

Atmospheric and Oceanic Fluid Dynamics

Fundamentals and Large-Scale Circulation



A high-angle aerial photograph of the ocean's surface, showing intricate, swirling patterns of white and light blue. These patterns represent large-scale atmospheric and oceanic circulation systems, such as cyclones and wind-driven surface currents. The image occupies the majority of the book cover.

Geoffrey K. Vallis

CAMBRIDGE

CAMBRIDGE

www.cambridge.org/9780521849692

This page intentionally left blank

Atmospheric and Oceanic Fluid Dynamics

Fundamentals and Large-Scale Circulation

Fluid dynamics is fundamental to our understanding of the atmosphere and oceans. Although many of the same principles of fluid dynamics apply to both the atmosphere and oceans, textbooks on the topic have tended to concentrate on either the atmosphere or the ocean, or on the theory of geophysical fluid dynamics (GFD). However, there is much to be said for a unified discussion, and this major new textbook provides a comprehensive, coherent treatment of all these topics. It is based on course notes that the author has developed over a number of years at Princeton and the University of California.

The first part of the book provides an introduction to the fundamentals of geophysical fluid dynamics, including discussions of rotation and stratification, the role of vorticity and potential vorticity, and scaling and approximations. The second part of the book discusses baroclinic and barotropic instabilities, wave-mean flow interactions and turbulence. The third and fourth parts discuss the general circulation of the atmosphere and ocean. Student problems and exercises, as well as bibliographic and historical notes, are included at the end of each chapter.

Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation will prove to be an invaluable graduate textbook on advanced courses in GFD, meteorology, atmospheric science, and oceanography, and will also be an excellent review volume for researchers. Additional resources are available at www.cambridge.org/9780521849692

GEOFFREY K. VALLIS is a senior scientist and professor in the Program in Atmospheric and Oceanic Sciences and NOAA's Geophysical Fluid Dynamics Laboratory at Princeton University. He is also an associate faculty member at the Program in Applied and Computational Mathematics, and a former professor at the University of California, Santa Cruz. Until recently he was editor of the *Journal of the Atmospheric Sciences*. His research interests include the general circulation of the ocean and atmosphere, turbulence theory, and climate dynamics. He has taught a wide range of topics at Princeton and the University of California, and he has published extensively in both the oceanographic and meteorological literature.

Pre-publication praise for *Atmospheric and Oceanic Fluid Dynamics*

“Geoff Vallis’ *Atmosphere and Ocean Dynamics* will become the standard text on modern large-scale atmosphere and ocean dynamics. It covers the field from the equations of motion to modern developments such as wave-mean flow interaction theory and theories for the global-scale circulations of atmospheres and oceans. There is no book of comparable comprehensiveness, spanning the needs of beginning graduate students and researchers alike.”

Tapio Schneider, California Institute of Technology

“This clearly written, self-contained new book is a modern treatment of atmospheric and oceanic dynamics. The book starts from classical concepts in fluid dynamics and thermodynamics and takes the reader to the frontier of current research. This is an accessible textbook for beginning students in meteorology, oceanography and climate sciences. Mature researchers will welcome this work as a stimulating resource. This is also the only textbook on geophysical fluid dynamics with a comprehensive collection of problems; these cement the material and expand it to a more advanced level. Highly recommended!”

Paola Cessi, Scripps Institution of Oceanography, University of California, San Diego

“Vallis provides a cohesive view of GFD that smoothly blends classic results with modern interpretations. The book strikes an ideal balance between mathematical rigor and physical intuition, and between atmosphere- and ocean-relevant applications. The use of a hierarchy of models is particularly welcome. Each physical phenomenon is modeled with the right degree of complexity, and the reader is introduced to the value of the hierarchy at an early stage. Well-designed homework problems spanning a broad range of difficulty make the book very appropriate for use in introductory courses in GFD.”

Adam Sobel, Lamont-Doherty Earth Observatory, Columbia University

“I have adopted this text for my course in Atmosphere–Ocean Dynamics because the ideas are clearly presented and up-to-date. The text provides the flexibility for the instructor to choose among a variety of paths that take the student from the foundations of the subject to current research topics. For me as a researcher, the text is satisfying because it presents a unified view of the ideas that underlie the modern theory of large scale atmospheric and oceanic circulations.”

Paul J. Kushner, University of Toronto

“The large-scale circulation in the atmosphere–ocean system is maintained by small-scale turbulent motions that interact with large-scale radiative processes. The first half of the book introduces the basic theories of large-scale atmosphere–ocean flows and of small-scale turbulent motions. In the second half, the two theories are brought together to explain how the interactions of motions on different scales maintain the global-scale climate. The emphasis on turbulent motions and their effect on larger scales makes this book a gem in the GFD literature. Finally, we have a textbook that is up to date with our current understanding of the climate system.”

Raffaele Ferrari, Massachusetts Institute of Technology

ATMOSPHERIC AND OCEANIC FLUID DYNAMICS

Fundamentals and Large-scale Circulation

G E O F F R E Y K . V A L L I S
Princeton University, New Jersey



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press

The Edinburgh Building, Cambridge CB2 8RU, UK

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

www.cambridge.org

Information on this title: www.cambridge.org/9780521849692

© G. Vallis 2006

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

First published in print format 2006

ISBN-13 978-0-511-34879-2 eBook (EBL)

ISBN-10 0-511-34879-7 eBook (EBL)

ISBN-13 978-0-521-84969-2 hardback

ISBN-10 0-521-84969-1 hardback

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

To my parents, Jim and Doreen Vallis.

Contents

An asterisk indicates more advanced material that may be omitted on a first reading. A dagger indicates material that is still a topic of research or that is not settled.

Preface	<i>page</i>	xix
Notation		xxiv
Part I FUNDAMENTALS OF GEOPHYSICAL FLUID DYNAMICS		1
1 Equations of Motion		3
1.1 Time Derivatives for Fluids		3
1.1.1 Field and material viewpoints		3
1.1.2 The material derivative of a fluid property		4
1.1.3 Material derivative of a volume		6
1.2 The Mass Continuity Equation		7
1.2.1 An Eulerian derivation		7
1.2.2 Mass continuity via the material derivative		9
1.2.3 A general continuity equation		11
1.3 The Momentum Equation		11
1.3.1 Advection		12
1.3.2 The pressure force		12
1.3.3 Viscosity and diffusion		13
1.3.4 Hydrostatic balance		13
1.4 The Equation of State		14
1.5 Thermodynamic Relations		16
1.5.1 A few fundamentals		16
1.5.2 Various thermodynamic relations		18
1.6 Thermodynamic Equations for Fluids		22

1.6.1	Thermodynamic equation for an ideal gas	23
1.6.2	* Thermodynamic equation for liquids	26
1.7	* More Thermodynamics of Liquids	31
1.7.1	Potential temperature, potential density and entropy	31
1.7.2	* Thermodynamic properties of seawater	33
1.8	Sound Waves	37
1.9	Compressible and Incompressible Flow	38
1.9.1	Constant density fluids	38
1.9.2	Incompressible flows	39
1.10	The Energy Budget	40
1.10.1	Constant density fluid	40
1.10.2	Variable density fluids	42
1.10.3	Viscous effects	43
1.11	An Introduction to Non-Dimensionalization and Scaling	43
1.11.1	The Reynolds number	44

2 Effects of Rotation and Stratification

2.1	Equations in a Rotating Frame	51
2.1.1	Rate of change of a vector	52
2.1.2	Velocity and acceleration in a rotating frame	53
2.1.3	Momentum equation in a rotating frame	54
2.1.4	Mass and tracer conservation in a rotating frame	54
2.2	Equations of Motion in Spherical Coordinates	55
2.2.1	* The centrifugal force and spherical coordinates	55
2.2.2	Some identities in spherical coordinates	57
2.2.3	Equations of motion	60
2.2.4	The primitive equations	61
2.2.5	Primitive equations in vector form	62
2.2.6	The vector invariant form of the momentum equation	63
2.2.7	Angular momentum	64
2.3	Cartesian Approximations: The Tangent Plane	66
2.3.1	The f-plane	66
2.3.2	The beta-plane approximation	67
2.4	The Boussinesq Approximation	67
2.4.1	Variation of density in the ocean	68
2.4.2	The Boussinesq equations	68
2.4.3	Energetics of the Boussinesq system	72
2.5	The Anelastic Approximation	73
2.5.1	Preliminaries	73
2.5.2	The momentum equation	74
2.5.3	Mass conservation	75
2.5.4	Thermodynamic equation	76
2.5.5	* Energetics of the anelastic equations	76
2.6	Changing Vertical Coordinate	77
2.6.1	General relations	77
2.6.2	Pressure coordinates	78

2.6.3	Log-pressure coordinates	80
2.7	Scaling for Hydrostatic Balance	80
2.7.1	Preliminaries	80
2.7.2	Scaling and the aspect ratio	81
2.7.3	* Effects of stratification on hydrostatic balance	82
2.7.4	Hydrostasy in the ocean and atmosphere	84
2.8	Geostrophic and Thermal Wind Balance	85
2.8.1	The Rossby number	85
2.8.2	Geostrophic balance	86
2.8.3	Taylor-Proudman effect	88
2.8.4	Thermal wind balance	89
2.8.5	* Effects of rotation on hydrostatic balance	91
2.9	Static Instability and the Parcel Method	91
2.9.1	A simple special case: a density-conserving fluid	92
2.9.2	The general case: using potential density	93
2.9.3	Lapse rates in dry and moist atmospheres	95
2.10	Gravity Waves	98
2.10.1	Gravity waves and convection in a Boussinesq fluid	98
2.11	* Acoustic-Gravity Waves in an Ideal Gas	100
2.11.1	Interpretation	101
2.12	The Ekman Layer	104
2.12.1	Equations of motion and scaling	105
2.12.2	Integral properties of the Ekman layer	107
2.12.3	Explicit solutions. I: a bottom boundary layer	109
2.12.4	Explicit solutions. II: the upper ocean	112
2.12.5	Observations of the Ekman layer	113
2.12.6	* Frictional parameterization of the Ekman layer	114
3	Shallow Water Systems and Isentropic Coordinates	123
3.1	Dynamics of a Single, Shallow Layer	123
3.1.1	Momentum equations	124
3.1.2	Mass continuity equation	125
3.1.3	A rigid lid	127
3.1.4	Stretching and the vertical velocity	128
3.1.5	Analogy with compressible flow	129
3.2	Reduced Gravity Equations	129
3.2.1	Pressure gradient in the active layer	130
3.3	Multi-Layer Shallow Water Equations	131
3.3.1	Reduced-gravity multi-layer equation	133
3.4	Geostrophic Balance and Thermal wind	134
3.5	Form Drag	135
3.6	Conservation Properties of Shallow Water Systems	136
3.6.1	Potential vorticity: a material invariant	136
3.6.2	Energy conservation: an integral invariant	139
3.7	Shallow Water Waves	140
3.7.1	Non-rotating shallow water waves	140

3.7.2	Rotating shallow water (Poincaré) waves	141
3.7.3	Kelvin waves	143
3.8	Geostrophic Adjustment	144
3.8.1	Non-rotating flow	145
3.8.2	Rotating flow	146
3.8.3	* Energetics of adjustment	148
3.8.4	* General initial conditions	149
3.8.5	A variational perspective	151
3.9	Isentropic Coordinates	152
3.9.1	A hydrostatic Boussinesq fluid	152
3.9.2	A hydrostatic ideal gas	153
3.9.3	* Analogy to shallow water equations	154
3.10	Available Potential Energy	155
3.10.1	A Boussinesq fluid	156
3.10.2	An ideal gas	158
3.10.3	Use, interpretation, and the atmosphere and ocean	159

4	Vorticity and Potential Vorticity	163
4.1	Vorticity and Circulation	163
4.1.1	Preliminaries	163
4.1.2	Simple axisymmetric examples	164
4.2	The Vorticity Equation	165
4.2.1	Two-dimensional flow	167
4.3	Vorticity and Circulation Theorems	168
4.3.1	The ‘frozen-in’ property of vorticity	168
4.3.2	Kelvin’s circulation theorem	171
4.3.3	Baroclinic flow and the solenoidal term	173
4.3.4	Circulation in a rotating frame	173
4.3.5	The circulation theorem for hydrostatic flow	174
4.4	Vorticity Equation in a Rotating Frame	175
4.4.1	The circulation theorem and the beta effect	175
4.4.2	The vertical component of the vorticity equation	176
4.5	Potential Vorticity Conservation	178
4.5.1	PV conservation from the circulation theorem	178
4.5.2	PV conservation from the frozen-in property	180
4.5.3	PV conservation: an algebraic derivation	182
4.5.4	Effects of salinity and moisture	183
4.5.5	Effects of rotation, and summary remarks	183
4.6	* Potential Vorticity in the Shallow Water System	184
4.6.1	Using Kelvin’s theorem	184
4.6.2	Using an appropriate scalar field	185
4.7	Potential Vorticity in Approximate, Stratified Models	186
4.7.1	The Boussinesq equations	186
4.7.2	The hydrostatic equations	187
4.7.3	Potential vorticity on isentropic surfaces	187
4.8	* The Impermeability of Isentropes to Potential Vorticity	188

4.8.1	Interpretation and application	190
5	Simplified Equations for Ocean and Atmosphere	197
5.1	Geostrophic Scaling	198
5.1.1	Scaling in the shallow water equations	198
5.1.2	Geostrophic scaling in the stratified equations	200
5.2	The Planetary-Geostrophic Equations	203
5.2.1	Using the shallow water equations	203
5.2.2	Planetary-geostrophic equations for stratified flow	205
5.3	The Shallow Water Quasi-Geostrophic Equations	207
5.3.1	Single-layer shallow water quasi-geostrophic equations	207
5.3.2	Two-layer and multi-layer quasi-geostrophic systems	211
5.3.3	† Non-asymptotic and intermediate models	214
5.4	The Continuously Stratified Quasi-Geostrophic System	215
5.4.1	Scaling and assumptions	215
5.4.2	Asymptotics	216
5.4.3	Buoyancy advection at the surface	219
5.4.4	Quasi-geostrophy in pressure coordinates	220
5.4.5	The two-level quasi-geostrophic system	221
5.5	* Quasi-geostrophy and Ertel Potential Vorticity	224
5.5.1	* Using height coordinates	224
5.5.2	Using isentropic coordinates	225
5.6	* Energetics of Quasi-Geostrophy	226
5.6.1	Conversion between APE and KE	227
5.6.2	Energetics of two-layer flows	228
5.6.3	Enstrophy conservation	229
5.7	Rossby Waves	229
5.7.1	Waves in a single layer	229
5.7.2	Rossby waves in two layers	232
5.8	* Rossby Waves in Stratified Quasi-Geostrophic Flow	234
5.8.1	Setting up the problem	234
5.8.2	Wave motion	235
	Appendix: Wave Kinematics, Group Velocity and Phase Speed	236
5.A.1	Kinematics and definitions	236
5.A.2	Wave propagation	237
5.A.3	Meaning of group velocity	239
Part II	INSTABILITIES, WAVE-MEAN FLOW INTERACTION AND TURBULENCE	245
6	Barotropic and Baroclinic Instability	247
6.1	Kelvin-Helmholtz Instability	248
6.2	Instability of Parallel Shear Flow	250
6.2.1	Piecewise linear flows	251

6.2.2	Kelvin–Helmholtz instability, revisited	253
6.2.3	Edge waves	254
6.2.4	Interacting edge waves producing instability	254
6.3	Necessary Conditions for Instability	258
6.3.1	Rayleigh’s criterion	258
6.3.2	Fjørtoft’s criterion	260
6.4	Baroclinic Instability	261
6.4.1	A physical picture	261
6.4.2	Linearized quasi-geostrophic equations	263
6.4.3	Necessary conditions for baroclinic instability	264
6.5	The Eady Problem	265
6.5.1	The linearized problem	266
6.5.2	Atmospheric and oceanic parameters	268
6.6	Two-Layer Baroclinic Instability	271
6.6.1	Posing the problem	271
6.6.2	The solution	272
6.7	An Informal View of the Mechanism of Baroclinic Instability	277
6.7.1	The two-layer model	278
6.7.2	Interacting edge waves in the Eady problem	280
6.8	* The Energetics of Linear Baroclinic Instability	282
6.9	* Beta, Shear and Stratification in a Continuous Model	284
6.9.1	Scaling arguments for growth rates, scales and depth	285
6.9.2	Some numerical calculations	287
7	Wave–Mean Flow Interaction	295
7.1	Quasi-geostrophic Preliminaries	296
7.1.1	Potential vorticity flux in the linear equations	297
7.2	The Eliassen–Palm Flux	298
7.2.1	The Eliassen–Palm relation	299
7.2.2	The group velocity property	300
7.2.3	* The orthogonality of modes	302
7.3	The Transformed Eulerian Mean	304
7.3.1	Quasi-geostrophic form	304
7.3.2	The TEM in isentropic coordinates	306
7.3.3	Residual and thickness-weighted circulation	307
7.3.4	* The TEM in the primitive equations	309
7.4	The Non-acceleration Result	314
7.4.1	A derivation from the potential vorticity equation	314
7.4.2	Using TEM to give the non-acceleration result	316
7.4.3	The EP flux and form drag	317
7.5	Influence of Eddies on the Mean Flow in the Eady Problem	319
7.5.1	Formulation	319
7.5.2	Solution	321
7.5.3	The two-level problem	324
7.6	* Necessary Conditions for Instability	324
7.6.1	Stability conditions from pseudomomentum conservation	325

7.6.2	Inclusion of boundary terms	325
7.7	* Necessary Conditions for Instability: Use of Pseudoenergy	327
7.7.1	Two-dimensional flow	327
7.7.2	* Stratified quasi-geostrophic flow	330
7.7.3	* Applications to baroclinic instability	331
8	Basic Theory of Incompressible Turbulence	337
8.1	The Fundamental Problem of Turbulence	338
8.1.1	The closure problem	338
8.1.2	Triad interactions in turbulence	339
8.2	The Kolmogorov Theory	341
8.2.1	The physical picture	341
8.2.2	Inertial-range theory	342
8.2.3	* Another expression of the inertial-range scaling argument	348
8.2.4	A final note on our assumptions	349
8.3	Two-Dimensional Turbulence	349
8.3.1	Energy and enstrophy transfer	351
8.3.2	Inertial ranges in two-dimensional turbulence	354
8.3.3	† More about the phenomenology	357
8.3.4	Numerical illustrations	360
8.4	Predictability of Turbulence	361
8.4.1	Low-dimensional chaos and unpredictability	362
8.4.2	* Predictability of a turbulent flow	363
8.4.3	Implications and weather predictability	365
8.5	* Spectra of Passive Tracers	366
8.5.1	Examples of tracer spectra	367
9	Geostrophic Turbulence and Baroclinic Eddies	377
9.1	Effects of Differential Rotation	377
9.1.1	The wave-turbulence cross-over	378
9.1.2	Generation of zonal flows and jets	380
9.1.3	† Joint effect of β and friction	382
9.2	Stratified Geostrophic Turbulence	384
9.2.1	An analogue to two-dimensional flow	384
9.2.2	Two-layer geostrophic turbulence	385
9.2.3	Phenomenology of two-layer turbulence	388
9.3	† A Scaling Theory for Geostrophic Turbulence	391
9.3.1	Preliminaries	392
9.3.2	Scaling properties	393
9.3.3	The halting scale and the β -effect	395
9.4	† Phenomenology of Baroclinic Eddies in the Atmosphere and Ocean	395
9.4.1	The magnitude and scale of baroclinic eddies	396
9.4.2	Baroclinic eddies and their lifecycle in the atmosphere	397
9.4.3	Baroclinic eddies and their lifecycle in the ocean	400

10 Turbulent Diffusion and Eddy Transport	407
10.1 Diffusive Transport	408
10.1.1 An explicit example	409
10.2 Turbulent Diffusion	409
10.2.1 Simple theory	409
10.2.2 * An anisotropic generalization	413
10.2.3 Discussion	415
10.3 Two-Particle Diffusivity	415
10.3.1 Large particle separation	416
10.3.2 Separation within the inertial range	417
10.4 Mixing Length Theory	419
10.4.1 Requirements for turbulent diffusion	421
10.4.2 A macroscopic perspective	422
10.5 Homogenization of a Scalar that is Adveected and Diffused	423
10.5.1 Non-existence of extrema	423
10.5.2 Homogenization in two-dimensional flow	424
10.6 † Transport by Baroclinic Eddies	425
10.6.1 Symmetric and antisymmetric diffusivity tensors	426
10.6.2 * Diffusion with the symmetric tensor	426
10.6.3 * The skew flux	427
10.6.4 The story so far	429
10.7 † Eddy Diffusion in the Atmosphere and Ocean	430
10.7.1 Preliminaries	430
10.7.2 Magnitude of the eddy diffusivity	430
10.7.3 * Structure: the symmetric transport tensor	432
10.7.4 * Structure: the antisymmetric transport tensor	435
10.7.5 Examples	437
10.8 † Thickness Diffusion	440
10.8.1 Equations of motion	440
10.8.2 Diffusive thickness transport	442
10.9 † Eddy Transport and the Transformed Eulerian Mean	443
10.9.1 Potential vorticity diffusion	443
Part III LARGE-SCALE ATMOSPHERIC CIRCULATION	449
11 The Overturning Circulation: Hadley and Ferrel Cells	451
11.1 Basic Features of the Atmosphere	452
11.1.1 The radiative equilibrium distribution	452
11.1.2 Observed wind and temperature fields	453
11.1.3 Meridional overturning circulation	456
11.1.4 Summary	457
11.2 A Steady Model of the Hadley Cell	457
11.2.1 Assumptions	457
11.2.2 Dynamics	458

11.2.3 Thermodynamics	460
11.2.4 Zonal wind	462
11.2.5 Properties of solution	463
11.2.6 Strength of the circulation	464
11.2.7 † Effects of moisture	465
11.2.8 The radiative equilibrium solution	466
11.3 A Shallow Water Model of the Hadley Cell	468
11.3.1 Momentum balance	468
11.3.2 Thermodynamic balance	468
11.4 † Asymmetry Around the Equator	469
11.5 Eddies, Viscosity and the Hadley Cell	473
11.5.1 Qualitative considerations	473
11.5.2 An idealized eddy-driven model	474
11.6 The Hadley Cell: Summary and Numerical Solutions	477
11.7 The Ferrel Cell	480
12 Zonally Averaged Mid-Latitude Atmospheric Circulation	485
12.1 Surface Westerlies and the Maintenance of a Barotropic Jet	486
12.1.1 Observations and motivation	486
12.1.2 The mechanism of jet production	487
12.1.3 A numerical example	495
12.2 Layered Models of the Mid-latitude Circulation	496
12.2.1 A single-layer model	497
12.2.2 A two-layer model	503
12.2.3 Dynamics of the two-layer model	507
12.3 † Eddy Fluxes and an Example of a Closed Model	513
12.3.1 Equations for a closed model	513
12.3.2 * Eddy fluxes and necessary conditions for instability	514
12.4 A Stratified Model and the Real Atmosphere	516
12.4.1 Potential vorticity and its fluxes	516
12.4.2 Overturning circulation	522
12.5 † The Tropopause and the Stratification of the Atmosphere	522
12.5.1 A radiative-convective model	526
12.5.2 Radiative and dynamical constraints	528
12.6 † Baroclinic eddies and Potential Vorticity Transport	529
12.6.1 A linear argument	530
12.6.2 Mixing potential vorticity and baroclinic adjustment	530
12.6.3 Diffusive transport of potential vorticity	532
12.7 † Extratropical Convection and the Ventilated Troposphere	534
Appendix: TEM for the Primitive Equations in Spherical Coordinates	536
13 Planetary Waves and the Stratosphere	541
13.1 Forced and Stationary Rossby Waves	542
13.1.1 A simple one-layer case	542
13.1.2 Application to Earth's atmosphere	543

13.1.3 * One-dimensional Rossby wave trains	545
13.1.4 The adequacy of linear theory	548
13.2 * Meridional Propagation and Dispersion	549
13.2.1 Ray tracing	549
13.2.2 Rossby waves and Rossby rays	550
13.2.3 Application to an idealized atmosphere	553
13.3 * Vertical Propagation of Rossby Waves in a Stratified Medium	554
13.3.1 Model formulation	554
13.3.2 Model solution	555
13.3.3 Properties of the solution	559
13.4 * Effects of Thermal Forcing	560
13.4.1 Thermodynamic balances	561
13.4.2 Properties of the solution	562
13.4.3 Numerical solutions	563
13.5 Stratospheric Dynamics	566
13.5.1 A descriptive overview	566
13.5.2 † Dynamics of the overturning circulation	569
13.5.3 † The polar vortex and the quasi-horizontal circulation	575

Part IV LARGE-SCALE OCEANIC CIRCULATION	581
--	------------

14 Wind-Driven Gyres	583
14.1 The Depth Integrated Wind-Driven Circulation	585
14.1.1 The Stommel model	586
14.1.2 Alternative formulations	587
14.1.3 Approximate solution of Stommel model	589
14.2 Using Viscosity Instead of Drag	593
14.3 Zonal Boundary Layers	597
14.4 * The Nonlinear Problem	599
14.4.1 A perturbative approach	600
14.4.2 A numerical approach	600
14.5 * Inertial Solutions	601
14.5.1 Roles of friction and inertia	603
14.5.2 Attempting an inertial western boundary solution	604
14.5.3 A fully inertial approach: the Fofonoff model	606
14.6 Topographic Effects on Western Boundary Currents	608
14.6.1 Homogeneous model	608
14.6.2 Advective dynamics	609
14.6.3 Bottom pressure stress and form drag	611
14.7 * Vertical Structure of the Wind-Driven Circulation	613
14.7.1 A two-layer quasi-geostrophic Model	613
14.7.2 The functional relationship between ψ and q	616
14.8 * A Model with Continuous Stratification	619
14.8.1 Depth of the wind's influence	619
14.8.2 The complete solution	620

15 The Buoyancy-Driven Ocean Circulation	627
15.1 Sideways Convection	629
15.1.1 Two-dimensional convection	630
15.1.2 † Phenomenology of the overturning circulation	633
15.2 The Maintenance of Sideways Convection	634
15.2.1 The energy budget	635
15.2.2 Conditions for maintaining a thermally-driven circulation	635
15.2.3 Surface fluxes and non-turbulent flow at small diffusivities	637
15.2.4 The importance of mechanical forcing	639
15.3 Simple Box Models	640
15.3.1 A two-box model	640
15.3.2 * More boxes	644
15.4 A Laboratory Model of the Abyssal Circulation	646
15.4.1 Set-up of the laboratory model	646
15.4.2 Dynamics of flow in the tank	647
15.5 A Model for Oceanic Abyssal Flow	650
15.5.1 Completing the solution	652
15.5.2 Application to the ocean	653
15.5.3 A two-hemisphere model	655
15.6 * A Shallow Water Model of the Abyssal Flow	656
15.6.1 Potential vorticity and poleward interior flow	657
15.6.2 The solution	658
15.7 Scaling for the Buoyancy-Driven Circulation	659
15.7.1 Summary remarks on the Stommel–Arons model	661
16 The Wind- and Buoyancy-Driven Ocean Circulation	667
16.1 The Main Thermocline: an Introduction	667
16.1.1 A simple kinematic model	668
16.2 Scaling and Simple Dynamics of the Main Thermocline	670
16.2.1 An advective scale	671
16.2.2 A diffusive scale	672
16.2.3 Summary of the physical picture	673
16.3 The Internal Thermocline	674
16.3.1 The M equation	674
16.3.2 * Boundary-layer analysis	676
16.4 The Ventilated Thermocline	681
16.4.1 A reduced gravity, single-layer model	682
16.4.2 A two-layer model	683
16.4.3 The shadow zone	686
16.4.4 † The western pool	688
16.5 † A Model of Deep Wind-Driven Overturning	691
16.5.1 A single-hemisphere model	693
16.5.2 A cross-equatorial wind-driven deep circulation	697
16.6 † Flow in a Channel and the Antarctic Circumpolar Current	700
16.6.1 Steady and eddyng flow	701

16.6.2 Vertically integrated momentum balance	702
16.6.3 Form drag and baroclinic eddies	703
16.6.4 † An idealized adiabatic model	708
16.6.5 Form stress and Ekman stress at the ocean bottom	709
16.6.6 Differences between gyres and channels	710
Appendix: Miscellaneous Relationships in a Layered Model	710
16.A.1 Hydrostatic balance	711
16.A.2 Geostrophic and thermal wind balance	711
16.A.3 Explicit cases	712
References	717
Index	738

We must be ignorant of much, if we would know anything.
Cardinal John Newman (1801–1890).

Preface

THIS IS A BOOK on the fluid dynamics of the atmosphere and ocean, with an emphasis on the fundamentals and on the large-scale circulation, the latter meaning flows from the scale of the first deformation radius (a few tens of kilometres in the ocean, several hundred kilometres in the atmosphere) to the global scale. The book is primarily a textbook; it is designed to be accessible to students and could be used as a text for graduate courses. It may be also useful as an introduction to the field for scientists in other areas and as a reference for researchers in the field, and some aspects of the book have the flavour of a research monograph.

Atmospheric and oceanic fluid dynamics (AOFD) is a fascinating field, and simultaneously both pure and applied. It is a pure field because it is intimately tied to some of the most fundamental and unsolved problems in fluid dynamics — problems in turbulence and wave-mean flow interaction, problems in chaos and predictability, and problems in the general circulation itself. Yet it is applied because the climate and weather so profoundly affect the human condition, and so a great deal of effort goes into making predictions — indeed the practice of weather forecasting is a remarkable example of a successful applied science, in spite of the natural limitations to predictability that are now reasonably well understood. The field is plainly important, for we live in the atmosphere and the ocean covers about two-thirds of the Earth. It is also very broad, encompassing such diverse topics as the general circulation, gyres, boundary layers, waves, convection and turbulence. My goal in this book is to present a coherent selection of these topics, concentrating on the foundations but without shying away from the boundaries of active areas of research — for a book that limits itself to what is absolutely settled would, I think, be rather dry, a quality best reserved for martinis and humour.

AOFD is closely related to the field of *geophysical fluid dynamics* (GFD). The latter can be, depending on one's point of view, both a larger and a smaller field than the former. It is larger because GFD, in its broadest meaning, includes not just the fluid dynamics of the Earth's atmosphere and ocean, but also the fluid dynamics of such things as the Earth's interior, volcanoes, lava flows and planetary atmospheres; it is the fluid mechanics of all

things geophysical. But at the same time the appellation ‘GFD’ implies a certain austerity, and the subject is often seen as the one that provides the fundamental principles and language for understanding geophysical flows without being suffocated by the overwhelming detail of the real world. In this book we are guided by the ascetic spirit of GFD, and my hope is that the reader will gain a solid grounding in the fundamentals, motivated by and with an appreciation for the problems of the real world.

The book is an outgrowth of various courses that I have taught over the years, mainly at Princeton University but also at the University of California and at summer schools or similar in Boulder and Kyoto. There are four parts to the book: fundamentals of geophysical fluid dynamics; instabilities, wave-mean flow interaction and turbulence; atmospheric circulation; and ocean circulation. Each corresponds, very roughly, to a one-term graduate course, although parts could also be used for undergraduates. Limitations enforced both by the need to keep the book coherent and focused, and my own expertise or (especially) lack thereof, naturally limit the choice of topics. In particular the chapters on the circulation focus on the steady and statistically steady large-scale circulation and postpone a number of important topics are omitted — tropical and equatorial dynamics, many of the effects of moisture on atmospheric circulation, the spin-up of the ocean circulation, atmospheric and oceanic tides, the quasi-biennial oscillation, and so on. I have however — and at no extra charge, mind you — discussed the large-scale circulation of both atmosphere and ocean. The similarities and differences between the two systems are, I believe, so instructive that even if one’s interest is solely in one, there is much to be gained by studying the other. The references at the end of the book are representative and not exhaustive, and almost certainly disproportionately represent articles written in English and those with which I happen to be familiar. For the benefit of the reader interested in exploring the development of the subject I have included references to a number of historical articles, even when the presentation given does not draw from them. If there are other references that are particularly relevant I trust the reader will inform me.

I have tried to keep the overall treatment of topics as straightforward and as clear as I know how. In particular, I have tried to be as explicit as possible in my explanations, even at the risk of descending from pedagogy into pedantry. Relatedly, there is a certain amount of repetition between sections, and this serves both to emphasize the important things and to keep chapters reasonably self-contained. The chapters are of course intellectually linked, for example, heat transport in the atmosphere depends on baroclinic instability, but hopefully the reader already familiar with the latter will be able to read about the former without too much cross-referencing. The treatment generally is fairly physical and phenomenological, and rigour in the mathematical sense is absent; I treat the derivatives of integrals and of infinitesimal quantities rather informally, for example.

The figures (many in colour) may all be downloaded from the CUP web site associated with this book. An asterisk, *, next to a section heading means that the section may be omitted on first reading; although normally uncontroversial, it may contain advanced material that is not essential for subsequent sections. A dagger, †, next to a section heading means that the section discusses topics of research. Very roughly speaking, one might interpret as asterisk as indicating there is advanced manipulation of the equations, whereas a dagger might indicate there is approximation of the equations, or an interpretation that is not universally regarded as settled; *caveat emptor*. There is some arbitrariness in such markings, especially where the section deals with a well understood model of a poorly understood

reality. Sections so-marked may be regarded as providing an introduction to the literature, rather than a complete or finished treatment, and the section may also require knowledge of material that appears later in the book. If the asterisk or dagger is applied to a section it applies to all the subsections within, and if a dagger or asterisk appears within a section that is already marked, the warning is even more emphatic. Reading the endnotes at the end of each chapter is not needed in order to follow the arguments in the main text. Problems marked with black diamonds may, like similarly marked ski-slopes, be difficult, and I do not know the solutions to all of them. Good answers to some of them may be publishable and I would appreciate hearing about any such work. I would also appreciate any comments on the material presented in the text. *Qui docet discit.*

Finally, I should say that this book owes its existence in part to my own hubris and selfishness: hubris to think that others might wish to read what I have written, and selfishness because the enjoyable task of writing such a book masquerades as work.

Summary of Contents

The chapters within each part, and the sections within each chapter, form *logical* units, and do not necessarily directly correspond to a single lecture or set number of lectures.

Part I. Fundamentals of geophysical fluid dynamics

Chapter 1 is a brief introduction to fluid dynamics and the basic equations of motion, assuming no prior knowledge of the field. Readers with prior knowledge of fluid dynamics might skim it lightly, concentrating on those aspects unique to the atmosphere or ocean.

Chapter 2 introduces the effects of stratification and rotation, these being the two main effects that most differentiate AOFD from other branches of fluid dynamics. Fundamental topics such as the primitive equations, the Boussinesq equations, and Ekman layers are introduced here, and these form the foundation for the rest of the book.

Chapter 3 focuses on the shallow water equations. Many of the principles of geophysical fluid dynamics have their simplest expression in the shallow water equations, because the effects of stratification are either eliminated or much simplified. The equations thus provide a relatively gentle introduction to the field.

Chapter 4 discusses vorticity and potential vorticity. Potential vorticity plays an especially important role in large-scale, rotating and stratified flows, and its conservation provides the basis for the equation sets of chapter 5.

Chapter 5 derives simplified equation sets for large-scale flows, in particular the quasi-geostrophic and planetary-geostrophic equation sets, and introduces a simple application, Rossby waves. Much of our theoretical understanding of the large-scale circulation has arisen through the use of these equations.

Part II. Instabilities, wave-mean flow interaction and turbulence

Chapter 6 covers barotropic and baroclinic instability, the latter being the instability that gives rise to weather — and therefore being, perhaps, the form of hydrodynamic instability that most affects the human condition.

Chapter 7 provides an introduction to the important topic, albeit one that is regarded as difficult, of wave-mean flow interaction. That is, how do the waves and instabilities affect the mean flow in which they propagate?

Chapter 8 and 9 are on the statistical theory of turbulence. Chapter 8 introduces the basic concepts of two- and three-dimensional turbulence, and chapter 9 applies similar ideas to geostrophic turbulence.

Chapter 10 discusses turbulent diffusion — a hoary subject, both used and abused, yet one that plays a central role in our thinking about the transport properties of eddies in the atmosphere and ocean.

Part III. Large-scale atmospheric circulation

Chapter 11 is mostly concerned with the dynamics of the Hadley Cell, and, rather descriptively, with the Ferrel Cell.

Chapter 12 addresses the mid-latitude circulation. The goal of this chapter is to provide the basis of an understanding of such topics as the surface westerly winds, the dynamics of the Ferrel Cell, the stratification of the atmosphere and the height of the tropopause. These are still active topics of research, so we cannot always be definitive, although many of the underlying principles are now established, and our discussion emphasizes the fundamentals that have, I think, permanent value.

Chapter 13 discusses stationary waves in the atmosphere, mainly produced by the interaction of the zonal wind with mountains and land-sea temperature contrasts. It also discusses the vertical propagation of Rossby waves, and how these induce a stratospheric circulation.

Part IV. Large-scale oceanic circulation

Chapter 14 discusses the wind-driven circulation, in particular the ocean gyres, either ignoring buoyancy effects or assuming that buoyancy forcing acts primarily to set up a stratification that can be taken as a given.

Chapter 15 discusses the buoyancy-driven circulation, largely neglecting the effects of wind forcing.

Chapter 16 addresses the combined effects of wind and buoyancy forcing in setting up the stratification and in driving both the predominantly horizontal gyral circulation and the overturning circulation. As with many sections of Part III, many of these topics are still being actively researched, although, again, many of the underlying principles have, I think, now been established and have permanent value.

Acknowledgements

This book would not have been possible without the input, criticism, encouragement and advice of a large number of people, from students to senior scientists. Students at Princeton University, New York University, Columbia University, California Institute of Technology, MIT and the University of Toronto have used earlier versions of the text in various courses, and I am grateful for the feedback received from the instructors and students about what works and what doesn't, as well as for numerous detailed comments. Parts of the first few chapters and many of the problems draw on notes prepared over the years for a graduate class at Princeton University taught by Steve Garner, Isaac Held, Yoshio Kurihara, Paul Kushner and me. Steve Garner has been notably generous with his time and ideas with respect to this material.

I would like to thank following individuals for providing comments on the text or for

conversational input: Alistair Adcroft, Brian Arbic, Roger Berlind, Pavel Berloff, Thomas Birner, Paola Cessi, Sorin Codoban, Agatha de Boer, Roland de Szoeke, Raffaele Ferrari, Baylor Fox-Kemper, Dargan Frierson, Steve Griffies, Brian Hoskins, Huei-Ping Huang, Chris Hughes, Kosuke Ito and his fellow students in Kyoto, Laura Jackson, Martin Juckes, Allan Kaufman, Samar Khatiwala, Andrew Kositsky, Paul Kushner, Joe LaCasce, Trevor McDougall, Paul O'Gorman, Tim Palmer, David Pearson, Lorenzo Polvani, Cathy Raphael, Adam Scaife, John Scinocca, Rob Scott, Tiffany Shaw, Sabrina Speich, Robbie Toggweiler, Yue-Kin Tsang, Ross Tulloch, Eli Tziperman, Jacques Vanneste, Chris Walker, Ric Williams, Michal Ziemiński and Pablo Zurita-Gotor. Some of the above provided detailed comments on several sections or chapters, and I am very grateful to them; a number of other people provided figures and these are acknowledged where they appear. Anders Persson and Roger Samelson provided a number of useful historical remarks and interpretations, and Michael McIntyre (see also McIntyre 1997) emphasized the importance of being as explicit as possible in my explanations. I am particularly indebted to Tapio Schneider, Adam Sobel, Jürgen Theiss and Andy White who read and gave detailed comments on many chapters of text, to Isaac Held for many conversations over the years on all aspects of atmospheric and oceanic fluid dynamics, to Shafer Smith for both comments and code, to Ed Gerber for several calculations and figures, and to Bill Young for very useful input on thermodynamics and the Boussinesq equations. I would also like to thank Matt Lloyd of CUP for his encouragement, my copy-editor, Louise Staples, for her light but careful touch, and Anna Valerio for efficient secretarial assistance. Needless to say, I take full responsibility for the errors, both scientific and presentational, that undoubtedly remain.

My first exposure to the field came as a graduate student in the fecund atmosphere of Atmospheric Physics Group at Imperial College in the late 1970s and early 1980s, and I'd like to thank everyone who was there at that time. You know who you are. I would also like to thank everyone at the Geophysical Fluid Dynamics Laboratory (GFDL) and in Princeton University's Atmospheric and Oceanic Science program for creating the pleasant and stimulating work environment that made it possible to write this book. I am grateful to the staff and scientists at the Plymouth Marine Laboratory and the U. K. Meteorological Office for their hospitality during a sabbatical visit. Finally, I would like to thank the National Science Foundation and the National Oceanic and Atmospheric Administration for providing financial support over the years, and Bess for support of a much greater kind.

Note on second printing

I have taken advantage the opportunity afforded by a second printing of this book to correct a number of typographic errors and make a number of small corrections to the text throughout. The pagination and equation numbering are mostly unaltered, although some microtypographic improvements have changed some of the line breaks. I am grateful to many readers who have sent in comments and corrections, and I would particularly like to acknowledge Roger Berlind for his exceptionally detailed and perceptive comments on the entire book. Additional resources, including downloadable figures and solutions to many of the end-of-chapter exercises, are available at www.vallisbook.org and www.cambridge.org/9780521849692.

NOTATION

The list below contains only the more important variables, or instances of non-obvious notation. Distinct meanings are separated with a semi-colon. Variables are normally set in italics, constants (e.g., π) in roman (i.e., upright), differential operators in roman, vectors in bold, and tensors in bold sans serif. Thus, vector variables are in bold italics, vector constants (e.g., unit vectors) in bold roman, and tensor variables are in bold slanting sans serif. Physical units are set in roman. A subscript denotes a derivative only if the subscript is a coordinate, such as x, y, z or t ; a subscript 0 generally denotes a constant reference value (e.g., ρ_0). The components of a vector are denoted by superscripts.

Variable	Description
b	Buoyancy, $-g\delta\rho/\rho_0$ or $g\delta\theta/\tilde{\theta}$.
c_g	Group velocity, (c_g^x, c_g^y, c_g^z) .
c_p	Phase speed; heat capacity at constant pressure.
c_v	Heat capacity constant volume.
c_s	Sound speed.
f, f_0	Coriolis parameter, and its reference value.
\mathbf{g}, g	Vector acceleration due to gravity, magnitude of \mathbf{g} .
h	Layer thickness (in shallow water equations).
$\mathbf{i}, \mathbf{j}, \mathbf{k}$	Unit vectors in (x, y, z) directions.
i	An integer index.
i	Square root of -1 .
\mathbf{k}	Wave vector, with components (k, l, m) or (k^x, k^y, k^z) .
k_d	Wave number corresponding to deformation radius.
L_d	Deformation radius.
L, H	Horizontal length scale, vertical (height) scale.
m	Angular momentum about the Earth's axis of rotation.
M	Montgomery function, $M = c_p T + \Phi$.
N	Buoyancy, or Brunt–Väisälä, frequency.
p	Pressure.
Pr	Prandtl ratio, f_0/N .
q	Quasi-geostrophic potential vorticity.
Q	Potential vorticity (in particular Ertel PV).
\dot{Q}	Rate of heating.
Ra	Rayleigh number.
Re	Real part of expression.
Re	Reynolds number, UL/ν .
Ro	Rossby number, U/fL .
S	Salinity; source term on right-hand side of an evolution equation.
S_o, \mathbf{S}_o	Solenoidal term, solenoidal vector.
T	Temperature.
t	Time.
\mathbf{u}	Two-dimensional (horizontal) velocity, (u, v) .
\mathbf{v}	Three-dimensional velocity, (u, v, z) .
x, y, z	Cartesian coordinates, usually in zonal, meridional and vertical directions.
Z	Log-pressure, $-H \log p/p_R$. Usually, $H = 7.5$ km and $p_R = 10^5$ Pa.

Variable	Description
\mathcal{A}	Wave activity.
α	Inverse density, or specific volume; aspect ratio.
β	Rate of change of f with latitude, $\partial f / \partial y$.
β_T, β_S	Coefficient of expansion with respect to temperature, salinity.
ϵ	Generic small parameter (epsilon).
ε	Cascade or dissipation rate of energy (varepsilon).
η	Specific entropy; perturbation height; enstrophy cascade or dissipation rate.
\mathcal{F}	Eliassen Palm flux, $(\mathcal{F}^y, \mathcal{F}^z)$.
γ	Vorticity gradient, $\beta - u_{yy}$; the ratio c_p/c_v .
Γ	Lapse rate.
κ	Diffusivity; the ratio R/c_p .
\mathcal{K}	Kolmogorov or Kolmogorov-like constant.
Λ	Shear, e.g., $\partial U / \partial z$.
μ	Viscosity.
ν	Kinematic viscosity, μ/ρ .
v	Meridional component of velocity.
ϕ	Pressure divided by density, p/ρ ; passive tracer.
Φ	Geopotential, usually gz .
Π	Exner function, $\Pi = c_p T / \theta = c_p (p/p_R)^{R/c_p}$.
ω	Vorticity.
Ω, Ω	Rotation rate of Earth and associated vector.
ψ	Streamfunction.
ρ	Density.
ρ_θ	Potential density.
σ	Layer thickness, $\partial z / \partial \theta$; Prandtl number ν/κ ; measure of density, $\rho - 1000$.
τ	Stress vector, often wind stress.
$\tilde{\tau}$	Kinematic stress, $\tilde{\tau}/\rho$.
τ	Zonal component or magnitude of wind stress; eddy turnover time.
θ	Potential temperature.
ϑ, λ	Latitude, longitude.
ζ	Vertical component of vorticity.
$\left(\frac{\partial a}{\partial b}\right)_c$	Derivative of a with respect to b at constant c .
$\frac{\partial a}{\partial b} \Big _{a=c}$	Derivative of a with respect to b evaluated at $a = c$.
∇_a	Gradient operator at constant value of coordinate a , e.g., $\nabla_z = \mathbf{i} \partial_x + \mathbf{j} \partial_y$.
$\nabla_a \cdot$	Divergence operator at constant value of coordinate a , e.g., $\nabla_z \cdot = (\mathbf{i} \partial_x + \mathbf{j} \partial_y) \cdot$.
∇^\perp	Perpendicular gradient, $\nabla^\perp \phi \equiv \mathbf{k} \times \nabla \phi$.
curl_z	Vertical component of $\nabla \times$ operator, $\text{curl}_z \mathbf{A} = \mathbf{k} \cdot \nabla \times \mathbf{A} = \partial_x A^y - \partial_y A^x$.
$\frac{D}{Dt}$	Material derivative (generic).
$\frac{D_g}{Dt}$	Material derivative using geostrophic velocity, for example $\partial / \partial t + \mathbf{u}_g \cdot \nabla$.
$\frac{D_3}{Dt}, \frac{D_2}{Dt}$	Material derivative in three dimensions and in two dimensions, for example $\partial / \partial t + \mathbf{v} \cdot \nabla$ and $\partial / \partial t + \mathbf{u} \cdot \nabla$ respectively.

Part I

FUNDAMENTALS OF GEOPHYSICAL FLUID DYNAMICS

Are you sitting comfortably? Then I'll begin.

Julia Lang, *Listen With Mother*, BBC radio program, 1950–1982.

CHAPTER ONE

Equations of Motion

THIS CHAPTER establishes the fundamental governing equations of motion for a fluid, with particular attention to the fluids of the Earth's atmosphere and ocean. Our approach in many places is quite informal, and the interested reader may consult the references given for more detail.

1.1 TIME DERIVATIVES FOR FLUIDS

The equations of motion of fluid mechanics differ from those of rigid-body mechanics because fluids form a continuum, and because fluids flow and deform. Thus, even though both classical solid and fluid media are governed by the same relatively simple physical laws (Newton's laws and the laws of thermodynamics), the expression of these laws differs between the two. To determine the equations of motion for fluids we must clearly establish what the time derivative of some property of a fluid actually means, and that is the subject of this section.

1.1.1 Field and material viewpoints

In solid-body mechanics one is normally concerned with the position and momentum of identifiable objects — the angular velocity of a spinning top or the motions of the planets around the Sun are two well-worn examples. The position and velocity of a particular object is then computed as a function of time by formulating equations of the form

$$\frac{dx_i}{dt} = F(\{x_i\}, t), \quad (1.1)$$

where $\{x_i\}$ is the set of positions and velocities of all the interacting objects and the operator F on the right-hand side is formulated using Newton's laws of motion. For example, two massive point objects interacting via their gravitational field obey

$$\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i, \quad \frac{d\mathbf{v}_i}{dt} = \frac{Gm_j}{(\mathbf{r}_i - \mathbf{r}_j)^2} \hat{\mathbf{r}}_{i,j}, \quad i = 1, 2; j = 3 - i. \quad (1.2)$$

We thereby predict the positions, \mathbf{r}_i , and velocities, \mathbf{v}_i , of the objects given their masses, m_i , and the gravitational constant G , and where $\hat{\mathbf{r}}_{i,j}$ is a unit vector directed from \mathbf{r}_i to \mathbf{r}_j .

In fluid dynamics such a procedure would lead to an analysis of fluid motions in terms of the positions and momenta of different fluid parcels, each identified by some label, which might simply be their position at an initial time. We call this a *material* point of view, because we are concerned with identifiable pieces of material; it is also sometimes called a *Lagrangian* view, after J.-L. Lagrange. The procedure is perfectly acceptable in principle, and if followed would provide a complete description of the fluid dynamical system. However, from a practical point of view it is much more than we need, and it would be extremely complicated to implement. Instead, for most problems we would like to know what the values of velocity, density and so on are at *fixed points* in space as time passes. (A weather forecast we might care about tells us how warm it will be where we live, and if we are given that we do not particularly care where a fluid parcel comes from, or where it subsequently goes.) Since the fluid is a continuum, this knowledge is equivalent to knowing how the fields of the dynamical variables evolve in space and time, and this is often known as the *field* or *Eulerian* viewpoint, after L. Euler.¹ Thus, whereas in the material view we consider the time evolution of identifiable fluid elements, in the field view we consider the time evolution of the fluid field from a particular frame of reference. That is, we seek evolution equations of the general form

$$\frac{\partial}{\partial t} \varphi(x, y, z, t) = G(\varphi, x, y, z, t), \quad (1.3)$$

where the field $\varphi(x, y, z, t)$ represents all the dynamical variables (velocity, density, temperature, etc.) and G is some operator to be determined from Newton's laws of motion and appropriate thermodynamic laws.

