space. Thus, that it is never necessary to compute products of spectral functions.

To illustrate this method it is useful to rewrite the barotropic vorticity equation in the form

$$\frac{\partial \nabla^2 \psi}{\partial t} = -\frac{1}{a^2}\left[2\Omega \frac{\partial \psi}{\partial \lambda} + A\left(\lambda, \mu\right)\right] \tag{13.45}$$

where

$$A(\lambda, \mu) \equiv \left[-\frac{\partial \psi}{\partial \mu}\frac{\partial \nabla^2 \psi}{\partial \lambda} + \frac{\partial \psi}{\partial \lambda}\frac{\partial \nabla^2 \psi}{\partial \mu}\right] \tag{13.46}$$

Substituting from (13.40) into (13.45) then yields for the spherical harmonic coefficients

$$d\psi_\gamma / dt = i\nu_\gamma \psi_\gamma + A_\gamma \left[n\left(n+1\right)\right]^{-1} \tag{13.47}$$

where A_γ is the γ component of the transform of $A(\lambda, \mu)$:

$$A_\gamma = \frac{1}{4\pi}\int_0^{2\pi}\int_{-1}^{+1} A\left(\lambda, \mu\right) Y_\gamma^* d\lambda d\mu \tag{13.48}$$

The transform method utilizes the fact that if the sum $\gamma = (n, m)$ is taken over a finite number of modes, the integral in the transform (13.48) can be evaluated exactly by numerical quadrature—that is, by summing appropriately weighted values of $A(\lambda, \mu)$ evaluated at the grid points (λ_j, μ_k) of a latitude–longitude grid mesh. Computation of the distribution of $A(\lambda, \mu)$ for all grid points can be carried out without need to introduce finite differences for derivatives by noting that we can express the advection term in the form

$$A\left(\lambda_j, \mu_k\right) = \left(1 - \mu^2\right)^{-1}\left[F_1 F_2 + F_3 F_4\right] \tag{13.49}$$

where

$$\begin{aligned} F_1 &= -\left(1 - \mu^2\right)\partial \psi / \partial \mu, \quad & F_2 &= \partial \nabla^2 \psi / \partial \lambda \\ F_3 &= \left(1 - \mu^2\right)\partial \nabla^2 \psi / \partial \mu, \quad & F_4 &= \partial \psi / \partial \lambda \end{aligned}$$

The quantities $F_1 - F_4$ can be computed exactly for each grid point using the spectral coefficients ψ_γ and the known differential properties of the spherical harmonics. For example,

$$F_4 = \partial \psi / \partial \lambda = \sum_\gamma i m \psi_\gamma Y_\gamma (\lambda_j, \mu_k)$$

Once these terms have been computed for all grid points, $A(\lambda, \mu)$ can be computed at the grid points by forming the products $F_1 F_2$ and $F_3 F_4$; no finite difference approximations to the derivatives are required in this procedure. The transform (13.48) is then evaluated by numerical quadrature to compute the spherical harmonic components A_γ. Finally, (13.47) can be stepped ahead by a time increment δt to obtain new estimates of the spherical harmonic components of the streamfunction. The whole process is then repeated until the desired forecast period is reached. The steps in forecasting with the barotropic vorticity equation using the spectral transform method are summarized in schematic form in Figure 13.4.

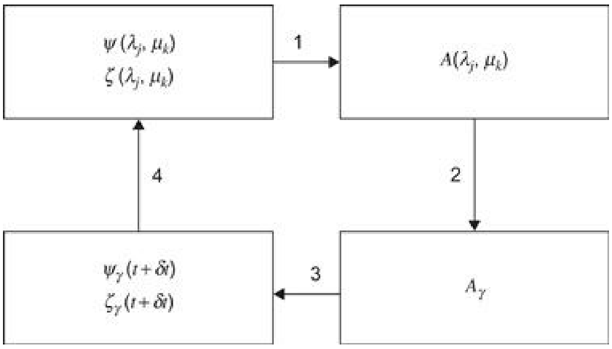

$\psi(\lambda_j, \mu_k)$
$\zeta(\lambda_j, \mu_k)$
1
$A(\lambda_j, \mu_k)$
4
2
$\psi_\gamma(t+\delta t)$
$\zeta_\gamma(t+\delta t)$
3
A_γ

Figure 13.4 Steps in the prediction cycle for the spectral transform method of solution of the barotropic vorticity equation.

13.5 Primitive Equation Models

Modern numerical forecast models are based on a formulation of the dynamical equations, referred to as the *primitive equations*, which is essentially the formulation proposed by Richardson. The primitive equations differ from the complete momentum equations (2.19), (2.20), and (2.21) in that the vertical momentum equation is replaced by the hydrostatic approximation, and the small terms in the horizontal momentum equations given in columns C and D in Table 2.1 are neglected. In most models some version of the σ-coordinate system introduced in Section 10.3.1 is used, and the vertical dependence is represented by dividing the atmosphere into a number of levels and utilizing finite difference expressions for vertical derivatives. Both finite differencing and the spectral method have been used for horizontal discretization in operational primitive equation forecast models.

Grid point models utilize the finite difference method described in Section 13.2.1. Most grid point models utilize a *grid-staggering* approach to improve the accuracy of certain fields. Two examples of where the momentum field is staggered relative to the thermodynamic variables are shown in Figure 13.5. The Arakawa-C grid (a) is staggered so that the horizontal divergence ($\partial u/\partial x + \partial v/\partial y$) is located in the center of the grid cell, whereas the Arakawa-D grid (b) is staggered so that the vertical vorticity $\partial v/\partial x - \partial u/\partial y$ is located in the center of the grid cell.

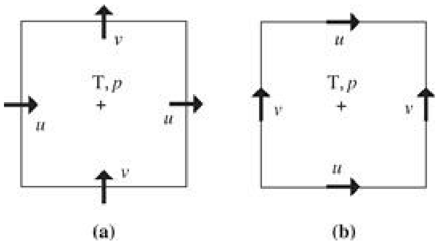


Figure 13.5 Horizontal grid staggering schemes for the Arakawa-C (a) and Arakawa-D (b) grids. Thermodynamic (T, p) and other scalar fields are located at the center of the grid cell; horizontal velocity components u and v are located along the grid

edge.

13.5.1 Spectral Models

Most operational forecast centers utilize spectral models as their primary global forecast models, whereas grid point models are generally employed for fine- scale limited area models. Operational spectral models employ a primitive equation version of the spectral transform method described in Section 13.4.3. In this method the values of all meteorological fields are available in both spectral and grid point domains at each time step. Vorticity and divergence are employed as predictive variables rather than u and v. Physical computations involving such processes as radiative heating and cooling, condensation and precipitation, and convective overturning are calculated in physical space on the grid mesh, whereas differential dynamical quantities, such as pressure gradients and velocity gradients, are evaluated exactly in spectral space. This combination preserves the simplicity of the grid point representation for physical processes that are "local" in nature, while retaining the superior accuracy of the spectral method for dynamical calculations.

Two spectral truncations are typically employed in global spectral modes. The first, *triangular truncation*, refers to the fact that on a plot of m versus n (see 13.38) the retained modes occupy a triangular area. The second spectral truncation is *rhomboidal truncation*, which has $N = |m| + M$. Both of these truncations are shown schematically in Figure 13.6. In triangular truncation, the horizontal resolution in the zonal and meridional directions is nearly equal. In rhomboidal truncation, however, the latitudinal resolution is the same for every zonal wave number. Rhomboidal truncation has some advantages for low-resolution models, but at high resolution the triangular truncation appears to be superior.

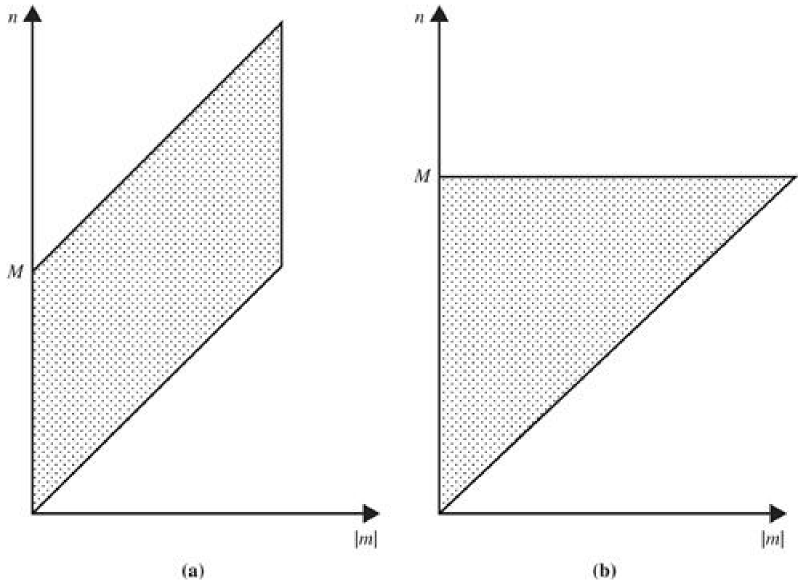


Figure 13.6 Regions of wave number space (n, m) in which spectral components are retained for the rhomboidal truncation (a) and triangular truncation (b).

After Simmons and Bengtsson, 1984. Used with permission.

13.5.2 Physical Parameterizations

The physical processes included in modern operational forecast models are generally the same as those included in general circulation models and were shown schematically in Figure 10.21.

Inclusion of such processes as boundary layer fluxes, vertical mixing by dry and moist convection, formation of clouds and precipitation, and the interaction of cloud and radiation fields requires that the relevant subgrid-scale processes be represented in terms of model-predicted fields. The approximation of unresolved processes in terms of resolved variables is referred to as*parameterization*; it is probably the most difficult and controversial area of weather and climate modeling.

Perhaps the most important physical process that must be parameterized is convection. Vertical heat transfer by convection is essential in maintaining the observed tropospheric lapse rate and moisture distribution. The simplest way to mimic this effect of unresolved convective motions is through *convective adjustment*. In a simple version of this method, the relative humidity and lapse rate in each grid column are examined at the end of each time step. If the lapse rate is superadiabatic, the temperature profile is adjusted to dry static neutrality in a manner that conserves energy; if the column is conditionally unstable, and the humidity exceeds a specified value, the column is adjusted to moist static neutrality. More sophisticated schemes utilize the fact that moist convection is dependent on low-level moisture convergence, and they include a moisture budget as part of the parameterization.

Various methods are used to relate clouds to the resolved humidity, temperature, and wind fields. None is completely satisfactory. In some models the model-predicted cloud fields are used only in the portion of the model concerned with precipitation, and cloud–radiation interactions are parameterized using specified cloud climatologies.

13.6 Data Assimilation

Since forecasting is an initial value problem, observations are needed to establish the initial condition, which is called an analysis. Observations are distributed inhomogeneously in space and time, in contrast to numerical models, which have regularly spaced grids and forecasts that begin at specific times. Data assimilation is a technique that maps the inhomogeneous observations onto a regular grid, and also accounts for the errors in observations. As the name suggests, observations are assimilated into the numerical model, and the model is used to transmit the information from observations to a future time when observations are again available.

For this reason, data assimilation involves blending observations with estimates of the observations from short-range forecasts. Since the short-range forecasts start from relatively accurate initial conditions, the error in the model-estimated observations is so small that, for operational forecast systems, only about 15% of the analysis comes from millions of observations; the other 85% derives from the model, which has accumulated the information from all observations before that time. An essential aspect of data assimilation is that nothing is known exactly and the problem must be treated probabilistically.

13.6.1 Data Assimilation for a Single Variable

To define the data assimilation problem, and different techniques used to solve it, we begin with a single scalar before moving on to multiple variables. Given a scalar variable x, assume that there are two estimates of its true value: an observation, y, and a prior estimate, or background, x_b. The background may come from different sources, but mainly it comes from a short-range forecast. We wish to estimate the true value of x given the background and observation. The mathematical transcription of this statement involves a conditional probability

$$p(x_a) = p(x|y) \tag{13.50}$$

In words, the probability density of the analysis, x_a, is equal to the conditional probability density of the true state, x, given observation y (see Appendix G for background on conditional probability). The background information comes in through Bayes' Rule, which allows the conditional probability density to be "inverted" so that

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{13.51}$$

where, in this case, $p(y)$ is just a scaling constant that may be neglected.[3] Since $p(y|x)$ is a function of the second argument, x, it is a *likelihood function* for the observation, given x. A more complete description is given in Appendix G, but the basic idea can be illustrated by flipping a coin to determine the likelihood that the coin is fair (e.g., determine the likelihood that the coin is fair from observing three consecutive tosses that all yield heads). This problem is the reverse of *knowing* that the coin is fair and determining the probability of an outcome conditioned on that knowledge (e.g., the probability of three heads in a row for a fair coin).

At this point the probability density functions are general, but for problems with many variables, a simplification is needed. The main simplification is that the variables are distributed normally (or Gaussian), which for a scalar reduces the probability density description to the mean and variance,

$$\begin{aligned} p(y|x) &= c_1 e^{-\frac{1}{2}\left(\frac{y-x}{\sigma_y}\right)^2} \\ p(x) &= c_2 e^{-\frac{1}{2}\left(\frac{x-x_b}{\sigma_b}\right)^2} \end{aligned} \tag{13.52}$$

where c_1 and c_2 are constants. The error variances of the observation and background are σ_y^2 and σ_b^2, respectively. Using (13.52) and (13.51) in (13.50) gives the analysis probability density

$$p(x_a) = Ce^{-\frac{1}{2}\left(\frac{x-x_b}{\sigma_b}\right)^2} e^{-\frac{1}{2}\left(\frac{y-x}{\sigma_y}\right)^2} \tag{13.53}$$

This result involves a product of Gaussians, which may be simplified to a new Gaussian distribution. To do so it proves useful to consider the following metric, J, of the probability density:

$$J(x) \equiv -\log\left[p(x)\right] = \frac{1}{2}\frac{(x-x_b)^2}{\sigma_b^2} + \frac{1}{2}\frac{(y-x)^2}{\sigma_y^2} - \log(C) \tag{13.54}$$

Metric J, the "cost function," is a quadratic measure of the misfit between the true value of the scalar and the two estimates we have, with smaller values of J indicating a better fit. Finding the best fit involves solving for the minimum value of J by taking $\partial J/\partial x$ and setting the result to zero. Solving for x at this point yields the analysis value, x_a,

$$x_a = \left(\frac{\sigma_y^2}{\sigma_y^2 + \sigma_b^2}\right) x_b + \left(\frac{\sigma_b^2}{\sigma_y^2 + \sigma_b^2}\right) y \tag{13.55}$$

This result reveals that the best estimate of the true state is given by a linear combination[4] of the prior estimate and the observation, with weights determined by the errors in these quantities. In the limit of vanishing error in the observation ($\sigma_y^2 \to 0$) the weight on the prior estimate approaches zero and the weight on the observation approaches one. The error variance for x_a is given by

(13.56)

$$\sigma_a^2 = \frac{\sigma_b^2}{1+\left(\frac{\sigma_b}{\sigma_y}\right)^2} = \frac{\sigma_y^2}{1+\left(\frac{\sigma_y}{\sigma_b}\right)^2} < \sigma_b^2, \sigma_y^2$$

which shows that the error in the analysis is always less than the error in *both* the prior estimate and the observation. Thus the analysis is Gaussian distributed with mean x_a and variance σ_a^2.

The generalization from a scalar to vectors in the case of many analysis variables and many observations is facilitated by rewriting (13.55) as

(13.57)

$$x_a = x_b + K(y - x_b)$$

and (13.56) as

(13.58)

$$\sigma_a^2 = (1 - K)\,\sigma_b^2$$

where

(13.59)

$$K = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_y^2}$$

From (13.57) we see that the analysis may be viewed as a linear combination of the prior estimate and the *innovation*, or new information, contained in the observation. If the observation gives the same value as our prior estimate, nothing has been learned, and there is no need to adjust the prior estimate. When there is new information, its weight relative to the background is controlled by the *gain* weighting factor, *K*, which is given by the ratio of the variance in the prior to the variance in the innovation.[5] *K* has an upper limit of one for $\sigma_y^2 \to 0$ and a lower limit of zero for $\sigma_b^2 \to 0$. Notice in (13.58) that, in the case of a perfect observation, the error in the prior is reduced to zero in the analysis.