Although the field viewpoint will often turn out to be the most practically useful, the material description is invaluable both in deriving the equations and in the subsequent insight it frequently provides. This is because the important quantities from a fundamental point of view are often those which are associated with a given fluid element: it is these which directly enter Newton's laws of motion and the thermodynamic equations. It is thus important to have a relationship between the rate of change of quantities associated with a given fluid element and the local rate of change of a field. The material or advective derivative provides this relationship.

1.1.2 The material derivative of a fluid property

A *fluid element* is an infinitesimal, indivisible, piece of fluid — effectively a very small fluid parcel of fixed mass. The material derivative is the rate of change of a property (such as temperature or momentum) of a particular fluid element or finite mass; that is to say, it is the total time derivative of a property of a piece of fluid. It is also known as the 'substantive derivative' (the derivative associated with a parcel of fluid substance),

the ‘advection derivative’ (because the fluid property is being advected), the ‘convective derivative’ (convection is a slightly old-fashioned name for advection, still used in some fields), or the ‘Lagrangian derivative’.

Let us suppose that a fluid is characterized by a (given) velocity field $\mathbf{v}(\mathbf{x}, t)$, which determines its velocity throughout. Let us also suppose that it has another property ϕ , and let us seek an expression for the rate of change of ϕ of a fluid element. Since ϕ is changing in time and in space we use the chain rule,

$$\delta\phi = \frac{\partial\phi}{\partial t}\delta t + \frac{\partial\phi}{\partial x}\delta x + \frac{\partial\phi}{\partial y}\delta y + \frac{\partial\phi}{\partial z}\delta z = \frac{\partial\phi}{\partial t}\delta t + \delta\mathbf{x} \cdot \nabla\phi. \quad (1.4)$$

This is true in general for any δt , δx , etc. The total time derivative is then

$$\frac{d\phi}{dt} = \frac{\partial\phi}{\partial t} + \frac{d\mathbf{x}}{dt} \cdot \nabla\phi. \quad (1.5)$$

If this equation is to represent a material derivative we must identify the time derivative in the second term on the right-hand side with the rate of change of position of a fluid element, namely its velocity. Hence, the material derivative of the property ϕ is

$$\frac{d\phi}{dt} = \frac{\partial\phi}{\partial t} + \mathbf{v} \cdot \nabla\phi. \quad (1.6)$$

The right-hand side expresses the material derivative in terms of the local rate of change of ϕ plus a contribution arising from the spatial variation of ϕ , experienced only as the fluid parcel moves. Because the material derivative is so common, and to distinguish it from other derivatives, we denote it by the operator D/Dt . Thus, the material derivative of the field ϕ is

$$\frac{D\phi}{Dt} = \frac{\partial\phi}{\partial t} + (\mathbf{v} \cdot \nabla)\phi$$

. (1.7)

The brackets in the last term of this equation are helpful in reminding us that $(\mathbf{v} \cdot \nabla)$ is an operator acting on ϕ .

Material derivative of vector field

The material derivative may act on a vector field \mathbf{b} , in which case

$$\frac{D\mathbf{b}}{Dt} = \frac{\partial\mathbf{b}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{b}. \quad (1.8)$$

In Cartesian coordinates this is

$$\frac{D\mathbf{b}}{Dt} = \frac{\partial\mathbf{b}}{\partial t} + u \frac{\partial\mathbf{b}}{\partial x} + v \frac{\partial\mathbf{b}}{\partial y} + w \frac{\partial\mathbf{b}}{\partial z}, \quad (1.9)$$

and for a particular component of \mathbf{b} ,

$$\frac{Db^x}{Dt} = \frac{\partial b^x}{\partial t} + u \frac{\partial b^x}{\partial x} + v \frac{\partial b^x}{\partial y} + w \frac{\partial b^x}{\partial z}, \quad (1.10)$$

or, in Cartesian tensor notation,

$$\frac{Db_i}{Dt} = \frac{\partial b_i}{\partial t} + v_j \frac{\partial b_i}{\partial x_j} = \frac{\partial b_i}{\partial t} + v_j \partial_j b_i. \quad (1.11)$$

where the subscripts denote the Cartesian components and repeated indices are summed. In coordinate systems other than Cartesian the advective derivative of a vector is not simply the sum of the advective derivative of its components, because the coordinate vectors change direction with position; this will be important when we deal with spherical coordinates. Finally, we note that the advective derivative of the position of a fluid element, \mathbf{r} say, is its velocity, and this may easily be checked by explicitly evaluating $D\mathbf{r}/Dt$.

1.1.3 Material derivative of a volume

The volume that a given, unchanging, mass of fluid occupies is deformed and advected by the fluid motion, and there is no particular reason why it should remain constant. Indeed, the volume will change as a result of the movement of each element of its bounding material surface, and will in general change if there is a non-zero normal component of the velocity at the fluid surface. That is, if the volume of some fluid is $\int dV$, then

$$\frac{D}{Dt} \int_V dV = \int_S \mathbf{v} \cdot d\mathbf{S}, \quad (1.12)$$

where the subscript V indicates that the integral is a definite integral over some finite volume V , although the limits of the integral will be functions of time if the volume is changing. The integral on the right-hand side is over the closed surface, S , bounding the volume. Although intuitively apparent (to some), this expression may be derived more formally using Leibnitz's formula for the rate of change of an integral whose limits are changing (problem 1.2). Using the divergence theorem on the right-hand side, (1.12) becomes

$$\frac{D}{Dt} \int_V dV = \int_V \nabla \cdot \mathbf{v} dV. \quad (1.13)$$

The rate of change of the volume of an infinitesimal fluid element of volume ΔV is obtained by taking the limit of this expression as the volume tends to zero, giving

$$\lim_{\Delta V \rightarrow 0} \frac{1}{\Delta V} \frac{D\Delta V}{Dt} = \nabla \cdot \mathbf{v}. \quad (1.14)$$

We will often write such expressions informally as

$$\frac{D\Delta V}{Dt} = \Delta V \nabla \cdot \mathbf{v}, \quad (1.15)$$

with the limit implied.

Consider now the material derivative of some fluid property, ξ say, multiplied by the volume of a fluid element, ΔV . Such a derivative arises when ξ is the amount per unit volume of ξ -substance — it might, for example, be mass density or the amount of a dye per unit volume. Then we have

$$\frac{D}{Dt} (\xi \Delta V) = \xi \frac{D\Delta V}{Dt} + \Delta V \frac{D\xi}{Dt}. \quad (1.16)$$

Using (1.15) this becomes

$$\frac{D}{Dt} (\xi \Delta V) = \Delta V \left(\xi \nabla \cdot \mathbf{v} + \frac{D\xi}{Dt} \right), \quad (1.17)$$

and the analogous result for a finite fluid volume is just

$$\frac{D}{Dt} \int_V \xi dV = \int_V \left(\xi \nabla \cdot \mathbf{v} + \frac{D\xi}{Dt} \right) dV. \quad (1.18)$$

This expression is to be contrasted with the Eulerian derivative for which the volume, and so the limits of integration, are fixed and we have

$$\frac{d}{dt} \int_V \xi dV = \int_V \frac{\partial \xi}{\partial t} dV. \quad (1.19)$$

Now consider the material derivative of a fluid property φ multiplied by the mass of a fluid element, $\rho \Delta V$, where ρ is the fluid density. Such a derivative arises when φ is the amount of φ -substance per unit mass (note, for example, that the momentum of a fluid element is $\rho \mathbf{v} \Delta V$). The material derivative of $\varphi \rho \Delta V$ is given by

$$\frac{D}{Dt} (\varphi \rho \Delta V) = \rho \Delta V \frac{D\varphi}{Dt} + \varphi \frac{D}{Dt} (\rho \Delta V) \quad (1.20)$$

But $\rho \Delta V$ is just the mass of the fluid element, and that is constant — that is how a fluid element is defined. Thus the second term on the right-hand side vanishes and

$$\frac{D}{Dt} (\varphi \rho \Delta V) = \rho \Delta V \frac{D\varphi}{Dt} \quad \text{and} \quad \frac{D}{Dt} \int_V \varphi \rho dV = \int_V \rho \frac{D\varphi}{Dt} dV, \quad (1.21a,b)$$

where (1.21b) applies to a finite volume. That expression may also be derived more formally using Leibnitz's formula for the material derivative of an integral, and the result also holds when φ is a vector. The result is quite different from the corresponding Eulerian derivative, in which the volume is kept fixed; in that case we have:

$$\frac{d}{dt} \int_V \varphi \rho dV = \int_V \frac{\partial}{\partial t} (\varphi \rho) dV. \quad (1.22)$$

Various material and Eulerian derivatives are summarized in the shaded box on the following page.

1.2 THE MASS CONTINUITY EQUATION

In classical mechanics mass is absolutely conserved, and in solid-body mechanics we normally do not need an explicit equation of mass conservation. However, in fluid mechanics fluid flows into and away from regions, and fluid density may change, and an equation that explicitly accounts for the flow of mass is one of the 'equations of motion' of the fluid.

1.2.1 An Eulerian derivation

We will first derive the mass conservation equation from an Eulerian point of view; that is to say, our reference frame is fixed in space and the fluid flows through it.

Material and Eulerian Derivatives

The material derivative of a scalar (ϕ) and a vector (\mathbf{b}) field are given by:

$$\frac{D\phi}{Dt} = \frac{\partial\phi}{\partial t} + \mathbf{v} \cdot \nabla \phi, \quad \frac{D\mathbf{b}}{Dt} = \frac{\partial\mathbf{b}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{b}. \quad (\text{D.1})$$

Various material derivatives of integrals are:

$$\frac{D}{Dt} \int_V \phi \, dV = \int_V \left(\frac{D\phi}{Dt} + \phi \nabla \cdot \mathbf{v} \right) \, dV = \int_V \left(\frac{\partial\phi}{\partial t} + \nabla \cdot (\phi \mathbf{v}) \right) \, dV, \quad (\text{D.2})$$

$$\frac{D}{Dt} \int_V \, dV = \int_V \nabla \cdot \mathbf{v} \, dV, \quad (\text{D.3})$$

$$\frac{D}{Dt} \int_V \rho \phi \, dV = \int_V \rho \frac{D\phi}{Dt} \, dV. \quad (\text{D.4})$$

These formulae also hold if ϕ is a vector. The Eulerian derivative of an integral is:

$$\frac{d}{dt} \int_V \phi \, dV = \int_V \frac{\partial\phi}{\partial t} \, dV, \quad (\text{D.5})$$

so that

$$\frac{d}{dt} \int_V \, dV = 0 \quad \text{and} \quad \frac{d}{dt} \int_V \rho \phi \, dV = \int_V \frac{\partial \rho \phi}{\partial t} \, dV. \quad (\text{D.6})$$

Cartesian derivation

Consider an infinitesimal rectangular parallelepiped (i.e., a cuboid) control volume, $\Delta V = \Delta x \Delta y \Delta z$ that is fixed in space, as in Fig. 1.1. Fluid moves into or out of the volume through its surface, including through its faces in the y - z plane of area $\Delta A = \Delta y \Delta z$ at coordinates x and $x + \Delta x$. The accumulation of fluid within the control volume due to motion in the x -direction is evidently

$$\Delta y \Delta z [(\rho u)(x, y, z) - (\rho u)(x + \Delta x, y, z)] = -\frac{\partial(\rho u)}{\partial x} \Big|_{x, y, z} \Delta x \Delta y \Delta z. \quad (\text{1.23})$$

To this must be added the effects of motion in the y - and z -directions, namely

$$-\left[\frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} \right] \Delta x \Delta y \Delta z. \quad (\text{1.24})$$

This net accumulation of fluid must be accompanied by a corresponding increase of fluid mass within the control volume. This is

$$\frac{\partial}{\partial t} (\text{density} \times \text{volume}) = \Delta x \Delta y \Delta z \frac{\partial \rho}{\partial t}, \quad (\text{1.25})$$

because the volume is constant. Thus, because mass is conserved, (1.23), (1.24) and (1.25) give

$$\Delta x \Delta y \Delta z \left[\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} \right] = 0. \quad (\text{1.26})$$

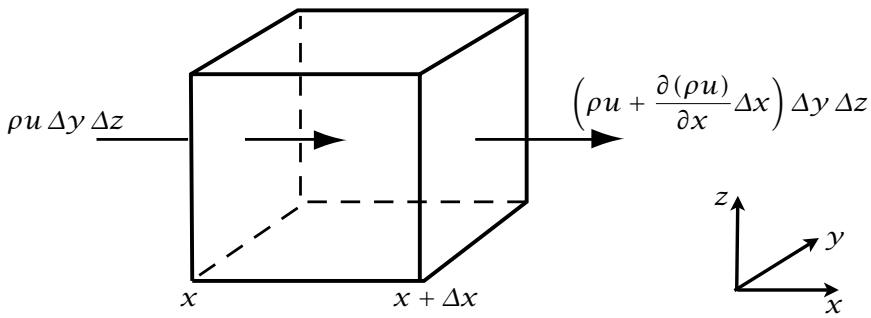


Fig. 1.1 Mass conservation in an Eulerian cuboid control volume.

Because the control volume is arbitrary the quantity in square brackets must be zero and we have the *mass continuity equation*:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \quad (1.27)$$

Vector derivation

Consider an arbitrary control volume V bounded by a surface S , fixed in space, with by convention the direction of \mathbf{S} being toward the outside of V , as in Fig. 1.2. The rate of fluid loss due to flow through the closed surface S is then given by

$$\text{fluid loss} = \int_S \rho \mathbf{v} \cdot d\mathbf{S} = \int_V \nabla \cdot (\rho \mathbf{v}) dV, \quad (1.28)$$

using the divergence theorem. This must be balanced by a change in the mass M of the fluid within the control volume, which, since its volume is fixed, implies a density change. That is

$$\text{fluid loss} = -\frac{dM}{dt} = -\frac{d}{dt} \int_V \rho dV = -\int_V \frac{\partial \rho}{\partial t} dV. \quad (1.29)$$

Equating (1.28) and (1.29) yields

$$\int_V \left[\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) \right] dV = 0. \quad (1.30)$$

Because the volume is arbitrary, the integrand must vanish and we recover (1.27).

1.2.2 Mass continuity via the material derivative

We now derive the mass continuity equation (1.27) from a material perspective. This is the most fundamental approach of all since the principle of mass conservation states simply that the mass of a given element of fluid is, by definition of the element, constant. Thus, consider a small mass of fluid of density ρ and volume ΔV . Then conservation of mass may be represented by

$$\frac{D}{Dt}(\rho \Delta V) = 0. \quad (1.31)$$

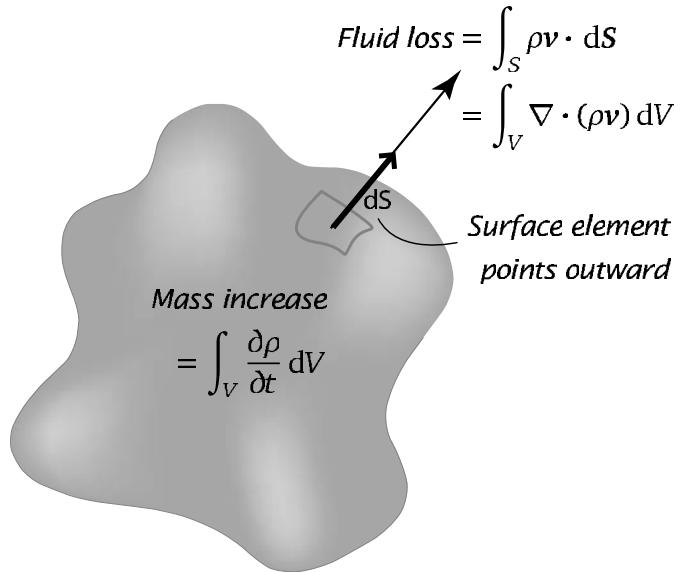


Fig. 1.2 Mass conservation in an arbitrary Eulerian control volume V bounded by a surface S . The mass gain, $\int_V (\partial \rho / \partial t) dV$ is equal to the mass flowing into the volume, $-\int_S (\rho \mathbf{v}) \cdot d\mathbf{S} = -\int_V \nabla \cdot (\rho \mathbf{v}) dV$.

Both the density and the volume of the parcel may change, so

$$\Delta V \frac{D\rho}{Dt} + \rho \frac{D\Delta V}{Dt} = \Delta V \left(\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} \right) = 0, \quad (1.32)$$

where the second expression follows using (1.15). Since the volume element is arbitrary, the term in brackets must vanish and

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0. \quad (1.33)$$

After expansion of the first term this becomes identical to (1.27). This result may be derived more formally by rewriting (1.31) as the integral expression

$$\frac{D}{Dt} \int_V \rho dV = 0. \quad (1.34)$$

Expanding the derivative using (1.18) gives

$$\frac{D}{Dt} \int_V \rho dV = \int_V \left(\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} \right) dV = 0. \quad (1.35)$$

Because the volume over which the integral is taken is arbitrary the integrand itself must vanish and we recover (1.33). Summarizing, equivalent partial differential equation representing conservation of mass are:

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0, \quad \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0$$

(1.36a,b)

1.2.3 A general continuity equation

The derivation of continuity equation for a general scalar property of a fluid is similar to that for density, except that there may be an external source or sink, and potentially a means of transferring the property from one location to another than by fluid motion, for example by diffusion. If ξ is the amount of some property of the fluid per unit volume (which we will call the *concentration* of the property), and if the net effect per unit volume of all non-conservative processes is denoted by $Q_v[\xi]$, then the continuity equation for concentration may be written:

$$\frac{D}{Dt}(\xi \Delta V) = Q_v[\xi] \Delta V. \quad (1.37)$$

Expanding the left-hand side and using (1.15) we obtain

$$\frac{D\xi}{Dt} + \xi \nabla \cdot \mathbf{v} = Q_v[\xi], \quad (1.38)$$

or equivalently

$$\frac{\partial \xi}{\partial t} + \nabla \cdot (\xi \mathbf{v}) = Q_v[\xi]. \quad (1.39)$$

If we are interested in a tracer that is normally measured per unit mass of fluid (which is typical when considering thermodynamic quantities) then the conservation equation would be written

$$\frac{D}{Dt}(\varphi \rho \Delta V) = Q_m[\varphi] \rho \Delta V, \quad (1.40)$$

where φ is the tracer *mixing ratio* — that is, the amount of tracer per unit fluid mass — and $Q_m[\varphi]$ represents non-conservative sources per unit mass. Then, since $\rho \Delta V$ is constant we obtain

$$\frac{D\varphi}{Dt} = Q_m[\varphi]. \quad (1.41)$$

The source term $Q_m[\varphi]$ is evidently equal to the rate of change of φ of a fluid element. When this is so, we write it simply as $\dot{\varphi}$, so that

$$\frac{D\varphi}{Dt} = \dot{\varphi}. \quad (1.42)$$

A tracer obeying (1.42) with $\dot{\varphi} = 0$ is said to be *materially conserved*. If a tracer is materially conserved except for the effects of non-conservative sources then it is sometimes said to be ‘semi-materially conserved’ or ‘adiabatically conserved’.

1.3 THE MOMENTUM EQUATION

The momentum equation is a partial differential equation that describes how the velocity or momentum of a fluid responds to internal and imposed forces. We will derive it using material methods and informally deducing the terms representing the pressure, gravitational and viscous forces.

1.3.1 Advection

Let $\mathbf{m}(x, y, z, t)$ be the momentum-density field (momentum per unit volume) of the fluid. Thus, $\mathbf{m} = \rho \mathbf{v}$ and the total momentum of a volume of fluid is given by the volume integral $\int_V \mathbf{m} dV$. Now, for a fluid the rate of change of a momentum of an identifiable fluid mass is given by the material derivative, and by Newton's second law this is equal to the force acting on it. Thus,

$$\frac{D}{Dt} \int_V \rho \mathbf{v} dV = \int_V \mathbf{F} dV, \quad (1.43)$$

where \mathbf{F} is the force per unit volume. Now, using (1.21b) (with φ replaced by \mathbf{v}) to transform the left-hand side of (1.43), we obtain

$$\int_V \left(\rho \frac{D\mathbf{v}}{Dt} - \mathbf{F} \right) dV = 0. \quad (1.44)$$

Because the volume is arbitrary the integrand itself must vanish and we obtain

$$\rho \frac{D\mathbf{v}}{Dt} = \mathbf{F}, \quad \text{or} \quad \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = \frac{\mathbf{F}}{\rho}, \quad (1.45a,b)$$

having used (1.8) to expand the material derivative.

We have thus obtained an expression for how a fluid accelerates if subject to known forces. These forces are however not all external to the fluid itself; a stress arises from the direct contact between one fluid parcel and another, giving rise to pressure and viscous forces, sometimes referred to as *contact* forces. Because a complete treatment of these would be very lengthy, and is available elsewhere, we treat both of these very informally and intuitively.

1.3.2 The pressure force

Within or at the boundary of a fluid the pressure is the normal force per unit area due to the collective action of molecular motion. Thus

$$d\hat{\mathbf{F}}_p = -p d\mathbf{S}, \quad (1.46)$$

where p is the pressure, $\hat{\mathbf{F}}_p$ is the pressure force and $d\mathbf{S}$ an infinitesimal surface element. If we grant ourselves this intuitive notion, it is a simple matter to assess the influence of pressure on a fluid, for the pressure force on a volume of fluid is the integral of the pressure over its boundary and so

$$\hat{\mathbf{F}}_p = - \int_S p d\mathbf{S}. \quad (1.47)$$

The minus sign arises because the pressure force is directed inwards, whereas \mathbf{S} is a vector normal to the surface and directed outwards. Applying a form of the divergence theorem to the right-hand side gives

$$\hat{\mathbf{F}}_p = - \int_V \nabla p dV, \quad (1.48)$$

where the volume V is bounded by the surface S . The pressure force per unit volume, \mathbf{F}_p , is therefore just $-\nabla p$, and inserting this into (1.45b) we obtain

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{\nabla p}{\rho} + \mathbf{F}', \quad (1.49)$$

where \mathbf{F}' , equal to $(\mathbf{F} - \mathbf{F}_p)/\rho$, represents viscous and body forces, per unit mass.

	μ (kg m ⁻¹ s ⁻¹)	ν (m ² s ⁻¹)
Air	1.8×10^{-5}	1.5×10^{-5}
Water	1.1×10^{-3}	1.1×10^{-6}
Mercury	1.6×10^{-3}	1.2×10^{-7}

Table 1.1 Experimental values of viscosity for air, water and mercury at room temperature and pressure.

1.3.3 Viscosity and diffusion

Viscosity, like pressure, is a force due to the internal motion of molecules. The effects of viscosity are apparent in many situations — the flow of treacle or volcanic lava are obvious examples. In other situations, for example large-scale flow the atmosphere, viscosity is to a first approximation negligible. However, for a constant density fluid viscosity is the *only* way that energy may be removed from the fluid, so that if energy is being added in some way viscosity must ultimately become important if the fluid is to reach an equilibrium where energy input equals energy dissipation. When tea is stirred in a cup, it is viscous effects that cause the fluid to eventually stop spinning after we have removed our spoon.

A number of textbooks² show that, for most Newtonian fluids, the viscous force per unit volume is approximately equal to $\mu \nabla^2 \mathbf{v}$, where μ is the viscosity. Although not exact, this is an extremely good approximation for most liquids and gases. With this term, the momentum equation becomes,

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{v} + \mathbf{F}_b \quad (1.50)$$

where $\nu \equiv \mu/\rho$ is the *kinematic viscosity*, and \mathbf{F}_b represents any body forces (per unit mass) such as gravity, \mathbf{g} . For gases, dimensional arguments suggest that the magnitude of ν should be given by

$$\nu \sim \text{mean free path} \times \text{mean molecular velocity}, \quad (1.51)$$

which for a typical molecular velocity of 300 m s^{-1} and a mean free path of $7 \times 10^{-8} \text{ m}$ gives the not unreasonable estimate of $2.1 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}$, within a factor of two of the experimental value (Table 1.1). Interestingly, the kinematic viscosity is less for water and mercury than it is for air.

1.3.4 Hydrostatic balance

The vertical component — the component parallel to the gravitational force, \mathbf{g} — of the momentum equation is

$$\frac{Dw}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g, \quad (1.52)$$

where w is the vertical component of the velocity and $\mathbf{g} = -g\mathbf{k}$. If the fluid is static the gravitational term is balanced by the pressure term and we have

$$\frac{\partial p}{\partial z} = -\rho g, \quad (1.53)$$

and this relation is known as *hydrostatic balance*, or hydrostasy. It is clear in this case that the pressure at a point is given by the weight of the fluid above it, provided $p = 0$ at the top of the fluid. It might also appear that (1.53) would be a good *approximation* to (1.52) provided vertical accelerations, Dw/Dt , are small compared to gravity, which is nearly always the case in the atmosphere and ocean. While this statement is true if we need only a reasonable approximate value of the pressure at a point or in a column, the satisfaction of this condition is *not* sufficient to ensure that (1.53) provides an accurate enough pressure to determine the horizontal pressure gradients responsible for producing motion. We return to this point in section 2.7.

1.4 THE EQUATION OF STATE

In three dimensions the momentum and continuity equations provide four equations, but contain five unknowns — three components of velocity, density and pressure. Obviously other equations are needed, and an *equation of state* is an expression that diagnostically relates the various thermodynamic variables to each other. The *conventional* equation of state (also called the thermal equation of state) is an expression that relates temperature, pressure, composition (the mass fraction of the various constituents) and density, and we may write it, rather generally, as

$$p = p(\rho, T, \mu_n), \quad (1.54)$$

where μ_n is mass fraction of the n th constituent. An equation of this form is not the most fundamental equation of state from a thermodynamic perspective, an issue we visit later, but it connects readily measurable quantities.

For an ideal gas (and the air in the Earth's atmosphere is very close to ideal) the conventional equation of state is

$$p = \rho RT, \quad (1.55)$$

where R is the gas constant for the gas in question and T is temperature. (R is related to the universal gas constant R_u by $R = R_u/m$, where m is the mean molecular weight of the constituents of the gas. Also, $R = nk$, where k is Boltzmann's constant and n is the number of molecules per unit mass.) For dry air, $R = 287 \text{ J kg}^{-1} \text{ K}^{-1}$. Air has virtually constant composition except for variations in water vapour content. A measure of this is the water vapour mixing ratio, $w = \rho_w/\rho_d$ where ρ_w and ρ_d are the densities of water vapour and dry air, respectively, and in the atmosphere w varies between 0 and 0.03. This variation makes the gas constant in the equation of state a weak function of the water vapour mixing ratio; that is, $p = \rho R_{\text{eff}} T$ where $R_{\text{eff}} = R_d(1 + wR_v/R_d)/(1 + w)$ where R_d and R_v are the gas constants of dry air and water vapour. Since $w \sim 0.01$ the variation of R_{eff} is quite small and is often ignored, especially in theoretical studies.³

For a liquid such as seawater no simple expression akin to (1.55) is easily derivable, and semi-empirical equations are usually resorted to. For pure water in a laboratory setting a reasonable approximation of the equation of state is $\rho = \rho_0[1 - \beta_T(T - T_0)]$, where β_T is a thermal expansion coefficient and ρ_0 and T_0 are constants. In the ocean the density is also significantly affected by pressure and dissolved salts: seawater is a solution of many ions in water — chloride ($\approx 1.9\%$ by weight) sodium (1%), sulfate (0.26%), magnesium (0.13%) and so on, with a total average concentration of about 35‰ (ppt, or parts per thousand). The ratios of the fractions of these salts are more-or-less constant throughout the ocean, and

their total concentration may be parameterized by a single measure, the *salinity*, S . Given this, the density of seawater is a function of three variables — pressure, temperature, and salinity — and we may write the conventional equation of state as

$$\alpha = \alpha(T, S, p), \quad (1.56)$$

where $\alpha = 1/\rho$ is the specific volume, or inverse density. For small variations around a reference value we have

$$d\alpha = \left(\frac{\partial \alpha}{\partial T}\right)_{S,p} dT + \left(\frac{\partial \alpha}{\partial S}\right)_{T,p} dS + \left(\frac{\partial \alpha}{\partial p}\right)_{T,S} dp = \alpha(\beta_T dT - \beta_S dS - \beta_p dp), \quad (1.57)$$

where the rightmost expression serves to define the thermal expansion coefficient β_T , the saline contraction coefficient β_S , and the compressibility coefficient (or inverse bulk modulus) β_p . In general these quantities are not constants, but for small variations around a reference state they may be treated as such and we have

$$\alpha = \alpha_0 \left[1 + \beta_T(T - T_0) - \beta_S(S - S_0) - \beta_p(p - p_0) \right]. \quad (1.58)$$

Typical values of these parameters, with variations typically encountered through the ocean, are: $\beta_T \approx 2(\pm 1.5) \times 10^{-4} \text{ K}^{-1}$ (values increase with both temperature and pressure), $\beta_S \approx 7.6(\pm 0.2) \times 10^{-4} \text{ ppt}^{-1}$, $\beta_p \approx 4.1(\pm 0.5) \times 10^{-10} \text{ Pa}^{-1}$. Since the variations around the mean density are small (1.58) implies that

$$\rho = \rho_0 \left[1 - \beta_T(T - T_0) + \beta_S(S - S_0) + \beta_p(p - p_0) \right]. \quad (1.59)$$

A linear equation of state for seawater is emphatically *not* accurate enough for quantitative oceanography; the β parameters in (1.58) themselves vary with pressure, temperature and (more weakly) salinity so introducing nonlinearities to the equation. The most important of these are captured by an equation of state of the form

$$\alpha = \alpha_0 \left[1 + \beta_T(1 + \gamma^* p)(T - T_0) + \frac{\beta_T^*}{2}(T - T_0)^2 - \beta_S(S - S_0) - \beta_p(p - p_0) \right]. \quad (1.60)$$

The starred constants β_T^* and γ^* capture the leading nonlinearities: γ^* is the *thermobaric parameter*, which determines the extent to which the thermal expansion depends on pressure, and β_T^* is the second thermal expansion coefficient. Even this equation of state has quantitative deficiencies and more complicated semi-empirical formulae are often used if high accuracy is needed.⁴ An illustration of the variation of density of seawater with temperature, salinity and pressure is illustrated in Fig. 1.3, which uses an accurate, semi-empirical equation of state. More discussion is to be found in section 1.7.2.

Clearly, the equation of state introduces, in general, a sixth unknown, temperature, and we will have to introduce another physical principle — the first law of thermodynamics or the principle of energy conservation — to obtain a complete set of equations. However, if the equation of state were such that it linked only density and pressure, without introducing another variable, then the equations would be complete; the simplest case of all is a constant density fluid for which the equation of state is just $\rho = \text{constant}$. A fluid for which the density is a function of pressure alone is called a *barotropic fluid*; otherwise, it is a *baroclinic fluid*. (In this context, ‘barotropic’ is a shortening of the original phrase ‘auto-barotropic’.) Equations of state of the form $p = C\rho^\gamma$, where γ is a constant, are sometimes called ‘polytropic’.

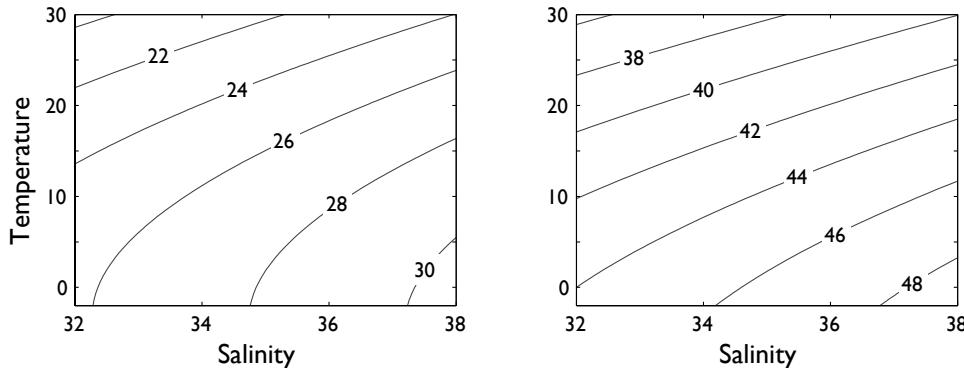


Fig. 1.3 A temperature–salinity diagram for seawater, calculated using an accurate empirical equation of state. Contours are $(\text{density} - 1000) \text{ kg m}^{-3}$, and the temperature is potential temperature, which in the deep ocean may be less than *in situ* temperature by a degree or so (see Fig. 1.4). Left panel: at sea-level ($p = 10^5 \text{ Pa} = 1000 \text{ mb}$). Right panel: at $p = 4 \times 10^7 \text{ Pa}$, a depth of about 4 km. Note that in both cases the contours are slightly convex.

1.5 THERMODYNAMIC RELATIONS

In this section we review a few aspects of thermodynamics. We provide neither a complete nor an *a priori* development of the subject; rather, we focus on aspects that are particularly relevant to fluid dynamics, and that are needed to derive a ‘thermodynamic equation’ for fluids.⁵ Readers whose interest is solely in an ideal gas or a simple Boussinesq fluid may skim this section, and then refer back to it as needed.

1.5.1 A few fundamentals

A fundamental postulate of thermodynamics is that the internal energy of a system in equilibrium is a function of its extensive properties volume, entropy, and the mass of its various constituents. (Extensive means that the property value is proportional to the amount of material present, as opposed to an intensive property such as temperature.) For our purposes it is more convenient to divide all of these quantities by the mass of fluid present, so expressing the internal energy per unit mass, I , as a function of the specific volume $\alpha = \rho^{-1}$, the specific entropy η , and the mass fractions of its various components. Our interest is in two-component fluids (dry air and water vapour, or water and salinity) so that we may parameterize the composition by a single parameter, S . Thus we have

$$I = I(\alpha, \eta, S) , \quad (1.61a)$$

or an equivalent equation for entropy,

$$\eta = \eta(I, \alpha, S) . \quad (1.61b)$$

Given the functional forms on the right-hand sides, either of these expressions constitutes a complete description of the macroscopic state of a system in equilibrium, and we call them

the *fundamental equation of state*. The conventional equation of state can be derived from (1.61), but not vice versa. The first differential of (1.61a) gives, formally,

$$dI = \left(\frac{\partial I}{\partial \alpha}\right)_{\eta, S} d\alpha + \left(\frac{\partial I}{\partial \eta}\right)_{\alpha, S} d\eta + \left(\frac{\partial I}{\partial S}\right)_{\alpha, \eta} dS. \quad (1.62)$$

We will now ascribe physical meaning to these differentials.

Conservation of energy states that the internal energy of a body may change because of work done by or on it, or because of a heat input, or because of a change in its chemical composition. We write this as

$$dI = \delta Q - \delta W + \delta C, \quad (1.63)$$

where δW is the work done *by* the body, δQ is the heat input *to* the body, and δC accounts for the change in internal energy caused by a change in its chemical composition (e.g., its salinity, or water vapour content), sometimes called the ‘chemical work’. The quantities on the right-hand side are imperfect differentials or infinitesimals: Q , W and C are not functions of the state of a body, and the internal energy cannot be regarded as the sum of a ‘heat’ and a ‘work’. We should think of heat and work as having meaning only as fluxes of energy, or rates of energy input, and not as amounts of energy; their sum changes the internal energy of a body, which *is* a function of its state. Equation (1.63) is sometimes called ‘the first law of thermodynamics’. Let us consider the causes of variations of the quantities on the right-hand side.

Heat Input: Thermodynamics provides a relationship between the heat input and the change in the entropy of a body, namely that in an (infinitesimal) quasi-static or reversible process, with constant composition,

$$T d\eta = \delta Q, \quad (1.64)$$

where η is the specific entropy of the body. The entropy is a function of the state of a body and is, by definition, an adiabatic invariant. As we are dealing with the amount of a quantity per unit mass, η is the specific entropy, although we will often refer to it just as the entropy. We may regard (1.63) as defining the heat input, δQ , by way of a statement of conservation of energy, and (1.64) then says that there is a function of state, the entropy, that changes by an amount equal to the heat input divided by the temperature.⁶

Work done: The work done by a body is equal to the pressure times the change in its volume. Thus, per unit mass, we have

$$\delta W = p d\alpha, \quad (1.65)$$

where $\alpha = 1/\rho$ is the specific volume of the fluid and p is the pressure.

Composition: The change in internal energy due to changes in composition is given by

$$\delta C = \mu dS, \quad (1.66)$$

where μ is the *chemical potential* of the solution. In the ocean compositional changes

(i.e., changes in salinity) arise through precipitation and evaporation at the surface, and molecular diffusion. When salinity does so change, the internal energy of a fluid parcel changes by (1.66), but in practice this change is usually small compared to other changes in internal energy. The most important effect of salinity is that it changes the density of seawater. In the atmosphere the composition of a parcel of air primarily mainly varies according to the amount of water vapour and liquid water in it; these variations cause corresponding changes in internal energy, but in the absence of phase-changes the internal energy variations are slight. The most important effect of water vapour is that when condensation or evaporation occurs, heat is released (or required) that provides an entropy source in (1.64).

Collecting equations (1.63) – (1.66) together we have

$$\boxed{dI = T d\eta - p d\alpha + \mu dS} \quad (1.67)$$

We refer to this as *the fundamental thermodynamic relation*. The fundamental equation of state, (1.61), describes the properties of a particular fluid, and the fundamental relation, (1.67), is a statement of the conservation of energy. Much of classical thermodynamics follows from these two expressions.

1.5.2 Various thermodynamic relations

From (1.62) and (1.67) it follows that

$$T = \left(\frac{\partial I}{\partial \eta} \right)_{\alpha,S}, \quad p = - \left(\frac{\partial I}{\partial \alpha} \right)_{\eta,S}, \quad \mu = \left(\frac{\partial I}{\partial S} \right)_{\eta,\alpha}. \quad (1.68a,b,c)$$

These may be regarded as the defining relations for these variables; because of the connection between (1.63) and (1.67) these are not just formal definitions, and the pressure and temperature so defined are indeed related to our intuitive concepts of these variables and to the motion of the fluid molecules. Note that if we write

$$d\eta = \frac{1}{T} dI + \frac{p}{T} d\alpha - \frac{\mu}{T} dS, \quad (1.69)$$

it is also clear that

$$p = T \left(\frac{\partial \eta}{\partial \alpha} \right)_{I,S}, \quad T^{-1} = \left(\frac{\partial \eta}{\partial I} \right)_{\alpha,S}, \quad \mu = -T \left(\frac{\partial \eta}{\partial S} \right)_{I,\alpha}. \quad (1.70a,b,c)$$

In the derivations below, we will, unless noted, suppose that the composition of a fluid parcel is fixed, and drop the suffix S on partial derivatives unless ambiguity might arise.

Because the right-hand side of (1.67) is equal to an exact differential, the second derivatives are independent of the order of differentiation. That is,

$$\frac{\partial^2 I}{\partial \eta \partial \alpha} = \frac{\partial^2 I}{\partial \alpha \partial \eta} \quad (1.71)$$

and therefore, using (1.68)

$$\left(\frac{\partial T}{\partial \alpha} \right)_n = - \left(\frac{\partial p}{\partial \eta} \right)_\alpha. \quad (1.72)$$

This is one of the *Maxwell relations*, which are a collection of four similar relations that follow directly from the fundamental thermodynamic relation (1.67) and simple relations between second derivatives. A couple of others will be useful.

Define the *enthalpy* of a fluid by

$$h \equiv I + p\alpha \quad (1.73)$$

then, for a parcel of constant composition, (1.67) becomes

$$dh = T d\eta + \alpha dp. \quad (1.74)$$

But h is a function only of η and p so that in general

$$dh = \left(\frac{\partial h}{\partial \eta} \right)_p d\eta + \left(\frac{\partial h}{\partial p} \right)_\eta dp. \quad (1.75)$$

Comparing the last two equations we have

$$T = \left(\frac{\partial h}{\partial \eta} \right)_p \quad \text{and} \quad \alpha = \left(\frac{\partial h}{\partial p} \right)_\eta. \quad (1.76)$$

Noting that

$$\frac{\partial^2 h}{\partial \eta \partial p} = \frac{\partial^2 h}{\partial p \partial \eta} \quad (1.77)$$

we evidently must have

$$\left(\frac{\partial T}{\partial p} \right)_\eta = \left(\frac{\partial \alpha}{\partial \eta} \right)_p, \quad (1.78)$$

and this is our second Maxwell relation.

To obtain the third, we write

$$dI = T d\eta - p d\alpha = d(T\eta) - \eta dT - d(p\alpha) + \alpha dp, \quad (1.79)$$

or

$$dG = -\eta dT + \alpha dp, \quad (1.80)$$

where $G \equiv I - T\eta + p\alpha$ is the *Gibbs function* (or ‘Gibbs free energy’ or ‘Gibbs potential’). Now, formally, we have

$$dG = \left(\frac{\partial G}{\partial T} \right)_p dT + \left(\frac{\partial G}{\partial p} \right)_T dp. \quad (1.81)$$

Comparing the last two equations we see that $\eta = -(\partial G / \partial T)_p$ and $\alpha = (\partial G / \partial p)_T$. Furthermore, because

$$\frac{\partial^2 G}{\partial p \partial T} = \frac{\partial^2 G}{\partial T \partial p} \quad (1.82)$$

we have our third Maxwell equation,

$$\left(\frac{\partial \eta}{\partial p} \right)_T = - \left(\frac{\partial \alpha}{\partial T} \right)_p. \quad (1.83)$$

Maxwell's Relations

The four Maxwell equations are:

$$\begin{aligned} \left(\frac{\partial T}{\partial \alpha}\right)_\eta &= - \left(\frac{\partial p}{\partial \eta}\right)_\alpha, & \left(\frac{\partial T}{\partial p}\right)_\eta &= \left(\frac{\partial \alpha}{\partial \eta}\right)_p, \\ \left(\frac{\partial \eta}{\partial p}\right)_T &= - \left(\frac{\partial \alpha}{\partial T}\right)_p, & \left(\frac{\partial \eta}{\partial \alpha}\right)_T &= \left(\frac{\partial p}{\partial T}\right)_\alpha. \end{aligned} \quad (\text{M.1})$$

These imply:

$$\frac{\partial(T, \eta)}{\partial(p, \alpha)} \equiv \left(\frac{\partial T}{\partial p}\right) \left(\frac{\partial \eta}{\partial \alpha}\right) - \left(\frac{\partial T}{\partial \alpha}\right) \left(\frac{\partial \eta}{\partial p}\right) = 0. \quad (\text{M.2})$$

The fourth Maxwell equation, whose derivation is left to the reader (make use of the Helmholtz free energy, $F \equiv I - T\eta$), is

$$\left(\frac{\partial \eta}{\partial \alpha}\right)_T = \left(\frac{\partial p}{\partial T}\right)_\alpha, \quad (1.84)$$

and all four Maxwell equations are summarized in the box above. All of them follow from the fundamental thermodynamic relation, (1.67), which is the real silver hammer of thermodynamics.

Fundamental equation of state

The fundamental equation of state (1.61) gives complete information about a fluid in thermodynamic equilibrium, and given this we can obtain expressions for the temperature, pressure and chemical potential using (1.68). These are also equations of state; however, each of them, taken individually, contains less information than the fundamental equation because a derivative has been taken. Equivalent to the fundamental equation of state are, using (1.74), an expression for the enthalpy as a function of its natural variables pressure, entropy and composition, or, using (1.80) the Gibbs function as a function of pressure, temperature and composition. Of these, the Gibbs function is the most practically useful because the pressure, temperature and composition may all be measured in the laboratory. Given the fundamental equation of state, the thermodynamic state of a body is fully specified by a knowledge of any two of p, ρ, T, η and I , plus its composition. The conventional equation of state, (1.54), is obtained by using (1.61a) to eliminate entropy from (1.68a) and (1.68b).

One simple fundamental equation of state is to take the internal energy to be a function of density and not entropy; that is, $I = I(\alpha)$. Bodies with such a property are called *homentropic*. Using (1.68), temperature and chemical potential have no role in the fluid dynamics and the density is a function of pressure alone — the defining property of a barotropic fluid. Neither water nor air are, in general, homentropic but under some circumstances the flow may be adiabatic and $p = p(\rho)$ (problem 1.10).

In an ideal gas the molecules do not interact except by elastic collisions, and the volume of the molecules is presumed to be negligible compared to the total volume they occupy.

The internal energy of the gas then depends only on temperature, and not on density. A *simple* ideal gas is an ideal gas for which the heat capacity is constant, so that

$$I = cT, \quad (1.85)$$

where c is a constant. Using this and the conventional ideal gas equation, $p = \rho RT$ (where R is also constant), along with the fundamental thermodynamic relation (1.67), we can infer the fundamental equation of state; however, we will defer that until we discuss potential temperature in section 1.6.1. A *general* ideal gas also obeys $p = \rho RT$, but it has heat capacities that may be a function of temperature (but only of temperature — see problem 1.12).

Internal energy and specific heats

We can obtain some useful relations between the internal energy and specific heat capacities, and some useful estimates of their values, by some simple manipulations of the fundamental thermodynamic relation. Assuming that the composition of the fluid is constant (1.67) is

$$T d\eta = dI + p d\alpha, \quad (1.86)$$

so that, taking I to be a function of α and T ,

$$T d\eta = \left(\frac{\partial I}{\partial T} \right)_\alpha dT + \left[\left(\frac{\partial I}{\partial \alpha} \right)_T + p \right] d\alpha. \quad (1.87)$$

From this, we see that the heat capacity at constant volume (i.e., constant α) c_v is given by

$$c_v \equiv T \left(\frac{\partial \eta}{\partial T} \right)_\alpha = \left(\frac{\partial I}{\partial T} \right)_\alpha. \quad (1.88)$$

Thus, c in (1.85) is equal to c_v .

Similarly, using (1.74) we have

$$T d\eta = dh - \alpha dp = \left(\frac{\partial h}{\partial T} \right)_p dT + \left[\left(\frac{\partial h}{\partial p} \right) - \alpha \right] dp. \quad (1.89)$$

The heat capacity at constant pressure, c_p , is then given by

$$c_p \equiv T \left(\frac{\partial \eta}{\partial T} \right)_p = \left(\frac{\partial h}{\partial T} \right)_p. \quad (1.90)$$

For later use, we define the ratios $\gamma \equiv c_p/c_v$ and $\kappa \equiv R/c_p$.

For an ideal gas $h = I + RT = T(c_v + R)$. But $c_p = (\partial h/\partial T)_p$, and hence $c_p = c_v + R$, and $(\gamma - 1)/\gamma = \kappa$. Statistical mechanics tells us that for a simple ideal gas the internal energy is equal to $kT/2$ per molecule, or $RT/2$ per unit mass, for each excited degree of freedom, where k is the Boltzmann constant and R the gas constant. The diatomic molecules N_2 and O_2 that comprise most of our atmosphere have two rotational and three translational degrees of freedom, so that $I \approx 5RT/2$, and so $c_v \approx 5R/2$ and $c_p \approx 7R/2$, both being constants. These are in fact very good approximations to the measured values for the Earth's atmosphere, and give $c_p \approx 10^3 \text{ J kg}^{-1} \text{ K}^{-1}$. The internal energy is simply $c_v T$ and the enthalpy is $c_p T$. For a liquid, especially one like seawater that contains dissolved salts, no such simple relations are possible: the heat capacities are functions of the state of the fluid, and the internal energy is a function of pressure (or density) as well as temperature.

1.6 THERMODYNAMIC EQUATIONS FOR FLUIDS

The thermodynamic relations — for example (1.67) — apply to identifiable bodies or systems; thus, the heat input affects the fluid parcel to which it is applied, and we can apply the material derivative to the above thermodynamic relations to obtain equations of motion for a moving fluid. But in doing so we make two assumptions.

- (i) That locally the fluid is in thermodynamic equilibrium. This means that, although the thermodynamic quantities like temperature, pressure and density vary in space and time, locally they are related by the thermodynamic relations such as the equation of state and Maxwell's relations.
- (ii) That macroscopic fluid motions are reversible and so not entropy producing. Thus, such effects as the viscous dissipation of energy, radiation, and conduction may produce entropy whereas the macroscopic fluid motion itself does not.

The first point requires that the temperature variation on the macroscopic scales must be slow enough that there can exist a volume that is small compared to the scale of macroscopic variations, so that temperature is effectively constant within it, but that is also sufficiently large to contain enough molecules so that macroscopic variables such as temperature have a proper meaning.

Now from (1.63) conservation of energy for a infinitesimal fluid parcel may be written as

$$dI = -p d\alpha + dQ_T \quad (1.91)$$

where $p d\alpha$ is the work done by the parcel and dQ_T is the total energy input to the parcel, including the contributions from a thermal flux and from changes in composition. Given the first assumption above, we may form the material derivative of (1.91) to obtain

$$\frac{DI}{Dt} + p \frac{D\alpha}{Dt} = \dot{Q}_T \quad (1.92)$$

where \dot{Q}_T is the rate of total energy input, per unit mass, with (as for dQ_T) possible contributions from thermal fluxes (including radiative heating, thermal diffusion and heat generated by viscous damping) and from diffusive fluxes of composition. (In general, both the thermal flux and the composition flux depend on the gradients of both temperature and composition, although in many circumstances the thermal flux is approximately determined by the temperature gradient, and the compositional flux by the gradient of composition.) Using the mass continuity equation in the form $D\alpha/Dt = \alpha \nabla \cdot \mathbf{v}$, we write (1.92) as

$$\frac{DI}{Dt} + p \alpha \nabla \cdot \mathbf{v} = \dot{Q}_T \quad . \quad (1.93)$$

This is the *internal energy equation* for a fluid.

Because \dot{Q}_T contains, in general, energy fluxes due to changes in composition, we need to know that composition. The composition of a fluid parcel is a carried with it as it moves, and changes only if there non-conservative sources and sinks, such as diffusive fluxes. Thus, and analogously to (1.42), the evolution of composition is determined by

$$\frac{DS}{Dt} = \dot{S}, \quad (1.94)$$

where \dot{S} represents all the non-conservative terms.

Rather than use an internal energy equation, we may use the fundamental thermodynamic relation to infer an evolution equation for entropy. Thus, forming the material derivative from (1.67) and using (1.93) and (1.94), we obtain the *entropy equation*

$$\frac{D\eta}{Dt} = \frac{1}{T} \dot{Q}_T - \frac{\mu}{T} \dot{S} \equiv \frac{1}{T} \dot{Q}, \quad (1.95)$$

where \dot{Q} is the heating rate per unit mass. (This equation could have been equivalently derived, by forming the material derivative of (1.64), along with assumption (ii) above.) It is important to realize that the entropy equation is not independent of the internal energy equation, but may be derived from it. The two equations are equivalently connected via the equation of state: if we use the internal energy equation we can in principle then calculate the entropy using the equation of state in the form $\eta = \eta(I, \alpha, S)$, or if we use the entropy equation we can then calculate the internal energy using $I = I(\eta, \alpha, S)$. Indeed, the internal energy equation and the entropy equation are both commonly referred to as ‘the thermodynamic equation’. In multi-component fluids irreversible diffusive processes provide a source of entropy, and the use of the internal energy equation is often more straightforward than the use of an entropy equation.⁷

Given evolution equations for composition and internal energy or entropy, and given the fundamental equation of state, we have, in principle, a complete set of equations for a fluid, as summarized in the shaded box on the following page. In practice, this form of the equations is not the most useful because, especially for liquids, the equation of state cannot be written in a simple and useful form that allows pressure and temperature to be diagnosed. Happily, for an ideal gas the equation of state does allow such a diagnosis, as we now discover.