A graphical illustration of these results is shown in Figure 13.7. Notice that, by assumption, the distribution in the observation (*thin solid line*) has less error than the prior (dashed line) and, as a result, the analysis (*thick solid line*) is weighted closer to the observation than the prior. Similarly, the distribution for the analysis has higher probability density than either the prior or the observation, because the analysis error variance is smaller as well. The thick gray curve shows the cost function, *J*, which reaches a minimum at the mean value of x_a.

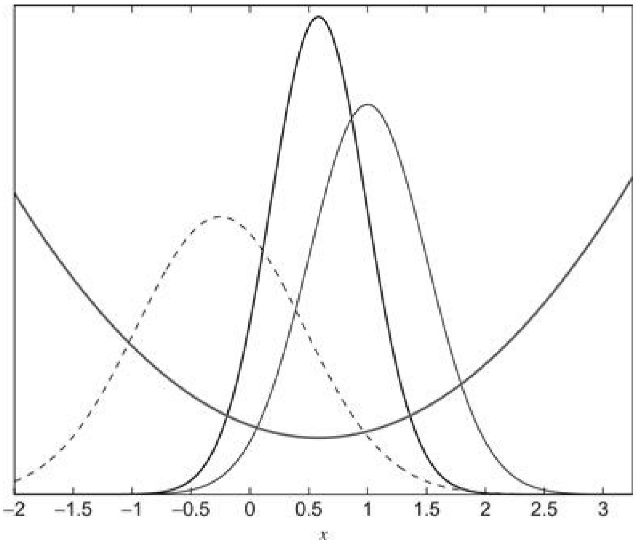


Figure 13.7 Data assimilation for scalar variable x assuming Gaussian error statistics. Prior estimate x_b, given by the *dashed line*, has mean −0.25 and variance 0.5. Observation y, given by the *thin solid line*, has mean 1.0 and variance 0.25. The analysis, x_a, given by the *thick solid line*, has mean 0.58 and variance 0.17. The parabolic *gray curve* denotes the cost function, J, which takes a minimum at the mean value of x_a.

13.6.2 Data Assimilation for Many Variables

The generalization to many variables involves a straightforward extension of the scalar case provided that matrix notation is used. Because the form of the equations is so similar to the scalar case, readers unfamiliar with matrix algebra should still be able to follow the discussion, if not the main points of the math as well.

The system is described by a *state vector*, **x**, which is a column vector containing all of the grid point (or spectral) variables. For many variables, the cost function (13.54) becomes

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^{\mathrm{T}}\,\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)+\frac{1}{2}(\mathbf{y}-\mathbf{H}\,\mathbf{x})^{\mathrm{T}}\,\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x}) \tag{13.60}$$

As before, J is a scalar measuring the misfit between the true state and both the prior estimate and the observations. The weight on each term is no longer determined by the variances, σ_b^2 and σ_y^2, but rather by the covariance matrices, **B** and **R**, for the prior and the observations, respectively. The diagonal elements of **B** contain the error variance of each variable in **x**, whereas the off-diagonal elements contain the covariance between the variables, which measure the linear relationships between the variables (e.g., the relationship between pressure and wind may obey gradient wind balance).

Similarly, the diagonal elements of **R** contain the error variance of each observation; the off-diagonal elements, the covariance between the observations. Superscript T denotes a transpose, turning column vectors into row vectors, so that the (inner) products in both terms yield scalars. Matrix **H** is the “observation operator," which maps the state variables to the observations. Simple examples of **H** involve linear interpolation of grid point values of, for example, temperature, to the observation location where a measurement was made. A more complicated example involves radiative transfer calculations through the entire model atmosphere to yield an estimate of a satellite radiance.

The minimum value of J again determines the analysis, although this time careful application of matrix calculus and algebra are required to yield the multivariate analog of (13.57):

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b) \tag{13.61}$$

The interpretation of (13.61) is identical to that of (13.57), except that now the weight is given by the *Kalman gain* matrix

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}\left[\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right]^{-1} \tag{13.62}$$

Although seemingly more complicated than (13.59), **K** may be interpreted similarly as the ratio of a prior covariance to an innovation covariance. The quantity $\mathbf{HBH}^{\mathrm{T}}$ measures the error covariance in the prior estimate of the observations, a direct analog of **R**.

The multivariate version of (13.58) involves the analysis error covariance, **A**, given by

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \tag{13.63}$$

where **I** is the identity matrix, and we see that the prior covariance is "reduced" by $\mathbf{I} - \mathbf{KH}$.

Analysis equations (13.61) and (13.63) represent one part of the Kalman filter, which is a recursive algorithm for estimating the state of a system sequentially in time. The second part of the Kalman filter includes a forecast from the analysis to a future time when observations are again available. In addition to a forecast of the prior estimate, (13.61) and (13.63) require a forecast of the error covariance **B**. Since the size of **B** is square in the number of variables, which can be on the order of 100 million for operational forecast systems, this calculation is prohibitively expensive. There are two leading approximations to deal with this problem: One gives variational assimilation, and the other gives the Ensemble Kalman filter (EnKF).

Rather than solving (13.61), which involves operations on enormous matrices including matrix inversion, in variational data assimilation (13.60) is solved by a direct search. Starting from the background, a descent strategy is pursued toward smaller J until changes become small within a tolerance. These strategies normally involve both J and its gradient, $\partial J/\partial \mathbf{x}$, which must be updated during the search. A common approximation in the variational approach is to assume a fixed covariance matrix for the prior, **B**, which eliminates the need to propagate a large matrix.

When the variational approach is performed at a single time, it is called three-dimensional variational assimilation (3DVAR). Augmenting the cost function (13.60) to include the misfit to observations over a window of time yields four-dimensional data assimilation (4DVAR). A discussion of the details is beyond the scope of this book, but 4DVAR requires substantially more resources than 3DVAR because the gradient descent involves an iteration over a period of time and specialized models (so-called "adjoints" of the forecast model), or ensembles, to accomplish this task. An example of a 4DVAR procedure, as used at the European Centre for Medium Range Weather Forecasts (ECMWF), is shown schematically in Figure 13.8. The ECMWF scheme utilizes a 12-h assimilation window extending from 9 h prior to 3 h after the analysis times, which are 00Z and 12Z. An assimilation cycle is begun by using a forecast from the initial state at the previous analysis time as the prior estimate. All available observations within the assimilation window are used to update the analysis.

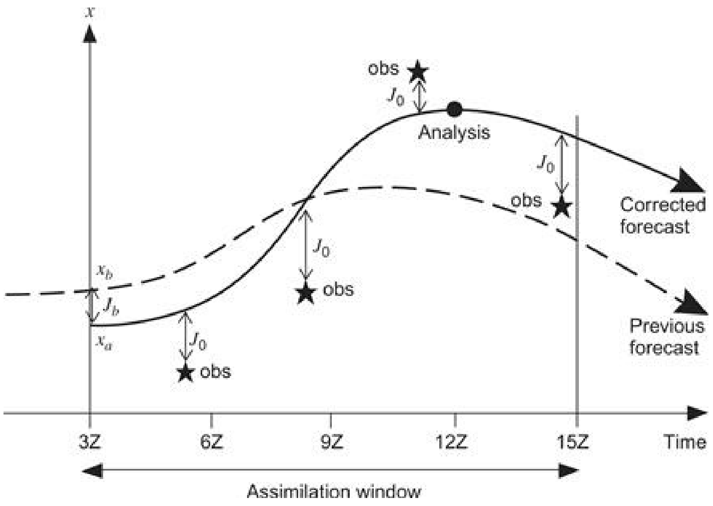


Figure 13.8 ECMWF data assimilation process. The *vertical axis* indicates the value of an atmospheric field variable. The *curve* labeled x_b is the first guess or "background" field, and x_a is the analysis field. J_b indicates the misfit between the background and analysis fields. *Stars* indicate the observations, and J_0 designates the "cost function," a measure of the misfit between the observations and the analysis, which along with J_b is minimized to produce the best estimate of the state of the atmosphere at the analysis time.

Courtesy ECMWF.

The EnKF approach deals directly with (13.61) and renders the calculation manageable either by processing the observations one at a time or by breaking the domain into smaller subdomains where the matrix operations are tractable. Furthermore, rather than propagate the large matrix **B**, the EnKF employs a sampling strategy, where the nonlinear forecast model is used to evolve each of a chosen number of ensemble members. This sample provides the basis to estimate **B**, taking care to deal with sampling error. There are two main strategies to updating the forecast ensemble with observations to produce an ensemble analysis. The first strategy, called "perturbed observations" adds random error consistent with the error covariance matrix, **R**, to the observations, so that every ensemble member assimilates slightly different observations. The second strategy, called "square root filters," uses the observations to update the ensemble mean forecast to an ensemble mean analysis, and then uses a modified version of (13.61) to update the deviations of each ensemble member from the mean. There are several, statistically equivalent, techniques that accomplish the square root update. In addition to the benefits of using state-dependent estimates of the prior covariance, **B**, the EnKF also has the attractive property that an ensemble analysis is immediately available for ensemble forecasting.

13.7 Predictability and Ensemble Forecasting

Atmospheric prediction is a fundamentally probabilistic problem. Errors in the observations and prior are nonzero and, as a result, errors in the analysis are as well. The problem of atmospheric prediction is to determine how these errors, as defined by a probability density function, evolve toward a future time. This probability density, p, obeys the *Liouville equation*,

$$\frac{Dp}{Dt} = -p\nabla \cdot \mathbf{F}(\mathbf{x}) \tag{13.64}$$

which expresses the conservation of total probability in direct analogy with mass conservation. The material derivative in this equation applies following the motion in *state-space* coordinates. The state-space velocity vector describes the change of the state vector, **x**, in time through a set of coordinates defined by a particular basis (e.g., spectral or grid point variables):

(13.65)

$$\frac{D\mathbf{x}}{Dt} = \mathbf{F}(\mathbf{x})$$

The velocity, **F**, is a vector-valued function that defines the dynamics of the system—for example, conservation of momentum, mass, and energy.

Having thus far made no approximations, it is a remarkable fact that, since the Liouville equation is a linear first-order PDE in *p*, there is an exact solution to the atmospheric predictability problem:

(13.66)

$$p[\mathbf{x}(t), t] \;=\; p[\mathbf{x}(t_0), t_0]e^{-\int_{t_0}^{t} \nabla \cdot \mathbf{F}(\mathbf{x})\, dt'}$$

Given an initial probability density function, $p\left[\mathbf{x}(t_0), t_0\right]$, the probability density at a future time, $p\left[\mathbf{x}(t), t\right]$, is determined by the divergence of state-space trajectories. When state-space trajectories diverge, $\nabla \cdot \mathbf{F} > 0$, probability density decreases exponentially fast. A **chaotic system** is characterized by a sensitivity to initial conditions, which is realized by diverging state-space trajectories (solutions become different exponentially fast). As a result, for chaotic systems such as the atmosphere, probability density approaches zero over the attractor, while the probability over state space remains unity.

The evolution of a probability density function with time is clearly illustrated for the scalar differential equation (Ehrendorfer, 2006)

(13.67)

$$\frac{dx}{dt} = x - x^3$$

An initially Gaussian distribution with a slightly positive mean shows that probability density quickly decreases and spreads in *x* as the solution diverges toward stable fixed points located at $x = -1$ and $x = 1$ (Figure 13.9a). State-space velocity (13.67) is positive (negative) for *x* positive (negative), so state-space trajectory divergence is maximized at $x = 0$, and becomes negative (convergence) for $|x| > 1/\sqrt{3}$. The solution ultimately converges at the fixed points with increasing probability density.

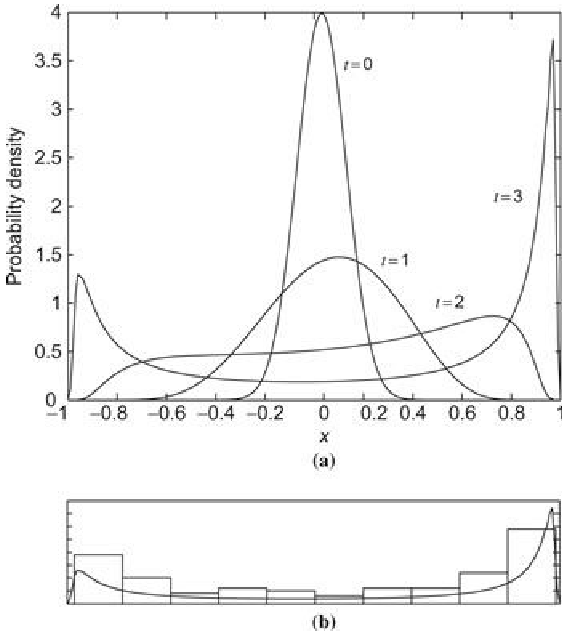


Figure 13.9 Time evolution of an initially Gaussian probability density function subject to the dynamics in (13.67) from numerically solving the Liouville equation (a). Populating a 100-member ensemble by randomly sampling from the initial probability density function, and solving (13.67) for the state at $t = 3$—summarized by a histogram in (b)—provides an estimate of the solution of the Liouville equation (*solid line*).

Unfortunately, numerical solution of (13.66) for any realistic problem is completely intractable. Consider discretizing the probability density by 100 bins for each state-space variable, as was done for Figure 13.9. For a three-variable system, we have a formidable calculation involving 1 million degrees of freedom, which becomes an intractable 10 billion for a five-variable system; obviously, solutions for a modern NWP system having 100 million variables is out of the question. This motivates a sampling strategy using an ensemble approach, whereby only a random selection of states from the initial distribution are integrated in time. An example is shown in the lower panel of Figure 13.9, where a 100-member ensemble is randomly drawn from the initial distribution, and then integrated in time. The ensemble-based histogram at time $t = 3$ provides a coarse, but useful, approximation to the Liouville solution (*solid line*).

Forecast accuracy is ultimately limited by sensitivity to the initial state, which is an essential characteristic of chaotic systems. A simple illustration of sensitivity to initial errors is given by the three-variable model of Lorenz (1963) (see Exercise M13.9). The sensitivity of forecasts in this system is illustrated in Figure 13.10, which shows the squared error in forecasts of fixed length (one nondimensional time unit) over the attractor. Errors are plotted on a logarithmic scale, so a wide range of errors is evident, with regions of large errors (*darker areas*) wrapping around the two fixed points toward the midpoint. Forecasts starting from a sensitive location are shown by the *solid curves*, which reveal that initially small differences grow suddenly as the two dynamical trajectories split in opposite directions. Such behavior is characteristic of a wide variety of systems governed by deterministic equations, including atmospheric flows.

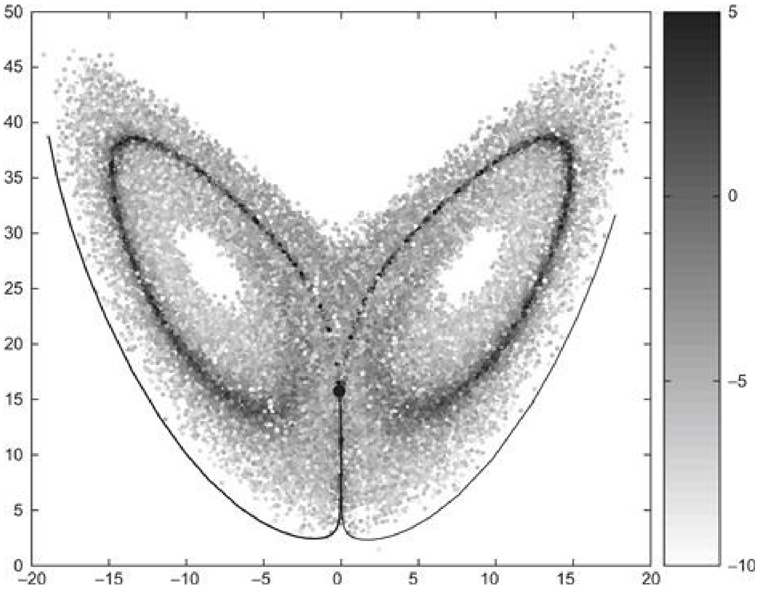


Figure 13.10 Sensitivity of forecasts to initial condition errors for the Lorenz (1963) attractor (forecasts are one time unit long). The natural logarithm of the errors is depicted in *gray-scale*, with *darker gray* indicating points that yield forecasts with larger errors. Solutions for the most sensitive location are given by *lines*.

Estimates of how this sort of error growth limits the inherent predictability of the atmosphere can be made using primitive equation forecast models. In these predictability experiments a “control” run is made using initial data corresponding to the observed flow at a given time. The initial data are then perturbed by introducing small random errors, as in Figure 13.10, and the model is run again. The growth of inherent error can then be estimated by comparing the second “forecast” run with the control. Results from a number of such studies indicate that the doubling time for the root mean square geopotential height error is about 2 to 3 days for small errors and somewhat greater for large errors. Thus, the theoretical limit for predictability on the synoptic scale is probably about 2 weeks.

Actual predictive skill with present deterministic models is, however, less than the theoretical limit imposed by inherent error growth. A number of factors are probably responsible for the failure of current models to reach the theoretical forecast skill limits. These include observational and analysis errors in initial data, inadequate model resolution, and unsatisfactory representation of physical processes involving radiation, clouds, precipitation, and boundary layer fluxes. An indication of the skill of the present generation of global forecast models is given in Figure 13.11, in which forecast skill is plotted in terms of the *anomaly correlation* of the 500-hPa geopotential height field, defined as the correlation between observed and predicted deviations from climatology. Subjective evaluations suggest that useful forecasts are obtained when the anomaly correlation is greater than 0.6. Thus, over the past 20 years the range for which ECMWF forecasts show useful skill has increased from 5 days to more than 7 days. There can, of course, be quite wide variations in skill from one situation to another reflecting variations in the degree of predictability from one atmospheric flow regime to another.

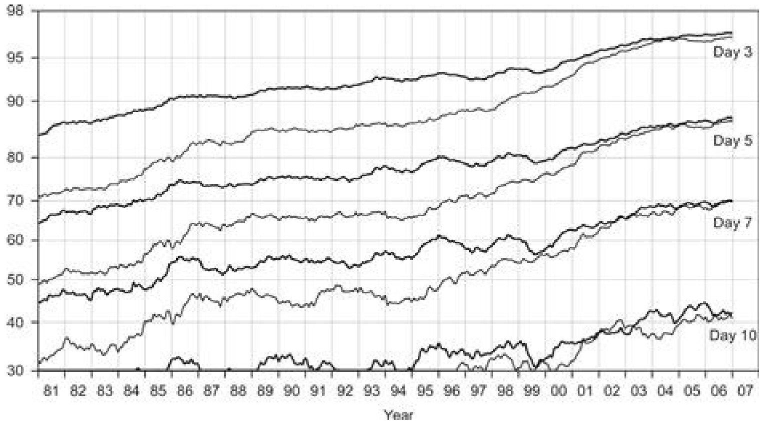


Figure 13.11 Anomaly correlation of 500-hPa height for 3-, 5-, 7-, and 10-day forecasts for the ECMWF operational model as a function of year. *Thick and thin lines* correspond to Northern and Southern Hemispheres, respectively. Note that the difference in skill between the two hemispheres has nearly disappeared in the last decade, a result of the successful assimilation of satellite data.

Courtesy ECMWF.

Ensembles provide the main technique for estimating the evolution of error in initial value problems. A few basic statistical results are available to evaluate the performance of ensemble forecast systems. For an ensemble of size *M*, the *ensemble spread*, *S*, about the ensemble mean is defined as

$$S = \frac{1}{M-1}\sum_{j=1}^{M}(u_j - \overline{u})^2 \tag{13.68}$$

where $\overline{u}$ is the ensemble mean and u_j is an ensemble member. This describes the variance about the ensemble mean. Averaging *S* over an infinite number of probability density functions, and over all states on the attractor (the "climate") gives

$$\langle S \rangle \equiv D \tag{13.69}$$

The error in ensemble mean is defined by

$$e_M = (\overline{u} - u_t)^2 \tag{13.70}$$

where u_t is the true (unknown) state. Again, averaging as above, it can be shown that (Murphy, 1989)

$$\langle e_M \rangle = \frac{M+1}{M}D \tag{13.71}$$

As a result, we see immediately that, for a deterministic forecast ($M = 1$), $\langle e_1 \rangle = 2D$.

Two main results derive from (13.69) and (13.71). The first is that the ratio of the error in the ensemble mean to a single deterministic forecast is

$$\frac{\langle e_M \rangle}{\langle e_1 \rangle} = \frac{M+1}{2M} \tag{13.72}$$

This clearly shows the benefit of averaging over errors with an ensemble: In the limit of large ensembles, the error in the ensemble mean is, on average, half of that for a single deterministic forecast (i.e., a randomly drawn ensemble member). The second main result involves the ratio of the error in the ensemble mean to the ensemble spread:

$$\frac{\langle e_M \rangle}{\langle S \rangle} = \frac{M+1}{M} \tag{13.73}$$

This shows that, except for small ensembles, the error in the ensemble mean should equal the spread in the ensemble. A common problem with ensemble forecast systems is that they have too little spread, so that (13.73) is larger than it would be for a well-calibrated system.

Suggested References

urran. Durran, *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*, is an excellent graduate-level textbook covering the numerical methods used in the atmospheric sciences.

alnay. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*, is an excellent graduate-level text on all aspects of modern numerical weather forecasting.

Problems

13.1. Show that for the barotropic vorticity equation on the Cartesian β plane (13.26) enstrophy and kinetic energy are conserved when averaged over the whole domain—that is, that the following integral constraints are satisfied:

$$\frac{d}{dt}\iint \frac{\zeta^2}{2} dx dy = 0, \qquad \frac{d}{dt}\iint \frac{\nabla\psi \cdot \nabla\psi}{2} dx dy = 0$$

Hint: To prove energy conservation, multiply (13.26) through by $-\psi$ and use the chain rule of differentiation.

13.2. Verify expression (13.31). Use periodic boundary conditions in both *x* and *y*.

13.3. The Euler backward method of finite differencing the advection equation is a two-step method consisting of a forward prediction step followed by a backward corrector step. In the notation of Section 13.2.2 the complete cycle is thus defined by

$$\begin{aligned} \hat{q}_m^* - \hat{q}_{m,s} &= -\frac{\sigma}{2}\left(\hat{q}_{m+1,s} - \hat{q}_{m-1,s}\right) \\ \hat{q}_{m,s+1} - \hat{q}_{m,s} &= -\frac{\sigma}{2}\left(\hat{q}_{m+1}^* - \hat{q}_{m-1}^*\right) \end{aligned}$$

where $\hat{q}_m^*$ is the first guess for time step $s + 1$. Use the method of Section 13.2.3 to determine the necessary condition for stability of this method.

13.4. Carry out truncation error analyses analogous to that of Table 13.1 for the centered difference approximation to the advection equation but for the cases $\sigma = 0.95$ and $\sigma = 0.25$.

13.5. Suppose that the streamfunction ψ is given by a single sinusoidal wave $\psi(x) = A\sin(kx)$. Find an expression for the error of the finite difference approximation

$$\frac{\partial^2 \psi}{\partial x^2} \approx \frac{\psi_{m+1} - 2\psi_m + \psi_{m-1}}{\delta x^2}$$

for $k\delta x = \pi/8, \pi/4, \pi/2,$ and π. Here $x = m\delta x$ with $m = 0, 1, 2, \ldots$.

13.6. Using the method given in Section 13.2.3, evaluate the computational *stability* of the following two finite difference approximations to the one-dimensional advection equation:

(a) $\hat{\zeta}_{m,s+1} - \hat{\zeta}_{m,s} = -\sigma\left(\hat{\zeta}_{m,s} - \hat{\zeta}_{m-1,s}\right)$

(b) $\hat{\zeta}_{m,s+1} - \hat{\zeta}_{m,s} = -\sigma\left(\hat{\zeta}_{m+1,s} - \hat{\zeta}_{m,s}\right)$

where $\sigma = c\delta t/\delta x > 0$. (The schemes labeled (a) and (b) are referred to as *upstream* and *downstream* differencing, respectively.) Show that scheme (a) damps the advected field and compute the fractional damping rate per time step for $\sigma = 0.25$ and $k\delta x = \pi/8$ for a field with the initial form $\zeta = \exp(ikx)$.

13.7. Using a staggered horizontal grid analogous to that shown in the left panel of Figure 13.5 (but for an equatorial *β*-plane geometry), express the linearized shallow water equations (11.27), (11.28), and (11.29) in finite difference form.

13.8. Verify the equality

$$\left(\frac{1 - i\tan\theta_p}{1 + i\tan\theta_p}\right) = \exp(-2i\theta_p)$$

given in (13.22).

13.9. Compute the ratio of the numerical phase speed to the true phase speed, c'/c, for the implicit differencing scheme of (13.19) for *p* = π, π/2,π/4,π/8, and π/16. Let $\sigma = 0.75$ and $\sigma = 1.25$. Compare your results to those of Table 13.1

13.10. Using the technique of Section 13.2.1, show that the following four-point difference formula for the first derivative is of fourth-order accuracy:

$$\psi'(x_0) \approx \frac{4}{3}\left(\frac{\psi(x_0 + \delta x) - \psi(x_0 - \delta x)}{2\delta x}\right) - \frac{1}{3}\left(\frac{\psi(x_0 + 2\delta x) - \psi(x_0 - 2\delta x)}{4\delta x}\right) \tag{13.74}$$

13.11. The Dufort–Frankel method for approximating the one-dimensional diffusion equation

$$\frac{\partial q}{\partial t} = K\frac{\partial^2 q}{\partial x^2}$$

can be expressed in the notation of Section 13.2.2 as

$$\hat{q}_{m,s+1} = \hat{q}_{m,s-1} + r\left[\hat{q}_{m+1,s} - \left(\hat{q}_{m,s+1} + \hat{q}_{m,s-1}\right) + \hat{q}_{m-1,s}\right]$$

where $r \equiv 2K\delta t/(\delta x)^2$. Show that this scheme is an explicit differencing scheme and that it is computationally stable for all values of δt.

13.12. Starting from (13.54), derive (13.55).

MATLAB Exercises

M13.1. The MATLAB script **finite_diff1.m** can be used to compare the centered difference first derivative formula (13.4) for the function $\psi(x) = \sin(\pi x/4)$ to the analytic expression $d\psi/dx = (\pi/4)\cos(\pi x/4)$ for various numbers of grid intervals in the domain $-4 \leq x \leq 4$. Graph the maximum error as a function of number of grid intervals, *ngrid*, in the range of 4 to 64. Carry out a similar analysis for the second derivative. (Note that first-order differencing is used next to the boundaries.)

M13.2. Modify the script of M13.1 to evaluate the error in the first derivative of the function $\tanh(x)$ using the same domain in x.

M13.3. The MATLAB script **advect_1.m** demonstrates the leapfrog differencing scheme for the one-dimensional advection equation as given in (13.8). By running the script for various grid intervals, find the dependence of phase error (degrees per wave period) on the number of grid intervals per wavelength for the range 4 to 64. Compare your results with Table 13.1.

M13.4. Modify the script **advect_1.m** by substituting the upstream differencing scheme described in Problem 13.6 and the Euler backward scheme described in Problem 13.3. Compare the accuracy in phase and amplitude of these two schemes with the leapfrog scheme of M13.3.

M13.5. The MATLAB script **advect_2.m** is similar to the script of M13.3 except that the initial tracer distribution is a localized positive definite pulse of width 0.25. Run the script for 100, 200, and 400 grid intervals. Explain why the pulse changes shape as it is advected. Using the results of Problem 13.10, modify the script to provide a fourth-order accurate approximation to the advection term and compare the accuracy of the fourth-order version to that of the second-order accurate system for the case with 400 grid intervals.

M13.6. The MATLAB script **advect_3.m** is a variant of the second-order accurate script of M13.5 in which the implicit differencing of (13.19) is utilized. Run this script with $\sigma = 1$ and 1.25 and 400 grid intervals and then try running **advect_2.m** with these values of σ. Give a qualitative explanation of these results referring to Problem 13.9.

M13.7. The MATLAB script **barotropic_model.m** can be used to solve a finite difference approximation to the barotropic vorticity equation, using the flux form of the nonlinear terms given in (13.30) and leapfrog time differencing. In this example the initial flow consists of a localized vortex and a constant zonal-mean flow. Run the model for a 10-day simulation using different values of the time step to determine the maximum time step *permitted* to maintain numerical stability. How does this compare with the CFL criterion (13.18)?

M13.8. The MATLAB script **Lorenz_model.m** gives an accurate numerical solution to the famous three-component Lorenz equations, which are often used to demonstrate sensitive dependence on initial conditions in chaotic systems (e.g., the atmosphere). The equations for the Lorenz model can be written as

$$dX/dt = -\sigma X + \sigma Y \quad (13.75)$$

$$dY/dt = -XZ + rX - Y \quad (13.76)$$

$$dZ/dt = XY - bZ \quad (13.77)$$

Here (*X, Y, Z*) may be regarded as a vector defining the "climate" and σ, r, and b are constants. Run the script, letting the initial value of $X = 10$ and verify that in the $X - Z$ plane the resulting trajectory of solution points has the well-known "butterfly wings" shape. Modify the code to save and plot the time histories of the variables and compare solutions in which the initial condition on X is increased and decreased by 0.1%. How long (in nondimensional time units) do the three solutions remain within 10% of each other?

M13.9. Modify the code in **Lorenz_model.m** to include a constant forcing $F = 10$ on the right side of the equation governing dX/dt. Describe how the character of the solution changes in this case. (See Palmer, 1993.)

M13.10. Use the results of Problem 13.7 to write a staggered grid version of the MATLAB script **forced_equatorial_mode2.m** (see Problem M11.6) for the shallow water model on the equatorial β plane. Set the staggered grid such that u and Φ are defined at the equator and v is defined at points $\delta y/2$ north and south of the equator. Compare the results of the staggered grid model with those of **forced_equatorial_mode2.m** when the latter has a grid spacing half that of the former.

M13.11. The MATLAB script **nonlinear_advect_diffuse.m** provides a numerical approximation to the one-dimensional nonlinear advection–diffusion equation

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} + K\frac{\partial^2 u}{\partial x^2}$$

with initial condition $u(x, 0) = \sin(2\pi x/Lx)$. The script uses leapfrog differencing for the advective term and forward differencing for the diffusion term. In the absence of diffusion, the flow would quickly evolve to a shock, but diffusion prevents this from occurring. Run the script and by varying δt determine the maximum time step for a stable solution. Modify the code by expressing the diffusion term in the Dufort–Frankel differencing scheme given in Problem 13.11. Determine the maximum time step allowed for stability in this case. How does the accuracy of the solution change as the time step is increased?

---

[1] See Richtmyer and Morton (1967).

[2] See Washington and Parkinson (1986) for a discussion of the properties of the Legendre function.

[3] The integral over $p(x_a)$, like all probability density functions, must be unity.

[4] The result is identical to the solution obtained using the method of linear least squares.

[5] The variance of the innovation is given by the sum of variances $\sigma_b^2 + \sigma_y^2$ provided that errors in the prior and observation are uncorrelated.