1.6.1 Thermodynamic equation for an ideal gas

For a dry ideal gas the internal energy is a function of temperature only and $dI = c_v dT$ (also see problems 1.12 and 1.14). The first law of thermodynamics becomes

$$dQ = c_v dT + p d\alpha \quad \text{or} \quad dQ = c_p dT - \alpha dp, \quad (1.96a,b)$$

where the second expression is derived using $\alpha = RT/p$ and $c_p - c_v = R$. Forming the material derivative of the above gives two forms of the internal energy equation:

$$c_v \frac{DT}{Dt} + p \frac{D\alpha}{Dt} = \dot{Q} \quad \text{or} \quad c_p \frac{DT}{Dt} - \frac{RT}{p} \frac{Dp}{Dt} = \dot{Q}. \quad (1.97a,b)$$

Using the mass continuity equation, (1.97a) may be written as

$$c_v \frac{DT}{Dt} + p \alpha \nabla \cdot \mathbf{v} = \dot{Q}. \quad (1.98)$$

Alternatively, using the ideal gas equation, we may eliminate T in favour of p and α and obtain

$$\frac{Dp}{Dt} + \gamma p \nabla \cdot \mathbf{v} = \dot{Q} \frac{\rho R}{c_v}. \quad (1.99)$$

Fundamental Equations of Motion of a Fluid

The following constitutes, in principle, a complete set of equations for an inviscid fluid heated at a rate \dot{Q} and whose composition, S , changes at a rate \dot{S} .

Evolution equations for velocity, density and composition:

$$\frac{D\mathbf{v}}{Dt} = -\frac{1}{\rho} \nabla p, \quad \frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0, \quad \frac{DS}{Dt} = \dot{S}. \quad (\text{F.1})$$

Internal energy equation or entropy equation:

$$\frac{DI}{Dt} - \frac{p}{\rho} \nabla \cdot \mathbf{v} = \dot{Q}_T, \quad \frac{D\eta}{Dt} = \frac{1}{T} \dot{Q}. \quad (\text{F.2})$$

where $\dot{Q}_T = \dot{Q} + \mu \dot{S}$ is the total rate of energy input.

Fundamental equation of state:

$$I = I(\rho, S, \eta). \quad (\text{F.3})$$

Diagnostic equations for temperature and pressure:

$$T = \left(\frac{\partial I}{\partial \eta} \right)_{\alpha, S}, \quad p = - \left(\frac{\partial I}{\partial \alpha} \right)_{\eta, S}. \quad (\text{F.4})$$

The Earth's atmosphere also contains water vapour with mixing ratio w , and an evolution equation for it takes the form

$$\frac{Dw}{Dt} = \dot{w}, \quad (1.100)$$

where \dot{w} represents the effects of condensation and evaporation. (An equation for liquid water may also be needed if cloudiness is to be modelled.) The main thermodynamic effect of water vapour occurs when water vapour condenses, or liquid water evaporates, and latent heat is released; this heating appears in thermodynamic equation, with $\dot{Q} = -L_c \dot{w}$ plus other effects, where L_c is the latent heat of condensation.

Potential temperature, potential density and entropy

We can use entropy instead of temperature for our thermodynamic equation, and this corresponds to using (1.95) instead of (1.93). This is effectively equivalent to deriving and using the fundamental equation of state for an ideal gas, although it turns out to be convenient to express entropy in terms of a temperature-like quantity. To proceed, we first note that when a fluid parcel changes pressure adiabatically, it will expand or contract and, using (1.96b), its temperature change is determined by

$$c_p dT = \alpha dp. \quad (1.101)$$

As this temperature change is not caused by diabatic effects (e.g., heating), it is useful to define a temperature-like quantity that changes *only* if diabatic effects are present. To this

end, we define the *potential temperature*, θ , to be the temperature that a fluid would have if moved adiabatically (and, in general, at constant composition) to some reference pressure (often taken to be 1000 hPa, which is close to the pressure at the Earth's surface). Thus, in adiabatic flow the potential temperature of a fluid parcel is conserved, essentially by definition, and

$$\frac{D\theta}{Dt} = 0. \quad (1.102)$$

For this equation to be useful we must be able to relate θ to the other thermodynamic variables. For an ideal gas we use (1.96b) and the equation of state to write the fundamental thermodynamic relation as

$$d\eta = c_p d\ln T - R d\ln p. \quad (1.103)$$

The definition of potential temperature then implies that

$$\int_T^\theta c_p d\ln T - \int_p^{p_R} R d\ln p = 0, \quad (1.104)$$

which, for constant c_p and R , may be solved to give

$$\theta = T \left(\frac{p_R}{p} \right)^\kappa \quad (1.105)$$

where p_R is the reference pressure and $\kappa = R/c_p$.

Note that as a consequence of its definition the potential temperature is related to entropy by

$$d\eta = c_p d\ln \theta \quad (1.106)$$

and, if c_p is constant,

$$\eta = c_p \ln \theta. \quad (1.107)$$

Equation (1.106) is in fact a general expression for the potential temperature of a fluid parcel (see section 1.7.1), but (1.107) applies only if c_p is constant, as it is, to a good approximation, in the Earth's atmosphere.

Forming the material derivative of (1.107) and using (1.95) we obtain

$c_p \frac{D\theta}{Dt} = \frac{\theta}{T} \dot{Q}, \quad (1.108)$

with θ given by (1.105). Equations (1.98), (1.99) and (1.108) are all equivalent forms of the thermodynamic equation for an ideal gas.

The *potential density*, ρ_θ , is the density that a fluid parcel would have if moved adiabatically and at constant composition to a reference pressure, p_R . If the equation of state is written as $\rho = f(p, T)$ then the potential density is just

$$\rho_\theta = f(p_R, \theta). \quad (1.109)$$

For an ideal gas we therefore have

$$\rho_\theta = \frac{p_R}{R\theta}; \quad (1.110)$$

that is, potential density is proportional to the inverse of potential temperature. We may also write (1.110) as

$$\rho_\theta = \rho \left(\frac{p_R}{p} \right)^{1/\gamma}. \quad (1.111)$$

Finally, for later use we note that for small variations around a reference state manipulation of the ideal gas equation gives

$$\frac{\delta \theta}{\theta} = \frac{\delta T}{T} - \kappa \frac{\delta p}{p} = \frac{1}{\gamma} \frac{\delta p}{p} - \frac{\delta \rho}{\rho}. \quad (1.112)$$

* The fundamental equation of state for an ideal gas

Equations (1.107) and (1.105) are closely related to the fundamental equation of state, and using $I = c_v T$ and the equation of state $p = \rho RT$, we can express the entropy explicitly in terms of the density and the internal energy, to wit

$$\boxed{\eta = c_v \ln I - R \ln \rho + \text{constant}}. \quad (1.113)$$

This is the fundamental equation of state for a simple ideal gas. If we were to begin with this, we could straightforwardly derive all the thermodynamic quantities of interest for a simple ideal gas: for example, using (1.70a) we immediately recover $P = \rho RT$, and from (1.70b) we obtain $I = c_v T$. Indeed, (1.113) could be used to define a simple ideal gas, but such an a priori definition may seem a little unmotivated. Of course the heat capacities must still be determined by experiment or by a kinetic theory — they are not given by the thermodynamics, and (1.113) holds only if they are constant.

1.6.2 * Thermodynamic equation for liquids

For a liquid such as seawater no simple exact equation of state exists. Thus, although (1.95) holds in general, an expression relating entropy to the other thermodynamic variables is still needed. For quantitative modelling and observational work such an equation of state must be accurate, but this entails a complicated, nonlinear expression that is, to most eyes, uninformative. On the other hand, for most theoretical work and for idealized models a simplified equation of state that captures the main qualitative effects will suffice. It is informative to proceed heuristically, and to begin with the entropy equation and make simplifications to it.

1. Entropy equation using pressure and density

If we regard η as a function of pressure and density (and salinity if appropriate) we obtain

$$\begin{aligned} T d\eta &= T \left(\frac{\partial \eta}{\partial \rho} \right)_{p,S} d\rho + T \left(\frac{\partial \eta}{\partial p} \right)_{\rho,S} dp + T \left(\frac{\partial \eta}{\partial S} \right)_{\rho,p} dS \\ &= T \left(\frac{\partial \eta}{\partial \rho} \right)_{p,S} d\rho - T \left(\frac{\partial \eta}{\partial \rho} \right)_{p,S} \left(\frac{\partial \rho}{\partial p} \right)_{\eta,S} dp + T \left(\frac{\partial \eta}{\partial S} \right)_{\rho,p} dS. \end{aligned} \quad (1.114)$$

From this, and using (1.95) and (1.94), we obtain for a moving fluid

$$T \left(\frac{\partial \eta}{\partial \rho} \right)_{p,S} \frac{D\rho}{Dt} - T \left(\frac{\partial \eta}{\partial \rho} \right)_{p,S} \left(\frac{\partial \rho}{\partial p} \right)_{\eta,S} \frac{Dp}{Dt} = \dot{Q} - T \left(\frac{\partial \eta}{\partial S} \right)_{\rho,p} \dot{S}. \quad (1.115)$$

But $(\partial p / \partial \rho)_{\eta, S} = c_s^2$ where c_s is the speed of sound (see section 1.8). This is a measurable quantity in a fluid, and often nearly constant, and so useful to keep in an equation. Then the thermodynamic equation may be written in the form

$$\boxed{\frac{D\rho}{Dt} - \frac{1}{c_s^2} \frac{Dp}{Dt} = Q[\rho]}, \quad (1.116)$$

where $Q[\rho] \equiv (\partial \rho / \partial \eta)_{p, S} \dot{Q} / T - (\partial \rho / \partial S)_{p, p} \dot{S}$ represents the effects of entropy and salinity source terms. This form of the thermodynamic equation is valid for both liquids and gases.

Approximations using pressure and density

The speed of sound in a fluid is related to its compressibility — the less compressible the fluid, the greater the sound speed (section 1.8). In a liquid, sound speed is often sufficiently high that the second term in (1.116) can be neglected, and the thermodynamic equation takes the simple form:

$$\frac{D\rho}{Dt} = Q[\rho]. \quad (1.117)$$

This equation is a very good approximation for many laboratory fluids. Note that this equation is a thermodynamic equation, arising from the principle of conservation of energy for a liquid. It is a very different equation from the mass conservation equation, which for compressible fluids is also an evolution equation for density.

In the ocean the enormous pressures resulting from columns of seawater kilometres deep mean that although the second term in (1.116) may be small, it is not negligible, and a better approximation results if we suppose that the pressure is given by the weight of the fluid above it — the hydrostatic approximation. In this case $dp = -\rho g dz$ and (1.116) becomes

$$\frac{D\rho}{Dt} + \frac{\rho g}{c_s^2} \frac{Dz}{Dt} = Q[\rho]. \quad (1.118)$$

In the second term the height field varies much more than the density field, so a good approximation is to replace ρ by a constant, ρ_0 , in this term only. Taking the speed of sound also to be constant gives

$$\frac{D}{Dt} \left(\rho + \frac{\rho_0 z}{H_\rho} \right) = Q[\rho], \quad (1.119)$$

where

$$H_\rho = c_s^2 / g \quad (1.120)$$

is the *density scale height* of the ocean. In water, $c_s \approx 1500 \text{ m s}^{-1}$ so that $H_\rho \approx 200 \text{ km}$. The quantity in brackets on the left-hand side of (1.119) is (in this approximation) the *potential density*, this being the density that a parcel would have if moved adiabatically and with constant composition to the reference height $z = 0$. The adiabatic lapse rate of density is the rate at which the density of a parcel changes when undergoing an adiabatic displacement. From (1.119) it is approximately

$$-\left(\frac{\partial \rho}{\partial z} \right)_\eta \approx \frac{\rho_0 g}{c_s^2} \approx 5 \text{ (kg m}^{-3}\text{) / km}, \quad (1.121)$$

so that if a parcel is moved adiabatically from the surface to the deep ocean (5 km depth, say) its density will increase by about 25 kg m^{-3} , a fractional change of about 1/40 or 2.5%.

II. Entropy equation using pressure and temperature

Taking entropy to be a function of pressure and temperature (and salinity if appropriate) we have

$$\begin{aligned} T d\eta &= T \left(\frac{\partial \eta}{\partial T} \right)_{p,S} dT + T \left(\frac{\partial \eta}{\partial p} \right)_{T,S} dp + T \left(\frac{\partial \eta}{\partial S} \right)_{T,p} dS \\ &= c_p dT + T \left(\frac{\partial \eta}{\partial p} \right)_{T,S} dp + T \left(\frac{\partial \eta}{\partial S} \right)_{T,p} dS. \end{aligned} \quad (1.122)$$

For a moving fluid, and using (1.95) and (1.94), this implies,

$$\frac{DT}{Dt} + \frac{T}{c_p} \left(\frac{\partial \eta}{\partial p} \right)_{T,S} \frac{Dp}{Dt} = Q[T], \quad (1.123)$$

where $Q[T] \equiv \dot{Q}/c_p - T c_p^{-1} \dot{S}(\partial \eta / \partial S)$. Now substitute the Maxwell relation (1.83) in the form

$$\left(\frac{\partial \eta}{\partial p} \right)_T = \frac{1}{\rho^2} \left(\frac{\partial \rho}{\partial T} \right)_p \quad (1.124)$$

to give

$$\frac{DT}{Dt} + \frac{T}{c_p \rho^2} \left(\frac{\partial \rho}{\partial T} \right)_p \frac{Dp}{Dt} = Q[T], \quad (1.125a)$$

or, equivalently,

$$\frac{DT}{Dt} - \frac{T}{c_p} \left(\frac{\partial \alpha}{\partial T} \right)_p \frac{Dp}{Dt} = Q[T]. \quad (1.125b)$$

The density and temperature are related through a measurable coefficient of thermal expansion β_T where

$$\left(\frac{\partial \rho}{\partial T} \right)_p = -\beta_T \rho. \quad (1.126)$$

Equation (1.125) then becomes

$$\frac{DT}{Dt} - \frac{\beta_T T}{c_p \rho} \frac{Dp}{Dt} = Q[T]. \quad (1.127)$$

This form of the thermodynamic equation is valid for both liquids and gases, and in an ideal gas $\beta_T = 1/T$.

Approximations using pressure and temperature

Liquids are characterized by a small coefficient of thermal expansion, and it is sometimes acceptable in laboratory fluids to neglect the second term on the left-hand side of (1.127). We then obtain an equation analogous to (1.117), namely

$$\frac{DT}{Dt} = Q[T]. \quad (1.128)$$

This approximation relies on the smallness of the coefficient of thermal expansion. A better approximation is to again suppose that the pressure in (1.127) varies according only to the weight of the fluid above it. Then $dp = -\rho g dz$ and (1.127) becomes

$$\frac{1}{T} \frac{DT}{Dt} + \frac{\beta_T g}{c_p} \frac{Dz}{Dt} = \frac{Q[T]}{T}. \quad (1.129)$$

For small variations of T , and if β_T is nearly constant, this simplifies to

$$\frac{D}{Dt} \left(T + \frac{T_0 z}{H_T} \right) = Q[T], \quad (1.130)$$

where

$$H_T = \frac{c_p}{\beta_T g} \quad (1.131)$$

is the *temperature scale height* of the fluid. The quantity $T + T_0 z / H_T$ is (in this approximation) the *potential temperature*, θ , so called because it is the temperature that a fluid at a depth z would have if moved adiabatically to a reference depth, here taken as $z = 0$. The temperature changes because of the work done by or on the fluid parcel as it expands or is compressed. That is,

$$\theta \approx T + \frac{\beta_T g T_0}{c_p} z. \quad (1.132)$$

In seawater the expansion coefficient β_T and c_p are functions of pressure and (1.132) is not good enough for quantitative calculations. With the approximate values for the ocean of $\beta_T \approx 2 \times 10^{-4} \text{ K}^{-1}$ and $c_p \approx 4 \times 10^3 \text{ J kg}^{-1} \text{ K}^{-1}$ we obtain $H_T \approx 2000 \text{ km}$.

The adiabatic lapse rate of temperature is the rate at which the temperature of a parcel changes in the vertical when undergoing an adiabatic displacement. From (1.129) it is

$$\Gamma_{ad} = - \left(\frac{\partial T}{\partial z} \right)_n = \frac{T g \beta_T}{c_p}. \quad (1.133)$$

In general it is a function of temperature, salinity and pressure, but it is a calculable quantity if β_T is known. With the oceanic values above, it is approximately 0.15 K km^{-1} . Equation (1.133) is not accurate enough for quantitative oceanography because the expansion coefficient is a function of pressure; nor is it a good measure of stability, because of the effects of salt.

It is noteworthy that the scale heights given by (1.120) and (1.131) differ so much. The first is due to the compressibility of seawater [and so related to c_s^2 , or β_p in (1.59)] whereas the second is due to the change of density with temperature [β_T in (1.59)], and is the distance over which the the difference between temperature and potential temperature changes by an amount equal to the temperature itself (i.e., by about 273 K). The two heights differ so much because the value of the thermal expansion coefficient is not directly tied to the compressibility — for example, fresh water at 4°C has a zero thermal expansion, and so would have an infinite temperature scale height, but its compressibility differs little from water at 20°C .

In the atmosphere the ideal gas relationship gives $\beta_T = 1/T$ and so

$$\Gamma_{ad} = \frac{g}{c_p}, \quad (1.134)$$

which is approximately 10 K km^{-1} . The only approximation involved in deriving this is the use of the hydrostatic relationship.

III. Entropy equation using density and temperature

Taking entropy to be a function of density and temperature (and salinity if appropriate) we have

$$T d\eta = T \left(\frac{\partial \eta}{\partial T} \right)_{\alpha, S} dT + T \left(\frac{\partial \eta}{\partial \alpha} \right)_{T, S} d\alpha + T \left(\frac{\partial \eta}{\partial S} \right)_{T, \alpha} dS$$

Forms of the Thermodynamic Equation

General form

$$\frac{DI}{Dt} + p \frac{D\alpha}{Dt} = \dot{Q}_T \quad \text{or} \quad \frac{DI}{Dt} + p \alpha \nabla \cdot \mathbf{v} = \dot{Q}_T, \quad (\text{T.1a,b})$$

where I is the internal energy and \dot{Q}_T is the rate of energy input, per unit mass. Equivalently, we may use an entropy or potential temperature equation,

$$\frac{D\eta}{Dt} = \frac{1}{T} \dot{Q} \quad \text{or} \quad c_p \frac{D \ln \theta}{Dt} \approx \frac{1}{T} \dot{Q}, \quad (\text{T.2a,b})$$

where η is the entropy, θ is the potential temperature, and \dot{Q} is the heating rate. For a fluid parcel of constant composition, $c_p d \ln \theta = d\eta$. Entropy or potential temperature must then be related to the other thermodynamic variables via an equation of state. [Equation (T.2b) is exactly true in the absence of compositional changes.]

Ideal gas

For an ideal gas $dI = c_v dT$, and the adiabatic thermodynamic equation may be written in the following equivalent, exact, forms:

$$\begin{aligned} c_p \frac{DT}{Dt} - \alpha \frac{Dp}{Dt} &= 0, & \frac{Dp}{Dt} + \gamma p \nabla \cdot \mathbf{v} &= 0, \\ c_v \frac{DT}{Dt} + p \alpha \nabla \cdot \mathbf{v} &= 0, & \frac{D\theta}{Dt} &= 0, \end{aligned} \quad (\text{T.3})$$

where $\theta = T(p_R/p)^\kappa$. The two expressions on the second line are usually the most useful in modelling and theoretical work.

Liquids

For liquids we may usefully write the adiabatic thermodynamic entropy equation as a conservation equation for potential temperature θ or potential density ρ_{pot} . For example:

$$\frac{D\theta}{Dt} = 0, \quad \theta \approx \begin{cases} T & \text{(approximately)} \\ T + (\beta_T g z / c_p) & \text{(with some thermal expansion),} \end{cases} \quad (\text{T.4a})$$

$$\frac{D\rho_{pot}}{Dt} = 0, \quad \rho_{pot} \approx \begin{cases} \rho & \text{(very approximately)} \\ \rho + (\rho_0 g z / c_s^2) & \text{(with some compression).} \end{cases} \quad (\text{T.4b})$$

Unlike (T.3), these are not equivalent forms. More accurate semi-empirical expressions that may also include saline effects must be used for quantitative applications. In principle, an internal energy equation may then be constructed using the equation of state; an oft-used alternative is to employ the Boussinesq approximation (discussed in section 2.4), which avoids consideration of internal energy.

$$= c_v \, dT + T \left(\frac{\partial \eta}{\partial \alpha} \right)_{T,S} \, d\alpha + T \left(\frac{\partial \eta}{\partial S} \right)_{T,\alpha} \, dS. \quad (1.135)$$

For a moving fluid this implies,

$$\frac{DT}{Dt} + \frac{T}{c_v} \left(\frac{\partial \eta}{\partial \alpha} \right)_{T,S} \frac{D\alpha}{Dt} = \hat{Q}[T], \quad (1.136)$$

where $\hat{Q}[T] \equiv \dot{Q}/c_v - T c_v^{-1} \dot{S}(\partial \eta / \partial S)$.

For an ideal gas, and using (1.84) (one of Maxwell's relations), (1.136) may be written as

$$c_v \frac{DT}{Dt} + p \frac{D\alpha}{Dt} = \dot{Q}, \quad (1.137)$$

so recovering the internal energy equation (1.97a). On the other hand, for a liquid of nearly constant density, the second term on the left-hand side of (1.136) is small, and also $c_p \approx c_v$, and we have to a first approximation $DT/Dt = \hat{Q}[T]$.

Various forms of the thermodynamic equation are summarized on page 30, and the complete equations of motion for a fluid are summarized on page 32. Also, note that for an ideal gas (1.116) and (1.127) are exactly equivalent to (1.98) or (1.99) (problem 1.11).

1.7 * MORE THERMODYNAMICS OF LIQUIDS

1.7.1 Potential temperature, potential density and entropy

For an ideal gas we were able to derive a simple thermodynamic equation for a single variable, potential temperature, which in turn could be simply related to the temperature and pressure and that was, furthermore, closely related to entropy. Potential temperature is, in fact, closely related to entropy for more general fluids, as we now see.

Potential temperature

The potential temperature is defined as the temperature that a parcel would have if moved adiabatically to a given reference pressure p_R , often taken as 10^5 Pa (or 1000 hPa, or 1000 mb, approximately the pressure at the sea-surface). Thus it may be calculated, at least in principle, by the integral

$$\theta(S, T, p, p_R) = T + \int_p^{p_R} \Gamma'_{ad}(S, T, p) \, dp, \quad (1.138)$$

where $\Gamma'_{ad} = (\partial T / \partial p)_\eta$. Such integrals may be hard to calculate, and if we know the equation of state in the form $\eta = \eta(S, T, p)$ then we can calculate potential temperature more directly because, by definition, potential temperature satisfies

$$\eta(S, T, p) = \eta(S, \theta, p_R). \quad (1.139)$$

Solving this for θ gives, in principle, $\theta = \theta(\eta, S, p_R) = \theta(T, S, p)$.

For a parcel of constant composition, changes in entropy are directly related to changes in potential temperature because, from the right-hand side of (1.139),

$$d\eta = \frac{\partial \eta(S, \theta, p_R)}{\partial \theta} d\theta. \quad (1.140)$$

The Equations of Motion of a Fluid

For dry air, or for a salt-free liquid, the equations of motion may be written as follows:

The *mass continuity equation*:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \quad (\text{EOM.1})$$

If density is constant this reduces to $\nabla \cdot \mathbf{v} = 0$.

The *momentum equation*:

$$\frac{D\mathbf{v}}{Dt} = -\frac{\nabla p}{\rho} + \nu \nabla^2 \mathbf{v} + \mathbf{F}, \quad (\text{EOM.2})$$

where \mathbf{F} represents the effects of body forces such as gravity and ν is the kinematic viscosity. If density is constant, or pressure is given as a function of density alone (e.g., $p = C\rho^\gamma$), then (EOM.1) and (EOM.2) form a complete system.

The *thermodynamic equation*:

$$\frac{DI}{Dt} + \frac{p}{\rho} \nabla \cdot \mathbf{v} = \dot{Q}, \quad \text{or} \quad \frac{D\theta}{Dt} = \frac{1}{c_p} \left(\frac{\theta}{T} \right) \dot{Q}, \quad (\text{EOM.3})$$

where \dot{Q} represents diabatic sources such as heating and diffusion.

An *equation of state*:

$$p = f(I, \rho) \quad \text{or} \quad p = g(\theta, \rho) \quad (\text{EOM.4})$$

where f and g are given functions. For example, for an ideal gas, $p = \rho RI/c_v$ and $p = p_R(T/\theta)^{1/\kappa}$ where $T = p/\rho R$.

The equations describing fluid motion are called the *Euler equations* if the viscous term is omitted, and the *Navier–Stokes equations* if viscosity is included.⁸ Sometimes these appellations are taken to mean only the momentum and mass conservation equations, and sometimes the Euler equations are restricted to a fluid of constant density.

Thus, if a fluid parcel moves adiabatically and at constant composition then $d\eta = 0$ and $d\theta = 0$. Furthermore, if we express entropy as a function of temperature and pressure then

$$T d\eta = T \left(\frac{\partial \eta}{\partial T} \right)_{pS} dT + T \left(\frac{\partial \eta}{\partial p} \right)_{TS} dp = c_p dT + T \left(\frac{\partial \eta}{\partial p} \right)_{TS} dp, \quad (1.141)$$

using the definition of c_p . If we evaluate this expression at the reference pressure, where $T = \theta$ and $dp = 0$, then we have $\theta d\eta = c_p(p_R, \theta) d\theta$, and therefore

$$d\eta = c_p(p_R, \theta) \frac{d\theta}{\theta}, \quad (1.142)$$

and $d\eta/d\theta = c_p(p_R, \theta)/\theta$. In the special case of constant c_p integration yields

$$\eta = c_p \ln \theta + \text{constant}, \quad (1.143)$$

as for a simple ideal gas (1.107). Given (1.142), the thermodynamic equation can be written, for a field of constant composition,

$$c_p \frac{D\theta}{Dt} = \frac{\theta}{T} \dot{Q}, \quad (1.144)$$

where the right-hand side represents heating. If composition is not constant, the right-hand side should also include any saline source terms and saline diffusion; however, such terms usually have a very small effect in the ocean. We can use (1.144) as a thermodynamic equation instead of (1.95), although we must now relate θ instead of η to the other state variables via an equation of state.

The notion of potential temperature is useful because it is connected to the actual temperature, with which we are familiar; roughly speaking, potential temperature is temperature plus a correction for the effects of compressibility. Entropy, on the other hand, may seem alien and unnecessarily exotic. However, the use of potential temperature brings no simplification to the equations of motion beyond that already afforded by the use of entropy as a thermodynamic variable. Indeed, for fluids of variable composition, the evolution equation for potential temperature is more complicated than that for entropy.⁹

Potential density

Potential density, ρ_θ , is the density that a parcel would have if it were moved adiabatically and with fixed composition to a given reference pressure, p_R , that is often, but not always, taken as 10^5 Pa, or 1 bar. If the equation of state is of the form $\rho = \rho(S, T, p)$ then by definition we have

$$\rho_\theta = \rho(S, \theta, p_R). \quad (1.145)$$

For a parcel moving adiabatically (i.e., fixed salinity and potential temperature) its potential density is therefore conserved. For an ideal gas (1.145) gives $\rho_\theta = p_R / (R\theta)$ [as in (1.110)] and potential density provides no more information than potential temperature. However, in the oceans potential density accounts for the effect of salinity on density and so is a much better measure of the static stability of a column of water than density itself. An approximate expression for ρ_θ is given by (1.119).

Because density is so nearly constant in the ocean, it is common to subtract the amount 1000 kg m^{-3} before quoting its value, and depending on whether we are referring to *in situ* density or the potential density the results are called σ_T ('sigma-tee') or σ_θ ('sigma-theta') respectively. Thus,

$$\sigma_T = \rho(p, T, S) - 1000, \quad \sigma_\theta = \rho(p_R, \theta, S) - 1000. \quad (1.146a,b)$$

If the potential density is referenced to a particular level, this is denoted by a subscript on the σ . Thus, σ_2 is the potential density referenced to 200 bars of pressure, or about 2 kilometres depth.

1.7.2 * Thermodynamic properties of seawater

In this section we illustrate how the thermodynamic variables for seawater — the conventional equation of state, an expression for potential temperature, and so on — can be obtained directly from the fundamental equation of state. Writing the fundamental equation

in the form $I = I(\eta, S, \alpha)$ is not practically useful, because the variables are not easily measured in the laboratory. However, if we cast the fundamental equation in terms of the Gibbs function, $G = I - T\eta + p\alpha$, then $dG = -\eta dT + \alpha dp + \mu dS$ and the independent variables are the familiar and measurable (T, S, p) .

Oceanographers go to great lengths to obtain an accurate equation of state and other physical properties of seawater, and we noted in section 1.4 that seawater has some nonlinear properties that, although small, are nevertheless important. A Gibbs function that accurately reproduces these properties is very complicated, but we can write down a model Gibbs function that, although not highly accurate except for small variations around a reference state, does capture and display the most important properties with a certain degree of economy and transparency.¹⁰ The function is

$$\begin{aligned} G = G_0 - \eta_0(T - T_0) + \mu_0(S - S_0) - c_{p0}T[\ln(T/T_0) - 1][1 + \beta_S^*(S - S_0)] \\ + \alpha_0(p - p_0) \left[1 + \beta_T(T - T_0) - \beta_S(S - S_0) - \frac{\beta_p}{2}(p - p_0) \right. \\ \left. + \frac{\beta_T\gamma^*}{2}(p - p_0)(T - T_0) + \frac{\beta_T^*}{2}(T - T_0)^2 \right]. \end{aligned} \quad (1.147)$$

In this equation the variables are G, T, S and p . The parameters [which all have subscripts or stars, with the starred parameters giving rise to nonlinear effects; refer also to (1.60)] are all constants that could in principle be determined in the laboratory with the help of the derived quantities like heat capacity, and their approximate values are given in Table 1.2. We will take $p_0 = 0$ and $\beta_p = \alpha_0/c_{s0}^2$, where c_{s0} is a reference sound speed. From (1.147) we may derive the following quantities of interest.

The conventional or thermal equation of state, $\alpha = (\partial G / \partial p)_{T,p}$:

$$\alpha = \alpha_0 \left[1 + \beta_T(1 + \gamma^* p)(T - T_0) + \frac{\beta_T^*}{2}(T - T_0)^2 - \beta_S(S - S_0) - \beta_p p \right]. \quad (1.148)$$

The entropy, $\eta = -(\partial G / \partial T)_{p,S}$:

$$\eta = \eta_0 + c_{p0} \ln \frac{T}{T_0} \left[1 + \beta_S^*(S - S_0) \right] - \alpha_0 p \left[\beta_T + \beta_T \gamma^* \frac{p}{2} + \beta_T^*(T - T_0) \right]. \quad (1.149)$$

The heat capacity, $c_p = T(\partial \eta / \partial T)_p$:

$$c_p = c_{p0} [1 + \beta_S^*(S - S_0)] - \alpha_0 p \beta_T^* T. \quad (1.150)$$

This is to a first approximation constant, varying mildly with salinity and more weakly with temperature and pressure.

The thermal expansion coefficient, $\hat{\beta}_T = \alpha^{-1}(\partial \alpha / \partial T)_{S,p}$:

$$\hat{\beta}_T = (\alpha_0 / \alpha) [\beta_T + \beta_T \gamma^* p + \beta_T^*(T - T_0)], \quad (1.151)$$

where α is given by (1.148).

The adiabatic lapse rate, $\Gamma = (\partial T / \partial p)_\eta$. Using (1.141) gives

$$\Gamma = \left(\frac{\partial T}{\partial p} \right)_\eta = \frac{T}{c_p} \left(\frac{\partial \alpha}{\partial T} \right)_{p,S} = \frac{T}{c_p} \alpha_0 [\beta_T(1 + \gamma^* p) + \beta_T^*(T - T_0)]. \quad (1.152)$$

where c_p is given by (1.150).

Parameter	Description	Value
ρ_0	Reference Density	$1.027 \times 10^3 \text{ kg m}^{-3}$
α_0	Reference Specific Volume	$9.738 \times 10^{-4} \text{ m}^3 \text{ kg}^{-1}$
T_0	Reference temperature	283 K
S_0	Reference salinity	35 psu $\approx 35\%$
c_{s0}	Reference sound speed	1490 m s^{-1}
β_T	First thermal expansion coefficient	$1.67 \times 10^{-4} \text{ K}^{-1}$
β_T^*	Second thermal expansion coefficient	$1.00 \times 10^{-5} \text{ K}^{-2}$
β_S	Haline contraction coefficient	$0.78 \times 10^{-3} \text{ psu}^{-1}$
β_p	Compressibility coefficient ($= \alpha_0/c_{s0}^2$)	$4.39 \times 10^{-10} \text{ m s}^2 \text{ kg}^{-1}$
γ^*	Thermobaric parameter ($\approx \gamma'^*$)	$1.1 \times 10^{-8} \text{ Pa}^{-1}$
c_{p0}	Specific heat capacity at constant pressure	$3986 \text{ J kg}^{-1} \text{ K}^{-1}$
β_S^*	Haline heat capacity coefficient	$1.5 \times 10^{-3} \text{ psu}^{-1}$

Table 1.2 Various thermodynamic and equation-of-state parameters for seawater, which may be used in the equations of state (1.60) and (1.147). The 'psu', or practical salinity unit, is approximately one part per thousand.

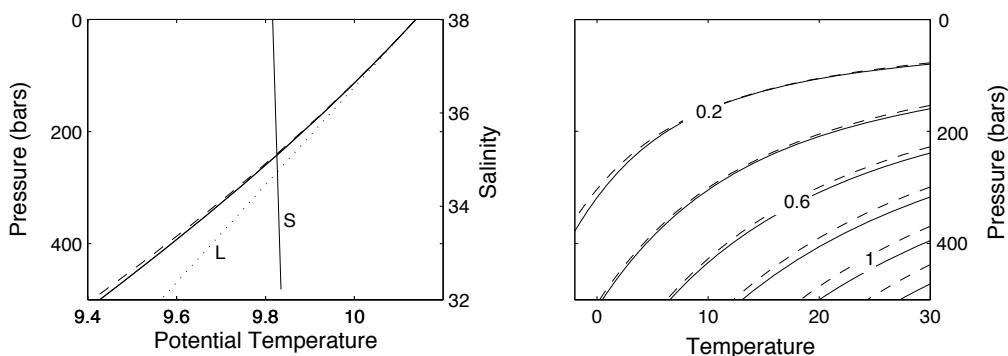


Fig. 1.4 Examples of the variation of potential temperature of seawater with pressure, temperature and salinity. Left panel: the sloping lines show potential temperature as a function of pressure at fixed salinity ($S = 35 \text{ psu}$) and temperature (10.14°C). The solid line is computed using an accurate, empirical equation of state, the almost-coincident dashed line uses the simpler expression (1.153) and the dotted line (labelled L) uses the linear expression (1.154c). The near vertical solid line, labelled S, shows the variation of potential temperature with salinity at fixed temperature and pressure. Right panel: Contours of the difference between temperature and potential temperature, ($T - \theta$) in the pressure-temperature plane, for $S = 35 \text{ psu}$. The solid lines use an accurate empirical formula, and the dashed lines use (1.153). The simpler equation can be improved locally, but not globally, by tuning the coefficients. (100 bars of pressure (10^7 Pa or 10 MPa) is approximately 1 km depth.)

Potential temperature and a simplified equation of state for numerical models

An expression for the potential temperature, θ , may be obtained by solving (1.139) for θ . In general, such an equation must be solved numerically, but for our equation of state, using (1.149) and taking $p_R = 0$, we find

$$\theta = T \exp \left\{ -\frac{\alpha_0 \beta_T p}{c'_p} \left[1 + \frac{1}{2} \gamma^* p + \frac{\beta_T^*}{\beta_T} (T - T_0) \right] \right\}. \quad (1.153)$$

where $c'_p = c_{p0} [1 + \beta_S^*(S - S_0)]$. Equation (1.153) is a relationship between T , θ and p analogous to (1.105) for an ideal gas. The exponent itself is small, the second and third terms in square brackets are small compared to unity, the deviations of both T and θ from T_0 are also presumed to be small, and c'_p is nearly constant. Taking advantage of all of this enables (1.153) to be rewritten, with increasing levels of approximation, as

$$T' \approx \frac{T_0 \alpha_0 \beta_T}{c_{p0}} p \left(1 + \frac{1}{2} \gamma^* p + T_0 \frac{\alpha_0 \beta_T^*}{c_{p0}} p \right) + \theta' \left(1 + T_0 \frac{\alpha_0 \beta_T^*}{c_{p0}} p \right) \quad (1.154a)$$

$$\approx \frac{T_0 \alpha_0 \beta_T}{c_{p0}} p \left(1 + \frac{1}{2} \gamma^* p \right) + \theta' \left(1 + T_0 \frac{\alpha_0 \beta_T^*}{c_{p0}} p \right) \quad (1.154b)$$

$$\approx \frac{T_0 \alpha_0 \beta_T}{c_{p0}} p + \theta', \quad (1.154c)$$

where $T' = T - T_0$ and $\theta' = \theta - T_0$. The last of the three, (1.154c), holds for a linear equation of state, and is useful for calculating approximate differences between temperature and potential temperature; making use of the hydrostatic approximation reveals that it is essentially the same as (1.132). Plots of the difference between temperature and potential temperature, using both a highly accurate empirical equation of state and using the simplified equation, (1.153), are given in Fig. 1.4, and some examples of the density variation of seawater are given in Fig. 1.5.

Numerical ocean models often use potential temperature as a thermodynamic variable. We then require an equation of state that gives density in terms of potential temperature, pressure, and salinity. Using (1.154b) in the equation of state (1.148) we find that

$$\alpha \approx \alpha_0 \left[1 - \frac{\alpha_0 p}{c_s'^2} + \beta_T (1 + \gamma'^* p) \theta' + \frac{1}{2} \beta_T^* \theta'^2 - \beta_S (S - S_0) \right], \quad (1.155)$$

where $\gamma'^* = \gamma^* + T_0 \beta_T^* \alpha_0 / c_{p0} \approx \gamma^*$ and $c_s'^{-2} = c_{s0}'^{-2} - \beta_T^2 T_0 / c_p \approx c_{s0}'^{-2}$ (γ^* and γ'^* differ by a few percent, and c_s^2 and $c_s'^2$ differ by only a few parts in a thousand). We may approximate (1.155) further by using the hydrostatic pressure instead of the actual pressure; thus, letting $p = -g(z - z_0) / \alpha_0$ where z_0 is the nominal value of z at which $p = 0$, we obtain

$$\alpha \approx \alpha_0 \left[1 + \frac{g(z - z_0)}{c_{s0}^2} + \beta_T \left(1 - \gamma'^* \frac{g(z - z_0)}{\alpha_0} \right) \theta' + \frac{\beta_T^*}{2} \theta'^2 - \beta_S (S - S_0) \right]. \quad (1.156)$$

Using z instead of p in the equation of state entails a slight loss of accuracy, but is necessary to ensure that the Boussinesq equations maintain good conservation properties, as discussed in sections 2.4 and 4.7.1.

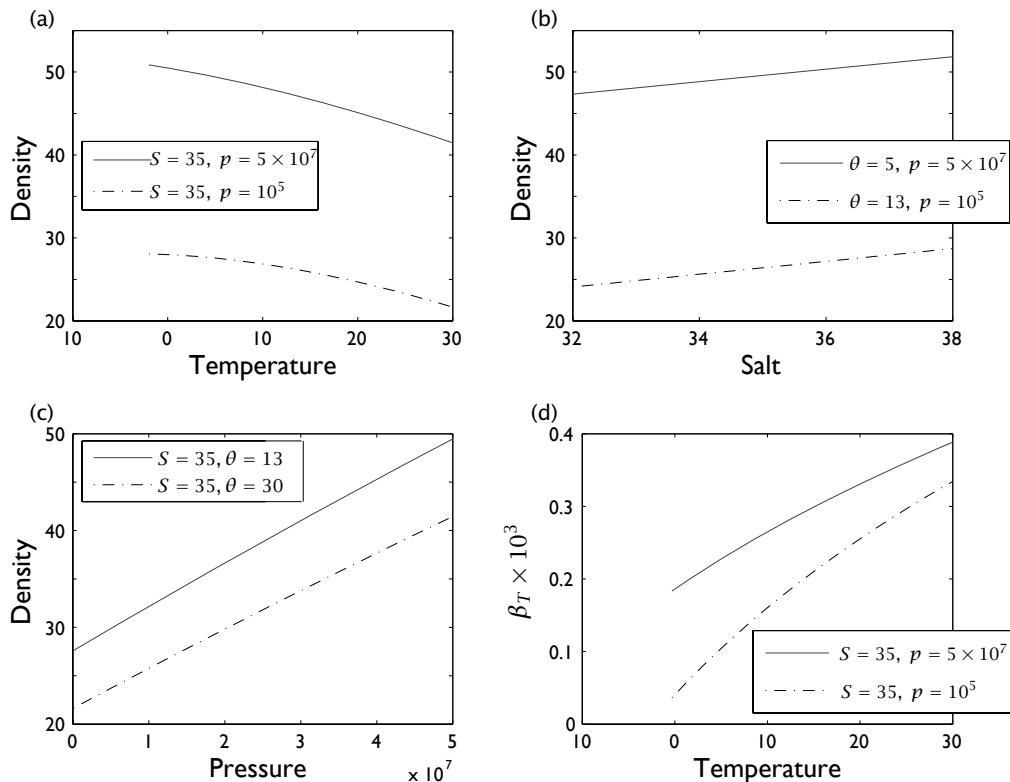


Fig. 1.5 Examples of the variation of density of seawater (minus 1000 kg m^{-3}) with (a) potential temperature ($^{\circ}\text{C}$); (b) salt (psu); and (c) pressure (Pa), for seawater. Panel (d), shows the thermal expansion coefficient, $\hat{\beta}_T = -\rho^{-1}(\partial\rho/\partial T)_{p,S} \text{ K}^{-1}$, for each of the two curves in panel (a).

1.8 SOUND WAVES

Full of sound and fury, signifying nothing.

William Shakespeare, *Macbeth*, c. 1606.

We now consider, rather briefly, one of the most common phenomena in fluid dynamics, yet one which is in most circumstances relatively unimportant for geophysical fluid dynamics — sound waves. Their unimportance stems from the fact that the pressure disturbance produced by sound waves is a tiny fraction of the ambient pressure and too small to affect the circulation. For example, the ambient surface pressure in the atmosphere is about 10^5 Pa and variations due to large-scale weather phenomena are about 10^3 Pa or larger, whereas sound waves of 70 dB (i.e., a loud conversation) produce pressure variations of about 0.06 Pa . ($1 \text{ dB} = 20 \log_{10}(p/p_r)$ where $p_r = 2 \times 10^{-5} \text{ Pa}$.)

The smallness of the disturbance produced by sound waves justifies a linearization of the equations of motion about a spatially uniform basic state that is a time-independent solution to the equations of motion. Thus, we write $\mathbf{v} = \mathbf{v}_0 + \mathbf{v}'$, $\rho = \rho_0 + \rho'$ (where a subscript 0 denotes a basic state and a prime denotes a perturbation) and so on, substitute in the

equations of motion, and neglect terms involving products of primed quantities. By choice of our reference frame we will simplify matters further by setting $\mathbf{v}_0 = 0$. The linearized momentum and mass conservation equations are then, respectively,

$$\rho_0 \frac{\partial \mathbf{v}'}{\partial t} = -\nabla p', \quad \frac{\partial \rho'}{\partial t} = -\rho_0 \nabla \cdot \mathbf{v}'. \quad (1.157a,b)$$

(Note that these linear equations do not in themselves determine the magnitude of the disturbance.) Now, sound waves are largely adiabatic. Thus,

$$\frac{dp}{dt} = \left(\frac{\partial p}{\partial \rho} \right)_\eta \frac{d\rho}{dt}, \quad (1.158)$$

where $(\partial p / \partial \rho)_\eta$ is the derivative at constant entropy, whose particular form is given by the equation of state for the fluid at hand. Then, from (1.157) and (1.158) we obtain a single equation for pressure,

$$\frac{\partial^2 p'}{\partial t^2} = c_s^2 \nabla^2 p', \quad (1.159)$$

where $c_s^2 = (\partial p / \partial \rho)_\eta$. Equation (1.159) is the classical wave equation; solutions propagate at a speed c_s , which may be identified as the speed of sound. For adiabatic flow in an ideal gas, manipulation of the equation of state leads to $p = C\rho^\gamma$, where $\gamma = c_p/c_v$, whence $c_s^2 = \gamma p / \rho = \gamma RT$. Values of γ typically range from 5/3 for a monatomic gas to 7/5 for a diatomic gas and so for air, which is almost entirely diatomic, we find $c_s \approx 350 \text{ m s}^{-1}$ at 300 K. In seawater no such theoretical approximation is easily available, but measurements show that $c_s \approx 1500 \text{ m s}^{-1}$.

1.9 COMPRESSIBLE AND INCOMPRESSIBLE FLOW

Although there are probably no fluids of truly constant density, in many cases the density of a fluid will vary so little that it is a very good approximation to consider the density effectively constant in the mass conservation equation. The fluid is then said to be incompressible. For example, in the Earth's oceans the density varies by less than 5%, even though the pressure at the ocean bottom, a few kilometres below the surface, is several hundred times the atmospheric pressure at the surface. Let us first consider how the mass conservation equation simplifies when density is truly constant, and then consider conditions under which treating density as constant is a good approximation.

1.9.1 Constant density fluids

If a fluid is strictly of constant density then the mass continuity equation, (1.36) (i.e., $D\rho/Dt + \rho \nabla \cdot \mathbf{v} = 0$), simplifies easily by neglecting all derivatives of density yielding

$$\boxed{\nabla \cdot \mathbf{v} = 0}. \quad (1.160)$$

The *prognostic* equation (1.36) has become a *diagnostic* equation (1.160), or a constraint to be satisfied by the velocity. A consequence of (1.160) is that the volume of each material

fluid element remains constant. To see this recall the expression for the conservation of mass in the form

$$\frac{D}{Dt}(\rho \Delta V) = 0. \quad (1.161)$$

If ρ is constant this reduces to an expression for volume conservation, $D\Delta V/Dt = 0$, whence (1.160) is recovered because $D\Delta V/Dt = \Delta V \nabla \cdot \mathbf{v}$.

1.9.2 Incompressible flows

An *incompressible fluid* is one in which the density of a given fluid element does not change. Thus, in the mass continuity equation, (1.36a), the material derivative of density is neglected and we recover (1.160). In reality no fluid is truly incompressible and for (1.160) to approximately hold we just require that

$$\left| \frac{D\rho}{Dt} \right| \ll \rho \left(\left| \frac{\partial u}{\partial x} \right| + \left| \frac{\partial v}{\partial y} \right| + \left| \frac{\partial w}{\partial z} \right| \right); \quad (1.162)$$

that is, the material derivative of density is much smaller than the individual terms constituting the divergence. As working definition we say that *in an incompressible fluid, density changes are so small that they have a negligible effect on the mass balance*. We do not need to assume that the densities of differing fluid elements are similar to each other, but in the ocean (and in most liquids) variations in density, $\delta\rho$, are in fact everywhere small compared to the mean density, ρ_0 . A sufficient condition for incompressibility, then, is that

$$\frac{\delta\rho}{\rho_0} \ll 1. \quad (1.163)$$

Note that the fact that $\nabla \cdot \mathbf{v} = 0$ does *not* imply that we may independently use $D\rho/Dt = 0$. Indeed for a liquid with an equation of state $\rho = \rho_0(1 - \beta_T(T - T_0))$ and a thermodynamic equation $c_p DT/Dt = \dot{Q}$ we have

$$\frac{D\rho}{Dt} = -\frac{\beta_T \rho_0}{c_p} \dot{Q}. \quad (1.164)$$

Furthermore, incompressibility does not necessarily imply the neglect of density variations in the momentum equation — it is only in the mass continuity equation that density variations are neglected.

Some conditions for incompressibility

The conditions under which incompressibility is a good approximation to the full mass continuity equation depend not only on the physical nature of the fluid but also on the flow itself. The condition that density is largely unaffected by pressure gives one necessary condition for the legitimate use of (1.160), as follows. First assume adiabatic flow, and omit the gravitational term. Then

$$\frac{dp}{dt} = \left(\frac{\partial p}{\partial \rho} \right)_n \frac{d\rho}{dt} = c_s^2 \frac{d\rho}{dt} \quad (1.165)$$

so that the density and pressure variations of a fluid parcel are related by

$$\delta p \sim c_s^2 \delta \rho. \quad (1.166)$$

From the momentum equation we estimate

$$\frac{U^2}{L} \sim \frac{1}{L} \frac{\delta p}{\rho_0}, \quad (1.167)$$

where U and L are typical velocities and lengths and where ρ_0 is a representative value of the density. Using (1.166) and (1.167) gives $U^2 \sim c_s^2 \delta \rho / \rho_0$. The incompressibility condition (1.163) then becomes

$$\frac{U^2}{c_s^2} \ll 1. \quad (1.168)$$

That is, for a flow to be incompressible the fluid velocities must be less than the speed of sound; that is, the Mach number, $M \equiv U/c_s$, must be small.

In the Earth's atmosphere it is apparent that density changes with height. To estimate how much density does change, let us first assume hydrostatic balance and an ideal gas, so that

$$\frac{1}{\rho} \frac{\partial p}{\partial z} = -g. \quad (1.169)$$

If (for simplicity) we also assume that atmosphere is isothermal then

$$\frac{\partial p}{\partial z} = \left(\frac{\partial p}{\partial \rho} \right)_T \frac{\partial \rho}{\partial z} = RT_0 \frac{\partial \rho}{\partial z}. \quad (1.170)$$

Using (1.169) and (1.170) gives

$$\rho = \rho_0 \exp(-z/H_\rho), \quad (1.171)$$

where $H_\rho = RT_0/g$ is the (density) *scale height* of the atmosphere. It is easy to see that density changes are negligible only if we concern ourselves with motion less than the scale height, so this is another necessary condition for incompressibility.

In the atmosphere, although the Mach number is small for most flows, vertical displacements often exceed the scale height and the flow cannot then be considered incompressible. In the ocean density changes from all causes are small and in most circumstances the ocean may be considered to contain an incompressible fluid. We return to this in the next chapter when we consider the Boussinesq equations.

1.10 THE ENERGY BUDGET

The total energy of a fluid includes the kinetic, potential and internal energies. Both fluid flow and pressure forces will, in general, move energy from place to place, but we nevertheless expect energy to be conserved in an enclosed volume. Let us now consider what form energy conservation takes in a fluid.

1.10.1 Constant density fluid

For a constant density fluid the momentum equation and the mass continuity equation $\nabla \cdot \mathbf{v} = 0$, are sufficient to completely determine the evolution of a system. The momentum equation is

$$\frac{D\mathbf{v}}{Dt} = -\nabla(\phi + \Phi) + \nu \nabla^2 \mathbf{v}, \quad (1.172)$$

where $\phi = p/\rho_0$ and Φ is the potential for any conservative force per unit mass (e.g., gz for a uniform gravitational field). We can rewrite the advective term on the left-hand side using the identity,

$$(\mathbf{v} \cdot \nabla) \mathbf{v} = -\mathbf{v} \times \boldsymbol{\omega} + \nabla(\mathbf{v}^2/2), \quad (1.173)$$

where $\boldsymbol{\omega} \equiv \nabla \times \mathbf{v}$ is the *vorticity*, discussed more in later chapters. Then, omitting viscosity, we have

$$\frac{\partial \mathbf{v}}{\partial t} + \boldsymbol{\omega} \times \mathbf{v} = -\nabla B, \quad (1.174)$$

where $B = (\phi + \Phi + \mathbf{v}^2/2)$ is the *Bernoulli function* for constant density flow. Consider for a moment steady flows ($\partial \mathbf{v} / \partial t = 0$). Streamlines are, by definition, parallel to \mathbf{v} everywhere, and the vector $\mathbf{v} \times \boldsymbol{\omega}$ is everywhere orthogonal to the streamlines, so that taking the dot product of the steady version of (1.174) with \mathbf{v} gives $\mathbf{v} \cdot \nabla B = 0$. That is, for steady flows the Bernoulli function is constant along a streamline, and $DB/Dt = 0$.

Reverting back to the time-varying case, take the dot product with \mathbf{v} and include the density to yield

$$\frac{1}{2} \frac{\partial \rho_0 \mathbf{v}^2}{\partial t} + \rho_0 \mathbf{v} \cdot (\boldsymbol{\omega} \times \mathbf{v}) = -\rho_0 \mathbf{v} \cdot \nabla B \quad (1.175)$$

The second term on the left-hand side vanishes identically. Defining the kinetic energy density K , or energy per unit volume, by $K = \rho_0 \mathbf{v}^2/2$, (1.175) becomes an expression for the rate of change of K ,

$$\frac{\partial K}{\partial t} + \nabla \cdot (\rho_0 \mathbf{v} B) = 0. \quad (1.176)$$

Because Φ is time-independent this may be written

$$\frac{\partial}{\partial t} \left[\rho_0 \left(\frac{1}{2} \mathbf{v}^2 + \Phi \right) \right] + \nabla \cdot \left[\rho_0 \mathbf{v} \left(\frac{1}{2} \mathbf{v}^2 + \Phi + \phi \right) \right] = 0. \quad (1.177)$$

or

$$\frac{\partial E}{\partial t} + \nabla \cdot [\mathbf{v}(E + p)] = 0. \quad (1.178)$$

where $E = K + \rho_0 \Phi$ is the total energy density (i.e, the total energy per unit volume). This has the form of a general conservation equation in which a local change in a quantity is balanced by the divergence of its flux. However, the energy flux, $\mathbf{v}(\rho_0 \mathbf{v}^2/2 + \rho_0 \Phi + \rho_0 \phi)$, is *not* simply the velocity times the energy density $\rho_0(\mathbf{v}^2/2 + \Phi)$; there is an additional term, $\mathbf{v}p$, that represents the energy transfer occurring when work is done by the fluid against the pressure force.

Now consider a volume through which there is no mass flux, for example a domain bounded by rigid walls. The rate of change of energy within that volume is then given by the integral of (1.175)

$$\frac{d\hat{K}}{dt} \equiv \frac{d}{dt} \int_V K dV = - \int_V \nabla \cdot (\rho_0 \mathbf{v} B) dV = - \int_S \rho_0 B \mathbf{v} \cdot d\mathbf{S} = 0, \quad (1.179)$$

using the divergence theorem, and where \hat{K} is the total kinetic energy. Thus, the total kinetic energy within the volume is conserved. Note that for a constant density fluid the gravitational potential energy, \hat{P} , is given by

$$\hat{P} = \int_V \rho_0 g z dV, \quad (1.180)$$

which is a constant, not affected by a rearrangement of the fluid. Thus, in a constant density fluid there is no exchange between kinetic energy and potential energy and the kinetic energy itself is conserved.

1.10.2 Variable density fluids

We start with the momentum equation

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla p - \rho \nabla \Phi, \quad (1.181)$$

and take its dot product with \mathbf{v} to obtain an equation for the evolution of kinetic energy,

$$\frac{1}{2} \rho \frac{D\mathbf{v}^2}{Dt} = -\mathbf{v} \cdot \nabla p - \rho \mathbf{v} \cdot \nabla \Phi = -\nabla \cdot (p\mathbf{v}) + p \nabla \cdot \mathbf{v} - \rho \mathbf{v} \cdot \nabla \Phi. \quad (1.182)$$

The internal energy equation for adiabatic flow is

$$\rho \frac{DI}{Dt} = -p \nabla \cdot \mathbf{v}. \quad (1.183)$$

Finally, and somewhat trivially, the potential energy density obeys

$$\rho \frac{D\Phi}{Dt} = \rho \mathbf{v} \cdot \nabla \Phi. \quad (1.184)$$

Adding (1.182), (1.183) and (1.184) we obtain

$$\rho \frac{D}{Dt} \left(\frac{1}{2} \mathbf{v}^2 + I + \Phi \right) = -\nabla \cdot (p\mathbf{v}), \quad (1.185)$$

which, on expanding the material derivative and using the mass conservation equation, becomes

$$\frac{\partial}{\partial t} \left[\rho \left(\frac{1}{2} \mathbf{v}^2 + I + \Phi \right) \right] + \nabla \cdot \left[\rho \mathbf{v} \left(\frac{1}{2} \mathbf{v}^2 + I + \Phi + p/\rho \right) \right] = 0. \quad (1.186)$$

This may be written

$$\frac{\partial E}{\partial t} + \nabla \cdot [\mathbf{v}(E + p)] = 0$$

(1.187)

where $E = \rho(\mathbf{v}^2/2 + I + \Phi)$ is the total energy per unit volume of the fluid. This is the energy equation for an unforced, inviscid and adiabatic, compressible fluid. Just as for the constant density case, the energy flux contains the term $p\mathbf{v}$ that represents the work done against pressure and, again, the second term vanishes when integrated over a closed domain with rigid boundaries, implying that the total energy is conserved. However, there can now be an exchange of energy between kinetic, potential and internal components. The quantity $\sigma = I + p\alpha + \Phi = h + \Phi$ is sometimes referred to as the dry static energy. However, it is not a component of the globally conserved total energy; the conserved energy contains only the quantity $I + \Phi$ plus the kinetic energy, and it is only the flux of static energy that affects the energetics. For an ideal gas we have $\sigma = c_v T + RT + \Phi = c_p T + \Phi$, and if the potential is caused by a uniform gravitational field then $\sigma = c_p T + gz$.

One other important point to note is that it is the divergent term, $\nabla \cdot \mathbf{v}$, that connects the kinetic energy equation, (1.182), and the internal energy equation, (1.183). In an incompressible fluid this term is absent, and the internal energy is divorced from the other components of energy. This consideration will be important when we consider the Boussinesq equations in section 2.4.

Bernoulli's theorem

For steady flow $\partial[]/\partial t = 0$ and $\nabla \cdot \rho \mathbf{v} = 0$ so that (1.186) may be written $\mathbf{v} \cdot \nabla B = 0$ where B is the Bernoulli function given by

$$B = \left(\frac{1}{2} \mathbf{v}^2 + I + \Phi + p/\rho \right) = \left(\frac{1}{2} \mathbf{v}^2 + h + \Phi \right), \quad (1.188)$$

and h is the enthalpy. Thus, for steady flow only,

$$\frac{DB}{Dt} = \mathbf{v} \cdot \nabla B = 0, \quad (1.189)$$

and the Bernoulli function is a constant along streamline; this result is commonly known as Bernoulli's theorem.¹¹ For an ideal gas in a constant gravitation field $B = \mathbf{v}^2/2 + c_p T + gz$. For adiabatic flow we also have $D\theta/Dt = 0$. Thus, steady flow is both along surfaces of constant θ and along surfaces of constant B , and the vector

$$\mathbf{l} = \nabla \theta \times \nabla B \quad (1.190)$$

is parallel to streamlines. A related result for unsteady flow is given in section 4.8.

1.10.3 Viscous effects

We might expect that viscosity will always act to reduce the kinetic energy of a flow, and we will demonstrate this for a constant density fluid. Retaining the viscous term in (1.172), the energy equation becomes

$$\frac{d\hat{E}}{dt} \equiv \frac{d}{dt} \int_V E \, dV = \mu \int_V \mathbf{v} \cdot \nabla^2 \mathbf{v} \, dV. \quad (1.191)$$

The right-hand side is negative definite. To see this we use the vector identity

$$\nabla \times (\nabla \times \mathbf{v}) = \nabla(\nabla \cdot \mathbf{v}) - \nabla^2 \mathbf{v}, \quad (1.192)$$

and because $\nabla \cdot \mathbf{v} = 0$ we have $\nabla^2 \mathbf{v} = -\nabla \times \boldsymbol{\omega}$, where $\boldsymbol{\omega} \equiv \nabla \times \mathbf{v}$. Thus,

$$\frac{d\hat{E}}{dt} = -\mu \int_V \mathbf{v} \cdot (\nabla \times \boldsymbol{\omega}) \, dV = -\mu \int_V \boldsymbol{\omega} \cdot (\nabla \times \mathbf{v}) \, dV = -\mu \int_V \boldsymbol{\omega}^2 \, dV, \quad (1.193)$$

after integrating by parts, providing $\mathbf{v} \times \boldsymbol{\omega}$ vanishes at the boundary. Thus, viscosity acts to extract kinetic energy from the flow. The loss of kinetic energy reappears as an irreversible warming of the fluid ('Joule heating'), and the total energy of the fluid is conserved, but this effect plays no role in a constant density fluid. The effect is normally locally small, at least in the Earth's oceans and atmosphere, although it is sometimes included in comprehensive General Circulation Models (GCMs).

1.11 AN INTRODUCTION TO NON-DIMENSIONALIZATION AND SCALING

The units we use to measure length, velocity and so on are irrelevant to the dynamics, and not necessarily the most appropriate units for a given problem. Rather, it is convenient

to express the equations of motion, so far as is possible, in so-called ‘non-dimensional’ variables, by which we mean expressing every variable (such as velocity) as the ratio of its value to some reference value. For velocity the reference could, for example, be the speed of light — but this would not be very helpful for fluid dynamical problems in the Earth’s atmosphere or oceans! Rather, we should choose the reference value as a natural one for a given flow, in order that, so far as possible, the non-dimensional variables are order-unity quantities, and doing this is called *scaling the equations*. Evidently, there is no reference velocity that is universally appropriate, and much of the art of fluid dynamics lies in choosing sensible scaling factors for the problem at hand. Non-dimensionalization plays an important role in fluid dynamics, and we introduce it here with a simple example.

1.11.1 The Reynolds number

Consider the constant-density momentum equation in Cartesian coordinates. If a typical velocity is U , a typical length is L , a typical time scale is T , and a typical value of the pressure deviation is Φ , then the approximate sizes of the various terms in the momentum equation are given by

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\nabla \phi + \nu \nabla^2 \mathbf{v} \quad (1.194a)$$

$$\frac{U}{T} \quad \frac{U^2}{L} \quad \sim \quad \frac{\Phi}{L} \quad \nu \frac{U}{L^2}. \quad (1.194b)$$

The ratio of the inertial terms to the viscous terms is $(U^2/L)/(\nu U/L^2) = UL/\nu$, and this is the *Reynolds number*.¹² More formally, we can non-dimensionalize the momentum equation by writing

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{U}, \quad \hat{\mathbf{x}} = \frac{\mathbf{x}}{L}, \quad \hat{t} = \frac{t}{T}, \quad \hat{\phi} = \frac{\phi}{\Phi}, \quad (1.195)$$

where the terms with hats on are *non-dimensional* values of the variables and the capitalized quantities are known as *scaling values*, and these are the approximate magnitudes of the variables. We choose the non-dimensionalization so that the non-dimensional variables are of order unity. Thus, for example, we choose U so that $u = \mathcal{O}(U)$, where this should be taken to mean that the magnitude of the variable u is approximately U , or that $u \sim U$, and we say that ‘ u scales like U ’. [This $\mathcal{O}()$ notation differs from the conventional mathematical meaning of ‘order’, in which $a = O(\epsilon^\alpha)$ represents a limit in which $a/\epsilon^\alpha \rightarrow \text{constant}$ as $\epsilon \rightarrow 0$.] Thus, if there are well-defined length and velocity scales in the problem, the non-dimensional variables are of order unity; that is, $\hat{u} = \mathcal{O}(1)$, and similarly for the other variables.

Because there are no external forces in this problem, appropriate scaling values for time and pressure are

$$T = \frac{L}{U}, \quad \Phi = U^2. \quad (1.196)$$

Substituting (1.195) and (1.196) into the momentum equation we obtain

$$\frac{U^2}{L} \left[\frac{\partial \hat{\mathbf{v}}}{\partial \hat{t}} + (\hat{\mathbf{v}} \cdot \nabla) \hat{\mathbf{v}} \right] = -\frac{U^2}{L} \nabla \hat{\phi} + \frac{\nu U}{L^2} \nabla^2 \hat{\mathbf{v}}, \quad (1.197)$$

where we use the convention that when ∇ operates on a non-dimensional variable it is a non-dimensional operator. Equation (1.197) simplifies to

$$\frac{\partial \hat{\mathbf{v}}}{\partial \hat{t}} + (\hat{\mathbf{v}} \cdot \nabla) \hat{\mathbf{v}} = -\nabla \hat{\phi} + \frac{1}{Re} \nabla^2 \hat{\mathbf{v}}, \quad (1.198)$$

where

$$Re \equiv \frac{UL}{\nu} \quad (1.199)$$

is, again, the Reynolds number. If we have chosen our length and velocity scales sensibly — that is, if we have scaled them properly — each variable in (1.198) is order unity, with the viscous term being multiplied by the parameter $1/Re$. There are two important conclusions:

- (i) The ratio of the importance of the inertial terms to the viscous terms is given by the *Reynolds number*, defined by (1.199). In the absence of other forces, such as those due to gravity and rotation, the Reynolds number is the only non-dimensional parameter explicitly appearing in the momentum equation. Hence its value, along with the boundary conditions, controls the behaviour of the system.
- (ii) More generally, by scaling the equations of motion appropriately the parameters determining the behaviour of the system become explicit. *Scaling the equations is intelligent non-dimensionalization.*

Notes

1 Joseph-Louis Lagrange (1736–1813) was a Franco-Italian, born and raised in Turin who then lived and worked mainly in Germany and France. He made notable contributions in analysis, number theory and mechanics and was recognized as one of the greatest mathematicians of the eighteenth century. He laid the foundations of the calculus of variations (to wit, the ‘Lagrange multiplier’) and first formulated the principle of least action, and his treatise *Mécanique Analytique* (1788) provides a unified analytic framework (it contains no diagrams, a feature emulated in Whittaker’s *Treatise on Analytical Dynamics*, 1927) for all Newtonian mechanics.

Leonard Euler (1707–1783), a Swiss mathematician who lived and worked for extended periods in Berlin and St. Petersburg, made important contributions in many areas of mathematics and mechanics, including the analytical treatment of algebra, the theory of equations, calculus, number theory and classical mechanics. He was the first to establish the form of the equations of motion of fluid mechanics, writing down both the field description of fluids and what we now call the material or advective derivative.

Truesdell (1954) points out that ‘Eulerian’ and ‘Lagrangian’, especially the latter, are inappropriate eponyms. The so-called Eulerian description was introduced by d’Alembert in 1749 and generalized by Euler in 1752, and the so-called Lagrangian description was introduced by Euler in 1759. The modern confusion evidently stems from a monograph by Dirichlet in 1860 that credits Euler in 1757 and Lagrange in 1788 for the respective methods. Clifford Truesdell (1919–2000) was a remarkable figure himself, known both for his own contributions to many areas of continuum mechanics and for his scholarly investigations on the history of mathematics and science; he also had a trenchant and at times pungent writing style. See Ball & James (2002) for more details.

- 2 For example Batchelor (1967).
- 3 R_d and R_v are related by the molecular weights of water and dry air, M_v and M_d , so that $\alpha \equiv R_d/R_v = M_v/M_d = 0.62$. Rather than allowing the gas constant to vary, meteorologists sometimes incorporate the variation of humidity into the definition of temperature, so that instead of $p = \rho R_{eff}T$ we use $p = \rho R_d T_v$, so defining the ‘virtual temperature’, T_v . It is easy to show that $T_v \approx [1 + w(\alpha^{-1} - 1)]T$. Atmospheric GCMs often use a virtual temperature.
- 4 The form of (1.60) was suggested by de Szoeke (2004). Accurate semi-empirical formulae are available from Mellor (1991), Jackett & McDougall (1995), Wright (1997) and (with particular attention to high accuracy) McDougall *et al.* (2002). The formulae of McDougall *et al.* are derived from the Gibbs function of Feistel & Hagen (1995); see Feistel (2003) for an updated version. The so-called ‘international equation of state of seawater’ (UNESCO 1981) is an empirical equation that fits measurements to an accuracy of order 10^{-5} (see Fofonoff 1985); however, it is hard to compute and the Feistel-Hagen formulation is as accurate, or more so, and more consistent. Wright’s formulae are used for Fig. 1.3 and Fig. 1.5. A formula of Bryden (1973) is used for Fig. 1.4.
- 5 For a development of thermodynamics from its fundamentals see, e.g., Callen (1985).
- 6 See Reif (1965) for more discussion.
- 7 Landau & Lifshitz (1987) have a longer discussion (their chapter 6), and also approach the energy equation in a different way.
- 8 Claude-Louis-Marie-Henri Navier (1785–1836) was a French civil engineer, professor at the École Polytechnique and later at the École des Ponts et Chaussée. He was an expert in road and bridge building (he developed the theory of suspension bridges) and, relatedly, made lasting theoretical contributions to the theory of elasticity, being the first to publish a set of general equations for the dynamics of an elastic solid. In fluid mechanics, he laid down the now-called *Navier–Stokes equations*, including the viscous terms, in 1822.
- George Gabriel Stokes (1819–1903). Irish born (in Skreen, County Sligo), he was a professor of mathematics at Cambridge from 1849 until his retirement. As well as having a role in the development of fluid mechanics, especially through his considerations of viscous effects, Stokes worked on the dynamics of elasticity, fluorescence, the wave theory of light, and was (rather ill-advisedly in hindsight) a proponent of the idea of an ether permeating all space.
- 9 Nevertheless, potential temperature and salinity are used as prognostic thermodynamic variables in many numerical ocean models. McDougall (2003) advocates using ‘potential enthalpy’ as a thermodynamic variable, because of its conservative properties and because it is a useful measure of heat content, suitably interpreted.
- 10 Drawing from de Szoeke (2004) and Young (2006). See also endnote 4 above. The accuracy of (1.147) with respect to saline variations could be further improved by adding small terms proportional to $TS \ln S$ and TS .
- 11 Bernoulli’s theorem was developed mainly by Daniel Bernoulli (1700–1782). It was based on earlier work on the conservation of energy that Daniel had done with his father, Johann Bernoulli (1667–1748), and so perhaps should be known as Bernoulli’s theorem. The two men fell out when Daniel was a young man, probably because of Johann’s jealousy of Daniel’s abilities, and subsequently had a very strained relationship. The Bernoulli family produced several (at least eight) talented mathematicians over three generations in the seventeenth and eighteenth centuries, and is often regarded as the most mathematically distinguished family of all time.
- 12 Osborne Reynolds (1842–1912) was an Irish born (Belfast) physicist who was professor of

engineering at Manchester University from 1868–1905. His early work was in electricity and magnetism, but he is now most famous for his work in hydrodynamics. The ‘Reynolds number’, which determines the ratio of inertial to viscous forces, and the ‘Reynolds stress’, which is the stress on the mean flow due to the fluctuating components, are both named after him. He was also one of the first scientists to think about the concept of group velocity.

Further reading

There are numerous books on hydrodynamics, some of them being:

Lamb, H., 1932. *Hydrodynamics*, 6th edn.

This is a classic text in the subject, although its notation is now too dated to make it really useful as an introduction. Another very well-known text is:

Batchelor, G. K. 1967. *An Introduction to Fluid Dynamics*.

Contains a very useful discussion of the fundamentals of fluid mechanics and a detailed derivation of the equations of motion.

Two other useful references are:

Tritton, D. J., 1988. *Physical Fluid Dynamics*, 2nd edn.

Kundu, P. & Cohen, I.M. 2002. *Fluid Mechanics*.

Both are introductions written at the advanced undergraduate/beginning graduate level, and are easier-going than Batchelor. Kundu and Cohen’s book has more material on geophysical fluid dynamics.

There are also numerous books on thermodynamics, two clear and useful ones being:

Reif, F., 1965. *Fundamentals of Statistical and Thermal Physics*,

Callen, H. B., 1985. *Thermodynamics and an Introduction to Thermostatistics*.

Reif’s book has become something of a classic, and Callen provides a rather more axiomatic approach.

Thermodynamic effects in fluids are discussed by:

Landau, L. D. & Lifshitz, E. M. 1987. *Fluid Mechanics*.

Salmon, R., 1998. *Lectures on Geophysical Fluid Dynamics*.

Salmon emphasizes the fundamentals, and Landau and Lifshitz provide a discussion of fluids with more than one component.

Problems

It is by the solution of problems that the investigator tests the temper of his steel; he finds new methods and new outlooks, and gains a wider and freer horizon.

David Hilbert (1862–1943), on presenting the Hilbert Problems.

1.1 For an infinitesimal volume, informally show that

$$\frac{D}{Dt}(\rho\varphi\Delta V) = \rho\Delta V\frac{D\varphi}{Dt}, \quad (P1.1)$$

where φ is some (differentiable) property of the fluid. Hence informally deduce that

$$\frac{D}{Dt}\int_V \rho\varphi \, dV = \int_V \rho \frac{D\varphi}{Dt} \, dV. \quad (P1.2)$$

- 1.2 Show that the derivative of an integral is given by

$$\frac{d}{dt} \int_{x_1(t)}^{x_2(t)} \varphi(x, t) dx = \int_{x_1}^{x_2} \frac{\partial \varphi}{\partial t} dx + \frac{dx_2}{dt} \varphi(x_2, t) - \frac{dx_1}{dt} \varphi(x_1, t). \quad (\text{P1.3})$$

By generalizing to three dimensions show that the material derivative of an integral of a fluid property is given by

$$\frac{D}{Dt} \int_V \varphi(\mathbf{x}, t) dV = \int_V \frac{\partial \varphi}{\partial t} dV + \int_S \varphi \mathbf{v} \cdot d\mathbf{S} = \int_V \left[\frac{\partial \varphi}{\partial t} + \nabla \cdot (\mathbf{v} \varphi) \right] dV, \quad (\text{P1.4})$$

where the surface integral (\int_S) is over the surface bounding the volume V . Hence deduce that

$$\frac{D}{Dt} \int_V \rho \varphi dV = \int_V \rho \frac{D\varphi}{Dt} dV. \quad (\text{P1.5})$$

- 1.3 Why is there no diffusion term in the mass continuity equation? Suppose that a fluid contains a binary mixture of dry air and water vapour. Show that the change in mass of a parcel of air due to the diffusion of water vapour is exactly balanced by the diffusion of dry air in the opposite direction.
- 1.4 By invoking Galilean invariance we can often choose, without loss of generality, the basic state for problems in sound waves to be such that $u_0 \equiv 0$. The perturbation velocity is then certainly larger than the basic state velocity. How can we then justify ignoring the nonlinear term in the perturbation equation, as the term $u' \partial u' / \partial x$ is certainly no smaller than the linear term $u_0 \partial u' / \partial x$?
- 1.5 What amplitude of sound wave is required for the nonlinear terms to become important? Is this achieved at a rock concert (120 dB), or near a jet aircraft that is taking off (160 dB)?
- 1.6 Using the observed value of molecular diffusion of heat in water, estimate how long it would take for a temperature anomaly to mix from the top of the ocean to the bottom, assuming that molecular diffusion alone is responsible. Comment on whether you think the real ocean has reached equilibrium after the last ice age (which ended about 12 000 years ago).
- 1.7 Consider the following flow:

$$u = \Gamma z, \quad v = V \sin[k(x - ct)],$$

where Γ , V , k and c are positive constants. (This is similar to the flow in the mid-latitude troposphere — an eastward flow increasing with height, with a transverse wave superimposed.) Suppose that $\Gamma z > c$ for the region of interest. Consider particles located along the $y = 0$ axis at $t = 0$, and compute their position at some later time t . Compare this with the *streamfunction* for the flow at the same time. (Hint: show that the meridional particle displacement is $\eta = \psi / (u - c)$, where ψ is the streamfunction and u and c are parameters.)

- 1.8 ♦ Consider the two-dimensional flow:

$$u = A(y) \sin \omega t, \quad v = A(y) \cos \omega t.$$

The time-mean of this at a fixed point flow is zero. If A is independent of y , then fluid parcels move clockwise in a circle. What is its radius? If A does depend on y , find an *approximate* expression for the average drift of a particle,

$$\lim_{t \rightarrow \infty} \frac{\mathbf{x}(\mathbf{a}, t)}{t}$$

where \mathbf{a} is a particle label and A is suitably ‘small’. Be precise about what small means.

- 1.9 (a) Suppose that a sealed, insulated container consists of two compartments, and that one of them is filled with an ideal gas and the other is a vacuum. The partition separating the compartments is removed. How does the temperature of the gas change? (Answer:

it stays the same. Explain.) Obtain an expression for the final potential temperature, in terms of the initial temperature of the gas and the volumes of the two compartments. Reconcile your answers with the first law of thermodynamics for an ideal gas, that

$$dQ = T d\eta = c_p \frac{d\theta}{\theta} = dI + dW = c_v dT + p d\alpha. \quad (\text{P1.6})$$

(b) A dry parcel that is ascending adiabatically through the atmosphere will generally cool as it moves to lower pressure and expands, and its potential temperature stays the same. How can this be consistent with your answer to part (a)?

- 1.10 Show that adiabatic flow in an ideal gas satisfies $p\rho^{-\gamma} = \text{constant}$, where $\gamma = c_p/c_v$.
- 1.11 (a) Show that for an ideal gas (1.116) is equivalent to (1.99). You may use the Maxwell relation $(\partial \alpha / \partial \eta)_p = (\partial T / \partial p)_\eta$.
 (b) Show that for an ideal gas (1.127) is equivalent to (1.98).
- 1.12 Show that it follows directly from the equation of state, $P = RT/\alpha$, that the internal energy of an ideal gas is a function of temperature only.

Solution: from (1.87) and $p = RT/\alpha$ we have

$$d\eta = \frac{1}{T} \left(\frac{\partial I}{\partial T} \right)_\alpha dT + \left[\frac{1}{T} \left(\frac{\partial I}{\partial \alpha} \right)_T + \frac{R}{\alpha} \right] d\alpha. \quad (\text{P1.7})$$

But, mathematically,

$$d\eta = \left(\frac{\partial \eta}{\partial T} \right)_\alpha dT + \left(\frac{\partial \eta}{\partial \alpha} \right)_T d\alpha. \quad (\text{P1.8})$$

Equating the coefficients of dT and $d\alpha$ in these two expressions gives

$$\left(\frac{\partial \eta}{\partial T} \right)_\alpha = \frac{1}{T} \left(\frac{\partial I}{\partial T} \right)_\alpha \quad \text{and} \quad \left(\frac{\partial \eta}{\partial \alpha} \right)_T = \frac{1}{T} \left(\frac{\partial I}{\partial \alpha} \right)_T + \frac{R}{\alpha}. \quad (\text{P1.9})$$

Noting that $\partial^2 \eta / (\partial \alpha \partial T) = \partial^2 \eta / (\partial T \partial \alpha)$ we obtain

$$\frac{\partial^2 I}{\partial \alpha \partial T} = \frac{\partial^2 I}{\partial T \partial \alpha} - \frac{1}{T} \left(\frac{\partial I}{\partial \alpha} \right)_T. \quad (\text{P1.10})$$

Thus, $(\partial I / \partial \alpha)_T = 0$. Because, in general, the internal energy may be considered either a function of temperature and density or temperature and pressure, this proves that for an ideal gas the internal energy is a function *only* of temperature.

- 1.13 Show that it follows directly from the equation of state $P = RT/\alpha$, that for an ideal gas the heat capacity at constant volume, c_v , is, at most, a function of temperature.
- 1.14 Show that for an ideal gas

$$T d\eta = c_v dT + p d\alpha. \quad (\text{P1.11})$$

and that its internal energy is given by $I = \int c_v dT$.

Solution: let us regard η as a function of T and α , where α is the specific volume $1/\rho$. Then

$$\begin{aligned} T d\eta &= T \left(\frac{\partial \eta}{\partial T} \right)_\alpha dT + T \left(\frac{\partial \eta}{\partial \alpha} \right)_T d\alpha \\ &= c_v dT + T \left(\frac{\partial \eta}{\partial \alpha} \right)_T d\alpha \end{aligned} \quad (\text{P1.12})$$

by definition of c_v . For an ideal gas the internal energy is a function of temperature alone (problem 1.12), so that using (1.70) the pressure of a fluid $p = T (\partial \eta / \partial \alpha)_I = T (\partial \eta / \partial \alpha)_T$ and (P1.12) becomes

$$T d\eta = c_v dT + p d\alpha \quad (\text{P1.13})$$

But, *in general*, the fundamental thermodynamic relation is

$$T d\eta = dI + p d\alpha. \quad (\text{P1.14})$$

The terms on the right-hand side of (P1.13) are identifiable as the change in the internal energy and the work done on a fluid, and so $dI = c_v dT$. The heat capacity need not necessarily be constant, although for air it very nearly is, but it must be a function of temperature only.

- 1.15 (a) Beginning with the expression for potential temperature for a simple ideal gas, $\theta = T(p_R/p)^\kappa$, where $\kappa = R/c_p$, show that

$$d\theta = (\theta/T)(dT - (\alpha/c_p)dp), \quad (\text{P1.15})$$

and that the first law of thermodynamics may be written as

$$dQ = T d\eta = c_p(T/\theta)d\theta. \quad (\text{P1.16})$$

(b) Verify that $d\eta = c_p(p_R, \theta)d\theta/\theta$, for a fluid parcel of constant composition that obeys (1.147).

- 1.16 Obtain an expression for the Gibbs function of a simple ideal gas, in terms of pressure and temperature.
- 1.17 From (1.113) derive the conventional equation of state for an ideal gas, and obtain expressions for the heat capacities.
- 1.18 A parcel of water is added to the ocean surface that is denser than any water in the ocean. Suppose the parcel sinks adiabatically to the ocean bottom. Estimate the change in temperature that the parcel undergoes, being explicit about the assumptions you make.
- 1.19 Consider an ocean at rest with known vertical profiles of potential temperature and salinity, $\theta(z)$ and $S(z)$. Suppose that we also know the equation of state in the form $\rho = \rho(\theta, S, p)$. Obtain an expression for the buoyancy frequency. Check your expression by substituting the equation of state for an ideal gas and recovering a known expression for the buoyancy frequency.
- 1.20 Consider a liquid, sitting in a container, with a free surface at the top (at $z = H$). The liquid obeys the equation of state $\rho = \rho_0[1 - \beta_T(T - T_0)]$, and its internal energy, I , is given by $I = c_p T$. Suppose that the fluid is heated, and its temperature rises uniformly by ΔT , and the free surface rises by a small amount ΔH . Obtain an expression for the ratio of the change in internal energy to the change in gravitational potential energy (GPE) of the ocean, and show that it is related to the scale height (1.131). If global warming increases the ocean temperature by 4 K, what is the ratio of the change of GPE to the change of I ? Estimate also the average rise in sea level due to thermal expansion.

Partial solution: the change in internal energy and in GPE are

$$\Delta I = c_p \rho_1 H_1 (T_2 - T_1), \quad \Delta GPE = \rho_1 g H_1 (H_2 - H_1)/2 = \rho_1 g H_1 \beta_T H_1 (T_2 - T_1)/2. \quad (\text{P1.17})$$

(Derive these. Use mass conservation where necessary. The subscripts 1 and 2 denote initial and final states.) Hence $\Delta GPE/\Delta I = g \beta_T H_1 / 2c_p$.

- 1.21 ♦ Obtain an expression, in terms of temperature and pressure, for the potential temperature of a van der Waals gas, with equation of state $(p + a/\alpha^2)(\alpha - b) = RT$, where a and b are constants. Show that it reduces to the expression for an ideal gas in the limit $a \rightarrow 0$, $b \rightarrow 0$.

If a body is moving in any direction, there is a force, arising from the Earth's rotation, which always deflects it to the right in the northern hemisphere, and to the left in the southern.

William Ferrel, *The influence of the Earth's rotation upon the relative motion of bodies near its surface*, 1858.

CHAPTER TWO

Effects of Rotation and Stratification

THE ATMOSPHERE AND OCEAN are shallow layers of fluid on a sphere in that their thickness or depth is much less than their horizontal extent. Furthermore, their motion is strongly influenced by two effects: rotation and stratification, the latter meaning that there is a mean vertical gradient of (potential) density that is often large compared with the horizontal gradient. Here we consider how the equations of motion are affected by these effects. First, we consider some elementary effects of rotation on a fluid and derive the Coriolis and centrifugal forces, and then we write down the equations of motion appropriate for motion on a sphere. Then we discuss some approximations to the equations of motion that are appropriate for large-scale flow in the ocean and atmosphere, in particular the hydrostatic and geostrophic approximations. Following this we discuss gravity waves, a particular kind of wave motion that is enabled by the presence of stratification, and finally we talk about how rotation leads to the production of certain types of boundary layers — Ekman layers — in rotating fluids.

2.1 THE EQUATIONS OF MOTION IN A ROTATING FRAME OF REFERENCE

Newton's second law of motion, that the acceleration on a body is proportional to the imposed force divided by the body's mass, applies in so-called inertial frames of reference. The Earth rotates with a period of almost 24 hours (23h 56m) relative to the distant stars, and the surface of the Earth manifestly is not, in that sense, an inertial frame. Nevertheless, because the surface of the Earth is moving (in fact at speeds of up to a few hundreds of metres per second) it is very convenient to describe the flow relative to the Earth's surface, rather than in some inertial frame. This necessitates recasting the equations into a form that is appropriate for a rotating frame of reference, and that is the subject of this section.

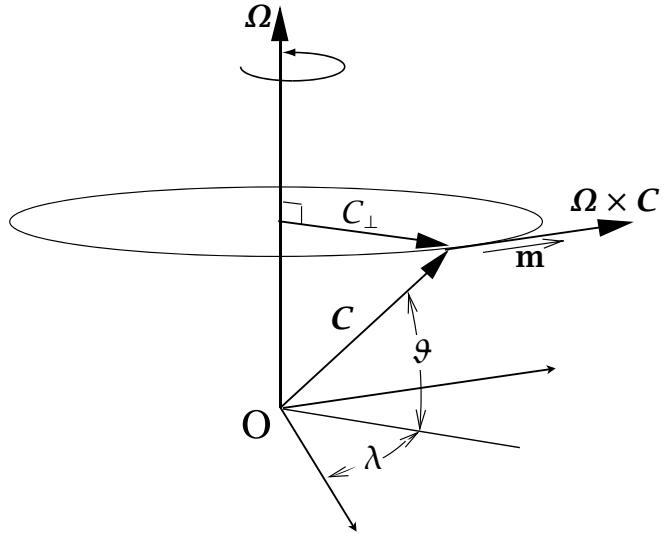


Fig. 2.1 A vector \mathbf{C} rotating at an angular velocity $\boldsymbol{\Omega}$. It appears to be a constant vector in the rotating frame, whereas in the inertial frame it evolves according to $(d\mathbf{C}/dt)_I = \boldsymbol{\Omega} \times \mathbf{C}$.

2.1.1 Rate of change of a vector

Consider first a vector \mathbf{C} of constant length rotating relative to an inertial frame at a constant angular velocity $\boldsymbol{\Omega}$. Then, in a frame rotating with that same angular velocity it appears stationary and constant. If in a small interval of time δt the vector \mathbf{C} rotates through a small angle $\delta\lambda$ then the change in \mathbf{C} , as perceived in the inertial frame, is given by (see Fig. 2.1)

$$\delta\mathbf{C} = |\mathbf{C}| \cos \vartheta \delta\lambda \mathbf{m}, \quad (2.1)$$

where the vector \mathbf{m} is the unit vector in the direction of change of \mathbf{C} , which is perpendicular to both \mathbf{C} and $\boldsymbol{\Omega}$. But the rate of change of the angle λ is just, by definition, the angular velocity so that $\delta\lambda = |\boldsymbol{\Omega}|\delta t$ and

$$\delta\mathbf{C} = |\mathbf{C}| |\boldsymbol{\Omega}| \sin \hat{\vartheta} \mathbf{m} \delta t = \boldsymbol{\Omega} \times \mathbf{C} \delta t. \quad (2.2)$$

using the definition of the vector cross product, where $\hat{\vartheta} = (\pi/2 - \vartheta)$ is the angle between $\boldsymbol{\Omega}$ and \mathbf{C} . Thus

$$\left(\frac{d\mathbf{C}}{dt} \right)_I = \boldsymbol{\Omega} \times \mathbf{C}, \quad (2.3)$$

where the left-hand side is the rate of change of \mathbf{C} as perceived in the inertial frame.

Now consider a vector \mathbf{B} that changes in the inertial frame. In a small time δt the change in \mathbf{B} as seen in the rotating frame is related to the change seen in the inertial frame by

$$(\delta\mathbf{B})_I = (\delta\mathbf{B})_R + (\delta\mathbf{B})_{rot}, \quad (2.4)$$

where the terms are, respectively, the change seen in the inertial frame, the change due to the vector itself changing as measured in the rotating frame, and the change due to the rotation. Using (2.2) $(\delta\mathbf{B})_{rot} = \boldsymbol{\Omega} \times \mathbf{B} \delta t$, and so the rates of change of the vector \mathbf{B} in the inertial and rotating frames are related by

$$\left(\frac{d\mathbf{B}}{dt} \right)_I = \left(\frac{d\mathbf{B}}{dt} \right)_R + \boldsymbol{\Omega} \times \mathbf{B}.$$

(2.5)

This relation applies to a vector \mathbf{B} that, as measured at any one time, is the same in both inertial and rotating frames.

2.1.2 Velocity and acceleration in a rotating frame

The velocity of a body is not measured to be the same in the inertial and rotating frames, so care must be taken when applying (2.5) to velocity. First apply (2.5) to \mathbf{r} , the position of a particle to obtain

$$\left(\frac{d\mathbf{r}}{dt} \right)_I = \left(\frac{d\mathbf{r}}{dt} \right)_R + \boldsymbol{\Omega} \times \mathbf{r} \quad (2.6)$$

or

$$\mathbf{v}_I = \mathbf{v}_R + \boldsymbol{\Omega} \times \mathbf{r}. \quad (2.7)$$

We refer to \mathbf{v}_R and \mathbf{v}_I as the relative and inertial velocity, respectively, and (2.7) relates the two. Apply (2.5) again, this time to the velocity \mathbf{v}_R to give

$$\left(\frac{d\mathbf{v}_R}{dt} \right)_I = \left(\frac{d\mathbf{v}_R}{dt} \right)_R + \boldsymbol{\Omega} \times \mathbf{v}_R, \quad (2.8)$$

or, using (2.7)

$$\left(\frac{d}{dt} (\mathbf{v}_I - \boldsymbol{\Omega} \times \mathbf{r}) \right)_I = \left(\frac{d\mathbf{v}_R}{dt} \right)_R + \boldsymbol{\Omega} \times \mathbf{v}_R, \quad (2.9)$$

or

$$\left(\frac{d\mathbf{v}_I}{dt} \right)_I = \left(\frac{d\mathbf{v}_R}{dt} \right)_R + \boldsymbol{\Omega} \times \mathbf{v}_R + \frac{d\boldsymbol{\Omega}}{dt} \times \mathbf{r} + \boldsymbol{\Omega} \times \left(\frac{d\mathbf{r}}{dt} \right)_I. \quad (2.10)$$

Then, noting that

$$\left(\frac{d\mathbf{r}}{dt} \right)_I = \left(\frac{d\mathbf{r}}{dt} \right)_R + \boldsymbol{\Omega} \times \mathbf{r} = (\mathbf{v}_R + \boldsymbol{\Omega} \times \mathbf{r}), \quad (2.11)$$

and assuming that the rate of rotation is constant, (2.10) becomes

$$\left(\frac{d\mathbf{v}_R}{dt} \right)_R = \left(\frac{d\mathbf{v}_I}{dt} \right)_I - 2\boldsymbol{\Omega} \times \mathbf{v}_R - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}). \quad (2.12)$$

This equation may be interpreted as follows. The term on the left-hand side is the rate of change of the relative velocity as measured in the rotating frame. The first term on the right-hand side is the rate of change of the inertial velocity as measured in the inertial frame (or, loosely, the inertial acceleration). Thus, by Newton's second law, it is equal to the force on a fluid parcel divided by its mass. The second and third terms on the right-hand side (including the minus signs) are the *Coriolis force* and the *centrifugal force* per unit mass. Neither of these are true forces — they may be thought of as quasi-forces (i.e., 'as if' forces); that is, when a body is observed from a rotating frame it seems to behave as if unseen forces are present that affect its motion. If (2.12) is written, as is common, with the terms $+2\boldsymbol{\Omega} \times \mathbf{v}_r$ and $+\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r})$ on the left-hand side then these terms should be referred to as the Coriolis and centrifugal *accelerations*.¹

Centrifugal force

If \mathbf{r}_\perp is the perpendicular distance from the axis of rotation (see Fig. 2.1 and substitute \mathbf{r} for \mathbf{C}), then, because $\boldsymbol{\Omega}$ is perpendicular to \mathbf{r}_\perp , $\boldsymbol{\Omega} \times \mathbf{r} = \boldsymbol{\Omega} \times \mathbf{r}_\perp$. Then, using the vector identity $\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}_\perp) = (\boldsymbol{\Omega} \cdot \mathbf{r}_\perp)\boldsymbol{\Omega} - (\boldsymbol{\Omega} \cdot \boldsymbol{\Omega})\mathbf{r}_\perp$ and noting that the first term is zero, we see that the centrifugal force per unit mass is just given by

$$\mathbf{F}_{ce} = -\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) = \boldsymbol{\Omega}^2 \mathbf{r}_\perp. \quad (2.13)$$

This may usefully be written as the gradient of a scalar potential,

$$\mathbf{F}_{ce} = -\nabla \Phi_{ce}. \quad (2.14)$$

where $\Phi_{ce} = -(\boldsymbol{\Omega}^2 r_\perp^2)/2 = -(\boldsymbol{\Omega} \times \mathbf{r}_\perp)^2/2$.

Coriolis force

The Coriolis force per unit mass is:

$$\mathbf{F}_{Co} = -2\boldsymbol{\Omega} \times \mathbf{v}_R. \quad (2.15)$$

It plays a central role in much of geophysical fluid dynamics and will be considered extensively later on. For now, we just note three basic properties.

- (i) There is no Coriolis force on bodies that are stationary in the rotating frame.
- (ii) The Coriolis force acts to deflect moving bodies at right angles to their direction of travel.
- (iii) The Coriolis force does no work on a body because it is perpendicular to the velocity, and so $\mathbf{v}_R \cdot (\boldsymbol{\Omega} \times \mathbf{v}_R) = 0$.

2.1.3 Momentum equation in a rotating frame

Since (2.12) simply relates the accelerations of a particle in the inertial and rotating frames, then in the rotating frame of reference the momentum equation may be written

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{v} = -\frac{1}{\rho} \nabla p - \nabla \Phi, \quad (2.16)$$

incorporating the centrifugal term into the potential, Φ . We have dropped the subscript R ; henceforth, unless we need to be explicit, all velocities without a subscript will be considered to be relative to the rotating frame.

2.1.4 Mass and tracer conservation in a rotating frame

Let ϕ be a scalar field that, in the inertial frame, obeys

$$\frac{D\phi}{Dt} + \phi \nabla \cdot \mathbf{v}_I = 0. \quad (2.17)$$

Now, observers in both the rotating and inertial frame measure the same value of ϕ . Further, $D\phi/Dt$ is simply the rate of change of ϕ associated with a material parcel, and therefore is reference frame invariant. Thus,

$$\left(\frac{D\phi}{Dt} \right)_R = \left(\frac{D\phi}{Dt} \right)_I, \quad (2.18)$$

where $(D\phi/Dt)_R = (\partial\phi/\partial t)_R + \mathbf{v}_R \cdot \nabla\phi$ and $(D\phi/Dt)_I = (\partial\phi/\partial t)_I + \mathbf{v}_I \cdot \nabla\phi$ and the local temporal derivatives $(\partial\phi/\partial t)_R$ and $(\partial\phi/\partial t)_I$ are evaluated at fixed locations in the rotating and inertial frames, respectively.

Further, since $\mathbf{v} = \mathbf{v}_I - \boldsymbol{\Omega} \times \mathbf{r}$, we have that

$$\nabla \cdot \mathbf{v}_I = \nabla \cdot (\mathbf{v}_I - \boldsymbol{\Omega} \times \mathbf{r}) = \nabla \cdot \mathbf{v}_R \quad (2.19)$$

since $\nabla \cdot (\boldsymbol{\Omega} \times \mathbf{r}) = 0$. Thus, using (2.18) and (2.19), (2.17) is equivalent to

$$\frac{D\phi}{Dt} + \phi \nabla \cdot \mathbf{v} = 0, \quad (2.20)$$

where all observables are measured in the *rotating* frame. Thus, the equation for the evolution of a scalar whose measured value is the same in rotating and inertial frames is unaltered by the presence of rotation. In particular, the mass conservation equation is unaltered by the presence of rotation.

Although we have taken (2.18) as true a priori, the individual components of the material derivative differ in the rotating and inertial frames. In particular

$$\left(\frac{\partial\phi}{\partial t} \right)_I = \left(\frac{\partial\phi}{\partial t} \right)_R - (\boldsymbol{\Omega} \times \mathbf{r}) \cdot \nabla\phi \quad (2.21)$$

because $\boldsymbol{\Omega} \times \mathbf{r}$ is the velocity, in the inertial frame, of a uniformly rotating body. Similarly,

$$\mathbf{v}_I \cdot \nabla\phi = (\mathbf{v}_R + \boldsymbol{\Omega} \times \mathbf{r}) \cdot \nabla\phi. \quad (2.22)$$

Adding the last two equations reprises and confirms (2.18).

2.2 EQUATIONS OF MOTION IN SPHERICAL COORDINATES

The Earth is very nearly spherical and it might appear obvious that we should cast our equations in spherical coordinates. Although this does turn out to be true, the presence of a centrifugal force causes some complications which we must first discuss. The reader who is willing ab initio to treat the Earth as a perfect sphere and to neglect the horizontal component of the centrifugal force may skip the next section.

2.2.1 * The centrifugal force and spherical coordinates

The centrifugal force is a potential force, like gravity, and so we may therefore define an 'effective gravity' equal to the sum of the true, or Newtonian, gravity and the centrifugal force. The Newtonian gravitational force is directed approximately toward the centre of the Earth, with small deviations due mainly to the Earth's oblateness. The line of action of the effective gravity will in general differ slightly from this, and therefore have a component in the 'horizontal' plane, that is the plane perpendicular to the radial direction. The magnitude of the centrifugal force is $\Omega^2 r_\perp$, and so the effective gravity is given by

$$\mathbf{g} \equiv \mathbf{g}_{\text{eff}} = \mathbf{g}_{\text{grav}} + \Omega^2 \mathbf{r}_\perp, \quad (2.23)$$

where \mathbf{g}_{grav} is the Newtonian gravitational force due to the gravitational attraction of the Earth and \mathbf{r}_\perp is normal to the rotation vector (in the direction \mathbf{C} in Fig. 2.2), with $r_\perp = r \cos \theta$.

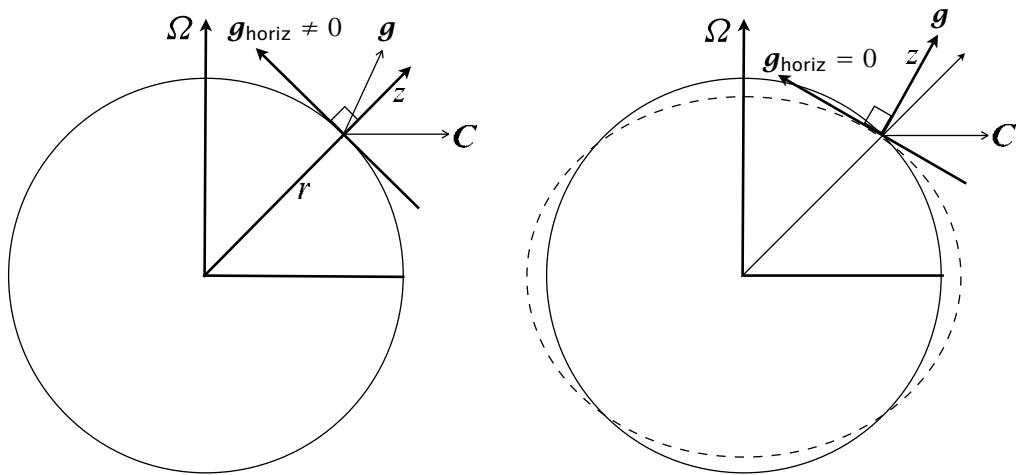


Fig. 2.2 Left: directions of forces and coordinates in true spherical geometry. \mathbf{g} is the effective gravity (including the centrifugal force, \mathbf{C}) and its horizontal component is evidently non-zero. Right: a modified coordinate system, in which the vertical direction is defined by the direction of \mathbf{g} , and so the horizontal component of \mathbf{g} is identically zero. The dashed line schematically indicates a surface of constant geopotential. The differences between the direction of \mathbf{g} and the direction of the radial coordinate, and between the sphere and the geopotential surface, are much exaggerated and in reality are similar to the thickness of the lines themselves.