Appendix A

Useful Constants and Parameters

Gravitational constant	$G = 6.673 \times 10^{-11}\ \mathrm{N\,m^2\,kg^{-2}}$
Gravity at sea level	$g_0 = 9.81\ \mathrm{m\,s^{-2}}$
Mean radius of Earth	$a = 6.37 \times 10^6\ \mathrm{m}$
Earth's angular speed of rotation	$\Omega = 7.292 \times 10^{-5}\ \mathrm{rad\,s^{-1}}$
Universal gas constant	$R^* = 8.314 \times 10^3\ \mathrm{J\,K^{-1}\,kmol^{-1}}$
Gas constant for dry air	$R = 287\ \mathrm{J\,K^{-1}\,kg^{-1}}$
Specific heat of dry air at constant pressure	$c_p = 1004\ \mathrm{J\,K^{-1}\,kg^{-1}}$
Specific heat of dry air at constant volume	$c_v = 717\ \mathrm{J\,K^{-1}\,kg^{-1}}$
Ratio of specific heats	$\gamma = c_p/c_v = 1.4$
Molecular weight of water	$m_v = 18.016\ \mathrm{kg\,kmol^{-1}}$
Latent heat of condensation at 0°C	$L_c = 2.5 \times 10^6\ \mathrm{J\,kg^{-1}}$
Mass of Earth	$M = 5.988 \times 10^{24}\ \mathrm{kg}$
Standard sea-level pressure	$p_0 = 1013.25\ \mathrm{hPa}$
Standard sea-level temperature	$T_0 = 288.15\ \mathrm{K}$
Standard sea-level density	$\rho_0 = 1.225\ \mathrm{kg\,m^{-3}}$

Appendix B

List of Symbols

Only the principal symbols are listed. Symbols formed by adding primes, overbars, or subscripted indices are not listed separately. Boldface type indicates vector quantities. Where symbols have more than one meaning, the section where the second meaning is first used is indicated in the list.

a	Radius of Earth
b	Buoyancy
c	Phase speed of a wave
c_p	Specific heat of dry air at constant pressure
c_{pv}	Specific heat of water vapor at constant pressure
c_v	Specific heat of dry air at constant volume
c_w	Specific heat of liquid water
d	Grid distance
e	Internal energy per unit mass
f	Coriolis parameter ($\equiv 2\Omega \sin\phi$)
g	Magnitude of gravity
$\mathbf{g}$	Gravity
$\mathbf{g}^*$	Gravitational acceleration
h	Depth of fluid layer; moist static energy (Section 2.9.1)
i	Square root of -1
$\mathbf{i}$	Unit vector along the x axis
$\mathbf{j}$	Unit vector along the y axis
$\mathbf{k}$	Unit vector along the z axis
k	Zonal wave number

l	Mixing length; meridional wave number
m	Mass element; vertical wave number; planetary wave number (Section 13.4)
m_v	Molecular weight of water
n	Distance in direction normal to a parcel trajectory; meridional index for equatorial waves (Section 11.4)
$\mathbf{n}$	Unit vector normal to a parcel trajectory
p	Pressure
p_s	Standard constant pressure; surface pressure in σ coordinates (Section 10.3.1)
q	Quasi-geostrophic potential vorticity; water vapor mixing ratio
q_s	Saturation mixing ratio
r	Radial distance in spherical coordinates
$\mathbf{r}$	Position vector
s	Generalized vertical coordinate; distance along a parcel trajectory (Section 3.2); entropy (Section 2.7); dry static energy (Section 2.9.1)
t	Time
$\mathbf{t}$	Unit vector parallel to a parcel trajectory
u^*	Friction velocity
u	x component of velocity (eastward)
v	y component of velocity (northward)
w	z component of velocity (upward)
w^*	Vertical motion in log-pressure system
x, y, z	Eastward, northward, and upward distance, respectively
z^*	Vertical coordinate in log-pressure system
$\mathbf{A}$	Arbitrary vector
A	Area
B	Convective available potential energy
C_d	Surface drag coefficient
D_e	Depth of Ekman layer

E	Evaporation rate
E_I	Internal energy
$\mathbf{F}$	Force; EP flux (Section 10.2)
$\mathbf{Fr}$	Frictional force
G	Universal gravitational constant; zonal force (Section 10.2)
H	Scale height
J	Diabatic heating rate
K	Total horizontal wavenumber; kinetic energy (Section 9.3)
K_m	Eddy viscosity coefficient
L	A length scale
L_c	Latent heat of condensation
M	Mass; mass convergence in Ekman layer (Section 8.4); absolute zonal momentum (Section 9.3); angular momentum (Section 10.3)
N	Buoyancy frequency
P	Available potential energy (Section 7.3); precipitation rate (Section 11.3)
$\mathbf{Q}$	Q vector
R	Gas constant for dry air; distance from the axis of rotation of Earth to a point on the surface of Earth (Section 1.3); diabatic energy generation rate (Section 10.4)
$\mathbf{R}$	Vector in the equatorial plane directed from the axis of rotation to a point on the surface of Earth
R^*	Universal gas constant
S_p	$\equiv -T\partial \ln\theta/\partial p$, stability parameter in pressure coordinates
T	Temperature
U	Horizontal velocity scale
V	Speed in natural coordinates
δV	Volume increment
$\mathbf{U}$	Three-dimensional velocity vector
$\mathbf{V}$	Horizontal velocity vector
W	Vertical motion scale

X	Zonal turbulent drag force
Z	Geopotential height
α	Specific volume
β	$\equiv df/dy$, variation of the Coriolis parameter with latitude; angular direction of the wind (Section 3.3)
γ	$\equiv c_p/c_v$, ratio of specific heats
ε	Rate of frictional energy dissipation
ζ	Vertical component of relative vorticity
η	Vertical component of absolute vorticity; weighting function for heating profile (Section 11.3)
θ	Potential temperature
$\dot{\theta}$	$\equiv D\theta/Dt$, vertical motion in isentropic coordinates
θ_e	Equivalent potential temperature
κ	$\equiv R/c_p$, ratio of gas constant to specific heat at constant pressure; Rayleigh friction coefficient (Section 10.6)
λ	Longitude, positive eastward
μ	Dynamic viscosity coefficient
ν	Angular frequency; kinematic viscosity (Section 1.2.2)
ρ	Density
σ	$\equiv -RT_0 p^{-1} d \ln \theta_0/dp$, standard atmosphere static stability parameter in isobaric coordinates
	$\equiv -p/p_s$, vertical coordinate in σ system (Section 10.3); “density” in isentropic coordinates (Section 4.6)
τ_d	Diffusive time scale
τ_E	Eddy stress
ϕ	Latitude
χ	Geopotential tendency; meridional streamfunction; tracer mixing ratio
ψ	Horizontal streamfunction
ω	Vertical velocity ($\equiv dp/dt$) in isobaric coordinates
$\boldsymbol{\omega}$	Vorticity vector

Γ	$\equiv -dT/dz$, lapse rate of temperature
Γ_d	Dry adiabatic lapse rate
Φ	Geopotential
Π	Ertel potential vorticity; Exner function
Θ	Potential temperature deviation
Ω	Angular speed of rotation of Earth
$\boldsymbol{\Omega}$	Angular velocity of Earth

Appendix C

Vector Analysis

C.1 Vector Identities

The following formulas may be shown to hold where Φ is an arbitrary scalar and **A** and **B** are arbitrary vectors.

$$\begin{aligned}
\nabla \times \nabla\Phi &= 0 \\
\nabla \cdot (\Phi\mathbf{A}) &= \Phi\nabla \cdot (\mathbf{A}) + \mathbf{A} \cdot \nabla\Phi \\
\nabla \times (\Phi\mathbf{A}) &= \nabla\Phi \times \mathbf{A} + \Phi(\nabla \times \mathbf{A}) \\
\nabla \cdot (\nabla \times \mathbf{A}) &= 0 \\
(\mathbf{A} \cdot \nabla)\mathbf{A} &= \frac{1}{2}\nabla(\mathbf{A} \cdot \mathbf{A}) - \mathbf{A} \times (\nabla \times \mathbf{A}) \\
\nabla \times (\mathbf{A} \times \mathbf{B}) &= \mathbf{A}(\nabla \cdot \mathbf{B}) - \mathbf{B}(\nabla \cdot \mathbf{A}) - (\mathbf{A} \cdot \nabla)\mathbf{B} + (\mathbf{B} \cdot \nabla)\mathbf{A} \\
\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) &= (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}
\end{aligned}$$

C.2 Integral Theorems

(a) Divergence theorem:

$$\int_A \mathbf{B} \cdot \mathbf{n}\, dA = \int_V \mathbf{V} \cdot \mathbf{B}\, dV$$

where V is a volume enclosed by surface A and **n** is a unit normal on A.
(b) Stokes's theorem:

$$\oint \mathbf{B} \cdot d\mathbf{l} = \int_A (\mathbf{V} \times \mathbf{B}) \cdot \mathbf{n}\, dA$$

where A is a surface bounded by the line traced by the position vector **l** and **n** is a unit normal of A.

C.3 Vector Operations in Various Coordinate Systems

(a) Cartesian coordinates: (x, y, z)

Coordinate	Symbol	Velocity component	Unit vector
Eastward	x	u	**i**
Northward	y	v	**j**
Upward	z	w	**k**

$$\nabla\Phi = \mathbf{i}\frac{\partial\Phi}{\partial x} + \mathbf{j}\frac{\partial\Phi}{\partial y} + \mathbf{k}\frac{\partial\Phi}{\partial z}$$

$$\nabla\cdot\mathbf{V} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$$

$$\mathbf{k}\cdot(\nabla\times\mathbf{V}) = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$$

$$\nabla_h^2\Phi = \frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2}$$

(b) Cylindrical coordinates: (r, λ, z)

Coordinate	Symbol	Velocity component	Unit vector
Radial	r	u	**i**
Azimuthal	λ	v	**j**
Upward	z	w	**k**

$$\nabla\Phi = \mathbf{i}\frac{\partial\Phi}{\partial r} + \mathbf{j}\frac{1}{r}\frac{\partial\Phi}{\partial\lambda} + \mathbf{k}\frac{\partial\Phi}{\partial z}$$

$$\nabla\cdot\mathbf{V} = \frac{1}{r}\frac{\partial(ru)}{\partial r} + \frac{1}{r}\frac{\partial v}{\partial\lambda}$$

$$\mathbf{k}\cdot(\nabla\times\mathbf{V}) = \frac{1}{r}\frac{\partial(rv)}{\partial r} - \frac{1}{r}\frac{\partial u}{\partial\lambda}$$

$$\nabla_h^2\Phi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\Phi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\Phi}{\partial\lambda^2}$$

(c) Spherical coordinates: (λ, ϕ, r)

Coordinate	Symbol	Velocity component	Unit vector
Longitude	λ	u	**i**
Latitude	ϕ	v	**j**
Height	r	w	**k**

$$\nabla\Phi = \frac{\mathbf{i}}{r\cos\phi}\frac{\partial\Phi}{\partial\lambda} + \mathbf{j}\frac{1}{r}\frac{\partial\Phi}{\partial\phi} + \mathbf{k}\frac{\partial\Phi}{\partial r}$$

$$\nabla\cdot\mathbf{V} = \frac{1}{r\cos\phi}\left[\frac{\partial u}{\partial\lambda} + \frac{\partial(v\cos\phi)}{\partial\phi}\right]$$

$$\mathbf{k}\cdot(\nabla\times\mathbf{V}) = \frac{1}{r\cos\phi}\left[\frac{\partial v}{\partial\lambda} - \frac{\partial(u\cos\phi)}{\partial\phi}\right]$$

$$\nabla_h^2\Phi = \frac{1}{r^2\cos^2\phi}\left[\frac{\partial^2\Phi}{\partial\lambda^2} + \cos\phi\frac{\partial}{\partial\phi}\left(\cos\phi\frac{\partial\Phi}{\partial\phi}\right)\right]$$

Appendix D

Moisture Variables

D.1 Equivalent Potential Temperature

A mathematical expression for θ_e can be derived by applying the first law of thermodynamics to a mixture of 1 kg dry air plus q kg of water vapor (q, called the *mixing ratio*, is usually expressed as grams of vapor per kilogram of dry air). If the parcel is not saturated, the dry air satisfies the energy equation

$$c_p dT - \frac{d(p-e)}{p-e} RT = 0 \tag{D.1}$$

and the water vapor satisfies

$$c_{pv} dT - \frac{de}{e}\frac{R^*}{m_v} T = 0 \tag{D.2}$$

where the motion is assumed to be adiabatic. Here, e is the partial pressure of the water vapor, c_{pv} is the specific heat at constant pressure for the vapor, R^* is the universal gas constant, and m_v is the molecular weight of water. If the parcel is saturated, then condensation of $-dq_s$ kg vapor per kilogram dry air will heat the mixture of air and vapor by an amount of heat that goes into the liquid water, and the saturated parcel must satisfy the energy equation

$$c_p dT + q_s c_{pv} dT - \frac{d(p-e_s)}{p-e_s} RT - q_s \frac{de_s}{e_s}\frac{R^*}{m_v} T = -L_c dq_s \tag{D.3}$$

where q_s and e_s are the saturation mixing ratio and vapor pressure, respectively. The quantity de_s/e_s may be expressed in terms of temperature using the Clausius–Clapeyron equation[1]

$$\frac{de_s}{dT} = \frac{m_v L_c e_s}{R^* T^2} \tag{D.4}$$

Substituting from (D.4) into (D.3) and rearranging terms, we obtain

$$-L_c d\left(\frac{q_s}{T}\right) = c_p \frac{dT}{T} - \frac{R d(p-e_s)}{p-e_s} + q_s c_{pv} \frac{dT}{T} \tag{D.5}$$

If we now define the potential temperature of the dry air θ_d, according to

$$c_p d \ln \theta_d = c_p d \ln T - R d \ln (p-e_s)$$

we can rewrite (D.5) as

$$-L_c d\left(\frac{q_s}{T}\right) = c_p d\ln\theta_d + q_s c_{pv} d\ln T \tag{D.6}$$

However, it may be shown that

$$dL_c/dT = c_{pv} - c_w \tag{D.7}$$

where c_w is the specific heat of liquid water. Combining (D.7) and (D.6) to eliminate c_{pv} yields

$$-d\left(\frac{L_c q_s}{T}\right) = c_p d\ln\theta_d + q_s c_w d\ln T \tag{D.8}$$

Neglecting the last term in (D.8), we may integrate from the originial state (p, T, q_s, e_s, θ_d) to a state where $q_s \rightarrow 0$. Therefore, the equivalent potential temperature of a saturated parcel is given by

$$\theta_e = \theta_d \exp\left(L_c q_s/c_p T\right) \approx \theta \exp\left(L_c q_s/c_p T\right) \tag{D.9}$$

Equation (D.9) may also be applied to an unsaturated parcel provided that the temperature used is the temperature that the parcel would have if brought to saturation by an adiabatic expansion.

D.2 Pseudoadiabatic Lapse Rate

Section 2.9.2 showed that from the first law of thermodynamics the lapse rate for a saturated parcel undergoing pseudoadiabatic ascent can be obtained from the formula

$$\frac{dT}{dz} + \frac{g}{c_p} = -\frac{L_c}{c_p}\left[\left(\frac{\partial q_s}{\partial T}\right)_p \frac{dT}{dz} - \left(\frac{\partial q_s}{\partial p}\right)_T \rho g\right] \tag{D.10}$$

Noting that $q_s \cong \varepsilon e_s/p$, where $\varepsilon = 0.622$ is the ratio of the molecular weight of water to that of dry air and utilizing (D.4), we can express the partial derivatives in (D.10) as

$$\left(\frac{\partial q_s}{\partial p}\right)_T \approx -\frac{q_s}{p} \quad \text{and} \quad \left(\frac{\partial q_s}{\partial T}\right)_p \approx \frac{\varepsilon}{p}\frac{\partial e_s}{\partial T} = \frac{\varepsilon^2 L_c e_s}{pRT^2} = \frac{\varepsilon L_c q_s}{RT^2}$$

Substitution into (D.10), and noting that $g/c_p = \Gamma_d$, then yields the desired result:

$$\Gamma_s \equiv -\frac{dT}{dz} = \Gamma_d \frac{\left[1 + L_c q_s/(RT)\right]}{\left[1 + \varepsilon L_c^2 q_s/\left(c_p R T^2\right)\right]}$$

---

[1] For a derivation, see Curry and Webster (1999, p. 108).

Appendix E

Standard Atmosphere Data

Geopotential Height versus Pressure

Pressure (hPa)	***Z* (km)**
1000	0.111
900	0.988
850	1.457
700	3.012
600	4.206
500	5.574
400	7.185
300	9.164
250	10.363
200	11.784
150	13.608
100	16.180
50	20.576
30	23.849
20	26.481
10	31.055

Standard Atmosphere Temperature, Pressure, and Density as a Function of Geopotential Height

Z (km)	Temperature (K)	Pressure (hPa)	Density (kg m^{-3})
0	288.15	1013.25	1.225
1	281.65	898.74	1.112
2	275.15	794.95	1.007
3	268.65	701.08	0.909
4	262.15	616.40	0.819
5	255.65	540.19	0.736
6	249.15	471.81	0.660
7	242.65	410.60	0.590
8	236.15	355.99	0.525
9	229.65	307.42	0.466
10	223.15	264.36	0.412
12	216.65	193.30	0.311
14	216.65	141.01	0.227
16	216.65	102.87	0.165
18	216.65	75.05	0.121
20	216.65	54.75	0.088
24	220.65	29.30	0.046
28	224.65	15.86	0.025
32	228.65	08.68	0.013

Appendix F

Symmetric Baroclinic Oscillations

A variant of the derivation of the Sawyer–Eliassen equation for forced transverse circulations in baroclinic zones (9.15) can be used to obtain an equation for free symmetric transverse oscillations, which can be used to derive an expression for the growth rate of unstable symmetric oscillations or the frequency of stable symmetric oscillations.

Suppose that the flow field is zonally symmetric so that $u_g = u_g(y,z)$ and $b = b(y,z)$. The ageostrophic (transverse) flow is given by the streamfunction $\psi(y,z)$ so that $v_a = -\partial\psi/\partial z;\ w_a = \partial\psi/\partial y$. Then from (9.12) $Q_2 = 0$ so that any transverse circulation is unforced and must arise from a departure from exact geostrophic balance. This can be simply represented by adding an acceleration term in the *y*-momentum equation so that (9.10) becomes

$$\frac{\partial}{\partial t}\left(-\frac{\partial^2\psi}{\partial z^2}\right)+f\frac{\partial u_g}{\partial z}+\frac{\partial b}{\partial y}=0 \tag{F.1}$$

Then, combining (9.11) and (9.13) as before and applying (F.1), we obtain

$$\frac{D}{Dt}\left[\frac{\partial}{\partial t}\left(\frac{\partial^2\psi}{\partial z^2}\right)\right]+N_s^2\frac{\partial^2\psi}{\partial y^2}+F^2\frac{\partial^2\psi}{\partial z^2}+2S^2\frac{\partial^2\psi}{\partial y\partial z}=0 \tag{F.2}$$

Neglecting terms quadratic in the streamfunction,

$$\frac{D}{Dt}=\frac{\partial}{\partial t}+v_a\frac{\partial}{\partial y}+w_a\frac{\partial}{\partial z}\approx\frac{\partial}{\partial t}$$

then yields the desired result:

$$\frac{\partial^2}{\partial t^2}\left(\frac{\partial^2\psi}{\partial z^2}\right)+N_s^2\frac{\partial^2\psi}{\partial y^2}+F^2\frac{\partial^2\psi}{\partial z^2}+2S^2\frac{\partial^2\psi}{\partial y\partial z}=0 \tag{F.3}$$

Appendix G

Conditional Probability and Likelihood

A random variable is defined by the result of a rule (e.g., a function) that associates a real number with each outcome in a sample space, S. Consider two discrete random variables, A and B, that are associated with events in the sample space. The probabilities of A and B, denoted by $P(A)$ and $P(B)$, respectively, take values between zero and one and, by definition, $P(S) = 1$. Conditional probability is denoted by $P(A \mid B)$ and read as "the probability of A given B." Given that B has occurred, the sample space shrinks from S to B.

We expect that $P(A \mid B)$ is proportional to $P(A \cap B)$, which is the intersection probability that both A and B occur. Since $P(B \mid B)$ must be unity and $P(B \cap B) = P(B)$, the constant of proportionality must be $1/P(B)$:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{G.1}$$

Similarly,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \tag{G.2}$$

Using (G.2) to replace $P(A \cap B)$ in (G.1) gives **Bayes's Theorem**,

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \tag{G.3}$$

Likelihood

With conditional probability, we are given specific information about an event that affects the probability of another event. For example, knowing that $B = b$ allows us to "update" the probability of A by $P(A \mid B = b)$ through (G.3). Conversely, if we view the conditional probability as a function of the *second* argument, B, we get a likelihood function

$$L(b \mid A) = \alpha P(A \mid B = b) \tag{G.4}$$

where α is a constant parameter. For example, consider a sequence of observations of a two-sided coin with outcomes "H" and "T." If we flip the coin once and observe "H," we can ask, "What is the likelihood that the true probability $P(H)$ is 0.5?" Figure G.1 shows that $L = 0.5$, and in fact the most likely value is $P(H) = 1$; note that $L(0) = 0$ since "H" has been observed and therefore must have a nonzero probability. Suppose flipping the coin again also gives "H." Maximum likelihood is still located at $P(H) = 1$, and the likelihood of smaller $P(H)$ is diminished. Remember that we can turn this around and ask, “If we *know* that $P(H) = 0.5$ (i.e., the coin is fair), what is the probability of observing ‘HH’?” This is conditional probability, which in this case is 0.25 (the same as $L(0.5 \mid HH)$. If, on the other hand, one of the observations is also “T,” then both $L(0)$ and $L(1)$ must be zero (Figure G.1). Note that, unlike conditional probability likelihood is not a probability since it may take values larger than unity and it does not integrate to one.

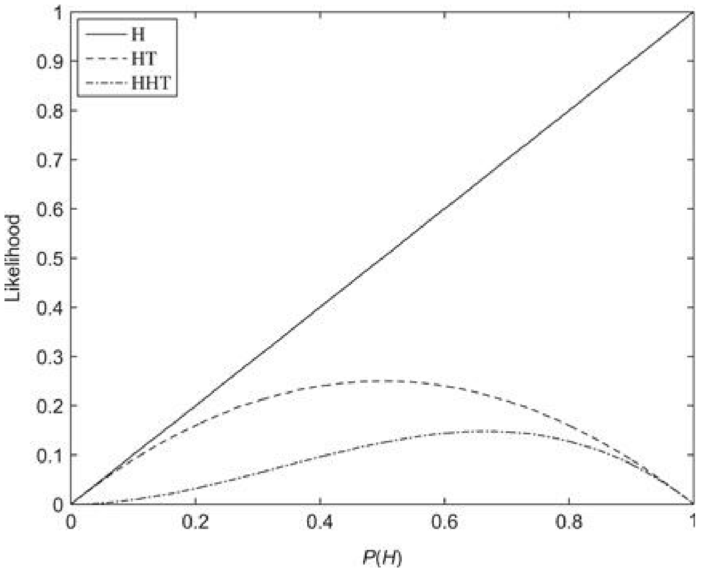


Figure G.1 Likelihood as a function of the probability of “H” on a two-sided coin given observations of heads “H” and tails “T.”

Bibliography

1. Acheson DJ. Elementary Fluid Dynamics New York: Oxford University Press; 1990.
2. Andrews DG, Holton JR, Leovy CB. Middle Atmosphere Dynamics Orlando: Academic Press; 1987.
3. Andrews DG, McIntyre ME. Planetary waves in horizontal and vertical shear: the generalized Eliassen–Palm relation and the mean zonal acceleration. *J Atmos Sci.* 1976;33:2031–2048.
4. Arakawa A, Schubert W. Interaction of a cumulus cloud ensemble with the large-scale environment Part I. *J Atmos Sci.* 1974;31:674–701.
5. Arya SP. Introduction to Micrometeorology second ed. Orlando: Academic Press; 2001.
6. Baer F. An alternate scale representation of atmospheric energy spectra. *J Atmos Sci.* 1972;29:649–664.
7. Battisti DS, Sarachik ES, Hirst AC. A consistent model for the large-scale steady surface atmospheric circulation in the tropics. *J Climate*. 1992;12:2956–2964.
8. Blackburn M. Interpretation of ageostrophic winds and implications for jet stream maintenance. *J Atmos Sci.* 1985;42:2604–2620.
9. Blackmon ML, Wallace JM, Lau N-C, Mullen SL. An observational study of the northern hemisphere wintertime circulation. *J Atmos Sci.* 1977;34:1040–1053.
10. Bluestein H. *Synoptic-Dynamic Meteorology in Midlatitudes*. Vol. II New York: Oxford University Press; 1993.
11. Bourne DE, Kendall PC. Vector Analysis Boston: Allyn & Bacon; 1968.
12. Brasseur G, Solomon S. Aeronomy of the Middle Atmosphere: Chemistry and Physics of the Stratosphere and Mesosphere second ed. Norwell, MA: Reidel; 1986.
13. Brown RA. A secondary flow model for the planetary boundary layer. *J Atmos Sci.* 1970;27:742–757.
14. Brown RA. Fluid Mechanics of the Atmosphere Orlando: Academic Press; 1991.
15. Chang CP. Westward propagating cloud patterns in the Tropical Pacific as seen from time-composite satellite photographs. *J Atmos Sci.* 1970;27:133–138.
16. Chang EKM. Downstream development of baroclinic waves as inferred from regression analysis. *J Atmos Sci.* 1993;50:2038–2053.
17. Chapman S, Lindzen RS. Atmospheric Tides: Thermal and Gravitational Dordrecht, Holland: Reidel; 1970.
18. Charney JG. The dynamics of long waves in a baroclinic westerly current. *J Meteor.* 1947;4:135–163.
19. Charney JG. On the scale of atmospheric motions. *Geofys Publ.* 1948;17(2):1–17.
20. Charney JG, DeVore JG. Multiple flow equilibria in the atmosphere and blocking. *J Atmos Sci.* 1979;36:1205–1216.
21. Charney JG, Eliassen A. A numerical method for predicting the perturbations of the middle latitude westerlies. *Tellus*. 1949;1(2):38–54.
22. Cunningham P, Keyser D. Dynamics of jet streaks in a stratified quasi-geostrophic atmosphere: steady-state representations. *Quart J Roy Meteor Soc.* 2006;130A:1579–1609.
23. Curry JA, Webster PJ. Thermodynamics of Atmospheres and Oceans San Diego: Academic Press; 1999.
24. Dai A, Wigley TML, Boville BA, Kiehl JT, Buja LE. Climates of the twentieth and twenty-first centuries simulated by the NCAR climate system model. *Climate*. 2001;14:485–519.
25. Dunkerton TJ. Middle atmosphere: quasi-biennial oscillation. In: Holton JR, Curry JA, Pyle JA, eds. *Encyclopedia of Atmospheric Sciences*. London: Academic Press; 2003.
26. Durran DR. Mountain waves and downslope winds. In: Blumen W, ed. *Atmospheric Processes over Complex Terrain*. American Meteorological Society 1990;59–82.
27. Durran DR. Is the Coriolis force really responsible for the inertial oscillation?. *Bull Am Meteorol Soc.* 1993;74(11):2179–

2184.

28. Durran DR. Numerical Methods for Wave Equations in Geophysical Fluid Dynamics New York: Springer; 1999.
29. Durran DR, Snellman LW. The diagnosis of synoptic-scale vertical motion in an operational environment. *Weather Forecasting*. 1987;1:17–31.
30. Eady ET. Long waves and cyclone waves. *Tellus*. 1949;1(3):33–52.
31. Eliassen A. Transverse circulations in frontal zones. In: Newton CW, Holopainen EO, eds. *Extratropical Cyclones, The Erik Palmén Memorial Volume*. Boston: American Meteorological Society; 1990;155–164.
32. Emanuel KA. Toward a general theory of hurricanes. *Am Sci.* 1988;76:370–379.
33. Emanuel KA. Atmospheric Convection New York: Oxford University Press; 1994.
34. Emanuel KA. Quasi-equilibrium thinking. In: Randall DA, ed. *General Circulation Model Development*. New York: Academic Press; 2000;225–255.
35. Fleming EL, Chandra S, Barnett JJ, Corey M. Zonal mean temperature, pressure, zonal wind and geopotential height as functions of latitude. *Adv Space Res.* 1990;10(12):11–59.
36. Garratt JR. The Atmospheric Boundary Layer Cambridge: Cambridge University Press; 1992.
37. Gill AE. Atmosphere-Ocean Dynamics New York: Academic Press; 1982.
38. Hakim GJ. Role of nonmodal growth and nonlinearity in cyclogenesis initial-value problems. *J Atmos Sci.* 2000;57:2951–2967.
39. Hakim GJ. Cyclogenesis. In: *Encyclopedia of the Atmospheric Sciences*. Boston: Elsevier; 2002.
40. Hamill TM. Ensemble-based atmospheric data assimilation. In: Palmer T, Hagedorn R, eds. *Predictability of Weather and Climate*. Cambridge, UK: Cambridge University; 2006;124–156.
41. Held IM. Stationary and quasi-stationary eddies in the extratropical troposphere: theory. In: Hoskins BJ, Pearce R, eds. *Large-Scale Dynamical Processes in the Atmosphere*. New York: Academic Press; 1983;127–168.
42. Hess SL. Introduction to Theoretical Meteorology New York: Holt; 1959.
43. Hide R. On the dynamics of rotating fluids and related topics in geophysical fluid mechanics. *Bull Am Meteorol Soc.* 1966;47:873–885.
44. Hildebrand FB. Advanced Calculus for Applications second ed. New York: Prentice Hall; 1976.
45. Holton JR. Meridional distribution of stratospheric trace constituents. *J Atmos Sci.* 1986;43:1238–1242.
46. Holton JR, Haynes PH, McIntyre ME, Douglass AR, Rood RB, Pfister L. Stratosphere-troposphere exchange. *Rev Geophys.* 1995;33:403–439.
47. Horel JD, Wallace JM. Planetary scale atmospheric phenomena associated with the southern oscillation. *Mon Wea Rev.* 1981;109:813–829.
48. Hoskins BJ. The geostrophic momentum approximation and the semi-geostrophic equations. *J Atmos Sci.* 1975;32:233–242.
49. Hoskins BJ. The mathematical theory of frontogenesis. *Annu Rev Fluid Mech.* 1982;14:131–151.
50. Hoskins BJ. Dynamical processes in the atmosphere and the use of models. *Quart J Roy Meteor Soc.* 1983;109:1–21.
51. Hoskins BJ, Bretherton FP. Atmospheric frontogenesis models: mathematical formulation and solution. *J Atmos Sci.* 1972;29:11–37.
52. Hoskins BJ, McIntyre ME, Robertson AW. On the use and significance of isentropic potential vorticity maps. *Quart J Roy Meteorol Soc.* 1985;111:877–946.
53. Houze Jr RA. Cloud Dynamics San Diego: Academic Press; 1993.
54. James IN. Introduction to Circulating Atmospheres Cambridge, UK: Cambridge University Press; 1994.
55. Kalnay E. Atmospheric Modeling, Data Assimilation and Predictability Cambridge, UK: Cambridge University Press; 2003.

56. Kepert JD, Wang Y. The dynamics of boundary layer jets within the tropical cyclone core Part II: Nonlinear enhancement. *J Atmos Sci.* 2001;58:2485–2501.

57. Klemp JB. Dynamics of tornadic thunderstorms. *Annu Rev Fluid Mech.* 1987;19:369–402.

58. Lackmann G. Midlatitude Synoptic Meteorology: Dynamics, Analysis, and Forecasting Boston: American Meteorological Society; 2012.

59. Lim GH, Holton JR, Wallace JM. The structure of the ageostrophic wind field in baroclinic waves. *J Atmos Sci.* 1991;48:1733–1745.

60. Lindzen RS, Batten ES, Kim JW. Oscillations in atmospheres with tops. *Mon Wea Rev.* 1968;96:133–140.

61. Lorenz EN. Energy and numerical weather prediction. *Tellus*. 1960;12:364–373.

62. Lorenz EN. The Nature and Theory of the General Circulation of the Atmosphere Geneva: World Meteorological Organization; 1967.

63. Lorenz EN. Some aspects of atmospheric predictability. In: Burridge DM, Källén E, eds. *Problems and Prospects in Long and Medium Range Weather Forecasting*. New York: Springer-Verlag; 1984;1–20.

64. Madden RA. Intraseasonal oscillation (Madden–Julian oscillation. In: Holton JR, Curry JA, Pyle JA, eds. *Encyclopedia of Atmospheric Sciences*. London: Academic Press; 2003;2334–2338.

65. Madden RA, Julian PR. Description of global-scale circulation cells in the tropics with a 40–50 day period. *Atmos Sci.* 1972;29:1109–1123.

66. Martin JE. Mid-Latitude Atmospheric Dynamics New York: John Wiley & Sons; 2006.

67. Matsuno T. Quasi-geostrophic motions in the equatorial area. *J Meteorol Soc Japan*. 1966;44:25–43.

68. Nappo CJ. An Introduction to Atmospheric Gravity Waves San Diego: Academic Press; 2002.

69. Naujokat B. An update of the observed quasi-biennial oscillation of the stratospheric winds over the tropics. *J Atmos Sci.* 1986;43:1873–1877.

70. Norton WA. Middle atmosphere: transport circulation. In: Holton JR, Curry JA, Pyle JA, eds. *Encyclopedia of Atmospheric Sciences*. London: Academic Press; 2003;1353–1358.

71. Oort, A.H., 1983. Global atmospheric circulation statistics, 1958–1973. *NOAA Professional Paper 14*, U. S. Government Printing Office, Washington, DC.

72. Oort AH, Peixoto JP. The annual cycle of the of the atmosphere on a planetary scale. *J Geophys Res.* 1974;79:2705–2719.

73. Oort AH, Peixoto JP. Global angular momentum and energy balance requirements from observations. *Adv Geophys.* 1983;25:355–490.