Both gravity and centrifugal force are potential forces and therefore we may define the *geopotential*, Φ , such that

$$\mathbf{g} = -\nabla\Phi. \quad (2.24)$$

Surfaces of constant Φ are not quite spherical because r_\perp , and hence the centrifugal force, vary with latitude (Fig. 2.2); this has certain ramifications, as we now discuss.

The components of the centrifugal force parallel and perpendicular to the radial direction are $\Omega^2 r \cos^2 \vartheta$ and $\Omega^2 r \cos \vartheta \sin \vartheta$. Newtonian gravity is much larger than either of these, and at the Earth's surface the ratio of centrifugal to gravitational terms is approximately, and no more than,

$$\alpha \approx \frac{\Omega^2 a}{g} \approx \frac{(7.27 \times 10^{-5})^2 \times 6.4 \times 10^6}{10} \approx 3 \times 10^{-3}. \quad (2.25)$$

(Note that at the equator and pole the horizontal component of the centrifugal force is zero and the effective gravity is aligned with Newtonian gravity.) The angle between \mathbf{g} and the line to the centre of the Earth is given by a similar expression and so is also small, typically around 3×10^{-3} radians. However, the horizontal component of the centrifugal force is still large compared to the Coriolis force, their ratio in mid-latitudes being given by

$$\frac{\text{horizontal centrifugal force}}{\text{Coriolis force}} \approx \frac{\Omega^2 a \cos \vartheta \sin \vartheta}{2\Omega u} \approx \frac{\Omega a}{4|u|} \approx 10, \quad (2.26)$$

using $u = 10 \text{ m s}^{-1}$. The centrifugal term therefore dominates over the Coriolis term, and

is largely balanced by a pressure gradient force. Thus, if we adhered to true spherical coordinates, both the horizontal and radial components of the momentum equation would be dominated by a static balance between a pressure gradient and gravity or centrifugal terms. Although in principle there is nothing wrong with writing the equations this way, it obscures the dynamical balances involving the Coriolis force and pressure that determine the large-scale horizontal flow.

A way around this problem is to use the direction of the geopotential force to *define* the vertical direction, and then for all geometric purposes to regard the surfaces of constant Φ as if they were true spheres.² The horizontal component of effective gravity is then identically zero, and we have traded a potentially large dynamical error for a very small geometric error. In fact, over time, the Earth has developed an equatorial bulge to compensate for and neutralize the centrifugal force, so that the effective gravity does act in a direction virtually normal to the Earth's surface; that is, the surface of the Earth is an oblate spheroid of nearly constant geopotential. The geopotential Φ is then a function of the vertical coordinate alone, and for many purposes we can just take $\Phi = gz$; that is, the direction normal to geopotential surfaces, the local vertical, is, in this approximation, taken to be the direction of increasing r in spherical coordinates. It is because the oblateness is very small (the polar diameter is about 12 714 km, whereas the equatorial diameter is about 12 756 km) that using spherical coordinates is a very accurate way to map the spheroid, and if the angle between effective gravity and a natural direction of the coordinate system were not small then more heroic measures would be called for.

If the solid Earth did not bulge at the equator, the *behaviour* of the atmosphere and ocean would differ significantly from that of the present system. For example, the surface of the ocean is nearly a geopotential surface, and if the solid Earth were exactly spherical then the ocean would become much deeper at low latitudes and the ocean basins would dry out completely at high latitudes. We could still choose to use the spherical coordinate system discussed above to describe the dynamics, but the shape of the surface of the solid Earth would have to be represented by a topography, with the topographic height increasing monotonically polewards nearly everywhere.

2.2.2 Some identities in spherical coordinates

The location of a point is given by the coordinates (λ, ϑ, r) where λ is the angular distance eastwards (i.e., longitude), ϑ is angular distance polewards (i.e., latitude) and r is the radial distance from the centre of the Earth — see Fig. 2.3. (In some other fields of study co-latitude is used as a spherical coordinate.) If a is the radius of the Earth, then we also define $z = r - a$. At a given location we may also define the Cartesian increments $(\delta x, \delta y, \delta z) = (r \cos \vartheta \delta \lambda, r \delta \vartheta, \delta r)$.

For a scalar quantity ϕ the material derivative in spherical coordinates is

$$\frac{D\phi}{Dt} = \frac{\partial \phi}{\partial t} + \frac{u}{r \cos \vartheta} \frac{\partial \phi}{\partial \lambda} + \frac{v}{r} \frac{\partial \phi}{\partial \vartheta} + w \frac{\partial \phi}{\partial r}, \quad (2.27)$$

where the velocity components corresponding to the coordinates (λ, ϑ, r) are

$$(u, v, w) \equiv \left(r \cos \vartheta \frac{D\lambda}{Dt}, r \frac{D\vartheta}{Dt}, \frac{Dr}{Dt} \right). \quad (2.28)$$

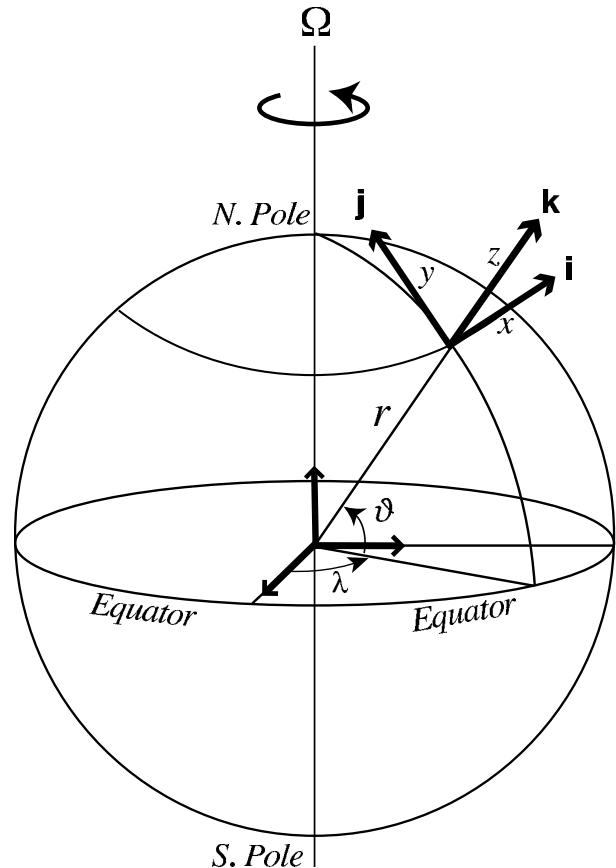


Fig. 2.3 The spherical coordinate system. The orthogonal unit vectors \mathbf{i} , \mathbf{j} and \mathbf{k} point in the direction of increasing longitude λ , latitude ϑ , and altitude z . Locally, one may apply a Cartesian system with variables x , y and z measuring distances along \mathbf{i} , \mathbf{j} and \mathbf{k} .

That is, u is the zonal velocity, v is the meridional velocity and w is the vertical velocity. If we define $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ to be the unit vectors in the direction of increasing (λ, ϑ, r) then

$$\mathbf{v} = \mathbf{i}u + \mathbf{j}v + \mathbf{k}w. \quad (2.29)$$

Note also that $D\mathbf{r}/Dt = D\mathbf{z}/Dt$.

The divergence of a vector $\mathbf{B} = \mathbf{i}B^\lambda + \mathbf{j}B^\vartheta + \mathbf{k}B^r$ is

$$\nabla \cdot \mathbf{B} = \frac{1}{\cos \vartheta} \left[\frac{1}{r} \frac{\partial B^\lambda}{\partial \lambda} + \frac{1}{r} \frac{\partial}{\partial \vartheta} (B^\vartheta \cos \vartheta) + \frac{\cos \vartheta}{r^2} \frac{\partial}{\partial r} (r^2 B^r) \right]. \quad (2.30)$$

The vector gradient of a scalar is:

$$\nabla \phi = \mathbf{i} \frac{1}{r \cos \vartheta} \frac{\partial \phi}{\partial \lambda} + \mathbf{j} \frac{1}{r} \frac{\partial \phi}{\partial \vartheta} + \mathbf{k} \frac{\partial \phi}{\partial r}. \quad (2.31)$$

The Laplacian of a scalar is:

$$\nabla^2 \phi \equiv \nabla \cdot \nabla \phi = \frac{1}{r^2 \cos \vartheta} \left[\frac{1}{\cos \vartheta} \frac{\partial^2 \phi}{\partial \lambda^2} + \frac{\partial}{\partial \vartheta} \left(\cos \vartheta \frac{\partial \phi}{\partial \vartheta} \right) + \cos \vartheta \frac{\partial}{\partial r} \left(r^2 \frac{\partial \phi}{\partial r} \right) \right]. \quad (2.32)$$

The curl of a vector is:

$$\text{curl } \mathbf{B} = \nabla \times \mathbf{B} = \frac{1}{r^2 \cos \vartheta} \begin{vmatrix} \mathbf{i} r \cos \vartheta & \mathbf{j} r & \mathbf{k} \\ \partial / \partial \lambda & \partial / \partial \vartheta & \partial / \partial r \\ B^\lambda r \cos \vartheta & B^\vartheta r & B^r \end{vmatrix}. \quad (2.33)$$

The vector Laplacian $\nabla^2 \mathbf{B}$ (used for example when calculating viscous terms in the momentum equation) may be obtained from the vector identity:

$$\nabla^2 \mathbf{B} = \nabla(\nabla \cdot \mathbf{B}) - \nabla \times (\nabla \times \mathbf{B}). \quad (2.34)$$

Only in Cartesian coordinates does this take the simple form:

$$\nabla^2 \mathbf{B} = \frac{\partial^2 \mathbf{B}}{\partial x^2} + \frac{\partial^2 \mathbf{B}}{\partial y^2} + \frac{\partial^2 \mathbf{B}}{\partial z^2}. \quad (2.35)$$

The expansion in spherical coordinates is, to most eyes, rather uninformative.

Rate of change of unit vectors

In spherical coordinates the defining unit vectors are \mathbf{i} , the unit vector pointing eastwards, parallel to a line of latitude; \mathbf{j} is the unit vector pointing polewards, parallel to a meridian; and \mathbf{k} , the unit vector pointing radially outward. The directions of these vectors change with location, and in fact this is the case in nearly all coordinate systems, with the notable exception of the Cartesian one, and thus their material derivative is not zero. One way to evaluate this is to consider geometrically how the coordinate axes change with position. Another way, and the way that we shall proceed, is to first obtain the effective rotation rate $\boldsymbol{\Omega}_{\text{flow}}$, relative to the Earth, of a unit vector as it moves with the flow, and then apply (2.3). Specifically, let the fluid velocity be $\mathbf{v} = (u, v, w)$. The meridional component, v , produces a displacement $r\delta\vartheta = v\delta t$, and this gives rise a local effective vector rotation rate around the local zonal axis of $-(v/r)\mathbf{i}$, the minus sign arising because a displacement in the direction of the north pole is produced by negative rotational displacement around the \mathbf{i} axis. Similarly, the zonal component, u , produces a displacement $\delta\lambda r \cos \vartheta = u\delta t$ and so an effective rotation rate, about the Earth's rotation axis, of $u/(r \cos \vartheta)$. Now, a rotation around the Earth's rotation axis may be written as (see Fig. 2.4)

$$\boldsymbol{\Omega} = \boldsymbol{\Omega}(\mathbf{j} \cos \vartheta + \mathbf{k} \sin \vartheta). \quad (2.36)$$

If the scalar rotation rate is not $\boldsymbol{\Omega}$ but is $u/(r \cos \vartheta)$, then the vector rotation rate is

$$\frac{u}{r \cos \vartheta} (\mathbf{j} \cos \vartheta + \mathbf{k} \sin \vartheta) = \mathbf{j} \frac{u}{r} + \mathbf{k} \frac{u \tan \vartheta}{r}. \quad (2.37)$$

Thus, the total rotation rate of a vector that moves with the flow is

$$\boldsymbol{\Omega}_{\text{flow}} = -\mathbf{i} \frac{v}{r} + \mathbf{j} \frac{u}{r} + \mathbf{k} \frac{u \tan \vartheta}{r}. \quad (2.38)$$

Applying (2.3) to (2.38), we find

$$\frac{D\mathbf{i}}{Dt} = \boldsymbol{\Omega}_{\text{flow}} \times \mathbf{i} = \frac{u}{r \cos \vartheta} (\mathbf{j} \sin \vartheta - \mathbf{k} \cos \vartheta), \quad (2.39a)$$

$$\frac{D\mathbf{j}}{Dt} = \boldsymbol{\Omega}_{\text{flow}} \times \mathbf{j} = -\mathbf{i} \frac{u}{r} \tan \vartheta - \mathbf{k} \frac{v}{r}, \quad (2.39b)$$

$$\frac{D\mathbf{k}}{Dt} = \boldsymbol{\Omega}_{\text{flow}} \times \mathbf{k} = \mathbf{i} \frac{u}{r} + \mathbf{j} \frac{v}{r}. \quad (2.39c)$$

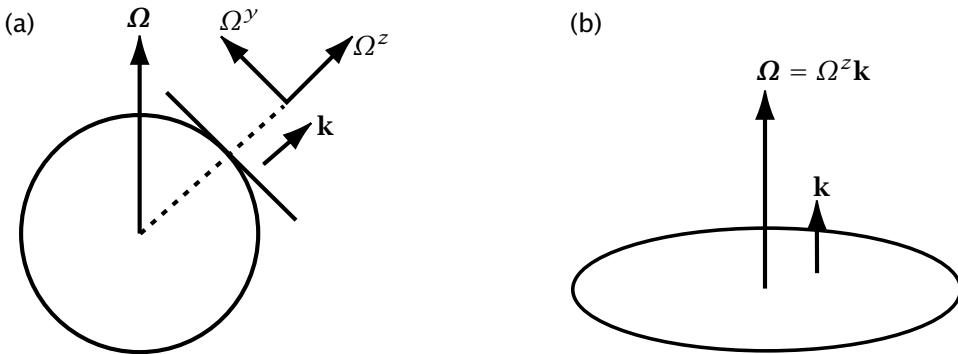


Fig. 2.4 (a) On the sphere the rotation vector Ω can be decomposed into two components, one in the local vertical and one in the local horizontal, pointing toward the pole. That is, $\Omega = \Omega_y \mathbf{j} + \Omega_z \mathbf{k}$ where $\Omega_y = \Omega \cos \theta$ and $\Omega_z = \Omega \sin \theta$. In geophysical fluid dynamics, the rotation vector in the local vertical is often the more important component in the horizontal momentum equations. On a rotating disk, (b), the rotation vector Ω is parallel to the local vertical \mathbf{k} .

2.2.3 Equations of motion

Mass Conservation and Thermodynamic Equation

The mass conservation equation, (1.36a), expanded in spherical co-ordinates, is

$$\frac{\partial \rho}{\partial t} + \frac{u}{r \cos \theta} \frac{\partial \rho}{\partial \lambda} + \frac{v}{r} \frac{\partial \rho}{\partial \theta} + w \frac{\partial \rho}{\partial r} + \frac{\rho}{r \cos \theta} \left[\frac{\partial u}{\partial \lambda} + \frac{\partial}{\partial \theta} (v \cos \theta) + \frac{1}{r} \frac{\partial}{\partial r} (w r^2 \cos \theta) \right] = 0. \quad (2.40)$$

Equivalently, using the form (1.36b), this is

$$\frac{\partial \rho}{\partial t} + \frac{1}{r \cos \theta} \frac{\partial (u \rho)}{\partial \lambda} + \frac{1}{r \cos \theta} \frac{\partial}{\partial \theta} (v \rho \cos \theta) + \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 w \rho) = 0. \quad (2.41)$$

The thermodynamic equation, (1.108), is a tracer advection equation. Thus, using (2.27), its (adiabatic) spherical coordinate form is

$$\frac{D\theta}{Dt} = \frac{\partial \theta}{\partial t} + \frac{u}{r \cos \theta} \frac{\partial \theta}{\partial \lambda} + \frac{v}{r} \frac{\partial \theta}{\partial \theta} + w \frac{\partial \theta}{\partial r} = 0, \quad (2.42)$$

and similarly for tracers such as water vapour or salt.

Momentum Equation

Recall that the inviscid momentum equation is:

$$\frac{D\mathbf{v}}{Dt} + 2\Omega \times \mathbf{v} = -\frac{1}{\rho} \nabla p - \nabla \Phi, \quad (2.43)$$

where Φ is the geopotential. In spherical coordinates the directions of the coordinate axes change with position and so the component expansion of (2.43) is

$$\frac{D\mathbf{v}}{Dt} = \frac{Du}{Dt} \mathbf{i} + \frac{Dv}{Dt} \mathbf{j} + \frac{Dw}{Dt} \mathbf{k} + u \frac{Di}{Dt} + v \frac{Dj}{Dt} + w \frac{Dk}{Dt} \quad (2.44a)$$

$$= \frac{Du}{Dt} \mathbf{i} + \frac{Dv}{Dt} \mathbf{j} + \frac{Dw}{Dt} \mathbf{k} + \boldsymbol{\Omega}_{\text{flow}} \times \mathbf{v}, \quad (2.44b)$$

using (2.39). Using either (2.44a) and the expressions for the rates of change of the unit vectors given in (2.39), or (2.44b) and the expression for $\boldsymbol{\Omega}_{\text{flow}}$ given in (2.38), (2.44) becomes

$$\begin{aligned} \frac{D\mathbf{v}}{Dt} &= \mathbf{i} \left(\frac{Du}{Dt} - \frac{uv \tan \vartheta}{r} + \frac{uw}{r} \right) + \mathbf{j} \left(\frac{Dv}{Dt} + \frac{u^2 \tan \vartheta}{r} + \frac{vw}{r} \right) \\ &\quad + \mathbf{k} \left(\frac{Dw}{Dt} - \frac{u^2 + v^2}{r} \right). \end{aligned} \quad (2.45)$$

Using the definition of a vector cross product the Coriolis term is:

$$\begin{aligned} 2\boldsymbol{\Omega} \times \mathbf{v} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 2\Omega \cos \vartheta & 2\Omega \sin \vartheta \\ u & v & w \end{vmatrix} \\ &= \mathbf{i} (2\Omega w \cos \vartheta - 2\Omega v \sin \vartheta) + \mathbf{j} 2\Omega u \sin \vartheta - \mathbf{k} 2\Omega u \cos \vartheta. \end{aligned} \quad (2.46)$$

Using (2.45) and (2.46), and the gradient operator given by (2.31), the momentum equation (2.43) becomes:

$$\frac{Du}{Dt} - \left(2\Omega + \frac{u}{r \cos \vartheta} \right) (v \sin \vartheta - w \cos \vartheta) = -\frac{1}{\rho r \cos \vartheta} \frac{\partial p}{\partial \lambda}, \quad (2.47a)$$

$$\frac{Dv}{Dt} + \frac{wv}{r} + \left(2\Omega + \frac{u}{r \cos \vartheta} \right) u \sin \vartheta = -\frac{1}{\rho r} \frac{\partial p}{\partial \vartheta}, \quad (2.47b)$$

$$\frac{Dw}{Dt} - \frac{u^2 + v^2}{r} - 2\Omega u \cos \vartheta = -\frac{1}{\rho} \frac{\partial p}{\partial r} - g. \quad (2.47c)$$

The terms involving Ω are called Coriolis terms, and the quadratic terms on the left-hand sides involving $1/r$ are often called metric terms.

2.2.4 The primitive equations

The so-called *primitive equations* of motion are simplifications of the above equations frequently used in atmospheric and oceanic modelling.³ Three related approximations are involved.

(i) *The hydrostatic approximation.* In the vertical momentum equation the gravitational term is assumed to be balanced by the pressure gradient term, so that

$$\frac{\partial p}{\partial z} = -\rho g. \quad (2.48)$$

The advection of vertical velocity, the Coriolis terms, and the metric term $(u^2 + v^2)/r$ are all neglected.

(ii) *The shallow-fluid approximation.* We write $r = a + z$ where the constant a is the radius of the Earth and z increases in the radial direction. The coordinate r is then replaced by a except where it used as the differentiating argument. Thus, for example,

$$\frac{1}{r^2} \frac{\partial(r^2 w)}{\partial r} \rightarrow \frac{\partial w}{\partial z}. \quad (2.49)$$

- (iii) *The traditional approximation.* Coriolis terms in the horizontal momentum equations involving the vertical velocity, and the still smaller metric terms uw/r and vw/r , are neglected.

The second and third of these approximations should be taken, or not, together, the underlying reason being that they both relate to the presumed small aspect ratio of the motion, so the approximations succeed or fail together. If we make one approximation but not the other then we are being asymptotically inconsistent, and angular momentum and energy conservation are not assured (see section 2.2.7 and problem 2.13). The hydrostatic approximation also depends on the small aspect ratio of the flow, but in a slightly different way. For large-scale flow in the terrestrial atmosphere and ocean all three approximations are in fact all very accurate approximations. We defer a more complete treatment until section 2.7, in part because a treatment of the hydrostatic approximation is done most easily in the context of the Boussinesq equations, derived in section 2.4.

Making these approximations, the momentum equations become

$$\frac{Du}{Dt} - 2\Omega \sin \vartheta v - \frac{uv}{a} \tan \vartheta = -\frac{1}{a\rho \cos \vartheta} \frac{\partial p}{\partial \lambda}, \quad (2.50a)$$

$$\frac{Dv}{Dt} + 2\Omega \sin \vartheta u + \frac{u^2 \tan \vartheta}{a} = -\frac{1}{\rho a} \frac{\partial p}{\partial \vartheta}, \quad (2.50b)$$

$$0 = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g, \quad (2.50c)$$

where

$$\frac{D}{Dt} = \left(\frac{\partial}{\partial t} + \frac{u}{a \cos \vartheta} \frac{\partial}{\partial \lambda} + \frac{v}{a} \frac{\partial}{\partial \vartheta} + w \frac{\partial}{\partial z} \right). \quad (2.51)$$

We note the ubiquity of the factor $2\Omega \sin \vartheta$, and take the opportunity to define the *Coriolis parameter*, $f \equiv 2\Omega \sin \vartheta$.

The corresponding mass conservation equation for a shallow fluid layer is:

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{u}{a \cos \vartheta} \frac{\partial \rho}{\partial \lambda} + \frac{v}{a} \frac{\partial \rho}{\partial \vartheta} + w \frac{\partial \rho}{\partial z} \\ + \rho \left[\frac{1}{a \cos \vartheta} \frac{\partial u}{\partial \lambda} + \frac{1}{a \cos \vartheta} \frac{\partial}{\partial \vartheta} (v \cos \vartheta) + \frac{\partial w}{\partial z} \right] = 0, \end{aligned} \quad (2.52)$$

or equivalently,

$$\frac{\partial \rho}{\partial t} + \frac{1}{a \cos \vartheta} \frac{\partial (u\rho)}{\partial \lambda} + \frac{1}{a \cos \vartheta} \frac{\partial}{\partial \vartheta} (v \rho \cos \vartheta) + \frac{\partial (w\rho)}{\partial z} = 0. \quad (2.53)$$

2.2.5 Primitive equations in vector form

The primitive equations may be written in a compact vector form provided we make a slight reinterpretation of the material derivative of the coordinate axes. Let $\mathbf{u} = u\mathbf{i} + v\mathbf{j} + 0\mathbf{k}$ be the horizontal velocity. The primitive equations (2.50a) and (2.50b) may be written as

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -\frac{1}{\rho} \nabla_z p, \quad (2.54)$$

where $\mathbf{f} = f\mathbf{k} = 2\Omega \sin \vartheta \mathbf{k}$ and $\nabla_z p = [(a \cos \vartheta)^{-1} \partial p / \partial \lambda, a^{-1} \partial p / \partial \vartheta]$, the gradient operator at constant z . In (2.54) the material derivative of the horizontal velocity is given by

$$\frac{D\mathbf{u}}{Dt} = \mathbf{i} \frac{Du}{Dt} + \mathbf{j} \frac{Dv}{Dt} + u \frac{Di}{Dt} + v \frac{Dj}{Dt}, \quad (2.55)$$

where instead of (2.39) we have

$$\frac{Di}{Dt} = \tilde{\Omega}_{flow} \times \mathbf{i} = \mathbf{j} \frac{u \tan \vartheta}{a}, \quad (2.56a)$$

$$\frac{Dj}{Dt} = \tilde{\Omega}_{flow} \times \mathbf{j} = -\mathbf{i} \frac{u \tan \vartheta}{a}, \quad (2.56b)$$

where $\tilde{\Omega}_{flow} = \mathbf{k} u \tan \vartheta / a$ [which is the vertical component of (2.38), with r replaced by a]. The advection of the horizontal wind \mathbf{u} is still by the three-dimensional velocity \mathbf{v} . The vertical momentum equation is the hydrostatic equation, (2.50c), and the mass conservation equation is

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0 \quad \text{or} \quad \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (2.57)$$

where D/Dt on a scalar is given by (2.51), and the second expression is written out in full in (2.53).

2.2.6 The vector invariant form of the momentum equation

The ‘vector invariant’ form of the momentum equation is so-called because it appears to take the same form in all coordinate systems — there is no advective derivative of the coordinate system to worry about. With the aid of the identity $(\mathbf{v} \cdot \nabla) \mathbf{v} = -\mathbf{v} \times \boldsymbol{\omega} + \nabla(\mathbf{v}^2/2)$, where $\boldsymbol{\omega} \equiv \nabla \times \mathbf{v}$ is the relative vorticity, the three-dimensional momentum equation may be written:

$$\frac{\partial \mathbf{v}}{\partial t} + (2\boldsymbol{\Omega} + \boldsymbol{\omega}) \times \mathbf{v} = -\frac{1}{\rho} \nabla p - \frac{1}{2} \nabla \mathbf{v}^2 + \mathbf{g}. \quad (2.58)$$

In spherical coordinates the relative vorticity is given by:

$$\begin{aligned} \boldsymbol{\omega} &= \nabla \times \mathbf{v} = \frac{1}{r^2 \cos \vartheta} \begin{vmatrix} \mathbf{i} r \cos \vartheta & \mathbf{j} r & \mathbf{k} \\ \partial / \partial \lambda & \partial / \partial \vartheta & \partial / \partial r \\ ur \cos \vartheta & rv & w \end{vmatrix} \\ &= \mathbf{i} \frac{1}{r} \left(\frac{\partial w}{\partial \vartheta} - \frac{\partial (rv)}{\partial r} \right) - \mathbf{j} \frac{1}{r \cos \vartheta} \left(\frac{\partial w}{\partial \lambda} - \frac{\partial}{\partial r} (ur \cos \vartheta) \right) \\ &\quad + \mathbf{k} \frac{1}{r \cos \vartheta} \left(\frac{\partial v}{\partial \lambda} - \frac{\partial}{\partial \vartheta} (u \cos \vartheta) \right). \end{aligned} \quad (2.59)$$

Making the traditional and shallow fluid approximations, the horizontal components of (2.58) may be written

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{f} + \mathbf{k} \zeta) \times \mathbf{u} + w \frac{\partial \mathbf{u}}{\partial z} = -\frac{1}{\rho} \nabla_z p - \frac{1}{2} \nabla \mathbf{u}^2, \quad (2.60)$$

where $\mathbf{u} = (u, v, 0)$, $\mathbf{f} = \mathbf{k} 2\Omega \sin \vartheta$, ∇_z is the horizontal gradient operator (the gradient at a constant value of z), and using (2.59), ζ is given by

$$\zeta = \frac{1}{a \cos \vartheta} \frac{\partial v}{\partial \lambda} - \frac{1}{a \cos \vartheta} \frac{\partial}{\partial \vartheta} (u \cos \vartheta) = \frac{1}{a \cos \vartheta} \frac{\partial v}{\partial \lambda} - \frac{1}{a} \frac{\partial u}{\partial \vartheta} + \frac{u}{a} \tan \vartheta. \quad (2.61)$$

The separate components of the momentum equation are given by:

$$\frac{\partial u}{\partial t} - (f + \zeta)v + w \frac{\partial u}{\partial z} = -\frac{1}{a\rho \cos \vartheta} \left(\frac{1}{\rho} \frac{\partial p}{\partial \lambda} + \frac{1}{2} \frac{\partial \mathbf{u}^2}{\partial \lambda} \right), \quad (2.62)$$

and

$$\frac{\partial v}{\partial t} + (f + \zeta)u + w \frac{\partial v}{\partial z} = -\frac{1}{a} \left(\frac{1}{\rho} \frac{\partial p}{\partial \vartheta} + \frac{1}{2} \frac{\partial \mathbf{u}^2}{\partial \vartheta} \right). \quad (2.63)$$

Related expressions are given in problem 2.3, and we treat vorticity at greater length in chapter 4.

2.2.7 Angular momentum

The zonal momentum equation can be usefully expressed as a statement about axial angular momentum; that is, angular momentum about the rotation axis. The zonal angular momentum per unit mass is the component of angular momentum in the direction of the axis of rotation and it is given by, without making any shallow atmosphere approximation,

$$m = (u + \Omega r \cos \vartheta)r \cos \vartheta. \quad (2.64)$$

The evolution equation for this quantity follows from the zonal momentum equation and has the simple form

$$\frac{Dm}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial \lambda}, \quad (2.65)$$

where the material derivative is

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \frac{u}{r \cos \vartheta} \frac{\partial}{\partial \lambda} + \frac{v}{r} \frac{\partial}{\partial \vartheta} + w \frac{\partial}{\partial r}. \quad (2.66)$$

Using the mass continuity equation, (2.65) can be written as

$$\frac{D\rho m}{Dt} + \rho m \nabla \cdot \mathbf{v} = -\frac{\partial p}{\partial \lambda} \quad (2.67)$$

or

$$\frac{\partial \rho m}{\partial t} + \frac{1}{r \cos \vartheta} \frac{\partial(\rho um)}{\partial \lambda} + \frac{1}{r \cos \vartheta} \frac{\partial}{\partial \vartheta} (\rho v m \cos \vartheta) + \frac{\partial}{\partial z} (\rho m w) = -\frac{\partial p}{\partial \lambda}. \quad (2.68)$$

This is an angular momentum conservation equation.

If the fluid is confined to a shallow layer near the surface of a sphere, then we may replace r , the radial coordinate, by a , the radius of the sphere, in the definition of m , and we define $\tilde{m} \equiv (u + \Omega a \cos \vartheta)a \cos \vartheta$. Then (2.65) is replaced by

$$\frac{D\tilde{m}}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial \lambda}, \quad (2.69)$$

where now

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \frac{u}{a \cos \vartheta} \frac{\partial}{\partial \lambda} + \frac{v}{a} \frac{\partial}{\partial \vartheta} + w \frac{\partial}{\partial z}. \quad (2.70)$$

Using mass continuity, (2.69) may be written as

$$\frac{\partial \rho \tilde{m}}{\partial t} + \frac{u}{a \cos \vartheta} \frac{\partial \tilde{m}}{\partial \lambda} + \frac{v}{a} \frac{\partial \tilde{m}}{\partial \vartheta} + w \frac{\partial \tilde{m}}{\partial z} = -\frac{1}{\rho} \frac{\partial p}{\partial \lambda}, \quad (2.71)$$

which is the appropriate angular momentum conservation equation for a shallow atmosphere.

* *From angular momentum to the spherical component equations*

An alternative way of deriving the three components of the momentum equation in spherical polar coordinates is to *begin* with (2.65) and the principle of conservation of energy. That is, we take the equations for conservation of angular momentum and energy as true a priori and demand that the forms of the momentum equation be constructed to satisfy these. Expanding the material derivative in (2.65), noting that $D\mathbf{r}/Dt = \mathbf{u}$ and $D\cos \theta/Dt = -(\mathbf{v}/r) \sin \theta$, immediately gives (2.47a). Multiplication by u then yields

$$u \frac{Du}{Dt} - 2\Omega uv \sin \theta + 2\Omega uw \cos \theta - \frac{u^2 v \tan \theta}{r} + \frac{u^2 w}{r} = -\frac{u}{\rho r \cos \theta} \frac{\partial p}{\partial \lambda}. \quad (2.72)$$

Now suppose that the meridional and vertical momentum equations are of the form

$$\frac{Dv}{Dt} + \text{Coriolis and metric terms} = -\frac{1}{\rho r} \frac{\partial p}{\partial \theta} \quad (2.73a)$$

$$\frac{Dw}{Dt} + \text{Coriolis and metric terms} = -\frac{1}{\rho r} \frac{\partial p}{\partial r}, \quad (2.73b)$$

but that we do not know what form the Coriolis and metric terms take. To determine that form, construct the kinetic energy equation by multiplying (2.73) by v and w , respectively. Now, the metric terms must vanish when we sum the resulting equations along with (2.72), so that (2.73a) must contain the Coriolis term $2\Omega u \sin \theta$ as well as the metric term $u^2 \tan \theta/r$, and (2.73b) must contain the term $-2\Omega u \cos \theta$ as well as the metric term u^2/r . But if (2.73b) contains the term u^2/r it must also contain the term v^2/r by isotropy, and therefore (2.73a) must also contain the term $v w/r$. In this way, (2.47) is precisely reproduced, although the sceptic might argue that the uniqueness of the form has not been demonstrated.

A particular advantage of this approach arises in determining the appropriate momentum equations that conserve angular momentum and energy in the shallow-fluid approximation. We begin with (2.69) and expand to obtain (2.50a). Multiplying by u gives

$$u \frac{Du}{Dt} - 2\Omega uv \sin \theta - \frac{u^2 v \tan \theta}{a} = -\frac{u}{\rho a \cos \theta} \frac{\partial p}{\partial \lambda}. \quad (2.74)$$

To ensure energy conservation, the meridional momentum equation must contain the Coriolis term $2\Omega u \sin \theta$ and the metric term $u^2 \tan \theta/a$, but the vertical momentum equation must have neither of the metric terms appearing in (2.47c). Thus we deduce the following equations:

$$\frac{Du}{Dt} - \left(2\Omega \sin \theta + \frac{u \tan \theta}{a} \right) v = -\frac{1}{\rho a \cos \theta} \frac{\partial p}{\partial \lambda}, \quad (2.75a)$$

$$\frac{Dv}{Dt} + \left(2\Omega \sin \theta + \frac{u \tan \theta}{a} \right) u = -\frac{1}{\rho a} \frac{\partial p}{\partial \theta}, \quad (2.75b)$$

$$\frac{Dw}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial r} - g. \quad (2.75c)$$

This equation set, when used in conjunction with the thermodynamic and mass continuity equations, conserves appropriate forms of angular momentum and energy. In the hydrostatic approximation the material derivative of w in (2.75c) is *additionally* neglected. Thus, the hydrostatic approximation is mathematically and physically consistent with the shallow-fluid

approximation, but it is an additional approximation with slightly different requirements that one may choose, rather than being required, to make. From an asymptotic perspective, the difference lies in the small parameter necessary for either approximation to hold, namely:

$$\text{shallow fluid and traditional approximations: } \gamma \equiv \frac{H}{a} \ll 1, \quad (2.76a)$$

$$\text{small aspect ratio for hydrostatic approximation: } \alpha \equiv \frac{H}{L} \ll 1, \quad (2.76b)$$

where L is the horizontal scale of the motion and a is the radius of the Earth. For hemispheric or global scale phenomena $L \sim a$ and the two approximations coincide. (Requirement (2.76b) for the hydrostatic approximation is derived in section 2.7.)

2.3 CARTESIAN APPROXIMATIONS: THE TANGENT PLANE

2.3.1 The f-plane

Although the rotation of the Earth is central for many dynamical phenomena, the sphericity of the Earth is not always so. This is especially true for phenomena on a scale somewhat smaller than global where the use of spherical coordinates becomes awkward, and it is more convenient to use a locally Cartesian representation of the equations. Referring to Fig. 2.4 we will define a plane tangent to the surface of the Earth at a latitude ϑ_0 , and then use a Cartesian coordinate system (x, y, z) to describe motion on that plane. For small excursions on the plane, $(x, y, z) \approx (a\lambda \cos \vartheta_0, a(\vartheta - \vartheta_0), z)$. Consistently, the velocity is $\mathbf{v} = (u, v, w)$, so that u, v and w are the components of the velocity *in the tangent plane*, in approximately in the east-west, north-south and vertical directions, respectively.

The momentum equations for flow in this plane are then

$$\frac{\partial u}{\partial t} + (\mathbf{v} \cdot \nabla) u + 2(\Omega^y w - \Omega^z v) = -\frac{1}{\rho} \frac{\partial p}{\partial x}, \quad (2.77a)$$

$$\frac{\partial v}{\partial t} + (\mathbf{v} \cdot \nabla) v + 2(\Omega^z u - \Omega^x w) = -\frac{1}{\rho} \frac{\partial p}{\partial y}, \quad (2.77b)$$

$$\frac{\partial w}{\partial t} + (\mathbf{v} \cdot \nabla) w + 2(\Omega^x v - \Omega^y u) = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g, \quad (2.77c)$$

where the rotation vector $\boldsymbol{\Omega} = \Omega^x \mathbf{i} + \Omega^y \mathbf{j} + \Omega^z \mathbf{k}$ and $\Omega^x = 0$, $\Omega^y = \Omega \cos \vartheta_0$ and $\Omega^z = \Omega \sin \vartheta_0$. If we make the traditional approximation, and so ignore the components of $\boldsymbol{\Omega}$ not in the direction of the local vertical, then

$$\frac{Du}{Dt} - f_0 v = -\frac{1}{\rho} \frac{\partial p}{\partial x}, \quad (2.78a)$$

$$\frac{Dv}{Dt} + f_0 u = -\frac{1}{\rho} \frac{\partial p}{\partial y}, \quad (2.78b)$$

$$\frac{Dw}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g. \quad (2.78c)$$

where $f_0 = 2\Omega^z = 2\Omega \sin \vartheta_0$. Defining the horizontal velocity vector $\mathbf{u} = (u, v, 0)$, the first two equations may also be written as

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f}_0 \times \mathbf{u} = -\frac{1}{\rho} \nabla_z p, \quad (2.79)$$

where $D\mathbf{u}/Dt = \partial\mathbf{u}/\partial t + \mathbf{v} \cdot \nabla \mathbf{u}$, $f_0 = 2\Omega \sin \vartheta_0 \mathbf{k}$, and \mathbf{k} is the direction perpendicular to the plane (it does not change its orientation with latitude). These equations are, evidently, exactly the same as the momentum equations in a system in which the rotation vector is aligned with the local vertical, as illustrated in the right-hand panel in Fig. 2.4 (on page 60). They will describe flow on the surface of a rotating sphere to a good approximation provided the flow is of limited latitudinal extent so that the effects of sphericity are unimportant; we have made what is known as the *f-plane* approximation since the Coriolis parameter is a constant. We may in addition make the hydrostatic approximation, in which case (2.78c) becomes the familiar $\partial p/\partial z = -\rho g$.

2.3.2 The beta-plane approximation

The magnitude of the vertical component of rotation varies with latitude, and this has important dynamical consequences. We can approximate this effect by allowing the effective rotation vector to vary. Thus, noting that, for small variations in latitude,

$$f = 2\Omega \sin \vartheta \approx 2\Omega \sin \vartheta_0 + 2\Omega(\vartheta - \vartheta_0) \cos \vartheta_0, \quad (2.80)$$

then on the tangent plane we may mimic this by allowing the Coriolis parameter to vary as

$f = f_0 + \beta y$

(2.81)

where $f_0 = 2\Omega \sin \vartheta_0$ and $\beta = \partial f/\partial y = (2\Omega \cos \vartheta_0)/a$. This important approximation is known as the *beta-plane*, or *β -plane*, approximation; it captures the the most important *dynamical* effects of sphericity, without the complicating *geometric* effects, which are not essential to describe many phenomena. The momentum equations (2.78) are unaltered except that f_0 is replaced by $f_0 + \beta y$ to represent a varying Coriolis parameter. Thus, sphericity combined with rotation is dynamically equivalent to a *differentially rotating* system. For future reference, we write down the β -plane horizontal momentum equations:

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -\frac{1}{\rho} \nabla_z p, \quad (2.82)$$

where $\mathbf{f} = (f_0 + \beta y)\hat{\mathbf{k}}$. In component form this equation becomes

$$\frac{Du}{Dt} - fv = -\frac{1}{\rho} \frac{\partial p}{\partial x}, \quad \frac{Dv}{Dt} + fu = -\frac{1}{\rho} \frac{\partial p}{\partial y}. \quad (2.83a,b)$$

The mass conservation, thermodynamic and hydrostatic equations in the β -plane approximation are the same as the usual Cartesian, *f*-plane, forms of those equations.

2.4 EQUATIONS FOR A STRATIFIED OCEAN: THE BOUSSINESQ APPROXIMATION

The density variations in the ocean are quite small compared to the mean density, and we may exploit this to derive somewhat simpler but still quite accurate equations of motion. Let us first examine how much density does vary in the ocean.

2.4.1 Variation of density in the ocean

The variations of density in the ocean are due to three effects: the compression of water by pressure (which we denote as $\Delta_p \rho$), the thermal expansion of water if its temperature changes ($\Delta_T \rho$), and the haline contraction if its salinity changes ($\Delta_S \rho$). How big are these? An appropriate equation of state to approximately evaluate these effects is the linear one

$$\rho = \rho_0 \left[1 - \beta_T (T - T_0) + \beta_S (S - S_0) + \frac{p}{\rho_0 c_s^2} \right], \quad (2.84)$$

where $\beta_T \approx 2 \times 10^{-4} \text{ K}^{-1}$, $\beta_S \approx 10^{-3} \text{ psu}^{-1}$ and $c_s \approx 1500 \text{ m s}^{-1}$ (see the table on page 35). The three effects may then be evaluated as follows.

Pressure compressibility. We have $\Delta_p \rho \approx \Delta p / c_s^2 \approx \rho_0 g H / c_s^2$, where H is the depth and the pressure change is quite accurately evaluated using the hydrostatic approximation. Thus,

$$\frac{|\Delta_p \rho|}{\rho_0} \ll 1 \quad \text{if} \quad \frac{gH}{c_s^2} \ll 1, \quad (2.85)$$

or if $H \ll c_s^2/g$. The quantity $c_s^2/g \approx 200 \text{ km}$ is the density scale height of the ocean. Thus, the pressure at the bottom of the ocean (say $H = 10 \text{ km}$ in the deep trenches), enormous as it is, is insufficient to compress the water enough to make a significant change in its density. Changes in density due to dynamical variations of pressure are small if the Mach number is small, and this is also the case.

Thermal expansion. We have $\Delta_T \rho \approx -\beta_T \rho_0 \Delta T$ and therefore

$$\frac{|\Delta_T \rho|}{\rho_0} \ll 1 \quad \text{if} \quad \beta_T \Delta T \ll 1. \quad (2.86)$$

For $\Delta T = 20 \text{ K}$, $\beta_T \Delta T \approx 4 \times 10^{-3}$, and evidently we would require temperature differences of order β_T^{-1} , or 5000 K to obtain order one variations in density.

Saline contraction. We have $\Delta_S \rho \approx \beta_S \rho_0 \Delta S$ and therefore

$$\frac{|\Delta_S \rho|}{\rho_0} \ll 1 \quad \text{if} \quad \beta_S \Delta S \ll 1. \quad (2.87)$$

As changes in salinity in the ocean rarely exceed 5 psu, for which $\beta_S \Delta S = 5 \times 10^{-3}$, the fractional change in the density of seawater is correspondingly very small.

Evidently, fractional density changes in the ocean are very small.

2.4.2 The Boussinesq equations

The *Boussinesq equations* are a set of equations that exploit the smallness of density variations in many liquids.⁴ To set notation we write

$$\rho = \rho_0 + \delta \rho(x, y, z, t) \quad (2.88a)$$

$$= \rho_0 + \hat{\rho}(z) + \rho'(x, y, z, t) \quad (2.88b)$$

$$= \tilde{\rho}(z) + \rho'(x, y, z, t), \quad (2.88c)$$

where ρ_0 is a constant and we assume that

$$|\hat{\rho}|, |\rho'|, |\delta\rho| \ll \rho_0. \quad (2.89)$$

We need not assume that $|\rho'| \ll |\hat{\rho}|$, but this is often the case in the ocean. To obtain the Boussinesq equations we will just use (2.88a), but (2.88c) will be useful for the anelastic equations considered later.

Associated with the reference density is a reference pressure that is defined to be in hydrostatic balance with it. That is,

$$p = p_0(z) + \delta p(x, y, z, t) \quad (2.90a)$$

$$= \tilde{p}(z) + p'(x, y, z, t), \quad (2.90b)$$

where $|\delta p| \ll p_0$, $|p'| \ll \tilde{p}$ and

$$\frac{dp_0}{dz} \equiv -g\rho_0, \quad \frac{d\tilde{p}}{dz} \equiv -g\tilde{\rho}. \quad (2.91a,b)$$

Note that $\nabla_z p = \nabla_z p' = \nabla_z \delta p$ and that $p_0 \approx \tilde{p}$ if $|\hat{\rho}| \ll \rho_0$.

Momentum equations

To obtain the Boussinesq equations we use $\rho = \rho_0 + \delta\rho$, and assume $\delta\rho/\rho_0$ is small. Without approximation, the momentum equation can be written as

$$(\rho_0 + \delta\rho) \left(\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{v} \right) = -\nabla \delta p - \frac{\partial p_0}{\partial z} \mathbf{k} - g(\rho_0 + \delta\rho) \mathbf{k}, \quad (2.92)$$

and using (2.91a) this becomes, again without approximation,

$$(\rho_0 + \delta\rho) \left(\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{v} \right) = -\nabla \delta p - g\delta\rho \mathbf{k}. \quad (2.93)$$

If density variations are small this becomes

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{v} = -\nabla \phi + b \mathbf{k} \quad , \quad (2.94)$$

where $\phi = \delta p/\rho_0$ and $b = -g\delta\rho/\rho_0$ is the *buoyancy*. Note that we should not and do not neglect the term $g\delta\rho$, for there is no reason to believe it to be small ($\delta\rho$ may be small, but g is big). Equation (2.94) is the momentum equation in the Boussinesq approximation, and it is common to say that the Boussinesq approximation ignores all variations of density of a fluid in the momentum equation, except when associated with the gravitational term.

For most large-scale motions in the ocean the *deviation* pressure and density fields are also approximately in hydrostatic balance, and in that case the vertical component of (2.94) becomes

$$\frac{\partial \phi}{\partial z} = b. \quad (2.95)$$

A condition for (2.95) to hold is that vertical accelerations are small *compared to* $g\delta\rho/\rho_0$, and not compared to the acceleration due to gravity itself. For more discussion of this point, see section 2.7.

Mass Conservation

The unapproximated mass conservation equation is

$$\frac{D\delta\rho}{Dt} + (\rho_0 + \delta\rho) \nabla \cdot \mathbf{v} = 0. \quad (2.96)$$

Provided that time scales advectively — that is to say that D/Dt scales in the same way as $\mathbf{v} \cdot \nabla$ — then we may approximate this equation by

$$\boxed{\nabla \cdot \mathbf{v} = 0}, \quad (2.97)$$

which is the same as that for a constant density fluid. This *absolutely does not* allow one to go back and use (2.96) to say that $D\delta\rho/Dt = 0$; the evolution of density is given by the thermodynamic equation in conjunction with an equation of state, and this should not be confused with the mass conservation equation. Note also that in eliminating the time-derivative of density we eliminate the possibility of sound waves.

Thermodynamic equation and equation of state

The Boussinesq equations are closed by the addition of an equation of state, a thermodynamic equation and, as appropriate, a salinity equation. Neglecting salinity for the moment, a useful starting point is to write the thermodynamic equation, (1.116), as

$$\frac{D\rho}{Dt} - \frac{1}{c_s^2} \frac{Dp}{Dt} = \frac{\dot{Q}}{(\partial\eta/\partial\rho)_p T} \approx -\dot{Q} \left(\frac{\rho_0 \beta_T}{c_p} \right) \quad (2.98)$$

using $(\partial\eta/\partial\rho)_p = (\partial\eta/\partial T)_p (\partial T/\partial\rho)_p \approx c_p/(T\rho_0\beta_T)$. Given the expansions (2.88a) and (2.90a), (2.98) can be written to a good approximation as

$$\frac{D\delta\rho}{Dt} - \frac{1}{c_s^2} \frac{Dp_0}{Dt} = -\dot{Q} \left(\frac{\rho_0 \beta_T}{c_p} \right), \quad (2.99)$$

or, using (2.91a),

$$\frac{D}{Dt} \left(\delta\rho + \frac{\rho_0 g}{c_s^2} z \right) = -\dot{Q} \left(\frac{\rho_0 \beta_T}{c_p} \right), \quad (2.100)$$

as in (1.119). The severest approximation to this is to neglect the second term in brackets on the left-hand side, and noting that $b = -g\delta\rho/\rho_0$ we obtain

$$\boxed{\frac{Db}{Dt} = \dot{b}}, \quad (2.101)$$

where $\dot{b} = g\beta_T\dot{Q}/c_p$. The momentum equation (2.94), mass continuity equation (2.97) and thermodynamic equation (2.101) then form a closed set, called the *simple Boussinesq equations*.

A somewhat more accurate approach is to include the compressibility of the fluid that is due to the hydrostatic pressure. From (2.100), the potential density is given by $\delta\rho_{\text{pot}} = \delta\rho + \rho_0 g z / c_s^2$, and so the *potential buoyancy*, that is the buoyancy based on potential density, is given by

$$b_\sigma \equiv -g \frac{\delta\rho_{\text{pot}}}{\rho_0} = -\frac{g}{\rho_0} \left(\delta\rho + \frac{\rho_0 g z}{c_s^2} \right) = b - g \frac{z}{H_\rho}, \quad (2.102)$$

where $H_\rho = c_s^2/g$. The thermodynamic equation, (2.100), may then be written

$$\frac{Db_\sigma}{Dt} = \dot{b}_\sigma, \quad (2.103)$$

where $\dot{b}_\sigma = \dot{b}$. Buoyancy itself is obtained from b_σ by the 'equation of state', $b = b_\sigma + gz/H_\rho$.