74. Ooyama K. Numerical simulation of the life cycle of tropical cyclones. *J Atmos Sci.* 1969;26:3–40.

75. Palmén E, Newton CW. Atmospheric Circulation Systems London: Academic Press; 1969.

76. Palmer TN. Extended-range atmospheric prediction and the Lorenz model. *Bull Am Meteor Soc.* 1993;74:49–65.

77. Panofsky HA, Dutton JA. Atmospheric Turbulence New York: Wiley; 1984.

78. Pedlosky J. Geophysical Fluid Dynamics second ed. New York: Springer-Verlag; 1987.

79. Philander SG. El Niño, La Niña, and the Southern Oscillation New York: Academic Press; 1990.

80. Phillips NA. The general circulation of the atmosphere: a numerical experiment. *Quart J Roy Meteorol Soc.* 1956;82:123–164.

81. Phillips NA. Geostrophic motion. *Rev Geophys.* 1963;1:123–176.

82. Pierrehumbert RT, Swanson KL. Baroclinic instability. *Annu Rev Fluid Mech.* 1995;27:419–467.

83. Plumb RA. The circulation of the middle atmosphere. *Aust Meteorol Mag.* 1982;30:107–121.

84. Randall DA. General Circulation Model Development San Diego: Academic Press; 2000.

85. Reed RJ, Norquist DC, Recker EE. The structure and properties of African wave disturbances as observed during Phase III

of GATE. *Mon Wea Rev.* 1977;105:317–333.

86. Richardson, L.F., 1922. Weather Prediction by Numerical Process. Cambridge University Press (reprinted by Dover, 1965).
87. Richtmyer RD, Morton KW. Difference Methods for Initial Value Problems second ed. New York: Wiley (Interscience); 1967.
88. Riehl H, Malkus JS. On the heat balance of the equatorial trough zone. *Geophysica*. 1958;6:503–538.
89. Roe GH, Baker MB. Why is climate sensitivity so unpredictable?. *Science*. 2007;318:629–632.
90. Salby ML. Fundamentals of Atmospheric Physics San Diego: Academic Press; 1996.
91. Salby ML, Hendon HH, Woodberry K, Tanaka K. Analysis of global cloud imagery from multiple satellites. *Bull Am Meteorol Soc.* 1991;72:467–480.
92. Sanders F, Hoskins BJ. An easy method for estimation of Q-vectors from weather maps. *Wea Forecasting.* 1990;5:346–353.
93. Sawyer JS. The vertical circulation at meteorological fronts and its relation to frontogenesis. *Proc Roy Soc A*. 1956;234:346–362.
94. Schneider T. The general circulation of the atmosphere. *Annu Rev Earth Planet Sci.* 2006;34:655–688.
95. Schubert, S., Park, C.-K., Higgins, W., Moorthi, S., Suarez, M., 1990. An atlas of ECMWF analyses (1980–1987) Part I — First moment quantities. NASA Technical Memorandum 100747.
96. Scorer RS. Natural Aerodynamics New York: Pergamon Press; 1958.
97. Shine KP. The middle atmosphere in the absence of dynamical heat fluxes. *Quart J Roy Meteor Soc.* 1987;113:603–633.
98. Simmons AJ, Bengtsson L. Atmospheric general circulation models: their design and use for climate studies. In: Houghton JT, ed. *The Global Climate*. Cambridge, UK: Cambridge University Press; 1984;37–62.
99. Simmons AJ, Burridge DM, Jarraud M, Girard C, Wergen W. The ECMWF medium-range prediction models: Development of the numerical formulations and the impact of increased resolution. *Meteorol Atmos Phys.* 1989;40:28–60.
100. Simmons AJ, Hollingsworth A. Some aspects of the improvement in skill of numerical weather prediction. *Quart J Roy Meteorol Soc.* 2002;128:647–678.
101. Sinclair PC. On the rotation of dust devils. *Bull Am Meteorol Soc.* 1965;46:388–391.
102. Smagorinsky J. The role of numerical modeling. *Bull Am Meteorol Soc.* 1967;46:89–93.
103. Smith RB. The influence of mountains on the atmosphere. *Adv Geophys.* 1979;21:87–230.
104. Stevens DE, Lindzen RS. Tropical wave–CISK with a moisture budget and cumulus friction. *J Atmos Sci.* 1978;35:940–961.
105. Stull RB. An Introduction to Boundary Layer Meteorology Boston: Kluwer Academic Publishers; 1988.
106. Thompson DWJ, Wallace JM, Hegerl GC. Annular modes in the extratropical circulation Part II: Trends. *J Climate.* 2000;13:1018–1036.
107. Thorpe AJ, Bishop CH. Potential vorticity and the electrostatics analogy: Ertel-Rossby formulation. *Quarti J Ref Meteorol Soc.* 1995;121:1477–1495.
108. Trenberth KE. General characteristics of El Niño–Southern Oscillation. In: Glantz M, Katz R, Nichols N, eds. *ENSO Teleconnections Linking Worldwide Climate Anomalies: Scientific Basis and Societal Impact.* Cambridge, UK: Cambridge University Press; 1991;13–42.
109. Turner JS. Buoyancy Effects in Fluids Cambridge: Cambridge University Press; 1973.
110. U.S. Government Printing Office. U S Standard Atmosphere, 1976 Washington, DC: U. S. Government Printing Office; 1976.
111. Vallis GK. Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation Cambridge, UK: Cambridge University Press; 2006.
112. Wallace JM. Spectral studies of tropospheric wave disturbances in the tropical Western Pacific. *Rev Geophys.* 1971;9:557–612.
113. Wallace JM. General circulation: overview. In: Holton JR, Curry JA, Pyle JA, eds. *Encyclopedia of Atmospheric Sciences*. London: Academic Press; 2003;821–829.

114. Wallace JM, Hobbs PV. Atmospheric Science: An Introductory Survey second ed. New York: Academic Press; 2006.

115. Wallace JM, Kousky VE. Observational evidence of Kelvin waves in the tropical stratosphere. *J Atmos Sci.* 1968;25:900–907.

116. Warsh KL, Echternacht KL, Garstang M. Structure of near-surface currents east of Barbados. *J Phys Oceanog.* 1971;1:123–129.

117. Washington WM, Parkinson CL. An Introduction to Three-Dimensional Climate Modeling Mill Valley, CA: University Science Books; 1986.

118. Webster PJ. The large-scale structure of the tropical atmosphere. In: Hoskins BJ, Pearc R, eds. *Large-Scale Dynamical Processes in the Atmosphere.* New York: Academic Press; 1983;235–275.

119. Webster PJ, Chang HR. Equatorial energy accumulation and emanation regions: impacts of a zonally varying basic state. *J Atmos Sci.* 1988;45:803–829.

120. Webster PJ, Fasullo J. Monsoon: dynamical theory. In: Holton JR, Curry JA, Pyle JA, eds. *Encyclopedia of Atmospheric Sciences*. London: Academic Press; 2003;1370–1386.

121. Williams J, Elder SA. Fluid Physics for Oceanographers and Physicists New York: Pergamon Press; 1989.

122. Williams, K.T., 1971: A statistical analysis of satellite-observed trade wind cloud clusters in the western North Pacific. Atmospheric Science Paper No. 161, Dept. of Atmospheric Science, Colorado State University, Fort Collins, CO.

123. Yanai M, Maruyama T. Stratospheric wave disturbances propagating over the equatorial Pacific. *J Meteor Soc Jpn.* 1966;44:291–294.

Index

Note: Page numbers followed by "*f*" indicate figures, "*t*" indicate tables and "*n*" indicate footnotes.

A

Absolute angular momentum, 342, 343

Absolute coordinate system, 96

Absolute momentum, 151, 292

Absolute vorticity, 100, 101, 109, 112, 116, 119*f*, 213
  advection of, 465
  conservation of, 119, 396
  of flow, 151
  geostrophic, 185

Acceleration
  centripetal, 10–11, 10*f*
  eddy-driven mean flow, 340
  gradient tangential, 74
  inertial, 127, 261
  net mean flow, 439*f*, 444
  wave-driven mean-flow, 433

Acoustic waves, 131, 136–139

Adiabatic cooling, 56, 331, 337, 365, 398–400

Adiabatic lapse rate, dry, 54

Adiabatic method, 53, 85, 87, 228

Adiabatic speed of sound, 139

Advection, temperature, 32, 196, 284
  vorticity, 183

Aerodynamic formula, 316

African wave disturbances, 384–386

AGCMs, *See* Atmospheric general circulation models

Ageostrophic circulation, 201–203, 203*f*, 289*f*
  cross-frontal, 287–290

Ageostrophic wind, 43, 85, 181–185

Air parcel and particle, 4
  Lagrangian motion of, 418–420

Air-sea interaction, 319

Along-front velocity in geostrophic balance, 286, 287

Amplification, 246

Angular momentum
  absolute, 342, 343
  circulation, 341–349
  oscillations, 17–18
  zonal mean, 345–349

Annular modes, 364

Anomaly correlation, 485, 485*f*

Antibaric flow, 77

Anticyclonic flow, 73, 74

APE, *See* Available potential energy

Arbitrary vector, 33

Asian monsoon, 386–387, 388*f*

Asymmetric atmospheric circulations, 389

Atmosphere, *See also* Middle atmosphere
  barotropic and baroclinic, 84
  data, standard, 507–508
  structure of static, 18–23
  turbulence, 256–258

Atmosphere–ocean coupling, 391

Atmospheric equatorial wave propagation, 401

Atmospheric general circulation models (AGCMs)
  development of, 367
  formulation of, 368
  physical processes and parameterizations, 369–370

Atmospheric motions, principles of, 31

Atmospheric tracers, motion of, 446

Available potential energy (APE), 227–229

B

Backing wind, 83, 83*f*

Balanced flow, 69–77

Balance equation, nonlinear, 397

Baric flow, 77

Baroclinic atmospheres, 84

Baroclinic disturbance, model of, 100*f*, 177, 204–206

Baroclinic instability, 177, 213–251
  continuous stratified atmosphere and, 234–245
  neutral modes, growth and propagation of, 245–251

  normal mode, 215–226
  Rayleigh theorem, 237–241
  symmetric, 290–294

Baroclinic (Ertel) potential vorticity equation, 120–122

Baroclinic Rossby wave, 180*f*, 219

Baroclinic waves, 177, 328, 355, 361
  energetics of, 227–234
  energy equations for two-layer model, 229–234
  potential energy, 227–229
  vertical motion in, 223–226, 225*f*

Barometer, 33

Barotropic atmospheres, 84

Barotropic fluids, vorticity and, 118–120

Barotropic instability, 214, 240

Barotropic potential vorticity, 118–120

Barotropic Rossby waves, 161–163

Barotropic vorticity equation, 119
  in finite difference, 455, 464–467
  in spherical coordinates, 468–470

Bayes’s theorem, 511

Bjerknes circulation theorem, 98

Blocking patterns, 361–362

Body forces, 4

Boundary condition, 89, 403
  Dirichlet, 190
  Neumann, 190
  no-slip, 255
  for pressure tendency, 195
  Robin, 190
  in two-layer model, 237

Boundary value problem, 367

Boussinesq approximation, 57–58, 146, 178, 181, 184, 255, 285, 288

β-plane, 160, 179, 206, 401, 464

Brunt–Väisälä frequency, 56*n*

Bulk aerodynamic formula, 262

Buoyancy frequency, 56

Buoyancy oscillations, 55, 144, 294

Buoyancy waves, *See* Gravity waves

Burger number, 187, 188

C

CAPE, *See* Convective available potential energy

Cartesian coordinate
  directions, 7
  form and vorticity equation, 104–106
  frame, 46
  system, 37, 200

Cartesian geometry, 401

CAT, *See* Clear air turbulence

Centered differences, 456, 457–458

Centrifugal force, 10–11, 14, 15, 318
  Coriolis force and, 72–73, 313

Centripetal acceleration, 10–11, 10*f*

CFL stability criterion, *See* Courant- Friedrichs-Levy stability criterion

Chaos, 364

Chaotic system, 482

Charney–DeVore model, 363*f*, 364

Charney, J. G., 454

Chemical timescale, 447–448

Chemical tracers, 447–448

Circulation, *See also* Mesoscale circulation; Tropical circulation
  ageostrophic, 201–203, 289*f*
  cross-frontal, 287–290
  diabatic, 325
  extratropical, structure of, 171–178
  Hadley, 326, 332
  secondary, 270–275, 284, 313
  theorem, 95–99
  Walker, 389–390
  zonal-mean circulation of middle atmosphere, 417–426

Circulation, general, 325
  angular momentum, 341–349
  climate regimes, 362–364
  conventional Eulerian mean, 330–337
  dynamical formulation, 368
  jet stream and storm tracks, 359–361
  Lorenz energy cycle, 349–355
  low-frequency variability, 325, 361–366
  monsoonal, 325

  nature of, 326–328
  numerical simulation of, 367–370
  quasi-stationary, 325
  stationary waves, 356–359
  transformed Eulerian mean, 337–339
  zonally averaged, 328–341

Clausius–Clapeyron equation, 59, 503

Clear air turbulence (CAT), 145

Climate feedbacks, 370–373

Climate models, 368, 371, 372

Climate variability, 1

Component equations in spherical coordinates, 37–41

Computational instability, 459, 460

Computational mode, 463

Computational stability, 458–460

Condensation heating, 398–400
  vorticity equation in, 396

Conditional instability, 62–64, 319, 321

Conditional instability of the second kind (CISK), 319

Conditional probability, 511–512

Confluent flow, 283

Conservation laws
  Boussinesq approximation, 57–58
  component equations in spherical coordinates, 37–41
  continuity equation, 45–49
  control volume, 31
  momentum equation in rotating coordinates, 35–37
  scale analysis of momentum equations, 41–42
  thermodynamic energy equation, 50–53
  thermodynamics of dry atmosphere, 53–57
  thermodynamics of moist atmosphere, 58–64
  total differentiation, 31–35

Conservation of energy, 50–53

Conservation of mass, 45–49

Continuity equation, 407
  Eulerian derivation, 46–47
  in isobaric coordinate system, 68–69, 207
  Lagrangian derivation, 47–48
  scale analysis of, 48–49

Continuum, approximation, 3–4

Control volume, 31

Convection, cumulus, 302–305, 319

Convective adjustment, 475

Convective available potential energy (CAPE), 302–303

Convective heating, 379

Convective storms, 306–312

Conventional Eulerian mean, 330–337, 418–420

Coordinates
  Cartesian, 500
  cylindrical, 500
  generalized, 22–23
  geostrophic, 287*n*
  inertial, 13
  isentropic, 120–121
  isobaric, 21, 206
  log-pressure, 235–237
  natural, 70–71
  sigma, 343–345
  spherical, 37–41, 501

Coriolis acceleration, 36

Coriolis force, 5, 9, 14–17, 16*f*, 43–44, 284, 318, 326, 331, 333, 335, 338
  causes of, 336–337
  and centrifugal force, 72–73, 313
  role in supercell dynamics, 306

Coriolis parameter, 16, 42, 105, 118, 160, 179, 206, 401, 406
  β-plane approximation, 160
  f-plane approximation, 179
  variation of, 422

Cost function, 477, 481*f*

Coupled climate models, 367

Coupled model intercomparison project (CMIP), 367

Courant–Friedrichs–Levy (CFL) stability criterion, 459–460

Courant number, 458

Covariance, 257, 479

Critical level, 222, 299

Critical surface, 429

Cross-frontal circulation, 287–290

Cumulus cells, 383

Cumulus convection, 302–305, 319

Curvature
  effect, 14–17
  radius of, 70, 72–73
  of streamlines and trajectories, 78, 80
  terms, 40
  vorticity, 103, 104

Cyclogenesis, 205–206, 213, 214, 215*f*

Cyclones
  extratropical, 177, 180*f*, 250
  surface, development of, 177, 180*f*, 243–244
  tropical, 312–321

Cyclonic flow, 73, 74

Cyclostrophic flow, 73–74
  force balance in, 75*f*

Cyclostrophic wind, 73–74

D

Dalton's Law, 59

Darwin–Tahiti pressure difference, 390, 391

Data assimilation, 478–481
  for many variables, 479–481
  for single variable, 476–478

Deflecting forces, 15

Deformation, radius of, 157, 243

Delayed oscillator model, 391

Derivative, total and local, 31–33

Diabatic circulation, 339

Diabatic cooling, 420, 450

Diabatic heating, 56, 57, 401, 417, 420, 450
  in equatorial regions, 389
  monsoons, 388
  in tropics, 377

Diagnostic relationship, 42–43

Dimensional homogeneity, 2

Dirichlet condition, 190

Discretization error, 463

Dispersion, 131–132

Disturbance vorticity, 224

Divergence, 8, 24*f*, 25

  equation, quasi-geostrophic, 186
  theorem, 499

Doppler shifting, 139, 429, 444

Downgradient diffusion, 8

Downslope windstorms, 299–302

Downward control, 422

Drag coefficient, 262

Dry air, 321

Dry static energy, 61

Dynamical instability, role of, 213

Dynamical timescale, 447–448

Dynamical tracers, 446–447

Dynamic meteorology, 1–4

E

Eady stability model, 235, 242, 243

Eady stability problem, 241–245

Eady wave, properties of, 244*f*

Earth system models (ESMs), 367

Eastward-propagating waves, 445

ECMWF, *See* European Centre for Medium-Range Weather Forecasts

Eddies, 328, 329
  diabatic process, 354
  geostrophic streamfunction, 363
  stress, 343
  viscosity, 264
  vorticity, 114

Eddy flux, 333*f*
  heat, 330, 335–336
  momentum, 330, 334*f*, 335–337, 335*f*, 336*f*
  potential vorticity, 329

Eddy kinetic energy, 352, 353, 388

Eddy motions, 347, 417

Eddy potential energy, 387–388

Eddy stress, 262*n*

Eddy transport, 425

Eddy viscosity coefficient, 264

e-folding time scale, 272

Ekman layer, 266–268, 267*n*, 390
  modified, 269–270

Ekman, V.W., 267

Eliassen-Palm (EP) flux, 338, 338*f*, 339
  convergence, 416, 424
  in wave activity, 433

Elliptical polarization, 150

Elliptic boundary value problem, 289, 294

El Niño-Southern Oscillation (ENSO), 390–392

Energetics of baroclinic waves, 227–234

Energy
  CAPE, 302–303
  conservation of, 50–53
  equations for two-layer model, 229–234
  equation, thermodynamic, 215, 216, 241
  flow in amplifying baroclinic wave, 235*f*
  Lorenz energy cycle, 349–355
  potential, 227–229
  transfer, 280
  turbulent kinetic, 259–261

EnKF, *See* Ensemble Kalman filter

Ensemble forecasting, 481–486

Ensemble Kalman filter (EnKF), 480

Ensemble spread, 486

Enstrophy, 466, 467

Entrainment, 303–305

Entropy form of first law of thermodynamics, 53, 54

EP flux, *See* Eliassen-Palm flux

Equations of motion, scale analysis of, 41–45

Equatorial β-plane, 401

Equatorial intraseasonal oscillation, 392

Equatorial Kelvin waves, 404–406, 405*f*, 406*f*

Equatorial motions, steady forced, 406–409

Equatorial stratosphere, waves in, 435–440

Equatorial wave disturbances, 381–384
  observed, 438–440
  schematic model for, 381, 382*f*
  theory, 401–406

Equivalent barotropic model, 454

Equivalent potential temperature, 59–61, 294, 304, 503–504

Equivalent static stability, 398

Ertel potential vorticity, 120–122, 175, 289
  conservation law, 184
  theorem, 111, 112, 114

Eulerian control volume, 31
  continuity equation, 45, 46*f*

Eulerian derivation, 46–47

Eulerian mean
  conventional, 330–337, 418–420
  in log-pressure coordinates, 349
  meridional streamfunction, 331*f*, 333*f*
  model, 355
  by poleward heat fluxes, 334
  transformed, 337–339, 420–424

European Centre for Medium-Range Weather Forecasts (ECMWF), 480

Exner function, 121

Explicit time differencing, 457–458

Extratropical circulation, structure of, 171–178

Extratropical cyclones, 250

Eyewall, 313

F

Ferrel cell, 335, 336

Field variables, 3–4

Finite difference approximations, horizontal divergence using, 86

Finite difference method, 455–456
  barotropic vorticity equation, 455, 464–467, 465*f*
  centered, 456, 457–458
  computational stability, 458–460
  explicit time differencing, 457–458
  implicit time differencing, 460–461

Finite Fourier series expansion, 455

First law of thermodynamics, *See* Thermodynamic energy equation

Flow
  balanced, 69–77
  baric, 77
  cyclostrophic, 73–74
  geostrophic, 71–72

gradient, 74–77
inertial, 72–73
time-averaged, 356–361

Fluid instabilities, 213

Flux, *See also* Eddy flux
gradient theory, 264
Richardson number, 260–261
turbulent momentum, 261, 262*n*

Forces
balance, 75, 76*f*
body, 4
centrifugal, 11
Coriolis, *See* Coriolis force
deflecting, 15
gravity, 8–9, 11–13, 12*f*
hydrostatic equilibrium, 19*f*
pressure gradient, 5–7, 7*f*
surface, 4
viscous, 6–8

Forecasting, *See* Numerical modeling and prediction

Form drag, 363

Four-dimensional data assimilation (4DVAR), 480

Fourier
coefficients, 131
component, 131
modes, 297
series, 130–131

f-plane approximation, 179

Free atmosphere, 255

Free transverse oscillations, 294

Frequency, intrinsic, 148

Friction, planetary boundary layer, 263

Friction velocity, 268

Frontal circulation, 280

Frontogenesis, 280–290

Fronts, 172, 280–290
gust, 306, 310

Froude number, 300, 301

G

Gas constant, 18

Gaussian functions, 406

Geocentric reference frame, 9

Geopotential
  constant, 11–12
  height, 19, 507*t*, 508*t*
  perturbation, 220–221
  tendency equation, 200

Geostrophic approximation, 42–44

Geostrophic balance, 42, 43
  adjustment to, 156–159
  along-front velocity in, 286, 287

Geostrophic coordinates, 287*n*

Geostrophic equilibrium, balance of forces for, 72, 72*f*

Geostrophic flow, 71–72

Geostrophic momentum approximation, 287

Geostrophic motion, 71, 284

Geostrophic streamfunction, 215

Geostrophic vorticity, 215, 220

Geostrophic wind, 42–43, 181, 206, 215, 286
  vertical shear of, 81, 81*f*

Gradient flow, 74–77

Gradient tangential acceleration, 74

Gradient wind approximation, 74–77

Gravitational constant, 8

Gravitational force, 8–9

Gravitational potential energy, 227

Gravity, 11–13, 12*f*

Gravity waves
  inertia, 152–156
  internal, 145–150
  modified by rotation, 150–156
  shallow water, 139–144
  topographic, 163–165

Grid point models, 473

Group velocity, 132–133, 132*f*

Gust front, 306, 310

H

Hadley circulation (cell), 326, 332, 333*f*, 343, 378, 381

Hadley, George, 326

Heat
  latent, 313, 319
  specific, 52, 53

Heating, condensation, 398–400

Height tendency equation, 194–197, 207

Helmholtz theorem, 396

Hermite polynomial, 403

High-index state, 362

Horizontal divergence, using finite difference approximations, 86

Horizontal momentum equation, 67–68, 236, 286
  approximate, 43
  in natural coordinate system, 71
  scale analysis of, 42, 42*t*

Horizontal shearing deformation, 281, 282*f*

Horizontal stretching deformation, 281, 282*f*

Horizontal velocity
  field, 43, 44, 48
  vector, 67

Hot towers, 378–379, 383–384

Hurricanes, 312
  development, 318–321
  dynamics of, 314–318

Hydrodynamic instability, 213–215

Hydrostatic approximation, 44–45

Hydrostatic balance, 18, 44, 55, 182, 206, 313

Hydrostatic equation, 18–20, 21, 115, 120, 236

Hydrostatic equilibrium, 19*f*, 44–45

Hydrostatic thermodynamic energy equation, 215

I

Ideal gas law, 54, 55, 58–59, 82

Implicit time differencing, 460–461

Incompressible fluid, 49

Induced flow, 188–193

Inertia circle, 73

Inertia gravity waves, 141, 152–156, 152*f*, 171

Inertial flow, 72–73

Inertial instability, 151, 213

Inertial motion, 9, 73

Inertial oscillation, 73, 150–151

Inertial reference frame, 33, 34
  law of motion (second) in, 35–37

Inflow of mass, 46, 47

Initial value approach, 214

Initial value problems, 367, 455

Instability
  baroclinic, 177, 213–251, 290–294
  barotropic, 214, 240
  computational, 459, 460
  conditional, 62–64, 321
  conditional of the second kind (CISK), 319
  dynamical, 213
  fluid, 213
  hydrodynamic, 213–215
  inertial, 151, 213
  Kelvin–Helmholtz, 280
  parcel, 213
  symmetric, 213–214
  wave, 214

Intergovernmental Panel on Climate Change (IPCC), 367

Internal gravity waves, 144–150, 425

Intertropical convergence zone (ITCZ), 378–381

Intrinsic frequency, 148

Inversion, potential vorticity, 188–189

Isentropic coordinates
  momentum equation in, 120
  system, 65

Isentropic vorticity, 121–122

Isobaric coordinates, 67–69
  QG equations in, 206

Isobaric coordinate system, 21, 106–107
  continuity equation in, 207
  equations of motion in, 120–121
  hydrostatic equation, 18–20, 21
  potential vorticity equation, 121

scale analysis and, 179
vorticity and, 106–107

Isobaric coordinate system, basic equations
continuity equation, 68–69
geostrophic wind using, 67–68
horizontal momentum equation, 67–68
thermodynamic energy equation, 69

Isolated ridges, flow over, 297–298

J

Jet streak, 192, 193*f*, 202

Jet stream, 172, 175, 359–361

K

Kalman gain matrix, 479

Kelvin–Helmholtz instability, 280

Kelvin's circulation theorem, 97, 110

Kelvin's theorem, 110

Kelvin waves, 404–406, 405*f*, 406*f*
in observational data, 438–440
vertically propagating, 436–437

Kinematics, 23–25
of frontogenesis, 281–285
method, 85–86

Kinematic viscosity coefficient, 7

Kinetic energy, 159, 228–229, 291, 303, 327, 355
See also Potential energy
eddy, 352, 353, 388
perturbation, 231, 233
power spectrum of, 74*f*
in thermodynamic energy, 50
turbulent, 259–261

Kronecker delta, 198

K theory, 264

L

Lagrangian control volume, 31, 50
as result of fluid motion, 48, 48*f*
in thermodynamic system, 50

Lagrangian derivation, 47–48

Lagrangian motion of air parcels, 418–420

La Niña, 390

Lapse rate of temperature, 54
  pseudoadiabatic, 504–505

Latitude dependence of unit vector, 38, 39*f*

Law of gravitation, 8

Law of motion
  Newton's first, 14
  Newton's second, 14, 33, 35–37

Lax equivalence theorem, 463

Leapfrog differencing method, 458

Lee wave, 298–299

Level of free convection (LFC), 60

Lifting condensation level (LCL), 60

Likelihood, 511–512

Linear interpolation, 224

Linearized equations, 136

Linear perturbation analysis, 217–223

Linear stability theory, 318

Liouville equation, 482, 483*f*

Local derivative, 31–33

Logarithmic wind profile, 269

Log-pressure coordinates, 235–237, 328–329

Longitudinal dependence of unit vector, 35*f*, 38, 39*f*

Longitudinally dependent time-averaged flow, 356–361

Longitudinal waves, 136–139

Long-lived tracers, 448, 449

Long-wave stabilization, 221

Lorenz energy cycle, 349–355

Lower-level cyclonic vorticity, 215*f*

Low-frequency variability, 325, 361–366

M

Madden–Julian oscillation (MJO), 392, 393*f*

Marginal stability, 222

Mass, conservation of, 45–49, 115

Mass flux, 49

Material derivative, 31

Maxwell relations, 314

Mean energy cycle, 354*f*

Mean zonal wind, 328

Measurements, SI (Systéme International) units of, 2–3

Mechanical energy, 52

Mechanical energy equation, 52

Meridional circulation
  Eulerian mean, 330–337
  residual, 337
  transformed Eulerian mean, 337–339

Meridional force balance for Kelvin mode, 406

Meridional wind, statistical analysis of, 177, 178*f*

Mesopause, 413

Mesoscale circulation
  convective storms, 306–312
  cross-frontal circulation, 287–290
  cumulus convection, 302–305, 319
  energy sources for, 279–280
  fronts and frontogenesis, 280–290
  hurricanes, 312–321
  mountain waves, 294–302
  symmetric baroclinic instability, 290–294

Mesosphere, 413. *See also* Middle atmosphere

Meteorology
  applications of, 33
  instabilities in, 214

Metric terms, 15

Middle atmosphere, 413
  conventional Eulerian mean, 418–420
  layers of, 413
  quasi-biennial oscillation, 440–445
  structure and circulation of, 413–417
  trace constituent transport, 446–450
  transformed Eulerian mean, 420–424
  vertically propagating planetary waves, 426–430
  warmings, sudden stratospheric, 430–435
  waves in equatorial stratosphere, 435–440
  zonal-mean circulation of, 417–426
  zonal-mean transport, 424–426

Midlatitude β-plane approximation, 160, 179

Mixing length hypothesis, 264–266

Mixing ratio, 59, 61, 503

MJO, *See* Madden–Julian oscillation

Moist static energy, 61

Moisture convergence, low-level, 401

Moisture variables, 503–505

Molecular diffusion, 255

Momentum equations, 407
  horizontal, 67–68
  isentropic coordinates and, 120
  planetary boundary layer, 261–270
  quasi-geostrophic, 186, 207
  in rotating coordinates, 35–37
  scale analysis of, 41–42
  in spherical coordinates, 37–41
  transformation of, 33

Monsoons, 325, 386–389

Montgomery streamfunction, 120

Motion
  and the Coriolis force, 16*f*
  equations of, 455–464
  geostrophic, 71
  inertial, 9
  Newton's first law of, 9
  scales of, 26*t*
  tropical, 392–398
  vertical, 84–87

Mountain lee waves, 145, 298–299

Mountain waves, 294–302

Multicell storms, 306

N

Natural coordinate system, 70–71
  vorticity in, 102–104

Neumann boundary condition, 190

Neutral curve, 221, 222*f*

Neutral modes
  downstream development, 250–251
  growth and propagation of, 245–250

Neutral stability, 242

  curve, 222*f*

Neutral waves, transient growth of, 247–250

Newtonian relaxation rate, 421

Newton’s laws
  of gravitation, 8
  of motion (first), 4
  of motion (second), 4, 9, 10, 14, 33, 35–37

Nonacceleration theorem, 433

Nondimensional geostrophic wind, 187*n*

Nonlinear balance equation, 397

Nonlinearity, 455

Nonlinear partial differential equations, 41

Normal mode method, 214
  baroclinic instability, 215–226

Northern Hemisphere, 430

Northern Hemisphere troposphere, 172

No-slip boundary condition, 255

Numerical dispersion, 464

Numerical modeling and prediction
  approximation of motion equations, 455–464
  barotropic vorticity equation in finite difference, 455, 464–467
  barotropic vorticity equation in spherical coordinates, 468–470
  data assimilation, 475–481
  finite differences, 455–456
  historical background, 453–455
  primitive equation models, 472–475
  purpose of, 453
  spectral method, 467–472

Numerical simulation of general circulation, 367–370

O

Observation operator, 479

Observed equatorial waves, 438–440

Oceanic equatorial wave propagation, 401

Omega equation, 87, 171, 198–200
  two-level model, 223

Operational forecast centers, 453

Orography, 172

Oscillations

  angular momentum, 17–18
  buoyancy, 55
  equatorial intraseasonal, 392
  quasi-biennial, 440–445
  southern, 390–392
  transverse, 137

P

Parameterization, 399, 475

Parcel dynamics in moist atmosphere, 60

Parcel instability, 213

Parcel method, 55, 59–60

Parcel motions
  for adiabatic planetary, 419, 419*f*
  on meridional plane, 420, 420*f*

Pendulum day, 73

Perturbation method
  geostrophic balance, adjustment to, 156–159
  linear, 217–223
  purpose of, 127–128
  waves, gravity, 145–150
  waves, properties of, 128–136
  waves, types of, 136–144

Perturbed observations, 481

Phase speed, 129, 132*f*, 219

Piecewise potential vorticity inversion, 188–193

Planetary boundary layer, 255
  atmospheric turbulence, 256–258
  depth of, 255–256
  Ekman layer, 266–268, 269–270
  flux gradient theory, 264
  friction, 263
  mixing length hypothesis, 264–266
  momentum equations, 261–270
  pumping, 271–272
  secondary circulation and spin down, 270–275
  surface layer, 268–269
  turbulent kinetic energy, 259–261
  well-mixed, 262–263