In many applications we may need to use a still more accurate equation of state. In that case (and see section 1.6.2) we replace (2.101) by the thermodynamic equations

$$\frac{D\theta}{Dt} = \dot{\theta}, \quad \frac{DS}{Dt} = \dot{S},$$

(2.104a,b)

where θ is the potential temperature and S is salinity, along with an equation of state. The equation of state has the general form $b = b(\theta, S, p)$, but to be consistent with the level of approximation in the other Boussinesq equations we replace p by the hydrostatic pressure calculated with the reference density, that is by $-\rho_0 gz$, and the equation of state then takes the general form

$$b = b(\theta, S, z).$$

(2.105)

An example of (2.105) is (1.156), taken with the definition of buoyancy $b = -g\delta\rho/\rho_0$. The closed set of equations (2.94), (2.97), (2.104) and (2.105) are the *general Boussinesq equations*. Using an accurate equation of state and the Boussinesq approximation is the procedure used in many comprehensive ocean general circulation models. The Boussinesq equations, which with the hydrostatic and traditional approximations are often considered to be the oceanic primitive equations, are summarized in the shaded box on the next page.

* Mean stratification and the buoyancy frequency

The processes that cause density to vary in the vertical often differ from those that cause it to vary in the horizontal. For this reason it is sometimes useful to write $\rho = \rho_0 + \hat{\rho}(z) + \rho'(x, y, z, t)$ and define $\tilde{b}(z) \equiv -g\hat{\rho}/\rho_0$ and $b' \equiv -g\rho'/\rho_0$. Using the hydrostatic equation to evaluate pressure, the thermodynamic equation (2.98) becomes, to a good approximation,

$$\frac{Db'}{Dt} + N^2 w = 0, \quad (2.106)$$

where D/Dt remains a three-dimensional operator and

$$N^2(z) = \left(\frac{d\tilde{b}}{dz} - \frac{g^2}{c_s^2} \right) = \frac{d\tilde{b}_\sigma}{dz}, \quad (2.107)$$

where $\tilde{b}_\sigma = \tilde{b} - gz/H_\rho$. The quantity N^2 is a measure of the mean stratification of the fluid, and is equal to the vertical gradient of the mean potential buoyancy. N is known as the buoyancy frequency, something we return to in section 2.9. Equations (2.106) and (2.107) also hold in the simple Boussinesq equations, but with $c_s^2 = \infty$.

Summary of Boussinesq Equations

The simple Boussinesq equations are, for an inviscid fluid:

$$\text{momentum equations: } \frac{D\mathbf{v}}{Dt} + \mathbf{f} \times \mathbf{v} = -\nabla\phi + b\mathbf{k}, \quad (B.1)$$

$$\text{mass conservation: } \nabla \cdot \mathbf{v} = 0, \quad (B.2)$$

$$\text{buoyancy equation: } \frac{Db}{Dt} = \dot{b}. \quad (B.3)$$

A more general form replaces the buoyancy equation by:

$$\text{thermodynamic equation: } \frac{D\theta}{Dt} = \dot{\theta}, \quad (B.4)$$

$$\text{salinity equation: } \frac{DS}{Dt} = \dot{S}, \quad (B.5)$$

$$\text{equation of state: } b = b(\theta, S, \phi). \quad (B.6)$$

Energy conservation is only assured if $b = b(\theta, S, z)$.

2.4.3 Energetics of the Boussinesq system

In a uniform gravitational field but with no other forcing or dissipation, we write the simple Boussinesq equations as

$$\frac{D\mathbf{v}}{Dt} + 2\Omega \times \mathbf{v} = b\mathbf{k} - \nabla\phi, \quad \nabla \cdot \mathbf{v} = 0, \quad \frac{Db}{Dt} = 0. \quad (2.108a,b,c)$$

From (2.108a) and (2.108b) the kinetic energy density evolution (cf. section 1.10) is given by

$$\frac{1}{2} \frac{D\mathbf{v}^2}{Dt} = bw - \nabla \cdot (\phi\mathbf{v}), \quad (2.109)$$

where the constant reference density ρ_0 is omitted. Let us now define the potential $\Phi \equiv -z$, so that $\nabla\Phi = -\mathbf{k}$ and

$$\frac{D\Phi}{Dt} = \nabla \cdot (\mathbf{v}\Phi) = -w, \quad (2.110)$$

and using this and (2.108c) gives

$$\frac{D}{Dt}(b\Phi) = -wb. \quad (2.111)$$

Adding (2.111) to (2.109) and expanding the material derivative gives

$$\frac{\partial}{\partial t} \left(\frac{1}{2} \mathbf{v}^2 + b\Phi \right) + \nabla \cdot \left[\mathbf{v} \left(\frac{1}{2} \mathbf{v}^2 + b\Phi + \phi \right) \right] = 0. \quad (2.112)$$

This constitutes an energy equation for the Boussinesq system, and may be compared to (1.186). (Also see problem 2.14.) The energy density (divided by ρ_0) is just $\mathbf{v}^2/2 + b\Phi$. What does the term $b\Phi$ represent? Its integral, multiplied by ρ_0 , is the potential energy of the

flow minus that of the basic state, or $\int g(\rho - \rho_0)z \, dz$. If there were a heating term on the right-hand side of (2.108c) this would directly provide a source of potential energy, rather than internal energy as in the compressible system. Because the fluid is incompressible, there is no conversion from kinetic and potential energy into internal energy.

** Energetics with a general equation of state*

Now consider the energetics of the general Boussinesq equations. Suppose first that we allow the equation of state to be a function of pressure; the equations of motion are then (2.108) except that (2.108c) is replaced by

$$\frac{D\theta}{Dt} = 0, \quad \frac{DS}{Dt} = 0, \quad b = b(\theta, S, \phi). \quad (2.113a,b,c)$$

A little algebraic experimentation will reveal that no energy conservation law of the form (2.112) generally exists for this system! The problem arises because, by requiring the fluid to be incompressible, we eliminate the proper conversion of internal energy to kinetic energy. However, if we use the approximation $b = b(\theta, S, z)$, the system does conserve an energy, as we now show.⁵

Define the potential, Π , as the integral of b at constant potential temperature and salinity; that is

$$\Pi(\theta, S, z) \equiv - \int_a^z b \, dz', \quad (2.114)$$

where a is any constant, so that $\partial\Pi/\partial z = -b$. Taking the material derivative of the left-hand side gives

$$\frac{D\Pi}{Dt} = \left(\frac{\partial\Pi}{\partial\theta} \right)_{S,z} \frac{D\theta}{Dt} + \left(\frac{\partial\Pi}{\partial S} \right)_{\theta,z} \frac{DS}{Dt} + \left(\frac{\partial\Pi}{\partial z} \right)_{\theta,S} \frac{Dz}{Dt} = -bw, \quad (2.115)$$

using (2.113a,b). Combining (2.115) and (2.109) gives

$$\frac{\partial}{\partial t} \left(\frac{1}{2} \mathbf{v}^2 + \Pi \right) + \nabla \cdot \left[\mathbf{v} \left(\frac{1}{2} \mathbf{v}^2 + \Pi + \phi \right) \right] = 0. \quad (2.116)$$

Thus, energetic consistency is maintained with an arbitrary equation of state, provided the pressure is replaced by a function of z . If b is not an explicit function of z in the equation of state, the conservation law is identical to (2.112).

2.5 EQUATIONS FOR A STRATIFIED ATMOSPHERE: THE ANELASTIC APPROXIMATION

2.5.1 Preliminaries

In the atmosphere the density varies significantly, especially in the vertical. However deviations of both ρ and p from a statically balanced state are often quite small, and the relative vertical variation of potential temperature is also small. We can usefully exploit these observations to give a somewhat simplified set of equations, useful both for theoretical and numerical analyses because sound waves are eliminated by way of an ‘anelastic’ approximation.⁶ To begin we set

$$\rho = \tilde{\rho}(z) + \delta\rho(x, y, z, t), \quad p = \tilde{p}(z) + \delta p(x, y, z, t), \quad (2.117a,b)$$

where we assume that $|\delta\rho| \ll |\tilde{\rho}|$ and we define \tilde{p} such that

$$\frac{\partial \tilde{p}}{\partial z} \equiv -g\tilde{\rho}(z). \quad (2.118)$$

The notation is similar to that for the Boussinesq case except that, importantly, the density basic state is now a (given) function of vertical coordinate. As with the Boussinesq case, the idea is to ignore dynamic variations of density (i.e., of $\delta\rho$) except where associated with gravity. First recall a couple of ideal gas relationships involving potential temperature, θ , and entropy s (divided by c_p , so $s \equiv \log \theta$), namely

$$s \equiv \log \theta = \log T - \frac{R}{c_p} \log p = \frac{1}{\gamma} \log p - \log \rho, \quad (2.119)$$

where $\gamma = c_p/c_v$, implying

$$\delta s = \frac{1}{\theta} \delta \theta = \frac{1}{\gamma} \frac{\delta p}{p} - \frac{\delta \rho}{\rho} \approx \frac{1}{\gamma} \frac{\delta p}{\tilde{p}} - \frac{\delta \rho}{\tilde{\rho}}. \quad (2.120)$$

Further, if $\tilde{s} \equiv \gamma^{-1} \log \tilde{p} - \log \tilde{\rho}$ then

$$\frac{d\tilde{s}}{dz} = \frac{1}{\gamma \tilde{p}} \frac{d\tilde{p}}{dz} - \frac{1}{\tilde{\rho}} \frac{d\tilde{\rho}}{dz} = -\frac{g\tilde{\rho}}{\gamma \tilde{p}} - \frac{1}{\tilde{\rho}} \frac{d\tilde{\rho}}{dz}. \quad (2.121)$$

In the atmosphere, the left-hand side is, typically, much smaller than either of the two terms on the right-hand side.

2.5.2 The momentum equation

The exact inviscid horizontal momentum equation is

$$(\tilde{\rho} + \rho') \left(\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} \right) = -\nabla_z \delta p. \quad (2.122)$$

Neglecting ρ' where it appears with $\tilde{\rho}$ leads to

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -\nabla_z \phi, \quad (2.123)$$

where $\phi = \delta p / \tilde{\rho}$, and this is similar to the corresponding equation in the Boussinesq approximation.

The vertical component of the inviscid momentum equation is, without approximation,

$$(\tilde{\rho} + \delta\rho) \frac{Dw}{Dt} = -\frac{\partial \tilde{p}}{\partial z} - \frac{\partial \delta p}{\partial z} - g\tilde{\rho} - g\delta\rho = -\frac{\partial \delta p}{\partial z} - g\delta\rho. \quad (2.124)$$

using (2.118). Neglecting $\delta\rho$ on the left-hand side we obtain

$$\frac{Dw}{Dt} = -\frac{1}{\tilde{\rho}} \frac{\partial \delta p}{\partial z} - g \frac{\delta \rho}{\tilde{\rho}} = -\frac{\partial}{\partial z} \left(\frac{\delta p}{\tilde{\rho}} \right) - \frac{\delta p}{\tilde{\rho}^2} \frac{\partial \tilde{\rho}}{\partial z} - g \frac{\delta \rho}{\tilde{\rho}}. \quad (2.125)$$

This is not a useful form for a gaseous atmosphere, since the variation of the mean density cannot be ignored. However, we may eliminate $\delta\rho$ in favour of δs using (2.120) to give

$$\frac{Dw}{Dt} = g\delta s - \frac{\partial}{\partial z} \left(\frac{\delta p}{\tilde{\rho}} \right) - \frac{g}{\gamma} \frac{\delta p}{\tilde{p}} - \frac{\delta p}{\tilde{\rho}^2} \frac{\partial \tilde{\rho}}{\partial z}, \quad (2.126)$$

and using (2.121) gives

$$\frac{Dw}{Dt} = g\delta s - \frac{\partial}{\partial z} \left(\frac{\delta p}{\tilde{\rho}} \right) + \frac{d\tilde{s}}{dz} \frac{\delta p}{\tilde{\rho}}. \quad (2.127)$$

What have these manipulations gained us? Two things:

- (i) The gravitational term now involves δs rather than $\delta \rho$ which enables a more direct connection with the thermodynamic equation.
- (ii) The potential temperature scale height (~ 100 km) in the atmosphere is much larger than the density scale height (~ 10 km), and so the last term in (2.127) is small.

The second item thus suggests that we choose our reference state to be one of constant potential temperature (see also problem 2.19). The term $d\tilde{s}/dz$ then vanishes and the vertical momentum equation becomes

$$\boxed{\frac{Dw}{Dt} = g\delta s - \frac{\partial \phi}{\partial z}}, \quad (2.128)$$

where $\phi = \delta p/\tilde{\rho}$ and $\delta s = \delta\theta/\theta_0$, where θ_0 is a constant. If we define a buoyancy by $b_a \equiv g\delta s = g\delta\theta/\theta_0$, then (2.123) and (2.128) have the same form as the Boussinesq momentum equations, but with a slightly different definition of buoyancy.

2.5.3 Mass conservation

Using (2.117a) the mass conservation equation may be written, without approximation, as

$$\frac{\partial \delta \rho}{\partial t} + \nabla \cdot [(\tilde{\rho} + \delta \rho) \mathbf{v}] = 0. \quad (2.129)$$

We neglect $\delta \rho$ where it appears with $\tilde{\rho}$ in the divergence term. Further, the local time derivative will be small if time itself is scaled advectively (i.e., $T \sim L/U$ and sound waves do not dominate), giving

$$\nabla \cdot \mathbf{u} + \frac{1}{\tilde{\rho}} \frac{\partial}{\partial z} (\tilde{\rho} w) = 0. \quad (2.130)$$

It is here that the eponymous ‘anelastic approximation’ arises: the elastic compressibility of the fluid is neglected, and this serves to eliminate sound waves. For reference, in spherical coordinates the equation is

$$\frac{1}{a \cos \theta} \frac{\partial u}{\partial \lambda} + \frac{1}{a \cos \theta} \frac{\partial}{\partial \theta} (v \cos \theta) + \frac{1}{\tilde{\rho}} \frac{\partial (w \tilde{\rho})}{\partial z} = 0. \quad (2.131)$$

In an ideal gas, the choice of constant potential temperature determines how the reference density $\tilde{\rho}$ varies with height. In some circumstances it is convenient to let $\tilde{\rho}$ be a constant, ρ_0 (effectively choosing a different equation of state), in which case the anelastic equations become identical to the Boussinesq equations, albeit with the buoyancy interpreted in terms of potential temperature in the former and density in the latter. Because of their similarity, the Boussinesq and anelastic approximations are sometimes referred to as the strong and weak Boussinesq approximations, respectively.

2.5.4 Thermodynamic equation

The thermodynamic equation for an ideal gas may be written

$$\frac{D \ln \theta}{Dt} = \frac{\dot{Q}}{T c_p}. \quad (2.132)$$

In the anelastic equations, $\theta = \tilde{\theta} + \delta\theta$, where $\tilde{\theta}$ is constant, and the thermodynamic equation is

$$\frac{D \delta s}{Dt} = \frac{\tilde{\theta}}{T c_p} \dot{Q}. \quad (2.133)$$

Summarizing, the complete set of anelastic equations, with rotation but with no dissipation or diabatic terms, is

$$\begin{aligned} \frac{D \mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{v} &= \mathbf{k} b_a - \nabla \phi \\ \frac{D b_a}{Dt} &= 0 \\ \nabla \cdot (\tilde{\rho} \mathbf{v}) &= 0 \end{aligned}, \quad (2.134a,b,c)$$

where $b_a = g\delta s = g\delta\theta/\tilde{\theta}$. The main difference between the anelastic and Boussinesq sets of equations is in the mass continuity equation, and when $\tilde{\rho} = \rho_0 = \text{constant}$ the two equation sets are formally identical. However, whereas the Boussinesq approximation is a very good one for ocean dynamics, the anelastic approximation is much less so for large-scale atmosphere flow: the constancy of the reference potential temperature state is not a particularly good approximation, and the deviations in density from its reference profile are not especially small, leading to inaccuracies in the momentum equation. Nevertheless, the anelastic equations have been used very productively in limited area 'large-eddy simulations' where one does not wish to make the hydrostatic approximation but where sound waves are unimportant.⁷ The equations also provide a good jumping-off point for theoretical studies and for the still simpler models of chapter 5.

2.5.5 * Energetics of the anelastic equations

Conservation of energy follows in much the same way as for the Boussinesq equations, except that $\tilde{\rho}$ enters. Take the dot product of (2.134a) with $\tilde{\rho} \mathbf{v}$ to obtain

$$\tilde{\rho} \frac{D}{Dt} \left(\frac{1}{2} \mathbf{v}^2 \right) = -\nabla \cdot (\phi \tilde{\rho} \mathbf{v}) + b_a \tilde{\rho} w. \quad (2.135)$$

Now, define a potential $\Phi(z)$ such that $\nabla \Phi = -\mathbf{k}$, and so

$$\tilde{\rho} \frac{D \Phi}{Dt} = -w \tilde{\rho}. \quad (2.136)$$

Combining this with the thermodynamic equation (2.134b) gives

$$\tilde{\rho} \frac{D(b_a \Phi)}{Dt} = -w b_a \tilde{\rho}. \quad (2.137)$$

Adding this to (2.135) gives

$$\tilde{\rho} \frac{D}{Dt} \left(\frac{1}{2} \mathbf{v}^2 + b_a \Phi \right) = -\nabla \cdot (\phi \tilde{\rho} \mathbf{v}), \quad (2.138)$$

or, expanding the material derivative,

$$\frac{\partial}{\partial t} \left[\tilde{\rho} \left(\frac{1}{2} \mathbf{v}^2 + b_a \Phi \right) \right] + \nabla \cdot \left[\tilde{\rho} \mathbf{v} \left(\frac{1}{2} \mathbf{v}^2 + b_a \Phi + \phi \right) \right] = 0. \quad (2.139)$$

This equation has the form

$$\frac{\partial E}{\partial t} + \nabla \cdot [\mathbf{v}(E + \tilde{\rho}\phi)] = 0, \quad (2.140)$$

where $E = \tilde{\rho}(\mathbf{v}^2/2 + b_a\Phi)$ is the energy density of the flow. This is a consistent energetic equation for the system, and when integrated over a closed domain the total energy is evidently conserved. The total energy density comprises the kinetic energy and a term $\tilde{\rho}b_a\Phi$, which is analogous to the potential energy of a simple Boussinesq system. However, it is not exactly equal to potential energy because b_a is the buoyancy based on potential temperature, not density; rather, the term combines contributions from both the internal energy and the potential energy into an enthalpy-like quantity.

2.6 CHANGING VERTICAL COORDINATE

Although using z as a vertical coordinate is a natural choice given our Cartesian worldview, it is not the only option, nor is it always the most useful one. Any variable that has a one-to-one correspondence with z in the vertical, so any variable that varies monotonically with z , could be used; pressure and, perhaps surprisingly, entropy, are common choices. In the atmosphere pressure almost always falls monotonically with height, and using it instead of z provides a useful simplification of the mass conservation and geostrophic relations, as well as a more direct connection with observations, which are often taken at fixed values of pressure. (In the ocean pressure coordinates are essentially almost the same as height coordinates, because density is almost constant.) Entropy seems an exotic vertical coordinate, but it is very useful in adiabatic flow and we consider it in chapter 3.

2.6.1 General relations

First consider a general vertical coordinate, ξ . Any variable Ψ that is a function of the coordinates (x, y, z, t) may be expressed instead in terms of (x, y, ξ, t) by considering z to be function of the independent variables (x, y, ξ, t) ; that is, we let $\Psi(x, y, \xi, t) = \Psi(x, y, z(x, y, \xi, t), t)$. Derivatives with respect to z and ξ are related by

$$\frac{\partial \Psi}{\partial \xi} = \frac{\partial \Psi}{\partial z} \frac{\partial z}{\partial \xi} \quad \text{and} \quad \frac{\partial \Psi}{\partial z} = \frac{\partial \Psi}{\partial \xi} \frac{\partial \xi}{\partial z}. \quad (2.141a,b)$$

Horizontal derivatives in the two coordinate systems are related by the chain rule,

$$\left(\frac{\partial \Psi}{\partial x} \right)_\xi = \left(\frac{\partial \Psi}{\partial x} \right)_z + \left(\frac{\partial z}{\partial x} \right)_\xi \frac{\partial \Psi}{\partial z}, \quad (2.142)$$

and similarly for time.

The material derivative in ξ coordinates may be derived by transforming the original expression in z coordinates using the chain rule, but because (x, y, t, ξ) are independent coordinates, and noting that the ‘vertical velocity’ in ξ coordinates is just $\dot{\xi}$ (i.e., $D\xi/Dt$, just as the vertical velocity in z coordinates is $w = Dz/Dt$), we can write down

$$\frac{D\Psi}{Dt} = \frac{\partial\Psi}{\partial t} \Big|_{x,y,\xi} + \mathbf{u} \cdot \nabla_\xi \Psi + \dot{\xi} \frac{\partial\Psi}{\partial\xi}, \quad (2.143)$$

where ∇_ξ is the gradient operator at constant ξ . The operator D/Dt is physically the same in z or ξ coordinates because it is the total derivative of some property of a fluid parcel, and this is independent of the coordinate system. However, the individual terms within it will differ between coordinate systems.

2.6.2 Pressure coordinates

Let us now transform the ideal gas primitive equations from height coordinates to pressure coordinates, (x, y, p, t) . In z coordinates the equations are

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -\frac{1}{\rho} \nabla p, \quad \frac{\partial p}{\partial z} = -\rho g, \quad (2.144a)$$

$$\frac{D\theta}{Dt} = 0, \quad \frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0, \quad (2.144b)$$

where $p = \rho RT$ and $\theta = T(p_R/p)^{R/c_p}$, and p_R is the reference pressure. These are respectively the horizontal momentum, hydrostatic, thermodynamic and mass continuity equations. The analogue of the vertical velocity is $\omega \equiv Dp/Dt$, and the advective derivative itself is given by

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_p + \omega \frac{\partial}{\partial p}. \quad (2.145)$$

To obtain an expression for the pressure force, now let $\xi = p$ in (2.142) and apply the relationship to p itself to give

$$0 = \left(\frac{\partial p}{\partial x} \right)_z + \left(\frac{\partial z}{\partial x} \right)_p \frac{\partial p}{\partial z}, \quad (2.146)$$

which, using the hydrostatic relationship, gives

$$\left(\frac{\partial p}{\partial x} \right)_z = \rho \left(\frac{\partial \Phi}{\partial x} \right)_p, \quad (2.147)$$

where $\Phi = gz$ is the *geopotential*. Thus, the horizontal pressure force in the momentum equations is

$$\frac{1}{\rho} \nabla_z p = \nabla_p \Phi, \quad (2.148)$$

where the subscripts on the gradient operator indicate that the horizontal derivatives are taken at constant z or constant p . Also, from (2.144a), the hydrostatic equation is just

$$\frac{\partial \Phi}{\partial p} = -\alpha. \quad (2.149)$$

The mass conservation equation simplifies attractively in pressure coordinates, if the hydrostatic approximation is used. Recall that the mass conservation equation can be derived from the material form

$$\frac{D}{Dt}(\rho \delta V) = 0, \quad (2.150)$$

where $\delta V = \delta x \delta y \delta z$ is a volume element. But by the hydrostatic relationship $\rho \delta z = (1/g) \delta p$ and thus

$$\frac{D}{Dt}(\delta x \delta y \delta p) = 0. \quad (2.151)$$

This is completely analogous to the expression for the material conservation of volume in an incompressible fluid, (1.15). Thus, without further ado, we write the mass conservation in pressure coordinates as

$$\nabla_p \cdot \mathbf{u} + \frac{\partial \omega}{\partial p} = 0, \quad (2.152)$$

where the horizontal derivative is taken at constant pressure. The primitive equations in pressure coordinates, equivalent to (2.144) in height coordinates, are thus:

$$\begin{aligned} \frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} &= -\nabla_p \Phi, & \frac{\partial \Phi}{\partial p} &= -\alpha \\ \frac{D\theta}{Dt} &= 0, & \nabla_p \cdot \mathbf{u} + \frac{\partial \omega}{\partial p} &= 0 \end{aligned} \quad (2.153)$$

where D/Dt is given by (2.145). The equation set is completed with the addition of the ideal gas equation and the definition of potential temperature. These equations are isomorphic to the hydrostatic general Boussinesq equations (see the shaded box on page 72) with $z \leftrightarrow -p$, $w \leftrightarrow -\omega$, $\phi \leftrightarrow \Phi$, $b \leftrightarrow \alpha$, and an equation of state $b = b(\theta, z) \leftrightarrow \alpha = \alpha(\theta, p)$. In an ideal gas, for example, $\alpha = (\partial R/p_R)(p_R/p)^{1/\gamma}$.

The main practical difficulty with the pressure-coordinate equations is the lower boundary condition. Using

$$w \equiv \frac{Dz}{Dt} = \frac{\partial z}{\partial t} + \mathbf{u} \cdot \nabla_p z + \omega \frac{\partial z}{\partial p}, \quad (2.154)$$

and (2.149), the boundary condition of $w = 0$ at $z = z_s$ becomes

$$\frac{\partial \Phi}{\partial t} + \mathbf{u} \cdot \nabla_p \Phi - \alpha \omega = 0 \quad (2.155)$$

at $p(x, y, z_s, t)$. In theoretical studies, it is common to assume that the lower boundary is in fact a constant pressure surface and simply assume that $\omega = 0$, or sometimes the condition $\omega = -\alpha^{-1} \partial \Phi / \partial t$ is used. For realistic studies (with general circulation models, say) the fact that the level $z = 0$ is not a coordinate surface must be properly accounted for. For this reason, and especially if the lower boundary is uneven because of the presence of topography, so-called *sigma coordinates* are sometimes used, in which the vertical coordinate is chosen so that the lower boundary is a coordinate surface. Sigma coordinates may use height itself as a vertical measure (typical in oceanic applications) or use pressure (typical in atmospheric applications). In the latter case the vertical coordinate is $\sigma = p/p_s$ where $p_s(x, y, t)$ is the surface pressure. The difficulty of applying (2.155) is replaced by a prognostic equation for the surface pressure, derived from the mass conservation equation (problem 2.24).

2.6.3 Log-pressure coordinates

A variant of pressure coordinates is *log-pressure* coordinates, in which the vertical coordinate is $Z = -H \ln(p/p_R)$ where p_R is a reference pressure (say 1000 mb) and H is a constant (for example the scale height RT_s/g) so that Z has units of length. (Uppercase letters are conventionally used for some variables in log-pressure coordinates, and these are not to be confused with scaling parameters.) The ‘vertical velocity’ for the system is now

$$W \equiv \frac{DZ}{Dt}, \quad (2.156)$$

and the advective derivative is

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_p + W \frac{\partial}{\partial Z}. \quad (2.157)$$

It is straightforward to show (problem 2.25) that the primitive equations of motion in these coordinates are:

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -\nabla_Z \Phi, \quad \frac{\partial \Phi}{\partial Z} = \frac{RT}{H}, \quad (2.158a)$$

$$\frac{D\theta}{Dt} = 0, \quad \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial W}{\partial Z} - \frac{W}{H} = 0. \quad (2.158b)$$

The last equation may be written $\nabla_Z \cdot \mathbf{u} + \rho_R^{-1} \partial(\rho_R W) / \partial z = 0$, where $\rho_R = \exp(-z/H)$, so giving a form similar to the mass conservation equation in the anelastic equations. Note that integrating the hydrostatic equation between two pressure levels gives, with $\Phi = gz$,

$$z(p_2) - z(p_1) = \frac{R}{g} \int_{p_1}^{p_2} T \, d \ln p. \quad (2.159)$$

Thus, the thickness of the layer is proportional to the average temperature of the layer.

2.7 SCALING FOR HYDROSTATIC BALANCE

In this section we consider one of the most fundamental balances in geophysical fluid dynamics, hydrostatic balance, and in the next section we consider another fundamental balance, geostrophic balance. The corresponding states, hydrostasy and geostrophy, are not exactly realized, but their approximate satisfaction has profound consequences on the behaviour of the atmosphere and ocean. We first encountered hydrostatic balance in section 1.3.4; we now look in more detail at the conditions required for it to hold.

2.7.1 Preliminaries

Consider the relative sizes of terms in (2.77c):

$$\frac{W}{T} + \frac{UW}{L} + \frac{W^2}{H} + \Omega U \sim \left| \frac{1}{\rho} \frac{\partial p}{\partial z} \right| + g. \quad (2.160)$$

For most large-scale motion in the atmosphere and ocean the terms on the right-hand side are orders of magnitude larger than those on the left, and therefore must be approximately

equal. Explicitly, suppose $W \sim 1 \text{ cm s}^{-1}$, $L \sim 10^5 \text{ m}$, $H \sim 10^3 \text{ m}$, $U \sim 10 \text{ m s}^{-1}$, $T = L/U$. Then by substituting into (2.160) it seems that the pressure term is the only one which could balance the gravitational term, and we are led to approximate (2.77c) by,

$$\frac{\partial p}{\partial z} = -\rho g. \quad (2.161)$$

This equation, which is a vertical momentum equation, is known as *hydrostatic balance*.

However, (2.161) is not always a useful equation! Let us suppose that the density is a constant, ρ_0 . We can then write the pressure as

$$p(x, y, z, t) = p_0(z) + p'(x, y, z, t), \quad (2.162)$$

where

$$\frac{\partial p_0}{\partial z} \equiv -\rho_0 g. \quad (2.163)$$

That is, p_0 and ρ_0 are in hydrostatic balance. The inviscid vertical momentum equation becomes, without approximation,

$$\frac{Dw}{Dt} = -\frac{1}{\rho_0} \frac{\partial p'}{\partial z}. \quad (2.164)$$

Thus, for constant density fluids, the gravitational term has no dynamical effect: there is no buoyancy force, and the pressure term in the horizontal momentum equations can be replaced by p' . Hydrostatic balance, and in particular (2.163), is certainly not an appropriate vertical momentum equation in this case. If the fluid is stratified, we should therefore subtract off the hydrostatic pressure associated with the mean density before we can determine whether hydrostasis is a useful *dynamical* approximation, accurate enough to determine the horizontal pressure gradients. This is automatic in the Boussinesq equations, where the vertical momentum equation is

$$\frac{Dw}{Dt} = -\frac{\partial \phi}{\partial z} + b, \quad (2.165)$$

and the hydrostatic balance of the basic state is already subtracted out. In the more general equation,

$$\frac{Dw}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g, \quad (2.166)$$

we need to compare the advective term on the left-hand side with the pressure variations arising from horizontal flow in order to determine whether hydrostasis is an appropriate vertical momentum equation. Nevertheless, if we only need to determine the pressure for use in an equation of state then we simply need to compare the sizes of the dynamical terms in (2.77c) with g itself, in order to determine whether a hydrostatic approximation will suffice.

2.7.2 Scaling and the aspect ratio

In a Boussinesq fluid we write the horizontal and vertical momentum equations as

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -\nabla \phi, \quad \frac{Dw}{Dt} = -\frac{\partial \phi}{\partial z} - b. \quad (2.167a,b)$$

With $f = 0$, (2.167a) implies the scaling

$$\phi \sim U^2. \quad (2.168)$$

If we use mass conservation, $\nabla_z \cdot \mathbf{u} + \partial w / \partial z = 0$, to scale vertical velocity then

$$w \sim W = \frac{H}{L} U = \alpha U, \quad (2.169)$$

where $\alpha \equiv H/L$ is the aspect ratio. The advective terms in the vertical momentum equation all scale as

$$\frac{Dw}{Dt} \sim \frac{UW}{L} = \frac{U^2 H}{L^2}. \quad (2.170)$$

Using (2.168) and (2.170) the ratio of the advective term to the pressure gradient term in the vertical momentum equations then scales as

$$\frac{|Dw/Dt|}{|\partial \phi / \partial z|} \sim \frac{U^2 H / L^2}{U^2 / H} \sim \left(\frac{H}{L}\right)^2. \quad (2.171)$$

Thus, the condition for hydrostasy, that $|Dw/Dt|/|\partial \phi / \partial z| \ll 1$, is:

$$\alpha^2 \equiv \left(\frac{H}{L}\right)^2 \ll 1.$$

(2.172)

The advective term in the vertical momentum may then be neglected. Thus, hydrostatic balance is a *small aspect ratio approximation*.

We can obtain the same result more formally by non-dimensionalizing the momentum equations. Using uppercase symbols to denote scaling values we write

$$(x, y) = L(\hat{x}, \hat{y}), \quad z = H\hat{z}, \quad \mathbf{u} = U\hat{\mathbf{u}}, \quad w = W\hat{w} = \frac{HU}{L}\hat{w},$$

$$t = T\hat{t} = \frac{L}{U}\hat{t}, \quad \phi = \Phi\hat{\phi} = U^2\hat{\phi}, \quad b = B\hat{b} = \frac{U^2}{H}\hat{b}, \quad (2.173)$$

where the hatted variables are non-dimensional and the scaling for w is suggested by the mass conservation equation, $\nabla_z \cdot \mathbf{u} + \partial w / \partial z = 0$. Substituting (2.173) into (2.167) (with $f = 0$) gives us the non-dimensional equations

$$\frac{D\hat{\mathbf{u}}}{Dt} = -\nabla \hat{\phi}, \quad \alpha^2 \frac{D\hat{w}}{Dt} = -\frac{\partial \hat{\phi}}{\partial \hat{z}} - \hat{b}, \quad (2.174a,b)$$

where $D/D\hat{t} = \partial / \partial \hat{t} + \hat{u}\partial / \partial \hat{x} + \hat{v}\partial / \partial \hat{y} + \hat{w}\partial / \partial \hat{z}$ and we use the convention that when ∇ operates on non-dimensional quantities the operator itself is non-dimensional. From (2.174b) it is clear that hydrostatic balance pertains when $\alpha^2 \ll 1$.

2.7.3 * Effects of stratification on hydrostatic balance

To include the effects of stratification we need to involve the thermodynamic equation, so let us first write down the complete set of non-rotating dimensional equations:

$$\frac{D\mathbf{u}}{Dt} = -\nabla_z \phi, \quad \frac{Dw}{Dt} = -\frac{\partial \phi}{\partial z} + b', \quad (2.175a,b)$$

$$\frac{D\hat{b}'}{Dt} + wN^2 = 0, \quad \nabla \cdot \mathbf{v} = 0. \quad (2.176a,b)$$

We have written, without approximation, $b = b'(x, y, z, t) + \tilde{b}(z)$, with $N^2 = d\tilde{b}/dz$; this separation is useful because the horizontal and vertical buoyancy variations may scale in different ways, and often N^2 may be regarded as given. (We have also redefined ϕ by subtracting off a static component in hydrostatic balance with \tilde{b} .) We non-dimensionalize (2.176) by first writing

$$(x, y) = L(\hat{x}, \hat{y}), \quad z = H\hat{z}, \quad \mathbf{u} = U\hat{\mathbf{u}}, \quad w = W\hat{w} = \epsilon \frac{HU}{L} \hat{w},$$

$$t = T\hat{t} = \frac{L}{U} \hat{t}, \quad \phi = U^2 \hat{\phi}, \quad b' = \Delta b \hat{b} = \frac{U^2}{H} \hat{b}', \quad N^2 = \bar{N}^2 \hat{N}^2,$$

$$(2.177)$$

where ϵ is, for the moment, undetermined, \bar{N} is a representative, constant, value of the buoyancy frequency and Δb scales only the horizontal buoyancy variations. Substituting (2.177) into (2.175) and (2.176) gives

$$\frac{D\hat{\mathbf{u}}}{Dt} = -\nabla_z \hat{\phi}, \quad \epsilon \alpha^2 \frac{D\hat{w}}{Dt} = -\frac{\partial \hat{\phi}}{\partial \hat{z}} + \hat{b}' \quad (2.178a,b)$$

$$\frac{U^2}{\bar{N}^2 H^2} \frac{D\hat{b}'}{Dt} + \epsilon \hat{w} \hat{N}^2 = 0, \quad \nabla \cdot \hat{\mathbf{u}} + \epsilon \frac{\partial \hat{w}}{\partial \hat{z}} = 0. \quad (2.179a,b)$$

where now $D/D\hat{t} = \partial/\partial\hat{t} + \hat{\mathbf{u}} \cdot \nabla_z + \epsilon \partial/\partial\hat{z}$. To obtain a non-trivial balance in (2.179a) we choose $\epsilon = U^2/(\bar{N}^2 H^2) \equiv Fr^2$, where *Fr* is the *Froude number*, a measure of the stratification of the flow. The vertical velocity then scales as

$$W = \frac{Fr U H}{L} \quad (2.180)$$

and if the flow is highly stratified the vertical velocity will be even smaller than a pure aspect ratio scaling might suggest. (There must, therefore, be some cancellation in horizontal divergence in the mass continuity equation; that is, $|\nabla_z \cdot \mathbf{u}| \ll U/L$.) With this choice of ϵ the non-dimensional Boussinesq equations may be written:

$$\frac{D\hat{\mathbf{u}}}{Dt} = -\nabla_z \hat{\phi}, \quad Fr^2 \alpha^2 \frac{D\hat{w}}{Dt} = -\frac{\partial \hat{\phi}}{\partial \hat{z}} + \hat{b}' \quad (2.181a,b)$$

$$\frac{D\hat{b}'}{Dt} + \hat{w} \hat{N}^2 = 0, \quad \nabla \cdot \hat{\mathbf{u}} + Fr^2 \frac{\partial \hat{w}}{\partial \hat{z}} = 0. \quad (2.182a,b)$$

The non-dimensional parameters in the system are the aspect ratio and the Froude number (in addition to \hat{N} , but by construction this is just an order one function of z). From (2.181b) condition for hydrostatic balance to hold is evidently that

$Fr^2 \alpha^2 \ll 1$

$$, \quad (2.183)$$

so generalizing the aspect ratio condition (2.172) to a stratified fluid. Because *Fr* is a measure of stratification, (2.183) formalizes our intuitive expectation that the more stratified a fluid

the more vertical motion is suppressed and therefore the more likely hydrostatic balance is to hold. Also note that (2.183) is equivalent to $U^2/(L^2N^2) \ll 1$.

Suppose we solve the hydrostatic equations; that is, we omit the advective derivative in the vertical momentum equation, and by numerical integration we obtain \mathbf{u} , w and b . This flow is the solution of the non-hydrostatic equations in the small aspect ratio limit. The solution never violates the scaling assumptions, even if w seems large, because we can always rescale the variables in order that condition (2.183) is satisfied.

Why bother with any of this scaling? Why not just say that hydrostatic balance holds when $|Dw/Dt| \ll |\partial\phi/\partial z|$? One reason is that we do not have a good idea of the value of w from direct measurements, and it may change significantly in different oceanic and atmospheric parameter regimes. On the other hand the Froude number and the aspect ratio are familiar non-dimensional parameters with a wide applicability in other contexts, and which we can control in a laboratory setting or estimate in the ocean or atmosphere. Still, in scaling theory it is common that ascertaining which parameters are to be regarded as given and which should be derived is a choice, rather than being set a priori.

2.7.4 Hydrostasy in the ocean and atmosphere

Is the hydrostatic approximation in fact a good one in the ocean and atmosphere?

In the ocean

For the large-scale ocean circulation, let $N \sim 10^{-2} \text{ s}^{-1}$, $U \sim 0.1 \text{ m s}^{-1}$ and $H \sim 1 \text{ km}$. Then $Fr = U/(NH) \sim 10^{-2} \ll 1$. Thus, $Fr^2\alpha^2 \ll 1$ even for unit aspect-ratio motion. In fact, for larger scale flow the aspect ratio is also small; for basin-scale flow $L \sim 10^6 \text{ m}$ and $Fr^2\alpha^2 \sim 0.01^2 \times 0.001^2 = 10^{-10}$ and hydrostatic balance is an extremely good approximation.

For intense convection, for example in the Labrador Sea, the hydrostatic approximation may be less appropriate, because the intense descending plumes may have an aspect ratio (H/L) of one or greater and the stratification is very weak. The hydrostatic condition then often becomes the requirement that the Froude number is small. Representative orders of magnitude are $U \sim W \sim 0.1 \text{ m s}^{-1}$, $H \sim 1 \text{ km}$ and $N \sim 10^{-3} \text{ s}^{-1}$ to 10^{-4} s^{-1} . For these values Fr ranges between 0.1 and 1, and at the upper end of this range hydrostatic balance is violated.

In the atmosphere

Over much of the troposphere $N \sim 10^{-2} \text{ s}^{-1}$ so that with $U = 10 \text{ m s}^{-1}$ and $H = 1 \text{ km}$ we find $Fr \sim 1$. Hydrostasy is then maintained because the aspect ratio H/L is much less than unity. For larger scale synoptic activity a larger vertical scale is appropriate, and with $H = 10 \text{ km}$ both the Froude number and the aspect ratio are much smaller than one; indeed with $L = 1000 \text{ km}$ we find $Fr^2\alpha^2 \sim 0.1^2 \times 0.1^2 = 10^{-4}$ and the flow is hydrostatic to a very good approximation indeed. However, for smaller scale atmospheric motions associated with fronts and, especially, convection, there can be little expectation that hydrostatic balance will be a good approximation.

For large-scale flows in both atmosphere and ocean, the conceptual simplifications afforded by the hydrostatic approximation can hardly be overemphasized.

Variable	Scaling symbol	Meaning	Atmos. value	Ocean value
(x, y)	L	Horizontal length scale	10^6 m	10^5 m
t	T	Time scale	$1 \text{ day} (10^5 \text{ s})$	$10 \text{ days} (10^6 \text{ s})$
(u, v)	U	Horizontal velocity	10 m s^{-1}	0.1 m s^{-1}
	Ro	Rossby number, U/fL	0.1	0.01

Table 2.1 Scales of large-scale flow in atmosphere and ocean. The choices given are representative of large-scale eddying motion in both systems.

2.8 GEOSTROPHIC AND THERMAL WIND BALANCE

We now consider the dominant dynamical balance in the horizontal components of the momentum equation. In the horizontal plane (meaning along geopotential surfaces) we find that the Coriolis term is much larger than the advective terms and the dominant balance is between it and the horizontal pressure force. This balance is called *geostrophic balance*, and it occurs when the Rossby number is small, as we now investigate.

2.8.1 The Rossby number

The *Rossby number* characterizes the importance of rotation in a fluid.⁸ It is, essentially, the ratio of the magnitude of the relative acceleration to the Coriolis acceleration, and it is of fundamental importance in geophysical fluid dynamics. It arises from a simple scaling of the horizontal momentum equation, namely

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{u} + \mathbf{f} \times \mathbf{u} = -\frac{1}{\rho} \nabla_z p, \quad (2.184a)$$

$$U^2/L \quad fU \quad (2.184b)$$

where U is the approximate magnitude of the horizontal velocity and L is a typical length scale over which that velocity varies. (We assume that $W/H \lesssim U/L$, so that vertical advection does not dominate the advection.) The ratio of the sizes of the advective and Coriolis terms is defined to be the Rossby number,

$$Ro \equiv \frac{U}{fL}. \quad (2.185)$$

If the Rossby number is small then rotation effects are important, and as the values in Table 2.1 indicate this is the case for large-scale flow in both ocean and atmosphere.

Another intuitive way to think about the Rossby number is in terms of time scales. The Rossby number based on a time scale is

$$Ro_T \equiv \frac{1}{fT}, \quad (2.186)$$

where T is a time scale associated with the dynamics at hand. If the time scale is an advective one, meaning that $T \sim L/U$, then this definition is equivalent to (2.185). Now, $f = 2\Omega \sin \vartheta$, where Ω is the angular velocity of the rotating frame and equal to $2\pi/T_p$ where T_p is the period of rotation (24 hours). Thus,

$$Ro_T = \frac{T_p}{4\pi T \sin \vartheta} = \frac{T_i}{T}, \quad (2.187)$$

where $T_i = 1/f$ is the ‘inertial time scale’, about three hours in mid-latitudes. Thus, for phenomena with time scales much longer than this, such as the motion of the Gulf Stream or a mid-latitude atmospheric weather system, the effects of the Earth’s rotation can be expected to be important, whereas a short-lived phenomena, such as a cumulus cloud or tornado, may be oblivious to such rotation. The expressions (2.185) and (2.186) are, of course, just approximate measures of the importance of rotation.

2.8.2 Geostrophic balance

If the Rossby number is sufficiently small in (2.184a) then the rotation term will dominate the nonlinear advection term, and if the time period of the motion scales advectively then the rotation term also dominates the local time derivative. The only term that can then balance the rotation term is the pressure term, and therefore we must have

$$\mathbf{f} \times \mathbf{u} \approx -\frac{1}{\rho} \nabla_z p, \quad (2.188)$$

or, in Cartesian component form

$$fu \approx -\frac{1}{\rho} \frac{\partial p}{\partial y}, \quad fv \approx \frac{1}{\rho} \frac{\partial p}{\partial x}. \quad (2.189)$$

This balance is known as *geostrophic balance*, and its consequences are profound, giving geophysical fluid dynamics a special place in the broader field of fluid dynamics. We *define* the geostrophic velocity by

$$fu_g \equiv -\frac{1}{\rho} \frac{\partial p}{\partial y}, \quad fv_g \equiv \frac{1}{\rho} \frac{\partial p}{\partial x},$$

(2.190)

and for low Rossby number flow $u \approx u_g$ and $v \approx v_g$. In spherical coordinates the geostrophic velocity is

$$fu_g = -\frac{1}{\rho a} \frac{\partial p}{\partial \vartheta}, \quad fv_g = \frac{1}{a \rho \cos \vartheta} \frac{\partial p}{\partial \lambda}, \quad (2.191)$$

where $f = 2\Omega \sin \vartheta$. Geostrophic balance has a number of immediate ramifications:

- ★ Geostrophic flow is parallel to lines of constant pressure (isobars). If $f > 0$ the flow is anticlockwise round a region of low pressure and clockwise around a region of high pressure (see Fig. 2.5).

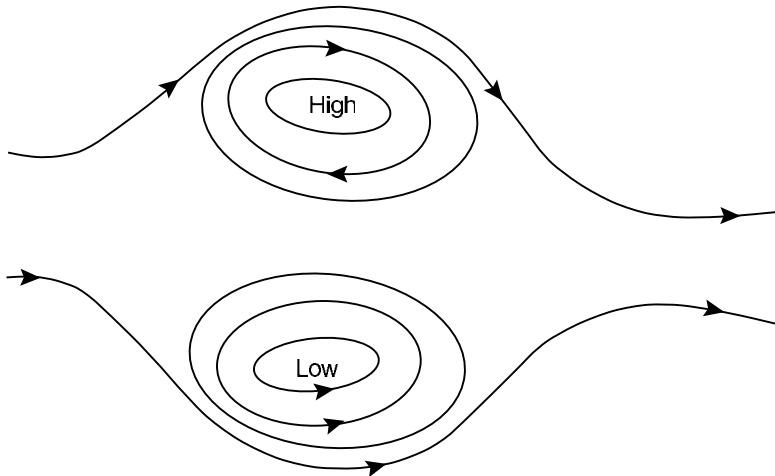


Fig. 2.5 Schematic of geostrophic flow with a positive value of the Coriolis parameter f . Flow is parallel to the lines of constant pressure (isobars). Cyclonic flow is anticlockwise around a low pressure region and anticyclonic flow is clockwise around a high. If f were negative, as in the Southern Hemisphere, (anti)cyclonic flow would be (anti)clockwise.

- ★ If the Coriolis force is constant and if the density does not vary in the horizontal the geostrophic flow is horizontally non-divergent and

$$\nabla_z \cdot \mathbf{u}_g = \frac{\partial u_g}{\partial x} + \frac{\partial v_g}{\partial y} = 0. \quad (2.192)$$

We may define the *geostrophic streamfunction*, ψ , by

$$\psi \equiv \frac{p}{f_0 \rho_0}, \quad (2.193)$$

whence

$$u_g = -\frac{\partial \psi}{\partial y}, \quad v_g = \frac{\partial \psi}{\partial x}. \quad (2.194)$$

The vertical component of vorticity, ζ , is then given by

$$\zeta = \mathbf{k} \cdot \nabla \times \mathbf{v} = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \nabla_z^2 \psi. \quad (2.195)$$

- ★ If the Coriolis parameter is not constant, then cross-differentiating (2.190) gives, for constant density geostrophic flow,

$$v_g \frac{\partial f}{\partial y} + f \nabla_z \cdot \mathbf{u}_g = 0, \quad (2.196)$$

which implies, using mass continuity,

$$\beta v_g = f \frac{\partial w}{\partial z}. \quad (2.197)$$

where $\beta \equiv \partial f / \partial y = 2\Omega \cos \theta / a$. This geostrophic vorticity balance is sometimes known as ‘Sverdrup balance’, although the latter expression is better restricted to the case when the vertical velocity results from external agents, and specifically a wind stress, as considered in chapter 14.

2.8.3 Taylor–Proudman effect

If $\beta = 0$, then (2.197) implies that the vertical velocity is not a function of height. In fact, in that case none of the components of velocity vary with height if density is also constant. To show this, in the limit of zero Rossby number we first write the three-dimensional momentum equation as

$$\mathbf{f}_0 \times \mathbf{v} = -\nabla \phi - \nabla \chi, \quad (2.198)$$

where $\mathbf{f}_0 = 2\mathbf{\Omega} = 2\Omega \mathbf{k}$, $\phi = p / \rho_0$, and $\nabla \chi$ represents other potential forces. If $\chi = gz$ then the vertical component of this equation represents hydrostatic balance, and the horizontal components represent geostrophic balance. On taking the curl of this equation, the terms on the right-hand side vanish and the left-hand side becomes

$$(\mathbf{f}_0 \cdot \nabla) \mathbf{v} - \mathbf{f}_0 \nabla \cdot \mathbf{v} - (\mathbf{v} \cdot \nabla) \mathbf{f}_0 + \mathbf{v} \nabla \cdot \mathbf{f}_0 = 0. \quad (2.199)$$

But $\nabla \cdot \mathbf{v} = 0$ by mass conservation, and because \mathbf{f}_0 is constant both $\nabla \cdot \mathbf{f}_0$ and $(\mathbf{v} \cdot \nabla) \mathbf{f}_0$ vanish. Thus

$$(\mathbf{f}_0 \cdot \nabla) \mathbf{v} = 0, \quad (2.200)$$

which, since $\mathbf{f}_0 = f_0 \mathbf{k}$, implies $f_0 \partial \mathbf{v} / \partial z = 0$, and in particular we have

$$\frac{\partial u}{\partial z} = 0, \quad \frac{\partial v}{\partial z} = 0, \quad \frac{\partial w}{\partial z} = 0. \quad (2.201)$$

A different presentation of this argument proceeds as follows. If the flow is exactly in geostrophic and hydrostatic balance then

$$v = \frac{1}{f_0} \frac{\partial \phi}{\partial x}, \quad u = -\frac{1}{f_0} \frac{\partial \phi}{\partial y}, \quad \frac{\partial \phi}{\partial z} = -g. \quad (2.202a,b,c)$$

Differentiating (2.202a,b) with respect to z , and using (2.202c) yields

$$\frac{\partial v}{\partial z} = \frac{-1}{f_0} \frac{\partial g}{\partial x} = 0, \quad \frac{\partial u}{\partial z} = \frac{1}{f_0} \frac{\partial g}{\partial y} = 0. \quad (2.203)$$

Noting that the geostrophic velocities are horizontally non-divergent ($\nabla_z \cdot \mathbf{u} = 0$), and using mass continuity then gives $\partial w / \partial z = 0$, as before.