Planetary vorticity, 101, 109, 207
  entropy and momentum, 114*f*

Planetary wave, *See* Rossby waves

Planetary wave number, 130

Poisson's equation, 54, 69, 345

Polar-frontal zone, 175

Polar stratosphere, warming of, 430, 431–432*f*

Positive cyclonic vorticity, 117

Positive eddy momentum flux, 348*f*

Potential energy, 158–159
  available, 227–229, 231, 233, 234, 291, 302–303
  eddy, 387–388
  gravitational, 227
  for hurricanes, 319
  of mean flow, 227
  perturbation, 232
  total, 227

Potential intensity, 317

Potential temperature, 53–54, 446
  advection of, 215*f*
  definition of, 110
  equivalent, 294, 304, 503–504

Potential vorticity (PV), 110–115, 143, 160*f*, 241
  anomaly, 214, 215*f*
  barotropic, 118–120
  conservation, 194–197
  equation, 121
  Ertel, 111, 175, 289
  evolution of, 242
  gradient, 235, 240
  inversion, 188–189
  piecewise inversion, 192
  quasi-geostrophic, 171, 183–186
  thinking, 187–197
  as a tracer, 447
  zonal mean, 340–341

Prandtl, L., 264

Predictability, 481–486

Predicting, *See* Numerical modeling and prediction

Pressure, as a vertical coordinates, 20–21

Pressure change, measuring, 33

Pressure gradient, 40

Pressure gradient force, 5–7, 5*f*, 22*f*

Pressure tendency
  equation, quasi-geostrophic, 195
  surface, 87–89

Primary circulation, 313

Primitive equation models, 472–475
  grid point models, 473
  physical parameterizations, 474–475
  spectral models, 473–474

Prognostic equations, approximate, 43–44

Propagating waves, 445

Pseudoadiabatic ascent, 60, 61, 378, 504

Pseudoadiabatic lapse rate, 61–62, 504–505

PV, *See* Potential vorticity

Q

Quadratic dispersion equation, 219

Quasi-biennial oscillation (QBO), 417, 440–445

Quasi-geostrophic approximation, 185

Quasi-geostrophic divergence equation, 186

Quasi-geostrophic energy cycle, 354

Quasi-geostrophic (QG) equations, 171
  derivation of, 178–183, 186
  isobaric form, 206–208
  potential vorticity derivation of, 183–186
  PV conservation, 194–197

Quasi-geostrophic height tendency equation, 194–197, 207

Quasi-geostrophic mass continuity equation, 185

Quasi-geostrophic model, 327, 454

Quasi-geostrophic momentum, 359
  equation, 186

Quasi-geostrophic perturbations, 233, 234

Quasi-geostrophic potential vorticity (QG PV), 171, 183–186
  inversion, 189
  scaling properties of, 187

Quasi-geostrophic pressure tendency equation, 195

Quasi-geostrophic scaling, 330

Quasi-geostrophic thermodynamic energy equation, 182, 185, 207

Quasi-geostrophic vertical motion equation, 198

Quasi-geostrophic vorticity equation, 183–186, 207, 213, 215

Quasi-Laplacian operator, 195, 198

Quasi-stationary circulation, 325

Quasi-stationary planetary waves in troposphere, 432

Q vector, 200–201, 202*f*, 207–208, 223–225
  in frontogenetic confluence, 283*f*

R

Radiative cooling, 417, 418

Radiative heating, 56, 414, 417, 418
  in tropical stratosphere, 424

Radius of curvature, 70, 72–73

Radius of deformation, 157, 243

Rayleigh theorem, 235, 237–241

Reference frames
  geocentric, 9–10
  inertial (absolute), 9
  noninertial, 9–18

Relative vorticity, 23, 100, 101, 109, 185, 187

Residual mean meridional streamfunction, 339

Residual meridional circulation, 337, 338

Residual vertical velocity, 337

Reversible process in thermodynamic, 53, 54

Reynolds averaging, 256–258

Rhomboidal truncation, 474

Richardson, L. F., 453, 454, 455, 472

Richardson number
  flux, 260–261
  mean-flow, 293

Right-moving storm, 310–312

Rossby adjustment problem, 157

Rossby critical velocity, 428

Rossby–gravity waves, 401–404, 442, 444
  in observational data, 438–440
  vertically propagating, 437–438

Rossby–Haurwitz waves, 470

Rossby number, 43–44, 74, 77, 112, 181, 188, 286, 395

Rossby radius of deformation, 157, 243

Rossby waves, 141, 159–161, 171, 220*n*
  baroclinic, 219
  barotropic, 161–163
  equatorial, 401–404, 404*f*
  linear, 426–428
  propagation, 221
  stationary, 356–359, 358*f*
  theory, 366, 366*f*
  topographic, 163–165
  vertical propagation, 426–430
  wavebreaking, 428–430
  zonal-mean flow, 362

Rotating coordinates
  momentum equation in, 35–37
  total differentiation of vector in, 33–35

Rotation
  circulation theorem, 95–99
  macroscopic measure of, 95
  in supercell thunderstorms, 306–310
  vorticity, 99–112

Roughness length, 269

S

Satellite cloud photos, 379, 380*f*
  longitude sections of, 381, 382*f*

Saturation mixing ratio, 61, 503

Saturation vapor pressure, 59

Sawyer–Eliassen equation, 288, 294

Scale analysis, 25–26
  of continuity equation, 48–49
  of momentum equations, 41–42
  isobaric coordinates and, 179
  of thermodynamic energy equation, 56–57
  of tropical motion, 392–398
  of vorticity equation, 107–110

Scorer parameter, 298, 299*f*

Sea breeze, 99

Sea surface temperature (SST), 361, 364–366, 377, 390–391

Secondary circulation, 270–275, 313
  associated with frontogenesis, 284

Semigeostrophic equations, 287

Semigeostrophic theory, 285–287

Semi-Lagrangian integration method, 462–463

Shallow water
  equations, 115–118
  gravity waves, 139–144

Shallow water theory, 435

Shearing stress, 7

Shear, vertical wind, 82

Shear vorticity, 103, 106

Sigma coordinates, 342, 343–345, 347

Single cell storms, 306

Sinking motion, 224

Sinusoidal topography, 294–297

SI (Systéme International) units of measurement, 2–3

Slantwise convection, 290

Solar heating in ozone layer, 414

Solenoidal term
  in circulation theorem, 97, 106, 107
  synoptic-scale motions, 107
  in vorticity equation, 106

Solid curves, 484

Solitary waves, 362

Sound waves, 136–139
  in equations of motion, 41

Southern oscillation, 390–392

Specific heat, 52, 53
  of water vapor, 504

Spectral method, 467–472

Spectral models, 473–474

Spectral transform method, 471–472, 472*f*

Spherical coordinates
  barotropic equation in, 468–470
  momentum equations in, 37–41

Spherical harmonics functions, 468, 469, 469*f*

Spin-down, Ekman, 272–275

Squall lines, 306

Square root filters, 481

SST, *See* Sea surface temperature

Stability
  computational, 458–460
  Eady problem, 241–245
  marginal, 222
  neutral, 242

Stably stratified, 55

Standard atmosphere data, 507–508

Standard density, 44

Standard pressure, 44

State potential vorticity gradient, 240

State vector, 479

Statically stable, 55

Static atmosphere, structure of, 18–23

Static stability, 54–56
  equivalent, 398
  normal conditions of, 222
  role of, 220

Stationary waves, 356–359

Steady circulations, dynamics of, 407

Steady forced equatorial motions, 406–409

Steady-state motions, 78

Steady-state storm, potential intensity formula for, 317

Steering level, 222

Stokes's theorem, 499
  relationship between vorticity and circulation, 102
  to solenoidal term, 106
  vorticity and circulation, 102

Storm tracks, 174, 359–361

Stratopause, 413

Stratosphere, 413, 414. *See also* Middle atmosphere

Streamfunction, 119
  geostrophic, 215

Streamlines, 78–81
  of meridional mass circulation, 422, 423*f*

Stretching deformation, 281, 282*f*

Stretching vorticity, 185

Subcritical flow, 300

Substantial derivative, 31

Sudden stratosperic warming, 430–435

Supercell thunderstorms, 306–310

Supercritical flow, 300

Surface cyclone, development of, 243–244

Surface forces, 4

Surface layer, 268–269

Surface pressure tendency, 87–89

Surf zone, 450

Symmetric baroclinic instability, 290–294

Symmetric instability, 213–214

Synoptic-scale motion, 41, 85, 86
  baroclinic disturbance, model of, 204–206
  baroclinic instability, continuous stratified atmosphere and, 234–245
  baroclinic instability, normal mode, 215–226
  continuity equation for, 48–49
  energetics of baroclinic waves, 227–234
  extratropical circulation, 171–178
  hydrodynamic instability, 213–215
  hydrostatic approximation for, 44
  neutral modes, growth and propagation of, 245–251
  potential vorticity thinking, 187–197
  quasi-geostrophic approximation, 178–183
  quasi-geostrophic vorticity equation, 183–186
  solenoidal term in, 107, 108
  vertical motion analysis, 197–204

Synoptic-scale pressure systems, 81

T

Taylor approximation, 7, 23

Taylor series, 5, 456

TEM, *See* Transformed Eulerian mean

Temperature
  advection, 32, 196, 284
  equivalent potential, 294, 304, 503–504
  lapse rate of, 54
  oscillation, 439
  potential, 53–54, 446
  pseudoadiabatic lapse rate, 504–505
  sea surface, 361, 364–366, 377

Thermal low, 387

Thermally direct cell, 332

Thermals, 302

Thermal wind, 33, 81–84, 280, 327, 330, 435
  equation, 82, 314

Thermocline, 390, 391

Thermodynamic energy equation, 50–53, 216, 241
  in condensation heating, 399
  hydrostatic, 215
  in isobaric coordinate system, 69, 207
  in log-pressure coordinates, 237
  quasi-geostrophic, 182, 185
  scale analysis of, 56–57

Thermodynamics of dry atmosphere, 53–57
  adiabatic lapse rate, 54
  potential temperature, 53–54
  static stability, 54–56

Thermodynamics of moist atmosphere, 58–64
  conditional instability, 62–64
  equivalent potential temperature, 58–64
  pseudoadiabatic lapse rate, 61–62

Three-dimensional variational assimilation (3DVAR), 480

Three dimensions, wave properties in, 133–135

Thunderstorms, 306–310

Timescale, dynamical and chemical, 447–448

Topographically forced waves, 430

Topographic waves, 163–165
  Rossby waves, 159–161

Total derivative, 31–33

Total differentiation, 31–33
  of vector in rotating system, 33–35

Tracers
  chemical, 447–448
  dynamical, 446–447
  long-lived, 448, 449

Trajectories, 78–81

Transformed Eulerian mean (TEM), 337–339, 342, 355, 420–424

Transient growth of neutral waves, 247–250

Transient wave, schematic of, 434, 434*f*

Transport barriers, 450

Transport process in stratosphere, 448–450

Transverse oscillations, 137

Trapezoidal implicit scheme, 460–461

Triangular truncation, 474

Tropical circulation
  African wave disturbances, 384–386
  condensation heating, 398–400
  El Niño and southern oscillation, 390–392
  equatorial intraseasonal oscillation, 392
  equatorial wave disturbances, 381–384
  equatorial wave theory, 401–406
  intertropical convergence zone, 378–381
  monsoons, 325, 386–389
  Walker circulation, 389–390

Tropical cyclones, 312–321

Tropical motion
  scale analysis of, 392–398
  steady forced equatorial, 406–409

Tropopause, 1, 113, 114, 172
  maps, 112–113

Troposphere, 1, 172, 413, 414

Truncation
  error, 463–464
  rhomboidal, 474
  triangular, 474

Turbulence, atmospheric, 256–258

Turbulent kinetic energy, 259–261

Two-layer model
  energy equations for, 229–234
  normal mode baroclinic instability, 215–226

Typhoons, 312–321

U

Upper-level anomaly, 214

V

Vapor pressure, 503

Vector analysis, 499–501

Vectorial momentum equation, 37

Velocity, 24*f*
  absolute, 36
  dispersion and group, 131–133

  friction, 268
  Rossby critical, 428
  vector, 32, 36

Vertical coordinates
  generalized, 22–23
  isentropic, 120
  isobaric, 21, 206
  log-pressure, 235–236
  pressure, 20–21
  sigma, 343–345

Vertically propagating waves
  Kelvin waves, 436–437
  planetary waves, 426–430
  Rossby–gravity waves, 437–438

Vertical momentum equation, 45
  in Boussinesq approximation, 57
  scale analysis of, 44, 44*t*

Vertical motion, 84–87, 197
  ageostrophic circulation, 201–203, 203*f*
  baroclinic waves and, 223–226, 225*f*
  omega equation, 198–200
  Q-vector, 200–201, 202*f*

Virtual temperature, 58

Viscosity coefficient, kinematic, 7

Viscosity, eddy, 264

Viscous force, 6–8

Von Karman’s constant, 269

Vorticity, 23, 24*f*, 25. *See also* under type of
  absolute, 100, 101, 119*f*, 213
  advection, 220, 223, 224, 225
  barotropic fluids and, 118–120
  barotropic potential, 118–120
  changes of, 220
  concentration of, 225, 226
  curvature, 103, 104
  definition of, 100
  generation, 105*f*
  geostrophic, 215
  lower-level cyclonic, 215*f*
  natural coordinates and, 102–104
  planetary, 207
  potential, *See* Potential vorticity
  relative, 100, 101, 185, 187

  shear, 103
  stretching, 185
  tendencies, 224–225, 250

Vorticity equation, 207, 216
  baroclinic (Ertel) potential, 120–122
  barotropic, 464–467
  barotropic potential, 118–120
  Cartesian coordinate form, 104–106
  isentropic, 121–122
  in isobaric coordinates, 106–107
  quasi-geostrophic, 213, 215
  scale analysis of, 107–110
  zonal mean potential, 340–341

W

Walker circulation, 389–390, 406

Walker, G. T, 389

Warm-air advection, 195

Warm core systems, 320*f*

Warming, sudden stratospheric, 430–435

Water vapor in atmospheric dynamics, 58

Wavebreaking, 428–430

Wave equation, 139

Wave instabilities, 214

Wave number, 129
  planetary, 130
  zonal, 130

Waves
  acoustic, 136–139
  African wave disturbances, 384–386
  baroclinic, 223–234, 328, 355, 361
  dispersive and group, 131–133
  eddy heat fluxes, 332
  equatorial Kelvin, 404–406, 405*f*, 406*f*
  in equatorial stratosphere, 435–440
  equatorial wave disturbances, 381–384
  equatorial wave theory, 401–406
  gravity, inertia, 152–156
  gravity, internal, 145–150
  gravity, shallow water, 139–144
  Kelvin, 404–406, 405*f*, 406*f*, 436–437
  longitudinal, 136–139

  mountain, 294–302
  mountain lee, 145, 298–299
  neutral, transient growth of, 247–250
  properties of, 128–136
  Rossby, 159–165, 356–359, 426–430
  Rossby–gravity, 401–404, 437–438
  Rossby–Haurwitz, 470
  solitary, 362
  sound, 136–139
  stationary, 356–359
  strategy, 135–136
  topographic, 163–165
  types of, 136–144

Wave speed
  Rossby, 162
  shallow water, 141
  two-layer, 219

Wave vector, 134, 134*f*

Weather forecasting, 455

Weather Prediction by Numerical Process (Richardson), 449

Well-mixed planetary boundary layer, 262–263, 263*f*

Wind
  ageostrophic, 85, 181, 185
  cyclostrophic, 73–74
  geostrophic, 42–43, 181, 215
  gradient approximation, 74–77
  irrotational, 396
  mean zonal, 328
  nondivergent, 396
  thermal, 81–84, 327, 330, 333, 335, 359, 361

Wind-induced surface heat exchange (WISHE), 319

Wind profile, logarithmic, 269

Windstorms
  downslope, 299–302
  hurricanes, 312–321

Z

Zonal force, 421
  in mesosphere, 424

Zonally averaged circulation, 328–341

Zonal mean angular momentum, 345–349

Zonal-mean circulation

  - analysis of, 339
  - maintenance of, 341
  - of middle atmosphere, 417–426
  - structure of, 325

Zonal-mean flow, 213, 325–326
  - Rossby waves, 362

Zonal mean potential vorticity equation, 340–341

Zonal-mean transport, 424–426
  - zonal wave number, 130

Zonal wind oscillations, 439