If there is a solid horizontal boundary anywhere in the fluid, for example at the surface, then $w = 0$ at that surface and thus $w = 0$ everywhere. Hence the motion occurs in planes that lie perpendicular to the axis of rotation, and the flow is effectively two dimensional. This result is known as the *Taylor–Proudman effect*, namely that for constant density flow in geostrophic and hydrostatic balance the vertical derivatives of the horizontal and the vertical velocities are zero.⁹ At zero Rossby number, if the vertical velocity is zero somewhere in the flow, it is zero everywhere in that vertical column; furthermore, the horizontal flow has no vertical shear, and the fluid moves like a slab. The effects of rotation have provided a *stiffening* of the fluid in the vertical.

In neither the atmosphere nor the ocean do we observe precisely such vertically coherent flow, mainly because of the effects of stratification. However, it is typical of geophysical fluid dynamics that the assumptions underlying a derivation are not fully satisfied, yet there are manifestations of it in real flow. Thus, one might have naïvely expected, because $\partial w / \partial z = -\nabla_z \cdot \mathbf{u}$, that the scales of the various variables would be related by $W/H \sim U/L$. However, if the flow is rapidly rotating we expect that the horizontal flow will be in near geostrophic balance and therefore nearly divergence free; thus $\nabla_z \cdot \mathbf{u} \ll U/L$, and $W \ll HU/L$.

2.8.4 Thermal wind balance

Thermal wind balance arises by combining the geostrophic and hydrostatic approximations, and this is most easily done in the context of the anelastic (or Boussinesq) equations, or in pressure coordinates. For the anelastic equations, geostrophic balance may be written

$$-fv_g = -\frac{\partial \phi}{\partial x} = -\frac{1}{a \cos \vartheta} \frac{\partial \phi}{\partial \lambda}, \quad fu_g = -\frac{\partial \phi}{\partial y} = -\frac{1}{a} \frac{\partial \phi}{\partial \vartheta}. \quad (2.204a,b)$$

Combining these relations with hydrostatic balance, $\partial \phi / \partial z = b$, gives

$$\begin{aligned} -f \frac{\partial v_g}{\partial z} &= -\frac{\partial b}{\partial x} = -\frac{1}{a \cos \lambda} \frac{\partial b}{\partial \lambda} \\ f \frac{\partial u_g}{\partial z} &= -\frac{\partial b}{\partial y} = -\frac{1}{a} \frac{\partial b}{\partial \vartheta} \end{aligned} . \quad (2.205a,b)$$

These equations represent *thermal wind balance*, and the vertical derivative of the geostrophic wind is the ‘thermal wind’. Eq. (2.205b) may be written in terms of the zonal angular momentum as

$$\frac{\partial m_g}{\partial z} = -\frac{a}{2\Omega \tan \vartheta} \frac{\partial b}{\partial y}, \quad (2.206)$$

where $m_g = (u_g + \Omega a \cos \vartheta) a \cos \vartheta$. Potentially more accurate than geostrophic balance is the so-called cyclostrophic or gradient-wind balance, which retains a centrifugal term in the momentum equation. Thus, we omit only the material derivative in the meridional momentum equation (2.50b) and obtain

$$2u\Omega \sin \vartheta + \frac{u^2}{a} \tan \vartheta \approx -\frac{\partial \phi}{\partial y} = -\frac{1}{a} \frac{\partial \phi}{\partial \vartheta}. \quad (2.207)$$

For large-scale flow this only differs significantly from geostrophic balance very close to the equator. Combining cyclostrophic and hydrostatic balance gives a modified thermal wind relation, and this takes a simple form when expressed in terms of angular momentum, namely

$$\frac{\partial m^2}{\partial z} \approx -\frac{a^3 \cos^3 \vartheta}{\sin \vartheta} \frac{\partial b}{\partial y}. \quad (2.208)$$

If the density or buoyancy is constant then there is no shear and (2.205) or (2.208) give the Taylor-Proudman result. But suppose that the temperature falls in the poleward direction. Then thermal wind balance implies that the (eastward) wind will increase with height — just as is observed in the atmosphere! In general, a vertical shear of the horizontal wind is associated with a horizontal temperature gradient, and this is one of the most simple and far-reaching effects in geophysical fluid dynamics. The underlying physical mechanism is illustrated in Fig. 2.6.

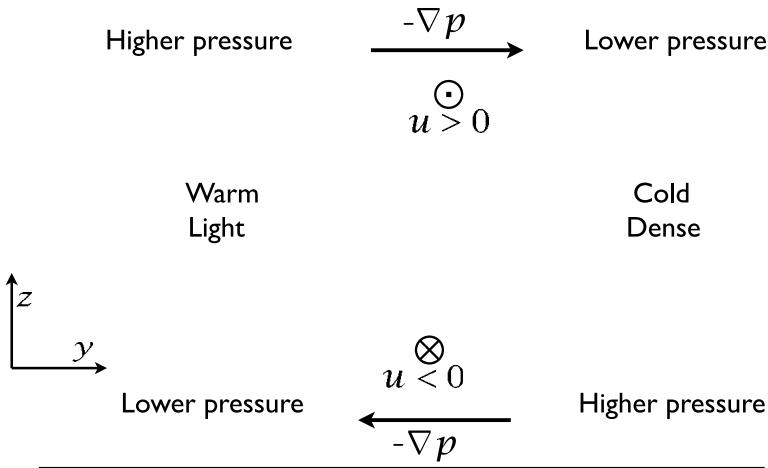


Fig. 2.6 The mechanism of thermal wind. A cold fluid is denser than a warm fluid, so by hydrostasy the vertical pressure gradient is greater where the fluid is cold. Thus, the pressure gradients form as shown, where ‘higher’ and ‘lower’ mean relative to the average at that height. The horizontal pressure gradients are balanced by the Coriolis force, producing (for $f > 0$) the horizontal winds shown (\otimes into the paper, and \odot out of the paper). Only the wind *shear* is given by the thermal wind.

Pressure coordinates

In pressure coordinates geostrophic balance is just

$$\mathbf{f} \times \mathbf{u}_g = -\nabla_p \Phi, \quad (2.209)$$

where Φ is the geopotential and ∇_p is the gradient operator taken at constant pressure. If f is constant, it follows from (2.209) that the geostrophic wind is non-divergent on pressure surfaces. Taking the vertical derivative of (2.209) (that is, its derivative with respect to p) and using the hydrostatic equation, $\partial \Phi / \partial p = -\alpha$, gives the thermal wind equation

$$\mathbf{f} \times \frac{\partial \mathbf{u}_g}{\partial p} = \nabla_p \alpha = \frac{R}{p} \nabla_p T, \quad (2.210)$$

where the last equality follows using the ideal gas equation and because the horizontal derivative is at constant pressure. In component form this is

$$-f \frac{\partial v_g}{\partial p} = \frac{R}{p} \frac{\partial T}{\partial x}, \quad f \frac{\partial u_g}{\partial p} = \frac{R}{p} \frac{\partial T}{\partial y}. \quad (2.211)$$

In log-pressure coordinates, with $Z = -H \ln(p/p_R)$, thermal wind is

$$\mathbf{f} \times \frac{\partial \mathbf{u}_g}{\partial Z} = -\frac{R}{H} \nabla_Z T. \quad (2.212)$$

The physical meaning in all these cases is the same: a horizontal temperature gradient, or a temperature gradient along an isobaric surface, is accompanied by a vertical shear of the horizontal wind.

2.8.5 * Effects of rotation on hydrostatic balance

Because rotation inhibits vertical motion, we might expect it to affect the requirements for hydrostasis. The simplest setting in which to see this is the rotating Boussinesq equations, (2.167). Let us non-dimensionalize these by writing

$$(x, y) = L(\hat{x}, \hat{y}), \quad z = H\hat{z}, \quad \mathbf{u} = U\hat{\mathbf{u}}, \quad t = T\hat{t} = \frac{U}{L}\hat{t}, \quad \mathbf{f} = f_0\hat{\mathbf{f}},$$

$$w = \frac{\beta HU}{f_0}\hat{w} = \hat{\beta}\frac{HU}{L}\hat{w}, \quad \phi = \Phi\hat{\phi} = f_0UL\hat{\phi}, \quad b = B\hat{b} = \frac{f_0uL}{H}\hat{b}, \quad (2.213)$$

where $\hat{\beta} \equiv \beta L/f_0$. (If \mathbf{f} is constant, then $\hat{\mathbf{f}}$ is a unit vector in the vertical direction.) These relations are the same as (2.173), except for the scaling for w , which is suggested by (2.197), and the scaling for ϕ and b' , which are suggested by geostrophic and thermal wind balance.

Substituting into (2.167) we obtain the following scaled momentum equations:

$$Ro \frac{D\hat{\mathbf{u}}}{Dt} + \hat{\mathbf{f}} \times \hat{\mathbf{u}} = -\nabla\hat{\phi}, \quad Ro\hat{\beta}\alpha^2 \frac{D\hat{w}}{Dt} = -\frac{\partial\hat{\phi}}{\partial z} - \hat{b}. \quad (2.214a,b)$$

Here, $D/D\hat{t} = \partial/\partial\hat{t} + \hat{\mathbf{u}} \cdot \nabla_z + \hat{\beta}\partial/\partial\hat{z}$ and $Ro = U/(f_0L)$. There are two notable aspects to these equations. First and most obviously, when $Ro \ll 1$, (2.214a) reduces to geostrophic balance, $\mathbf{f} \times \mathbf{u} = -\nabla\phi$. Second, the material derivative in (2.214b) is multiplied by three non-dimensional parameters, and we can understand the appearance of each as follows.

- (i) The aspect ratio dependence (α^2) arises in the same way as for non-rotating flows — that is, because of the presence of w and z in the vertical momentum equation as opposed to (u, v) and (x, y) in the horizontal equations.
- (ii) The Rossby number dependence (Ro) arises because in rotating flow the pressure gradient is balanced by the Coriolis force, which is Rossby number larger than the advective terms.
- (iii) The factor $\hat{\beta}$ arises because in rotating flow w is smaller than u by $\hat{\beta}$ times the aspect ratio.

The factor $Ro\hat{\beta}\alpha^2$ is very small for large-scale flow; the reader is invited to calculate representative values. Evidently, a rapidly rotating fluid is more likely to be in hydrostatic balance than a non-rotating fluid, other conditions being equal. The combined effects of rotation and stratification are, not surprisingly, quite subtle and we leave that topic for chapter 5.

2.9 STATIC INSTABILITY AND THE PARCEL METHOD

In this and the next couple of sections we consider how a fluid might oscillate if it were perturbed away from a resting state. Our focus is on vertical displacements, and the restoring force is gravity, and we will neglect the effects of rotation, and indeed initially we will neglect horizontal motion entirely. Given that, the simplest and most direct way to approach the problem is to consider from first principles the pressure and gravitational forces on a displaced parcel. To this end, consider a fluid initially at rest in a constant gravitational field, and therefore in hydrostatic balance. Suppose that a small parcel of the

fluid is adiabatically displaced upwards by the small distance δz , without altering the overall pressure field; that is, the fluid parcel instantly assumes the pressure of its environment. If after the displacement the parcel is lighter than its environment, it will accelerate upwards, because the upward pressure gradient force is now greater than the downward gravity force on the parcel; that is, the parcel is *buoyant* (a manifestation of Archimedes' principle) and the fluid is *statically unstable*. If on the other hand the fluid parcel finds itself heavier than its surroundings, the downward gravitational force will be greater than the upward pressure force and the fluid will sink back towards its original position and an oscillatory motion will develop. Such an equilibrium is *statically stable*. Using such simple 'parcel' arguments we will now develop criteria for the stability of the environmental profile.

2.9.1 A simple special case: a density-conserving fluid

Consider first the simple case of an incompressible fluid in which the density of the displaced parcel is conserved, that is $D\rho/Dt = 0$ (and refer to Fig. 2.7 setting $\rho_\theta = \rho$). If the environmental profile is $\tilde{\rho}(z)$ and the density of the parcel is ρ then a parcel displaced from a level z [where its density is $\tilde{\rho}(z)$] to a level $z + \delta z$ [where the density of the parcel is still $\tilde{\rho}(z)$] will find that its density then differs from its surroundings by the amount

$$\delta\rho = \rho(z + \delta z) - \tilde{\rho}(z + \delta z) = \tilde{\rho}(z) - \tilde{\rho}(z + \delta z) = -\frac{\partial\tilde{\rho}}{\partial z}\delta z. \quad (2.215)$$

The parcel will be heavier than its surroundings, and therefore the parcel displacement will be stable, if $\partial\tilde{\rho}/\partial z < 0$. Similarly, it will be unstable if $\partial\tilde{\rho}/\partial z > 0$. The upward force (per unit volume) on the displaced parcel is given by

$$F = -g\delta\rho = g\frac{\partial\tilde{\rho}}{\partial z}\delta z, \quad (2.216)$$

and thus Newton's second law implies that the motion of the parcel is determined by

$$\rho(z)\frac{\partial^2\delta z}{\partial t^2} = g\frac{\partial\tilde{\rho}}{\partial z}\delta z, \quad (2.217)$$

or

$$\frac{\partial^2\delta z}{\partial t^2} = \frac{g}{\tilde{\rho}}\frac{\partial\tilde{\rho}}{\partial z}\delta z = -N^2\delta z, \quad (2.218)$$

where

$$N^2 = -\frac{g}{\tilde{\rho}}\frac{\partial\tilde{\rho}}{\partial z} \quad (2.219)$$

is the *buoyancy frequency*, or the *Brunt-Väisälä frequency*, for this problem. If $N^2 > 0$ then a parcel displaced upward is heavier than its surroundings, and thus experiences a restoring force; the density profile is said to be stable and N is the frequency at which the fluid parcel oscillates. If $N^2 < 0$, the density profile is unstable and the parcel continues to ascend and convection ensues. In liquids it is often a good approximation to replace $\tilde{\rho}$ by ρ_0 in the denominator of (2.219).

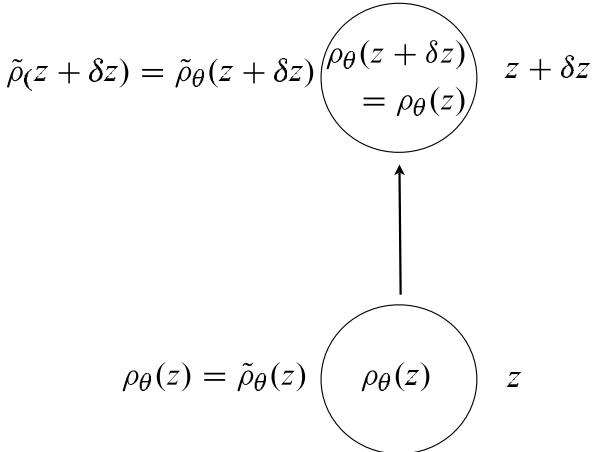


Fig. 2.7 A parcel is adiabatically displaced upward from level z to $z + \delta z$, preserving its potential density, which it takes from the environment at level z . If $z + \delta z$ is the reference level, the potential density there is equal to the actual density. The parcel's stability is determined by the difference between its density and the environmental density [see (2.220)]; if the difference is positive the displacement is stable, and conversely.

2.9.2 The general case: using potential density

More generally, in an adiabatic displacement it is *potential density*, ρ_θ , and not density itself that is materially conserved. Consider a parcel that is displaced adiabatically a vertical distance from z to $z + \delta z$; the parcel preserves its potential density, and let us use the pressure at level $z + \delta z$ as the reference level. The *in situ* density of the parcel at $z + \delta z$, namely $\rho(z + \delta z)$, is then equal to its potential density $\rho_\theta(z + \delta z)$ and, because ρ_θ is conserved, this is equal to the potential density of the environment at z , $\tilde{\rho}_\theta(z)$. The difference in *in situ* density between the parcel and the environment at $z + \delta z$, $\delta\rho$, is thus equal to the difference between the potential density of the environment at z and at $z + \delta z$. Putting this together (and see Fig. 2.7) we have

$$\begin{aligned}\delta\rho &= \rho(z + \delta z) - \tilde{\rho}(z + \delta z) = \rho_\theta(z + \delta z) - \tilde{\rho}_\theta(z + \delta z) \\ &= \rho_\theta(z) - \tilde{\rho}_\theta(z + \delta z) = \tilde{\rho}_\theta(z) - \tilde{\rho}_\theta(z + \delta z),\end{aligned}\quad (2.220)$$

and therefore

$$\delta\rho = -\frac{\partial \tilde{\rho}_\theta}{\partial z} \delta z, \quad (2.221)$$

where the derivative on the right-hand side is the environmental gradient of potential density. If the right-hand side is positive, the parcel is heavier than its surroundings and the displacement is stable. Thus, the conditions for stability are:

stability : instability :	$\frac{\partial \tilde{\rho}_\theta}{\partial z} < 0$ $\frac{\partial \tilde{\rho}_\theta}{\partial z} > 0$
------------------------------	---

(2.222a,b)

That is, *the stability of a parcel of fluid is determined by the gradient of the locally-referenced potential density*. The equation of motion of the fluid parcel is

$$\frac{\partial^2 \delta z}{\partial t^2} = \frac{g}{\rho} \left(\frac{\partial \tilde{\rho}_\theta}{\partial z} \right) \delta z = -N^2 \delta z, \quad (2.223)$$

where, noting that $\rho(z) = \tilde{\rho}_\theta(z)$ to within $O(\delta z)$,

$$N^2 = -\frac{g}{\tilde{\rho}_\theta} \left(\frac{\partial \tilde{\rho}_\theta}{\partial z} \right). \quad (2.224)$$

This is a general expression for the buoyancy frequency, true in both liquids and gases. It is important to realize that the quantity $\tilde{\rho}_\theta$ is the *locally-referenced* potential density of the environment, as will become more clear below.

An ideal gas

In the atmosphere potential density is related to potential temperature by $\rho_\theta = p_R/(\theta R)$. Using this in (2.224) gives

$$N^2 = \frac{g}{\tilde{\theta}} \left(\frac{\partial \tilde{\theta}}{\partial z} \right), \quad (2.225)$$

where $\tilde{\theta}$ refers to the environmental profile of potential temperature. The reference value p_R does not appear, and we are free to choose this value arbitrarily — the surface pressure is a common choice. The conditions for stability, (2.222), then correspond to $N^2 > 0$ for stability and $N^2 < 0$ for instability. In the troposphere (the lowest several kilometres of the atmosphere) the average N is about 0.01 s^{-1} , with a corresponding period, $(2\pi/N)$, of about 10 minutes. In the stratosphere (which lies above the troposphere) N^2 is a few times higher than this.

A liquid ocean

No simple, accurate, analytic expression is available for computing static stability in the ocean. If the ocean had no salt, then the potential density referenced to the surface would generally be a measure of the sign of stability of a fluid column, if not of the buoyancy frequency. However, in the presence of salinity, the surface-referenced potential density is not necessarily even a measure of the sign of stability, because the coefficients of compressibility β_T and β_S vary in different ways with pressure. To see this, suppose two neighbouring fluid elements at the surface have the same potential density, but different salinities and temperatures, and displace them both adiabatically to the deep ocean. Although their potential densities (referenced to the surface) are still equal, we can say little about their actual densities, and hence their stability relative to each other, without doing a detailed calculation because they will each have been compressed by different amounts. It is the profile of the *locally-referenced* potential density that determines the stability.

An approximate expression for stability that is sometimes useful arises by noting that in an adiabatic displacement

$$\delta\rho_\theta = \delta\rho - \frac{1}{c_s^2} \delta p = 0. \quad (2.226)$$

If the fluid is hydrostatic $\delta p = -\rho g \delta z$ so that if a parcel is displaced adiabatically its density changes according to

$$\left(\frac{\partial \rho}{\partial z} \right)_{\rho_\theta} = -\frac{\rho g}{c_s^2}. \quad (2.227)$$

If a parcel is displaced a distance δz upwards then the density difference between it and its new surroundings is

$$\delta\rho = - \left[\left(\frac{\partial\rho}{\partial z} \right)_{\rho_0} - \left(\frac{\partial\tilde{\rho}}{\partial z} \right) \right] \delta z = \left[\frac{\rho g}{c_s^2} + \left(\frac{\partial\tilde{\rho}}{\partial z} \right) \right] \delta z, \quad (2.228)$$

where the tilde again denotes the environmental field. It follows that the stratification is given by

$$N^2 = -g \left[\frac{g}{c_s^2} + \frac{1}{\tilde{\rho}} \left(\frac{\partial\tilde{\rho}}{\partial z} \right) \right].$$

(2.229)

This expression holds for both liquids and gases, and for ideal gases it is precisely the same as (2.225) (problem 2.9). In liquids, a good approximation is to use a reference value ρ_0 for the undifferentiated density in the denominator, whence (2.229) becomes equal to the Boussinesq expression (2.107). Typical values of N in the upper ocean where the density is changing most rapidly (i.e., in the pycnocline — ‘pycno’ for density, ‘cline’ for changing) are about 0.01 s^{-1} , falling to 0.001 s^{-1} in the more homogeneous abyssal ocean. These frequencies correspond to periods of about 10 and 100 minutes, respectively.

* Cabbeling

Cabbeling is an instability that arises because of the nonlinear equation of state of seawater. From Fig. 1.3 we see that the contours are slightly convex, bowing upwards, especially in the plot at sea level. Suppose we mix two parcels of water, each with the same density ($\sigma_\theta = 28$, say), but with different initial values of temperature and salinity. Then the resulting parcel of water will have a temperature and a salinity equal to the average of the two parcels, but its density will be *higher* than either of the two original parcels. In the appropriate circumstances such mixing may thus lead to a convective instability; this may, for example, be an important source of ‘bottom water’ formation in the Weddell Sea, off Antarctica.¹⁰

2.9.3 Lapse rates in dry and moist atmospheres

A dry ideal gas

The negative of the rate of change of the temperature in the vertical is known as the *temperature lapse rate*, or often just the *lapse rate*, and the lapse rate corresponding to $\partial\theta/\partial z = 0$ is called the *dry adiabatic lapse rate* and denoted Γ_d . Using $\theta = T(p_0/p)^{R/c_p}$ and $\partial p/\partial z = -\rho g$ we find that the lapse rate and the potential temperature lapse rate are related by

$$\frac{\partial T}{\partial z} = \frac{T}{\theta} \frac{\partial\theta}{\partial z} - \frac{g}{c_p}, \quad (2.230)$$

so that the dry adiabatic lapse rate is given by

$$\Gamma_d = \frac{g}{c_p}, \quad (2.231)$$

as in fact we derived in (1.134). (We use the subscript *d*, for dry, to differentiate it from the moist lapse rate considered below.) The conditions for static stability corresponding to

(2.222) are thus:

$$\boxed{\begin{aligned} \text{stability : } & \frac{\partial \tilde{\theta}}{\partial z} > 0; \quad \text{or} \quad -\frac{\partial \tilde{T}}{\partial z} < \Gamma_d \\ \text{instability : } & \frac{\partial \tilde{\theta}}{\partial z} < 0; \quad \text{or} \quad -\frac{\partial \tilde{T}}{\partial z} > \Gamma_d \end{aligned}}, \quad (2.232a,b)$$

where a tilde indicates that the values are those of the environment. The atmosphere is in fact generally stable by this criterion: the observed lapse rate, corresponding to an observed buoyancy frequency of about 10^{-2} s^{-1} , is often about 7 K km^{-1} , whereas a dry adiabatic lapse rate is about 10 K km^{-1} . Why the discrepancy? One reason, particularly important in the tropics, is that the atmosphere contains water vapour.

* Effects of water vapour on the lapse rate of an ideal gas

The amount of water vapour that can be contained in a given volume is an increasing function of temperature (with the presence or otherwise of dry air in that volume being largely irrelevant). Thus, if a parcel of water vapour is cooled, it will eventually become saturated and water vapour will condense into liquid water. A measure of the amount of water vapour in a unit volume is its partial pressure, and the partial pressure of water vapour at saturation, e_s , is given by the Clausius-Clapeyron equation,

$$\frac{de_s}{dT} = \frac{L_c e_s}{R_v T^2}, \quad (2.233)$$

where L_c is the latent heat of condensation or vapourization (per unit mass) and R_v is the gas constant for water vapour. If a parcel rises adiabatically it will cool, and at some height (known as the 'lifting condensation level', a function of its initial temperature and humidity only) the parcel will become saturated and any further ascent will cause the water vapour to condense. The ensuing condensational heating causes the temperature and buoyancy of the parcel to increase; the parcel thus rises further, causing more water vapour to condense, and so on, and the consequence of this is that an environmental profile that is stable if the air is dry may be unstable if saturated. Let us now derive an expression for the lapse rate of a saturated parcel that is ascending adiabatically apart from the effects of condensation.

Let w denote the mass of water vapour per unit mass of dry air, the mixing ratio, and let w_s be the saturation mixing ratio. ($w_s = \alpha e_s / (p - e_s) \approx \alpha_w e_s / p$ where $\alpha_w = 0.62$, the ratio of the mass of a water molecule to one of dry air.) The diabatic heating associated with condensation is then given by

$$Q_{cond} = -L_c \frac{Dw_s}{Dt}, \quad (2.234)$$

so that the thermodynamic equation is

$$c_p \frac{D \ln \theta}{Dt} = -\frac{L_c}{T} \frac{Dw_s}{Dt}, \quad (2.235)$$

or, in terms of p and T

$$c_p \frac{D \ln T}{Dt} - R \frac{D \ln P}{Dt} = -\frac{L_c}{T} \frac{Dw_s}{Dt}. \quad (2.236)$$

If these material derivatives are due to the parcel ascent then

$$\frac{d \ln T}{dz} - \frac{R}{c_p} \frac{d \ln p}{dz} = - \frac{L_c}{T c_p} \frac{dw_s}{dz}, \quad (2.237)$$

and using the hydrostatic relationship and the fact that w_s is a function of T and p we obtain

$$\frac{dT}{dz} + \frac{g}{c_p} = - \frac{L_c}{c_p} \left[\left(\frac{\partial w_s}{\partial T} \right)_p \frac{dT}{dz} - \left(\frac{\partial w_s}{\partial p} \right)_T \rho g \right]. \quad (2.238)$$

Solving for dT/dz , the lapse rate, Γ_s , of an ascending saturated parcel is given by

$$\Gamma_s = - \frac{dT}{dz} = \frac{g}{c_p} \frac{1 - \rho L_c (\partial w_s / \partial p)_T}{1 + (L_c / c_p) (\partial w_s / \partial T)_p} \approx \frac{g}{c_p} \frac{1 + L_c w_s / (RT)}{1 + L_c^2 w_s / (c_p RT^2)}. \quad (2.239)$$

where the last near equality follows with use of the Clausius–Clapeyron relation. The quantity Γ_s is variously called the *pseudoadiabatic* or *moist adiabatic* or *saturated adiabatic* lapse rate. Because g/c_p is the dry adiabatic lapse rate Γ_d , $\Gamma_s < \Gamma_d$, and values of Γ_s are typically around 6 K km^{-1} in the lower atmosphere; however, dw_s/dT is an increasing function of T so that Γ_s decreases with increasing temperature and can be as low as 3.5 K km^{-1} . For a saturated parcel, the stability conditions analogous to (2.232) are

$$\text{stability : } - \frac{\partial \tilde{T}}{\partial z} < \Gamma_s, \quad (2.240a)$$

$$\text{instability : } - \frac{\partial \tilde{T}}{\partial z} > \Gamma_s. \quad (2.240b)$$

where \tilde{T} is the environmental temperature. The observed environmental profile in convecting situations is often a combination of the dry adiabatic and moist adiabatic profiles: an unsaturated parcel that is unstable by the dry criterion will rise and cool following a dry adiabat, Γ_d , until it becomes saturated at the lifting condensation level, above which it will rise following a saturation adiabat, Γ_s . Such convection will proceed until the atmospheric column is stable and, especially in low latitudes, the lapse rate of the atmosphere is largely determined by such convective processes.

* Equivalent potential temperature

Suppose that all the moisture in a parcel of air condenses, and that all the heat released goes into heating the parcel. The *equivalent potential temperature*, θ_{eq} is the potential temperature that the parcel then achieves. We may obtain an approximate analytic expression for it by noting that the first law of thermodynamics, $dQ = T d\eta$, then implies, by definition of potential temperature,

$$-L_c dw = c_p T d \ln \theta, \quad (2.241)$$

where dw is the change in water vapour mixing ratio, so that a reduction of w via condensation leads to heating. Integrating gives, by definition of equivalent potential temperature,

$$- \int_w^0 \frac{L_c w}{c_p T} dw = \int_{\theta}^{\theta_{eq}} d \ln \theta, \quad (2.242)$$

and so, if T and L_c are assumed to be constant,

$$\theta_{eq} = \theta \exp\left(\frac{L_c w}{c_p T}\right). \quad (2.243)$$

The equivalent potential temperature so defined is approximately conserved during condensation, the approximation arising going from (2.242) to (2.243). It is a useful expression for diagnostic purposes, and in constructing theories of convection, but it is not accurate enough to use as a prognostic variable in a putatively realistic numerical model. The ‘equivalent temperature’ may be defined in terms of the equivalent potential temperature by

$$T_{eq} \equiv \theta_{eq} \left(\frac{p}{p_R}\right)^k. \quad (2.244)$$

2.10 GRAVITY WAVES

The parcel approach to oscillations and stability, while simple and direct, is divorced from the fluid-dynamical equations of motion, making it hard to include other effects such as rotation, or to explore the effects of possible differences between the hydrostatic and non-hydrostatic cases. To remedy this, we now use the equations of motion to analyse the motion resulting from a small disturbance.

2.10.1 Gravity waves and convection in a Boussinesq fluid

Let us consider a Boussinesq fluid, initially at rest, in which the buoyancy varies linearly with height and the buoyancy frequency, N , is a constant. Linearizing the equations of motion about this basic state gives the linear momentum equations,

$$\frac{\partial u'}{\partial t} = -\frac{\partial \phi'}{\partial x}, \quad \frac{\partial w'}{\partial t} = -\frac{\partial \phi'}{\partial z} + b', \quad (2.245a,b)$$

the mass continuity and thermodynamic equations,

$$\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} = 0, \quad \frac{\partial b'}{\partial t} + w' N^2 = 0, \quad (2.246a,b)$$

where for simplicity we assume that the flow is a function only of x and z . A little algebra gives a single equation for w' ,

$$\left[\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} \right) \frac{\partial^2}{\partial t^2} + N^2 \frac{\partial^2}{\partial x^2} \right] w' = 0. \quad (2.247)$$

Seeking solutions of the form $w' = \text{Re } W \exp[i(kx + mz - \omega t)]$ (where Re denotes the real part) yields the dispersion relationship for gravity waves:

$$\omega^2 = \frac{k^2 N^2}{k^2 + m^2}.$$

(2.248)

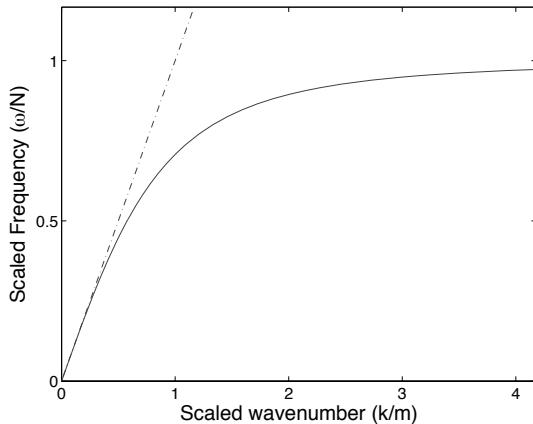


Fig. 2.8 Scaled frequency, ω/N , plotted as a function of scaled horizontal wavenumber, k/m , using the full dispersion relation of (2.248) (solid line, asymptoting to unit value for large k/m) and with the hydrostatic dispersion relation (2.252) (dashed line, tending to ∞ for large k/m).

The frequency (see Fig. 2.8) is thus always less than N , approaching N for small horizontal scales, $k \gg m$. If we neglect pressure perturbations, as in the parcel argument, then the two equations,

$$\frac{\partial w'}{\partial t} = b', \quad \frac{\partial b'}{\partial t} + w'N^2 = 0, \quad (2.249)$$

form a closed set, and give $\omega^2 = N^2$.

If the basic state density increases with height then $N^2 < 0$ and we expect this state to be unstable. Indeed, the disturbance grows exponentially according to $\exp(\sigma t)$ where

$$\sigma = i\omega = \frac{\pm k\tilde{N}}{(k^2 + m^2)^{1/2}}, \quad (2.250)$$

where $\tilde{N}^2 = -N^2$. Most convective activity in the ocean and atmosphere is, ultimately, related to an instability of this form, although of course there are many complicating issues — water vapour in the atmosphere, salt in the ocean, the effects of rotation and so forth.

Hydrostatic gravity waves and convection

Let us now suppose that the fluid satisfies the hydrostatic Boussinesq equations. The linearized two-dimensional equations of motion become

$$\frac{\partial u'}{\partial t} = -\frac{\partial \phi'}{\partial x}, \quad 0 = -\frac{\partial \phi'}{\partial z} + b', \quad (2.251a)$$

$$\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} = 0, \quad \frac{\partial b'}{\partial t} + w'N^2 = 0, \quad (2.251b)$$

where these are the horizontal and vertical momentum equations, the mass continuity equation and the thermodynamic equation respectively. A little algebra gives the dispersion relation,

$$\omega^2 = \frac{k^2 N^2}{m^2}. \quad (2.252)$$

The frequency and, if N^2 is negative the growth rate, is unbounded for as $k/m \rightarrow \infty$, and the hydrostatic approximation thus has quite unphysical behaviour for small horizontal scales (see also problem 2.11).¹¹

2.11 * ACOUSTIC-GRAVITY WAVES IN AN IDEAL GAS

We now consider wave motion in a stratified, compressible fluid such as the Earth's atmosphere. The complete problem is complicated and uninformative; we will specialize to the case of an isothermal, stationary atmosphere and ignore the effects of rotation and sphericity, but otherwise we will make few approximations. In this section we will denote the unperturbed state with a subscript 0 and the perturbed state with a prime (''); we will also omit many algebraic details. Because it is at rest, the basic state is in hydrostatic balance,

$$\frac{\partial p_0}{\partial z} = -\rho_0(z)g. \quad (2.253)$$

Ignoring variations in the y -direction for algebraic simplicity, the linearized equations of motion are:

$$u \text{ momentum:} \quad \rho_0 \frac{\partial u'}{\partial t} = -\frac{\partial p'}{\partial x} \quad (2.254a)$$

$$w \text{ momentum:} \quad \rho_0 \frac{\partial w'}{\partial t} = -\frac{\partial p'}{\partial z} - \rho' g \quad (2.254b)$$

$$\text{mass conservation:} \quad \frac{\partial \rho'}{\partial t} + w' \frac{\partial \rho_0}{\partial z} = -\rho_0 \left(\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} \right) \quad (2.254c)$$

$$\text{thermodynamic:} \quad \frac{\partial \theta'}{\partial t} + w' \frac{\partial \theta_0}{\partial z} = 0 \quad (2.254d)$$

$$\text{equation of state:} \quad \frac{\theta'}{\theta_0} + \frac{\rho'}{\rho_0} = \frac{1}{\gamma} \frac{p'}{p_0}. \quad (2.254e)$$

For an isothermal basic state we have $p_0 = \rho_0 RT_0$ where T_0 is a constant, so that $\rho_0 = \rho_s e^{-z/H}$ and $p_0 = p_s e^{-z/H}$ where $H = RT_0/g$. Further, using $\theta = T(p_s/p)^\kappa$ where $\kappa = R/c_p$, we have $\theta_0 = T_0 e^{\kappa z/H}$ and so $N^2 = \kappa g/H$. It is also convenient to use (1.99) on page 23 to rewrite the linear thermodynamic equation in the form

$$\frac{\partial p'}{\partial t} - w' \frac{p_0}{H} = -\gamma p_0 \left(\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} \right). \quad (2.254f)$$

Differentiating (2.254a) with respect to time and using (2.254f) leads to

$$\left(\frac{\partial^2}{\partial t^2} - c_s^2 \frac{\partial^2}{\partial x^2} \right) u' = c_s^2 \left(\frac{\partial}{\partial z} - \frac{1}{\gamma H} \right) \frac{\partial}{\partial x} w'. \quad (2.255a)$$

where $c_s^2 = (\partial p / \partial \rho)_\eta = \gamma RT_0 = \gamma p_0 / \rho_0$ is the square of the speed of sound, and $\gamma = c_p/c_v = 1/(1 - \kappa)$. Similarly, differentiating (2.254b) with respect to time and using (2.254c) and (2.254f) leads to

$$\left(\frac{\partial^2}{\partial t^2} - c_s^2 \left[\frac{\partial^2}{\partial z^2} - \frac{1}{H} \frac{\partial}{\partial z} \right] \right) w' = c_s^2 \left(\frac{\partial}{\partial z} - \frac{\kappa}{H} \right) \frac{\partial u' }{\partial x}, \quad (2.255b)$$

Equations (2.255a) and (2.255b) combine to give, after some cancellation,

$$\frac{\partial^4 w'}{\partial t^4} - c_s^2 \frac{\partial^2}{\partial t^2} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{H} \frac{\partial}{\partial z} \right) w' - c_s^2 \frac{\kappa g}{H} \frac{\partial^2 w'}{\partial x^2} = 0. \quad (2.256)$$

If we set $w' = W(x, z, t)e^{z/(2H)}$, so that $W = (\rho_0/\rho_s)^{1/2}w$, then the term with the single z -derivative is eliminated, giving

$$\frac{\partial^4 W}{\partial t^4} - c_s^2 \frac{\partial^2}{\partial t^2} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{4H^2} \right) W - c_s^2 \frac{\kappa g}{H} \frac{\partial^2 W}{\partial x^2} = 0. \quad (2.257)$$

Although superficially complicated, this equation has constant coefficients and we may seek wave-like solutions of the form

$$W = \operatorname{Re} \widetilde{W} e^{i(kx + mz - \omega t)}, \quad (2.258)$$

where \widetilde{W} is the complex wave amplitude. Using (2.258) in (2.257) leads to the dispersion relation for acoustic-gravity waves, namely

$$\omega^4 - c_s^2 \omega^2 \left(k^2 + m^2 + \frac{1}{4H^2} \right) + c_s^2 N^2 k^2 = 0, \quad (2.259)$$

with solution

$$\boxed{\omega^2 = \frac{1}{2} c_s^2 K^2 \left[1 \pm \left(1 - \frac{4N^2 k^2}{c_s^2 K^4} \right)^{1/2} \right]}, \quad (2.260)$$

where $K^2 = k^2 + m^2 + 1/(4H^2)$. (The factor $[1 - 4N^2 k^2/(c_s^2 K^4)]$ is always positive — see problem 2.26.) For an isothermal, ideal-gas atmosphere $4N^2 H^2/c_s^2 \approx 0.8$ and so this may be written

$$\frac{\omega^2}{N^2} \approx 2.5 \hat{K}^2 \left[1 \pm \left(1 - \frac{0.8 \hat{k}^2}{\hat{K}^4} \right)^{1/2} \right], \quad (2.261)$$

where $\hat{K}^2 = \hat{k}^2 + \hat{m}^2 + 1/4$, and $(\hat{k}, \hat{m}) = (kH, mH)$.

2.11.1 Interpretation

Acoustic and gravity waves

There are two branches of roots in (2.260), corresponding to acoustic waves (using the plus sign in the dispersion relation) and internal gravity waves (using the minus sign). These (and the Lamb wave, described below) are plotted in Fig. 2.9. If $4N^2 k^2/c_s^2 K^4 \ll 1$ then the two sets of waves are well separated. From (2.261) this is satisfied when

$$\frac{4\kappa}{\gamma} (kH)^2 \approx 0.8(kH)^2 \ll \left[(kH)^2 + (mH)^2 + \frac{1}{4} \right]^2; \quad (2.262)$$

that is, when *either* $mH \gg 1$ or $kH \gg 1$. The two roots of the dispersion relation are then

$$\omega_a^2 \approx c_s^2 K^2 = c_s^2 \left(k^2 + m^2 + \frac{1}{4H^2} \right) \quad (2.263)$$

and

$$\omega_g^2 \approx \frac{N^2 k^2}{k^2 + m^2 + 1/(4H^2)}, \quad (2.264)$$

corresponding to acoustic and gravity waves, respectively. The acoustic waves owe their existence to the presence of compressibility in the fluid, and they have no counterpart in the

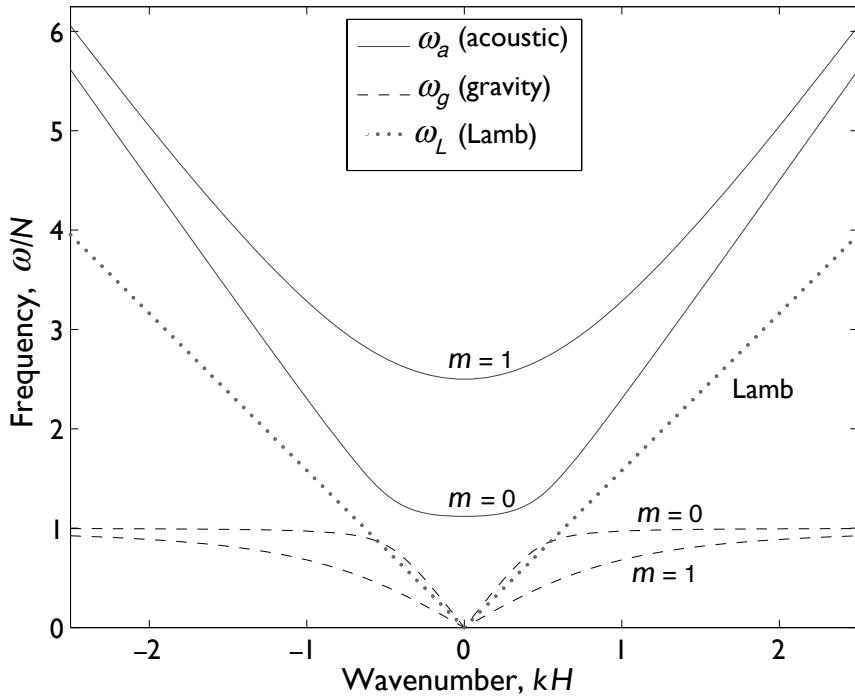


Fig. 2.9 Dispersion diagram for acoustic gravity waves in an isothermal atmosphere, calculated using (2.261). The frequency is given in units of the the buoyancy frequency N , and the wavenumbers are non-dimensionalized by the inverse of the scale height, H . The solid curves indicate acoustic waves, whose frequency is always higher than that of the corresponding Lamb wave at the same wavenumber (i.e., ck), and of the base acoustic frequency $\approx 1.12N$. The dashed curves indicate internal gravity waves, whose frequency asymptotes to N at small horizontal scales.

Boussinesq system. On the other hand, the internal gravity waves are just modified forms of those found in the Boussinesq system, and if we take the limit $(kH, mH) \rightarrow \infty$ then the gravity wave branch reduces to $\omega_g^2 = N^2 k^2 / (k^2 + m^2)$, which is the dispersion relationship for gravity waves in the Boussinesq approximation. We may consider this to be the limit of infinite scale height or (equivalently) the case in which wavelengths of the internal waves are sufficiently small that the fluid is essentially incompressible.

Vertical structure

Recall that $w' = W(x, z, t) e^{z/(2H)}$ and, by inspection of (2.255), u' has the same vertical structure. That is,

$$w' \propto e^{z/(2H)}, \quad u' \propto e^{z/(2H)}, \quad (2.265)$$

and the amplitude of the velocity field of the internal waves increases with height. The pressure and density perturbation amplitudes fall off with height, varying like

$$p' \propto e^{-z/(2H)}, \quad \rho' \propto e^{-z/(2H)}. \quad (2.266)$$

The kinetic energy of the perturbation, $\rho_0(u'^2 + w'^2)$ is *constant* with height, because $\rho_0 = \rho_s e^{-z/H}$.

Hydrostatic approximation and Lamb waves

Equations (2.255) also admit to a solution with $w' = 0$. We then have

$$\left(\frac{\partial^2}{\partial t^2} - c_s^2 \frac{\partial^2}{\partial x^2} \right) u' = 0 \quad \text{and} \quad \left(\frac{\partial}{\partial z} - \frac{\kappa}{H} \right) \frac{\partial u'}{\partial x} = 0, \quad (2.267)$$

and these have solutions of the form

$$u' = \text{Re } \tilde{U} e^{\kappa z/H} e^{i(kx - \omega t)}, \quad \omega = ck, \quad (2.268)$$

where \tilde{U} is the wave amplitude. These are horizontally propagating sound waves, known as *Lamb waves* after the hydrodynamicist Horace Lamb. Their velocity perturbation amplitude increases with height, but the pressure perturbation falls with height; that is

$$u' \propto e^{\kappa z/H} \approx e^{2z/(7H)}, \quad p' \propto e^{(\kappa-1)z/H} \approx e^{-5z/(7H)}. \quad (2.269)$$

Their kinetic energy density, $\rho_0 u'^2$, varies as

$$KE \propto e^{-z/H+2\kappa z/H} = e^{(2R-c_p)z/(c_p H)} = e^{(R-c_v)z/(c_p H)} \approx e^{-3z/(7H)} \quad (2.270)$$

for an ideal gas. (In a simple ideal gas, $c_v = nR/2$ where n is the number of excited degrees of freedom, 5 for a diatomic molecule.) The kinetic energy density thus falls away exponentially from the surface, and in this sense Lamb waves are an example of edge waves or surface-trapped waves.

Now consider the case in which we make the hydrostatic approximation *ab initio*, but do not restrict the perturbation to have $w' = 0$. The linearized equations are identical to (2.254), except that (2.254b) is replaced by

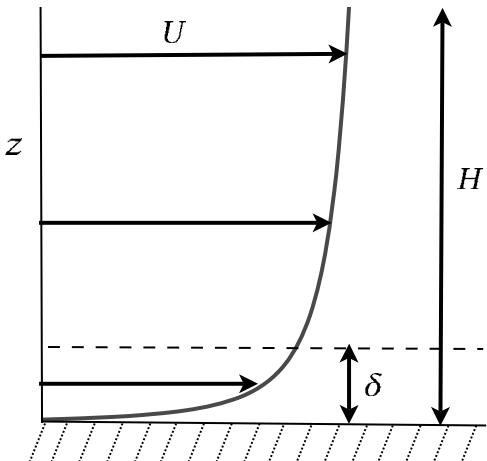
$$\frac{\partial p'}{\partial z} = -\rho' g. \quad (2.271)$$

The consequence of this is that first term $(\partial^2 w' / \partial t^2)$ in (2.255b) disappears, as do the first two terms in (2.256) [the terms $\partial^4 w' / \partial t^4 - c^2 (\partial^2 / \partial t^2) (\partial^2 w' / \partial x^2)$]. It is a simple matter to show that the dispersion relation is then

$$\omega^2 = \frac{N^2 k^2}{m^2 + 1/(4H^2)}. \quad (2.272)$$

These are long gravity waves, and may be compared with the corresponding Boussinesq result (2.252). Again, the frequency increases without bound as the horizontal wavelength diminishes. The Lamb wave, of course, still exists in the hydrostatic model, because (2.267) is still a valid solution. Thus, horizontally propagating sound waves still exist in hydrostatic (primitive equation) models, but vertically propagating sound waves do not — essentially because the term $\partial w / \partial t$ is absent from the vertical momentum equation.

Fig. 2.10 An idealized boundary layer. The values of a field, such as velocity, U , may vary rapidly in a boundary in order to satisfy the boundary conditions at a rigid surface. The parameter δ is a measure of the boundary layer thickness, and H is a typical scale of variation away from the boundary.



2.12 THE EKMAN LAYER

In the final topic of this chapter, we return to geostrophic flow and consider the effects of friction. The fluid fields in the interior of a domain are often set by different physical processes than those occurring at a boundary, and consequently often change rapidly in a thin *boundary layer*, as in Fig. 2.10. Such boundary layers nearly always involve one or both of viscosity and diffusion, because these appear in the terms of highest differential order in the equations of motion, and so are responsible for the number and type of boundary conditions that the equations must satisfy — for example, the presence of molecular viscosity leads to the condition that the tangential flow (as well as the normal flow) must vanish at a rigid surface.

In many boundary layers in non-rotating flow the dominant balance in the momentum equation is between the advective and viscous terms. In some contrast, in large-scale atmospheric and oceanic flow the effects of rotation are large, and this results in a boundary layer, known as the *Ekman layer*, in which the dominant balance is between Coriolis and frictional or stress terms.¹² Now, the direct effects of molecular viscosity and diffusion are nearly always negligible at distances more than a few millimetres away from a solid boundary, but it is inconceivable that the entire boundary layer between the free atmosphere (or free ocean) and the surface is only a few millimetres thick. Rather, in practice a balance occurs between the Coriolis terms and the forces due to the stress generated by small-scale turbulent motion, and this gives rise to a boundary layer that has a typical depth of a few tens to several hundreds of metres. Because the stress arises from the turbulence we cannot with confidence determine its precise form; thus, we should try to determine what general properties Ekman layers may have that are *independent* of the precise form of the friction.

The atmospheric Ekman layer occurs near the ground, and the stress at the ground itself is due to the surface wind (and its vertical variation). In the ocean the main Ekman layer is near the surface, and the stress at ocean surface is largely due to the presence of the overlying wind. There is also a weak Ekman layer at the bottom of the ocean, analogous to the atmospheric Ekman layer. To analyse all these layers, let us assume the following.

- ★ The Ekman layer is Boussinesq. This is a very good assumption for the ocean, and a reasonable one for the atmosphere if the boundary layer is not too deep.

- ★ The Ekman layer has a finite depth that is less than the total depth of the fluid, this depth being given by the level at which the frictional stresses essentially vanish. Within the Ekman layer, frictional terms are important, whereas geostrophic balance holds beyond it.
- ★ The nonlinear and time-dependent terms in the equations of motion are negligible, hydrostatic balance holds in the vertical, and buoyancy is constant, not varying in the horizontal.
- ★ As needed, we shall assume that friction can be parameterized by a viscous term of the form $\rho_0^{-1} \partial \boldsymbol{\tau} / \partial z = A \partial^2 \mathbf{u} / \partial z^2$, where A is constant and $\boldsymbol{\tau}$ is the stress. [In general, stress is a tensor, τ_{ij} , with an associated force given by $F_i = \partial \tau_{ij} / \partial x_j$, summing over the repeated index. It is common in geophysical fluid dynamics that the vertical derivative dominates, and in this case the force is $\mathbf{F} = \partial \boldsymbol{\tau} / \partial z$. We still use the word stress for $\boldsymbol{\tau}$, but it now refers to a vector whose derivative in a particular direction (z in this case) is the force on a fluid.] In laboratory settings A may be the molecular viscosity, whereas in the atmosphere and ocean it is a so-called *eddy viscosity*. (In turbulent flows momentum is transferred by the near-random motion of small parcels of fluid and, by analogy with the motion of molecules that produces a molecular viscosity, the associated stress is approximately represented, or parameterized, using a turbulent or eddy viscosity that may be orders of magnitude larger than the molecular one.)

2.12.1 Equations of motion and scaling

Frictional-geostrophic balance in the horizontal momentum equation is:

$$\mathbf{f} \times \mathbf{u} = -\nabla_z \phi + \frac{\partial \tilde{\boldsymbol{\tau}}}{\partial z}. \quad (2.273)$$

where $\tilde{\boldsymbol{\tau}} \equiv \boldsymbol{\tau} / \rho_0$ is the kinematic stress and $\mathbf{f} = f \mathbf{k}$, where the Coriolis parameter f is allowed to vary with latitude. If we model the stress with an eddy viscosity, (2.273) becomes

$$\mathbf{f} \times \mathbf{u} = -\nabla_z \phi + A \frac{\partial^2 \mathbf{u}}{\partial z^2}. \quad (2.274)$$

The vertical momentum equation is $\partial \phi / \partial z = b$, i.e., hydrostatic balance, and, because buoyancy is constant, we may without loss of generality write this as

$$\frac{\partial \phi}{\partial z} = 0. \quad (2.275)$$

The equation set is completed by the mass continuity equation, $\nabla \cdot \mathbf{v} = 0$.

The Ekman number

We non-dimensionalize the equations by setting

$$(u, v) = U(\hat{u}, \hat{v}), \quad (x, y) = L(\hat{x}, \hat{y}), \quad f = f_0 \hat{f}, \quad z = H \hat{z}, \quad \phi = \Phi \hat{\phi}, \quad (2.276)$$

where hatted variables are non-dimensional. H is a scaling for the height, and at this stage we will suppose it to be some height scale in the free atmosphere or ocean, not the height of

the Ekman layer itself. Geostrophic balance suggests that $\Phi = f_0 UL$. Substituting (2.276) into (2.274) we obtain

$$\hat{\mathbf{f}} \times \hat{\mathbf{u}} = -\hat{\nabla} \hat{\phi} + Ek \frac{\partial^2 \hat{\mathbf{u}}}{\partial \hat{z}^2}, \quad (2.277)$$

where the parameter

$$Ek \equiv \left(\frac{A}{f_0 H^2} \right), \quad (2.278)$$

is the *Ekman number*, and it determines the importance of frictional terms in the horizontal momentum equation. If $Ek \ll 1$ then the friction is small in the flow interior where $\hat{z} = \mathcal{O}(1)$. However, the friction term cannot necessarily be neglected in the boundary layer because it is of the highest differential order in the equation, and so determines the boundary conditions; if Ek is small the vertical scales become small and the second term on the right-hand side of (2.277) remains finite. The case when this term is simply omitted from the equation is therefore a *singular limit*, meaning that it differs from the case with $Ek \rightarrow 0$. If $Ek \geq 1$ friction is important everywhere, but it is usually the case that Ek is small for atmospheric and oceanic large-scale flow, and the interior flow is very nearly geostrophic. (In part this is because A itself is only large near a rigid surface where the presence of a shear creates turbulence and a significant eddy viscosity.)

Momentum balance in the Ekman layer

For definiteness, suppose the fluid lies above a rigid surface at $z = 0$. Sufficiently far away from the boundary the velocity field is known, and we suppose this flow to be in geostrophic balance. We then write the velocity field and the pressure field as the sum of the interior geostrophic part, plus a boundary layer correction:

$$\hat{\mathbf{u}} = \hat{\mathbf{u}}_g + \hat{\mathbf{u}}_E, \quad \hat{\phi} = \hat{\phi}_g + \hat{\phi}_E, \quad (2.279)$$

where the Ekman layer corrections, denoted with a subscript E , are negligible away from the boundary layer. Now, in the fluid interior we have, by hydrostatic balance, $\partial \hat{\phi}_g / \partial \hat{z} = 0$. In the boundary layer we still have $\partial \hat{\phi}_g / \partial \hat{z} = 0$ so that, to satisfy hydrostasy, $\partial \hat{\phi}_E / \partial \hat{z} = 0$. But because $\hat{\phi}_E$ vanishes away from the boundary we have $\hat{\phi}_E = 0$ everywhere. Thus, *there is no boundary layer in the pressure field*. Note that this is a much stronger result than saying that pressure is continuous, which is nearly always true in fluids; rather, it is a special result for Ekman layers.

Using (2.279) with $\hat{\phi}_E = 0$, the dimensional horizontal momentum equation (2.273) becomes, in the Ekman layer,

$$\mathbf{f} \times \mathbf{u}_E = \frac{\partial \tilde{\tau}}{\partial z}. \quad (2.280)$$

The dominant force balance in the Ekman layer is thus between the Coriolis force and the friction. We can determine the thickness of the Ekman layer if we model the stress with an eddy viscosity so that

$$\mathbf{f} \times \mathbf{u}_E = A \frac{\partial^2 \mathbf{u}_E}{\partial z^2}, \quad (2.281)$$

or, non-dimensionally,

$$\hat{\mathbf{f}} \times \hat{\mathbf{u}}_E = Ek \frac{\partial^2 \hat{\mathbf{u}}_E}{\partial \hat{z}^2}. \quad (2.282)$$

It is evident this equation can only be satisfied if $\hat{z} \neq \mathcal{O}(1)$, implying that H is not a proper scaling for z in the boundary layer. Rather, if the vertical scale in the Ekman layer is $\hat{\delta}$ (meaning $\hat{z} \sim \hat{\delta}$) we must have $\hat{\delta} \sim Ek^{1/2}$. In dimensional terms this means the thickness of the Ekman layer is

$$\delta = H\hat{\delta} = HEk^{1/2} \quad (2.283)$$

or

$$\boxed{\delta = \left(\frac{A}{f_0} \right)^{1/2}}. \quad (2.284)$$

[This estimate also emerges directly from (2.281).] Note that (2.283) can be written as

$$Ek = \left(\frac{\delta}{H} \right)^2. \quad (2.285)$$

That is, the Ekman number is equal to the square of the ratio of the depth of the Ekman layer to an interior depth scale of the fluid motion. In laboratory flows where A is the molecular viscosity we can thus estimate the Ekman layer thickness, and if we know the eddy viscosity of the ocean or atmosphere we can estimate their respective Ekman layer thicknesses. We can invert this argument and obtain an estimate of A if we know the Ekman layer depth. In the atmosphere, deviations from geostrophic balance are very small in the atmosphere above 1 km, and using this gives $A \approx 10^2 \text{ m}^2 \text{ s}^{-1}$. In the ocean Ekman depths are often 50 m or less, and eddy viscosities are about $0.1 \text{ m}^2 \text{ s}^{-1}$.

2.12.2 Integral properties of the Ekman layer

What can we deduce about the Ekman layer without specifying the detailed form of the frictional term? Using dimensional notation we recall frictional-geostrophic balance,

$$\mathbf{f} \times \mathbf{u} = -\nabla \phi + \frac{1}{\rho_0} \frac{\partial \boldsymbol{\tau}}{\partial z}, \quad (2.286)$$

where $\boldsymbol{\tau}$ is zero at the edge of the Ekman layer. In the Ekman layer itself we have

$$\mathbf{f} \times \mathbf{u}_E = \frac{1}{\rho_0} \frac{\partial \boldsymbol{\tau}}{\partial z}. \quad (2.287)$$

Consider either a top or bottom Ekman layer, and integrate over its thickness. From (2.287) we obtain

$$\mathbf{f} \times \mathbf{M}_E = \boldsymbol{\tau}_T - \boldsymbol{\tau}_B, \quad (2.288)$$

where

$$\mathbf{M}_E = \int_{EK} \rho_0 \mathbf{u}_E \, dz \quad (2.289)$$

is the ageostrophic mass transport in the Ekman layer, and $\boldsymbol{\tau}_T$ and $\boldsymbol{\tau}_B$ are the respective stresses at the top and the bottom of the Ekman layer at hand. The stress at the top (bottom) will be zero in a bottom (top) Ekman layer and therefore, from (2.288),

$$\boxed{\begin{aligned} \text{top Ekman layer:} \quad \mathbf{M}_E &= -\frac{1}{f} \mathbf{k} \times \boldsymbol{\tau}_T \\ \text{bottom Ekman layer:} \quad \mathbf{M}_E &= \frac{1}{f} \mathbf{k} \times \boldsymbol{\tau}_B \end{aligned}}. \quad (2.290a,b)$$

The transport is thus at right angles to the stress at the surface, and proportional to the magnitude of the stress. These properties have a simple physical explanation: integrated over the depth of the Ekman layer the surface stress must be balanced by the Coriolis force, which in turn acts at right angles to the mass transport. A consequence of (2.290) is that the mass transports in adjacent oceanic and atmospheric Ekman layers are equal and opposite, because the stress is continuous across the ocean-atmosphere interface. Equation (2.290a) is particularly useful in the ocean, where the stress at the surface is primarily due to the wind, and is largely independent of the interior oceanic flow. In the atmosphere, the surface stress mainly arises as a result of the interior atmospheric flow, and to calculate it we need to parameterize the stress in terms of the flow.

Finally, we obtain an expression for the vertical velocity induced by an Ekman layer. The mass conservation equation is

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0. \quad (2.291)$$

Integrating this over an Ekman layer gives

$$\frac{1}{\rho_0} \nabla \cdot \mathbf{M}_{To} = -(w_T - w_B), \quad (2.292)$$

where \mathbf{M}_{To} is the total (Ekman plus geostrophic) mass transport in the Ekman layer,

$$\mathbf{M}_{To} = \int_{Ek} \rho_0 \mathbf{u} \, dz = \int_{Ek} \rho_0 (\mathbf{u}_g + \mathbf{u}_E) \, dz \equiv \mathbf{M}_g + \mathbf{M}_E, \quad (2.293)$$

and w_T and w_B are the vertical velocities at the top and bottom of the Ekman layer; the former (latter) is zero in a top (bottom) Ekman layer. Equations (2.293) and (2.288) give

$$\mathbf{k} \times (\mathbf{M}_{To} - \mathbf{M}_g) = \frac{1}{f} (\boldsymbol{\tau}_T - \boldsymbol{\tau}_B). \quad (2.294)$$

Taking the curl of this (i.e., cross-differentiating) gives

$$\nabla \cdot (\mathbf{M}_{To} - \mathbf{M}_g) = \text{curl}_z [(\boldsymbol{\tau}_T - \boldsymbol{\tau}_B)/f], \quad (2.295)$$

where the curl_z operator on a vector \mathbf{A} is defined by $\text{curl}_z \mathbf{A} \equiv \partial_x A_y - \partial_y A_x$. Using (2.292) we obtain, for top and bottom Ekman layers respectively,

$$w_B = \frac{1}{\rho_0} \left(\text{curl}_z \frac{\boldsymbol{\tau}_T}{f} + \nabla \cdot \mathbf{M}_g \right), \quad w_T = \frac{1}{\rho_0} \left(\text{curl}_z \frac{\boldsymbol{\tau}_B}{f} - \nabla \cdot \mathbf{M}_g \right), \quad (2.296a,b)$$

where $\nabla \cdot \mathbf{M}_g = -(\beta/f) \mathbf{M}_g \cdot \mathbf{j}$ is the divergence of the geostrophic transport in the Ekman layer, and this is often small compared to the other terms in these equations. Thus, friction induces a vertical velocity at the edge of the Ekman layer, proportional to the curl of the stress at the surface, and this is perhaps the most used result in Ekman layer theory. Numerical models sometimes do not have the vertical resolution to explicitly resolve an Ekman layer, and (2.296) provides a means of *parameterizing* the Ekman layer in terms of resolved or known fields. It is particularly useful for the top Ekman layer in the ocean, where the stress can be regarded as a given function of the overlying wind.

2.12.3 Explicit solutions. I: a bottom boundary layer

We now assume that the frictional terms can be parameterized as an eddy viscosity and calculate the explicit form of the solution in the boundary layer. The frictional-geostrophic balance may be written as

$$\mathbf{f} \times (\mathbf{u} - \mathbf{u}_g) = A \frac{\partial^2 \mathbf{u}}{\partial z^2}, \quad (2.297a)$$

where

$$f(u_g, v_g) = \left(-\frac{\partial \phi}{\partial y}, \frac{\partial \phi}{\partial x} \right). \quad (2.297b)$$

We continue to assume there are no horizontal gradients of temperature, so that, via thermal wind, $\partial u_g / \partial z = \partial v_g / \partial z = 0$.

Boundary conditions and solution

Appropriate boundary conditions for a bottom Ekman layer are

$$\text{at } z = 0 : \quad u = 0, \quad v = 0 \quad \text{(the no slip condition)} \quad (2.298a)$$

$$\text{as } z \rightarrow \infty : \quad u = u_g, \quad v = v_g \quad \text{(a geostrophic interior).} \quad (2.298b)$$

Let us seek solutions to (2.297a) of the form

$$u = u_g + A_0 e^{\alpha z}, \quad v = v_g + B_0 e^{\alpha z}, \quad (2.299)$$

where A_0 and B_0 are constants. Substituting into (2.297a) gives two homogeneous algebraic equations

$$A_0 f - B_0 A \alpha^2 = 0, \quad -A_0 A \alpha^2 - B_0 f = 0. \quad (2.300a,b)$$

For non-trivial solutions the solvability condition $\alpha^4 = -f^2/A^2$ must hold, from which we find $\alpha = \pm(1 \pm i)\sqrt{f/2A}$. Using the boundary conditions we then obtain the solution

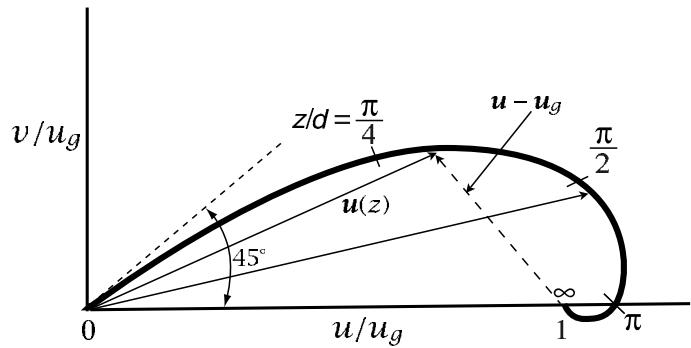
$$u = u_g - e^{-z/d} [u_g \cos(z/d) + v_g \sin(z/d)] \quad (2.301a)$$

$$v = v_g + e^{-z/d} [u_g \sin(z/d) - v_g \cos(z/d)], \quad (2.301b)$$

where $d = \sqrt{2A/f}$ is, within a constant factor, the depth of the Ekman layer obtained from scaling considerations. The solution decays exponentially from the surface with this e-folding scale, so that d is a good measure of the Ekman layer thickness. Note that the boundary layer correction depends on the interior flow, since the boundary layer serves to bring the flow to zero at the surface.

To illustrate the solution, suppose that the pressure force is directed in the y -direction (northwards), so that the geostrophic current is eastwards. Then the solution, the now-famous *Ekman spiral*, is plotted in Figs. 2.11 and 2.12. The wind falls to zero at the surface, and its direction just above the surface is northeastwards; that is, it is rotated by 45° to the left of its direction in the free atmosphere. Although this result is independent of the value of the frictional coefficients, it is dependent on the form of the friction chosen. The force balance in the Ekman layer is between the Coriolis force, the stress, and the pressure force. At the surface the Coriolis force is zero, and the balance is entirely between the northward pressure force and the southward stress force.

Fig. 2.11 The idealized Ekman layer solution in the lower atmosphere, plotted as a hodograph of the wind components: the arrows show the velocity vectors at a particular heights, and the curve traces out the continuous variation of the velocity. The values on the curve are of the non-dimensional variable z/d , where $d = (2A/f)^{1/2}$, and v_g is chosen to be zero.



Transport, force balance and vertical velocity

The cross-isobaric flow is given by (for $v_g = 0$)

$$V = \int_0^\infty v \, dz = \int_0^\infty u_g e^{-z/d} \sin(z/d) \, dz = \frac{u_g d}{2}. \quad (2.302)$$

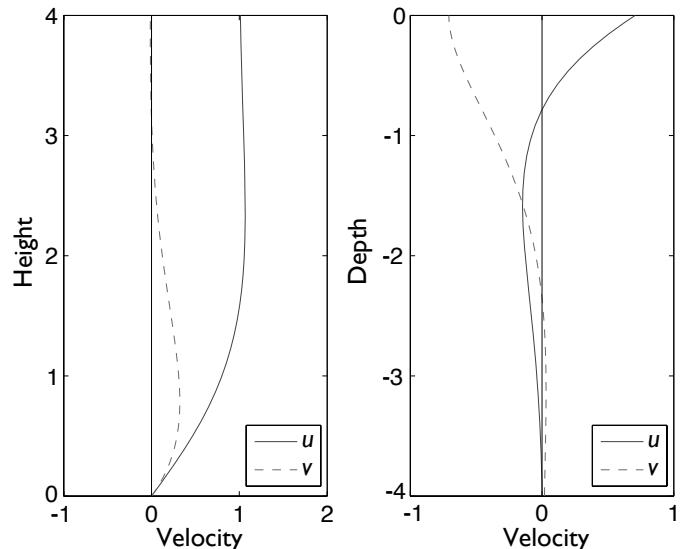
For positive f , this is to the left of the geostrophic flow — that is, down the pressure gradient. In the general case ($v_g \neq 0$) we obtain

$$V = \int_0^\infty (v - v_g) \, dz = \frac{d}{2}(u_g - v_g). \quad (2.303)$$

Similarly, the additional zonal transport produced by frictional effects are, for $v_g = 0$,

$$U = \int_0^\infty (u - u_g) \, dz = - \int_0^\infty e^{-z/d} \sin(z/d) \, dz = - \frac{u_g d}{2}, \quad (2.304)$$

Fig. 2.12 Solutions for a bottom Ekman layer with a given flow in the fluid interior (left), and for a top Ekman layer with a given surface stress (right), both with $d = 1$. On the left we have $u_g = 1$, $v_g = 0$. On the right we have $u_g = v_g = 0$, $\tilde{\tau}_y = 0$ and $\sqrt{2}\tilde{\tau}_x/(fd) = 1$.



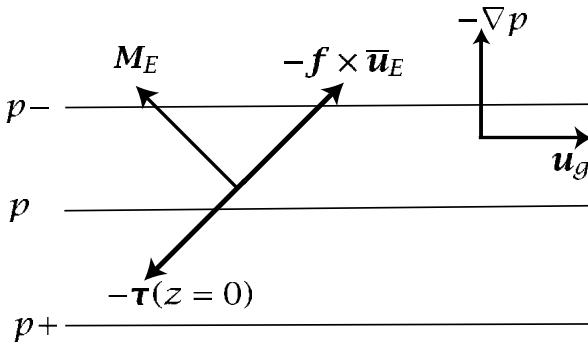


Fig. 2.13 A bottom Ekman layer, generated from an eastward geostrophic flow above it. An overbar denotes a vertical integral over the Ekman layer, so that $-\mathbf{f} \times \bar{\mathbf{u}}_E$ is the Coriolis force on the vertically integrated Ekman velocity. M_E is the frictionally induced boundary layer transport, and τ is the stress.

and in the general case

$$U = \int_0^\infty (u - u_g) dz = -\frac{d}{2} (u_g + v_g). \quad (2.305)$$

Thus, the total transport caused by frictional forces is

$$M_E = \frac{\rho_0 d}{2} \left[-\mathbf{i}(u_g + v_g) + \mathbf{j}(u_g - v_g) \right]. \quad (2.306)$$

The total stress at the bottom surface $z = 0$ induced by frictional forces is

$$\tilde{\tau}_B = A \frac{\partial \mathbf{u}}{\partial z} \Big|_{z=0} = \frac{A}{d} \left[\mathbf{i}(u_g - v_g) + \mathbf{j}(u_g + v_g) \right], \quad (2.307)$$

using the solution (2.301). Thus, using (2.306), (2.307) and $d^2 = 2A/f$, we see that the total frictionally induced transport in the Ekman layer is related to the stress at the surface by $M_E = (\mathbf{k} \times \tilde{\tau}_B)/f$, reprising the result of the previous more general analysis, (2.296). From (2.307), the stress is at an angle of 45° to the left of the velocity at the surface. (However, this result is not generally true for all forms of stress.) These properties are illustrated in Fig. 2.13.

The vertical velocity at the top of the Ekman layer, w_E , is obtained using (2.306) or (2.307). If f if constant we obtain

$$w_E = -\frac{1}{\rho_0} \nabla \cdot M_E = \frac{1}{f_0} \operatorname{curl}_z \tilde{\tau}_B = \frac{d}{2} \zeta_g, \quad (2.308)$$

where ζ_g is the vorticity of the geostrophic flow. Thus, the vertical velocity at the top of the Ekman layer, which arises because of the frictionally-induced divergence of the cross-isobaric flow in the Ekman layer, is proportional to the geostrophic vorticity in the free fluid and is proportional to the Ekman layer height $\sqrt{2A/f_0}$.

Another bottom boundary condition

In the analysis above we assumed a *no slip* condition at the surface, namely that the velocity tangential to the surface vanishes. This is certainly appropriate if A is a molecular velocity, but in a turbulent flow, where A is interpreted as an eddy viscosity, the flow very close to the surface may be far from zero. Then, unless we wish to explicitly calculate the flow in an additional very thin viscous boundary layer the no-slip condition may be inappropriate. An

alternative, slightly more general boundary condition is to suppose that the stress at the surface is given by

$$\boldsymbol{\tau} = \rho_0 C \mathbf{u}, \quad (2.309)$$

where C is a constant. The surface boundary condition is then

$$A \frac{\partial \mathbf{u}}{\partial z} = C \mathbf{u}. \quad (2.310)$$

If C is infinite we recover the no-slip condition. If $C = 0$, we have a condition of no stress at the surface, also known as a *free slip* condition. For intermediate values of C the boundary condition is known as a ‘mixed condition’. Evaluating the solution in these cases is left as an exercise for the reader (problem 2.28).

2.12.4 Explicit solutions. II: the upper ocean

Boundary conditions and solution

The wind provides a stress on the upper ocean, and the Ekman layer serves to communicate this to the oceanic interior. Appropriate boundary conditions are thus:

$$\text{at } z = 0 : \quad A \frac{\partial u}{\partial z} = \tilde{\tau}^x, \quad A \frac{\partial v}{\partial z} = \tilde{\tau}^y \quad (\text{a given surface stress}) \quad (2.311a)$$

$$\text{as } z \rightarrow -\infty : \quad u = u_g, \quad v = v_g \quad (\text{a geostrophic interior}) \quad (2.311b)$$

where $\tilde{\tau}$ is the given (kinematic) wind stress at the surface. Solutions to (2.297a) with (2.311) are found by the same methods as before, and are

$$u = u_g + \frac{\sqrt{2}}{fd} e^{z/d} [\tilde{\tau}^x \cos(z/d - \pi/4) - \tilde{\tau}^y \sin(z/d - \pi/4)], \quad (2.312)$$

and

$$v = v_g + \frac{\sqrt{2}}{fd} e^{z/d} [\tilde{\tau}^x \sin(z/d - \pi/4) + \tilde{\tau}^y \cos(z/d - \pi/4)]. \quad (2.313)$$

Note that the boundary layer correction depends only on the imposed surface stress, and not the interior flow itself. This is a consequence of the type of boundary conditions chosen, for in the absence of an imposed stress the boundary layer correction is zero — the interior flow already satisfies the gradient boundary condition at the top surface. Similar to the bottom boundary layer, the velocity vectors of the solution trace a diminishing spiral as they descend into the interior (Fig. 2.14, which is drawn for the Southern Hemisphere).

Transport, surface flow and vertical velocity

The transport induced by the surface stress is obtained by integrating (2.312) and (2.313) from the surface to $-\infty$. We explicitly find

$$U = \int_{-\infty}^0 (u - u_g) dz = \frac{\tilde{\tau}^y}{f}, \quad V = \int_{-\infty}^0 (v - v_g) dz = -\frac{\tilde{\tau}^x}{f}, \quad (2.314)$$

which indicates that the ageostrophic transport is perpendicular to the wind stress, as noted previously from more general considerations. Suppose that the surface wind is eastward; in this case $\tilde{\tau}^y = 0$ and the solutions immediately give

$$u(0) - u_g = \tilde{\tau}^x / fd, \quad v(0) - v_g = -\tilde{\tau}^x / fd. \quad (2.315)$$

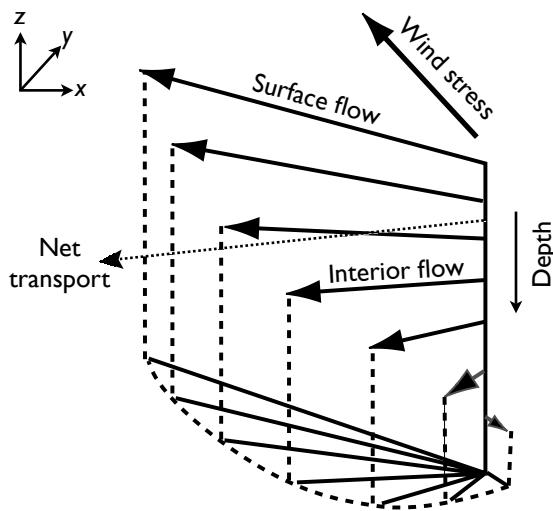


Fig. 2.14 An idealized Ekman spiral in a southern hemisphere ocean, driven by an imposed wind stress. A northern hemisphere spiral would be the reflection of this about the vertical axis. Such a clean spiral is rarely observed in the real ocean. The net transport is at right angles to the wind, independent of the detailed form of the friction. The angle of the surface flow is 45° to the wind only for a Newtonian viscosity.

Therefore the magnitudes of the frictional flow in the x and y directions are equal to each other, and the ageostrophic flow is 45° to the right (for $f > 0$) of the wind. This result depends on the form of the frictional parameterization, but not on the size of the viscosity.

At the edge of the Ekman layer the vertical velocity is given by (2.296), and so is proportional to the curl of the wind stress. (The second term on the right-hand side of (2.296) is the vertical velocity due to the divergence of the geostrophic flow, and is usually much smaller than the first term.) The production of a vertical velocity at the edge of the Ekman layer is one of most important effects of the layer, especially with regard to the large-scale circulation, for it provides an efficient means whereby surface fluxes are communicated to the interior flow (see Fig. 2.15).

2.12.5 Observations of the Ekman layer

Ekman layers — and in particular the Ekman spiral — are generally quite hard to observe, in either the ocean or atmosphere, both because of the presence of phenomena that are not included in the theory, and because of the technical difficulties of actually measuring the vector velocity profile, especially in the ocean. Ekman-layer theory does not take into account the effects of stratification or of inertial and gravity waves (section 2.10), nor does it account for the effects of convection or buoyancy-driven turbulence. If gravity waves are present, the instantaneous flow will be non-geostrophic and so time-averaging will be required to extract the geostrophic flow. If strong convection is present, the simple eddy-viscosity parameterizations used to derive the Ekman spiral will be rendered invalid, and the spiral Ekman profile cannot be expected to be observed in either atmosphere or ocean.

In the atmosphere, in convectively neutral cases, the Ekman profile can sometimes be qualitatively discerned. In convectively unstable situations the Ekman profile is generally not observed, but the flow is nevertheless cross-isobaric, from high pressure to low, consistent with the theory. (For most purposes, it is in any case the integral properties of the Ekman layer that is most important.) In the ocean, from about 1980 onwards improved instruments have made it possible to observe the vector current with depth, and to average that current

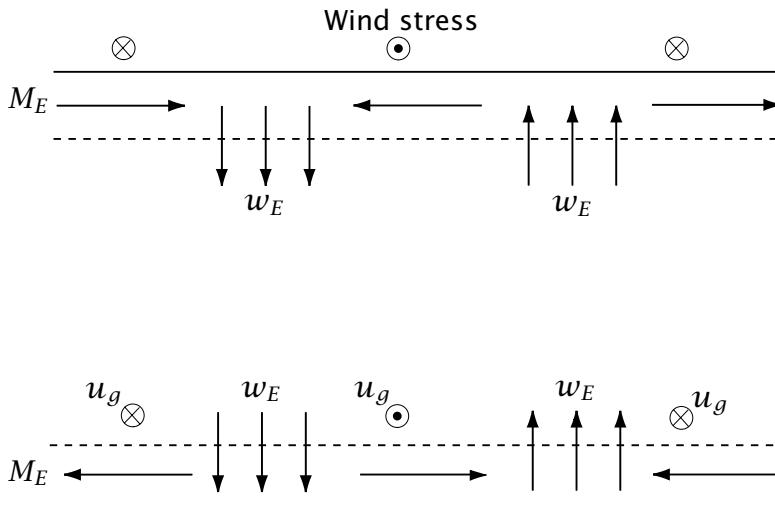


Fig. 2.15 Upper and lower Ekman layers. The upper Ekman layer in the ocean is primarily driven by an imposed wind stress, whereas the lower Ekman layer in the atmosphere or ocean largely results from the interaction of interior geostrophic velocity and a rigid lower surface. The upper part of figure shows the vertical Ekman ‘pumping’ velocities that result from the given wind stress, and the lower part of the figure shows the Ekman pumping velocities given the interior geostrophic flow.

and correlate it with the overlying wind, and a number of observations generally consistent with Ekman dynamics have emerged.¹³ There are some differences between observations and theory, and these can be ascribed to the effects of stratification (which causes a shallowing and flattening of the spiral), and to the interaction of the Ekman spiral with turbulence (and the inadequacy of the eddy-diffusivity parameterization). In spite of these differences, Ekman layer theory remains a remarkable and enduring foundation of geophysical fluid dynamics.

2.12.6 * Frictional parameterization of the Ekman layer

[Some readers will be reading these sections on Ekman layers after having been introduced to quasi-geostrophic theory; this section is for them. Other readers may return to this section after reading chapter 5, or take (2.316) on faith.]

Suppose that the free atmosphere is described by the quasi-geostrophic vorticity equation,

$$\frac{D\zeta_g}{Dt} = f_0 \frac{\partial w}{\partial z}, \quad (2.316)$$

where ζ_g is the geostrophic relative vorticity. Let us further model the atmosphere as a single homogeneous layer of thickness H lying above an Ekman layer of thickness $d \ll H$. If the vertical velocity is negligible at the top of the layer (at $z = H + d$) the equation of motion becomes

$$\frac{D\zeta_g}{Dt} = \frac{f_0 [w(H + d) - w(d)]}{H} = -\frac{f_0 d}{2H} \zeta_g \quad (2.317)$$

using (2.308). This equation shows that the Ekman layer acts as a *linear drag* on the interior flow, with a drag coefficient r equal to $f_0 d / 2H$ and with associated time scale T_{Ek} given by

$$T_{Ek} = \frac{2H}{f_0 d} = \frac{2H}{\sqrt{2f_0 A}}. \quad (2.318)$$

In the oceanic case the corresponding vorticity equation for the interior flow is

$$\frac{D\zeta_g}{Dt} = \frac{1}{H} \operatorname{curl}_z \tau_s, \quad (2.319)$$

where τ_s is the surface stress. The surface stress thus acts as if it were a body force on the interior flow, and neither the Coriolis parameter nor the depth of the Ekman layer explicitly appear in this formula.

The Ekman layer is a very efficient way of communicating surface stresses to the interior. To see this, suppose that eddy mixing were the sole mechanism of transferring stress from the surface to the fluid interior, and there were no Ekman layer. Then the time scale of spindown of the fluid would be given by using

$$\frac{d\zeta}{dt} = A \frac{\partial^2 \zeta}{\partial z^2}, \quad (2.320)$$

implying a turbulent spin-down time, T_{turb} , of

$$T_{turb} \sim \frac{H^2}{A}, \quad (2.321)$$

where H is the depth over which we require a spin-down. This is much longer than the spin-down of a fluid that has an Ekman layer, for we have

$$\frac{T_{turb}}{T_{Ek}} = \frac{(H^2/A)}{(2H/f_0 d)} = \frac{H}{d} \gg 1, \quad (2.322)$$

using $d = \sqrt{2A/f_0}$. The effects of friction are enhanced because of the presence of a secondary circulation confined to the Ekman layers (as in Fig. 2.15) in which the vertical scales are much smaller than those in the fluid interior and so where viscous effects become significant; these frictional stresses are then communicated to the fluid interior via the induced vertical velocities at the edge of the Ekman layers.

Notes

- 1 The distinction between Coriolis force and acceleration is not always made in the literature. For a fluid in geostrophic balance, one might either say that there is a balance between the pressure force and the Coriolis force, with no net acceleration, or that the pressure force produces a Coriolis acceleration. The descriptions are equivalent, because of Newton's second law, but should not be conflated.

The Coriolis forces is named after Gaspard Gustave de Coriolis (1792–1843), who introduced the force in the context of rotating mechanical systems (Coriolis 1832, 1835). See Persson (1998) for a historical account and interpretation.

- 2 Phillips (1973). There is also a related discussion in Stommel & Moore (1989).
- 3 Phillips (1966). See White (2003) for a review. In the early days of numerical modelling these equations were the most primitive — i.e., the least filtered — equations that could practically be integrated numerically. Associated with increasing computer power there is a tendency for comprehensive numerical models to use non-hydrostatic equations of motion that do not make the shallow-fluid or traditional approximations, and it is conceivable that the meaning of the word ‘primitive’ may evolve to accommodate them.
- 4 The Boussinesq approximation is named for Boussinesq (1903), although similar approximations were used earlier by Oberbeck (1879, 1888). Spiegel & Veronis (1960) give a physically based derivation for an ideal gas, and Mihaljan (1962) provides a somewhat more general asymptotic derivation. Mahrt (1986) discusses applicability to the atmosphere.
- 5 As pointed out to me by W.R. Young.
- 6 Various versions of anelastic equations exist — see Batchelor (1953a), Ogura & Phillips (1962), Gough (1969), Gilman & Glatzmaier (1981), Lipps & Hemler (1982), and Durran (1989), although not all have potential vorticity and energy conservation laws (Bannon 1995, 1996; Scinocca & Shepherd 1992). The system we derive is most similar to that of Ogura & Phillips (1962) and unpublished notes by J.S.A. Green. The connection between the Boussinesq and anelastic equations is discussed by, among others, Lilly (1996) and Ingersoll (2005).
- 7 A numerical model that includes sound waves must take very small timesteps in order to maintain numerical stability, in particular to satisfy the Courant–Friedrichs–Lewy (CFL) criterion. An alternative is to use an implicit timestepping scheme that effectively lets the numerics do the filtering of the sound waves, and this approach is favoured by many numerical modellers. If we make the hydrostatic approximation then all sound waves except those that propagate horizontally are eliminated, and there is little need, vis-à-vis the numerics, to also make the anelastic approximation.
- 8 The Rossby number is named for C.-G. Rossby (see endnote on page 241), but it was also used by Kibel (1940) and is sometimes called the Kibel or Rossby–Kibel number. The notion of geostrophic balance and so, implicitly, that of a small Rossby number, predates either Rossby or Kibel.
- 9 After Taylor (1921b) and Proudman (1916). The Taylor–Proudman effect is sometimes called the Taylor–Proudman ‘theorem’, but it is more usefully thought of as a physical effect, with manifestations even when the conditions for its satisfaction are not precisely met.
- 10 Foster (1972).
- 11 Many numerical models of the large-scale circulation in the atmosphere and ocean do make the hydrostatic approximation. In these models convection must be *parameterized*; otherwise, it would simply occur at the smallest scale available, namely the size of the numerical grid, and this type of unphysical behaviour should be avoided. Of course in non-hydrostatic models convection must also be parameterized if the horizontal resolution of the model is too coarse to properly resolve the convective scales. See also problem 2.11.
- 12 After Ekman (1905). The problem was posed to V. W. Ekman (1874–1954), a student of Vilhelm Bjerknes, by Fridtjof Nansen, the polar explorer and statesman, who wanted to understand the motion of pack ice and of his ship, the *Fram*, embedded in the ice.
- 13 For oceanic observations see Davis *et al.* (1981), Price *et al.* (1987), Rudnick & Weller (1993). For the atmosphere see, e.g., Nicholls (1985).

Further reading

Cushman-Roisin, B., 1994. *An Introduction to Geophysical Fluid Dynamics*.

Provides a compact introduction to a variety of topics in GFD.

Gill, A. E., 1982. *Atmosphere–Ocean Dynamics*.

A rich book, especially strong on equatorial dynamics and gravity wave motion.

Holton, J. R., 1992. *An Introduction to Dynamical Meteorology*.

A deservedly well-known textbook at the upper-division undergraduate/beginning graduate level.

Pedlosky, J., 1987. *Geophysical Fluid Dynamics*.

A primary reference for flow at low Rossby number. Although the book requires some effort, there is a handsome pay-off for those who study it closely.

White (2002) provides a clear and thorough summary of the equations of motion for meteorology, including the non-hydrostatic and primitive equations.

Zdunkowski, W. & Bott, A., 2003. *Dynamics of the Atmosphere: A Course in Theoretical Meteorology*.

Concentrates on the mathematical tools and equations of motion of dynamical meteorology.

Problems

2.1 Show that for an ideal gas in hydrostatic balance, changes in dry static energy ($M = c_p T + gz$) and potential temperature (θ) are related by $\delta M = c_p(T/\theta)\delta\theta$. (The quantity $c_p T/\theta$ is known as the ‘Exner function’, and is denoted Π .)

2.2 For an ideal gas in hydrostatic balance, show that:

- (a) The integral of the potential plus internal energy from the surface to the top of the atmosphere [$\int (P + I) dp$] is equal to its enthalpy;
- (b) $d\sigma/dz = c_p(T/\theta) d\theta/dz$, where $\sigma = I + p\alpha + \Phi$ is the dry static energy;
- (c) The following expressions for the pressure gradient force are all equal (even without hydrostatic balance):

$$-\frac{1}{\rho} \nabla p = -\theta \nabla \Pi = -\frac{c_s^2}{\rho \theta} \nabla(\rho \theta), \quad (\text{P2.1})$$

where $\Pi = c_p T/\theta$ is the Exner function.

(d) Show that item (a) also holds for a gas with an arbitrary equation of state.

2.3 Show that, without approximation, the unforced, inviscid momentum equation may be written in the forms

$$\frac{D\mathbf{v}}{Dt} = T \nabla \eta - \nabla(p\alpha + I) \quad (\text{P2.2})$$

and

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{\omega} \times \mathbf{v} = T \nabla \eta - \nabla B \quad (\text{P2.3})$$

where $\mathbf{\omega} = \nabla \times \mathbf{v}$, η is the specific entropy ($d\eta = c_p d \ln \theta$) and $B = I + \mathbf{v}^2/2 + p\alpha$ where I is the internal energy per unit mass.

Hint: First show that $T \nabla \eta = \nabla I + p \nabla \alpha$, and note also the vector identity $\mathbf{v} \times (\nabla \times \mathbf{v}) = \frac{1}{2} \nabla(\mathbf{v} \cdot \mathbf{v}) - (\mathbf{v} \cdot \nabla) \mathbf{v}$.

2.4 Consider two-dimensional fluid flow in a rotating frame of reference on the f -plane. Linearize the equations about a state of rest.

(a) Ignore the pressure term and determine the general solution to the resulting equations.

Show that the speed of fluid parcels is constant. Show that the trajectory of the fluid parcels is a circle with radius $|U|/f$, where $|U|$ is the fluid speed.

(b) What is the period of oscillation of a fluid parcel?

- (c) ♦ If parcels travel in straight lines in inertial frames, why is the answer to (b) not the same as the period of rotation of the frame of reference? [To answer this fully you need to understand the dynamics underlying inertial oscillations and inertia circles. See Durran (1993), Egger (1999) and Phillips (2000).]
- 2.5 A fluid at rest evidently satisfies the hydrostatic relation, which says that the pressure at the surface is given by the weight of the fluid above it. Now consider a *deep* atmosphere on a spherical planet. A unit cross-sectional area at the planet's surface lies beneath a column of fluid whose cross-section increases with height, because the total area of the atmosphere increases with distance away from the centre of the planet. Is the pressure at the surface still given by the hydrostatic relation, or is it greater than this because of the increased mass of fluid in the column? If it is still given by the hydrostatic relation, then the pressure at the surface, integrated over the entire area of the planet, is less than the total weight of the fluid; resolve this paradox. But if the pressure at the surface is greater than that implied by hydrostatic balance, explain how the hydrostatic relation fails.
- 2.6 In a self-gravitating spherical fluid, like a star, hydrostatic balance may be written
- $$\frac{\partial p}{\partial r} = -\frac{GM(r)}{r^2} \rho, \quad (\text{P2.4})$$
- where $M(r)$ is the mass interior to a sphere of radius r , and G is a constant. Obtain an expression for the pressure as a function of radius when the fluid (a) has constant density, and (b) is an isothermal ideal gas (if possible). The star is of radius a .
- 2.7 At what latitude is the angle between the direction of Newtonian gravity (due solely to the mass of the Earth) and that of effective gravity (Newtonian gravity plus centrifugal terms) the largest? At what latitudes, if any, is this angle zero?
- 2.8 ♦ Write the momentum equations in true spherical coordinates, including the centrifugal and gravitational terms. Show that for reasonable values of the wind, the dominant balance in the meridional component of this equation involves a balance between centrifugal and pressure gradient terms. Can this balance be subtracted out of the equations in a sensible way, so leaving a useful horizontal momentum equation that involves the Coriolis and acceleration terms? If so, obtain a closed set of equations for the flow this way. Discuss the pros and cons of this approach versus the geometric approximation discussed in section 2.2.1.
- 2.9 For an ideal gas show that the expressions (2.225) and (2.229) are equivalent.
- 2.10 Consider an ocean at rest with known vertical profiles of potential temperature and salinity, $\theta(z)$ and $S(z)$. Suppose we also know the equation of state in the form $\rho = \rho(\theta, S, p)$. Obtain an expression for the buoyancy frequency. Check your expression by substituting the equation of state for an ideal gas and recovering a known expression for the buoyancy frequency.
- 2.11 *Convection and its parameterization*
- (a) Consider a Boussinesq system in which the vertical momentum equation is modified by the parameter α to read
- $$\alpha^2 \frac{Dw}{Dt} = -\frac{\partial \phi}{\partial z} + b, \quad (\text{P2.5})$$
- and the other equations are unchanged. (If $\alpha = 0$ the system is hydrostatic, and if $\alpha = 1$ the system is the original one.) Linearize these equations about a state of rest and of constant stratification (as in section 2.10.1) and obtain the dispersion relation for the system, and plot it for various values of α , including 0 and 1. Show that for $\alpha > 1$ the system approaches its limiting frequency more rapidly than with $\alpha = 1$.
- (b) ♦ Argue that if $N^2 < 0$, convection in a system with $\alpha > 1$ generally occurs at a larger scale than with $\alpha = 1$. Show this explicitly by adding some diffusion or friction to the

right-hand sides of the equations of motion and obtaining the dispersion relation. You may do this approximately.

- 2.12 (a) The *geopotential height* is the height of a given pressure level. Show that in an atmosphere with a uniform lapse rate (i.e., $dT/dz = \Gamma = \text{constant}$) the geopotential height at a pressure p is given by

$$z = \frac{T_0}{\Gamma} \left[1 - \left(\frac{p_0}{p} \right)^{-R\Gamma/g} \right] \quad (\text{P2.6})$$

where T_0 is the temperature at $z = 0$.

- (b) In an isothermal atmosphere, obtain an expression for the geopotential height as function of pressure, and show that this is consistent with the expression (P2.6) in the appropriate limit.
- 2.13 Show that the inviscid, adiabatic, hydrostatic primitive equations for a compressible fluid conserve a form of energy (kinetic plus potential plus internal), and that the kinetic energy has no contribution from the vertical velocity. You may assume Cartesian geometry and a uniform gravitational field in the vertical direction.
- 2.14 Consider the simple Boussinesq equations, $D\mathbf{v}/Dt = -\nabla p + \mathbf{k}b + \nu \nabla^2 \mathbf{v}$, $\nabla \cdot \mathbf{v} = 0$, $Db/Dt = Q + \kappa \nabla^2 b$. Obtain an energy equation similar to (2.112), but now with the terms on the right-hand side that represent viscous and diabatic effects. Over a closed volume, show that the dissipation of kinetic energy is balanced by a buoyancy source. Also show that, in a statistically steady state, the heating must occur at a lower level than the cooling if a kinetic-energy dissipating circulation is to be maintained.
- 2.15 ♦ Suppose a fluid is contained in a closed container, with insulating sidewalls, and heated from below and cooled from above. The heating and cooling are adjusted so that there is no net energy flux into the fluid. Let us also suppose that any viscous dissipation of kinetic energy is returned as heating, so the total energy of the fluid is exactly constant. Suppose the fluid starts out at rest and at a uniform temperature, and the heating and cooling are then turned on. A very short time afterwards, the fluid is lighter at the bottom and heavier at the top; that is, its potential energy has increased. Where has this energy come from? Discuss this paradox for both a compressible fluid (e.g., an ideal gas) and for a simple Boussinesq fluid.
- 2.16 Consider a rapidly rotating (i.e., in near geostrophic balance) Boussinesq fluid on the f -plane.
- Show that the pressure divided by the density scales as $\phi \sim fUL$.
 - Show that the horizontal divergence of the geostrophic wind vanishes. Thus, argue that the scaling $W \sim UH/L$ is an *overestimate* for the magnitude of the vertical velocity. (Optional extra: obtain a scaling estimate for the magnitude of vertical velocity in rapidly rotating flow.)
 - Using these results, or otherwise, discuss whether hydrostatic balance is more or less likely to hold in a rotating flow than in non-rotating flow.
- 2.17 Estimate the size of the zonal wind 5 km above the surface in the mid-latitude atmosphere in summer and winter using (approximate) values for the meridional temperature gradient in the atmosphere. Also estimate the shear corresponding to the pole-equator temperature gradient in the ocean.
- 2.18 Using approximate but realistic values for the observed stratification, what is the buoyancy period for (a) the mid-latitude troposphere, (b) the stratosphere, (c) the oceanic thermocline, (d) the oceanic abyss?
- 2.19 Consider a dry, hydrostatic, ideal-gas atmosphere whose lapse rate is one of constant potential temperature. What is its vertical extent? That is, at what height does the density vanish? Is this a problem for the anelastic approximation discussed in the text?

- 2.20 Show that for an ideal gas, the expressions (2.229), (2.224), (2.225) are all equivalent, and express N^2 terms of the temperature lapse rate, $\partial T/\partial z$.
- 2.21 ♦ Calculate an approximate but reasonably accurate expression for the buoyancy equation for seawater. (From notes by R. de Szoeke)

Solution (i): the buoyancy frequency is given by

$$N^2 = -\frac{g}{\rho} \left(\frac{\partial \rho_{\text{pot}}}{\partial z} \right)_{\text{env}} = \frac{g}{\alpha} \left(\frac{\partial \alpha_{\text{pot}}}{\partial z} \right)_{\text{env}} = -\frac{g^2}{\alpha^2} \left(\frac{\partial \alpha_{\text{pot}}}{\partial p} \right)_{\text{env}} \quad (\text{P2.7})$$

where $\alpha_{\text{pot}} = \alpha(\theta, S, p_R)$ is the potential density, and p_R a reference pressure. From (1.155)

$$\alpha_{\text{pot}} = \alpha_0 \left[1 - \frac{\alpha_0}{c_0^2} p_R + \beta_T (1 + \gamma^* p_R) \theta' + \frac{1}{2} \beta_T^* \theta'^2 - \beta_S (S - S_0) \right]. \quad (\text{P2.8})$$

Using this and (P2.7) we obtain the buoyancy frequency,

$$N^2 = -\frac{g^2}{\alpha^2} \alpha_0 \left[\beta_T \left(1 + \gamma p_R + \frac{\beta_T^*}{\beta_T} \theta \right) \left(\frac{\partial \theta}{\partial p} \right)_{\text{env}} - \beta_S \left(\frac{\partial S}{\partial p} \right)_{\text{env}} \right], \quad (\text{P2.9})$$

although we must substitute local pressure for the reference pressure p_R . (Why?)

Solution (ii): the sound speed is given by

$$c_s^{-2} = -\frac{1}{\alpha^2} \left(\frac{\partial \alpha}{\partial p} \right)_{\theta, S} = \frac{1}{\alpha^2} \left(\frac{\alpha_0^2}{c_0^2} - \gamma \alpha_1 \theta \right) \quad (\text{P2.10})$$

and, using (P2.7) and (2.229) the square of the buoyancy frequency may be written

$$N^2 = \frac{g}{\alpha} \left(\frac{\partial \alpha}{\partial z} \right)_{\text{env}} - \frac{g^2}{c_s^2} = -\frac{g^2}{\alpha^2} \left[\left(\frac{\partial \alpha}{\partial p} \right)_{\text{env}} + \frac{\alpha^2}{c_s^2} \right]. \quad (\text{P2.11})$$

Using (1.155), (P2.10) and (P2.11) we recover (P2.9), although now with p explicitly in place of p_R .

- 2.22 (a) Use the chain rule to show that the horizontal gradients of a field in height coordinates and in ξ coordinates are related by

$$\nabla_z \Psi = \nabla_{\xi} \Psi - (\partial \Psi / \partial \xi) (\partial \xi / \partial z) \nabla_{\xi} z. \quad (\text{P2.12})$$

- (b) Show that w , the vertical velocity in height coordinates, may be expressed in ξ coordinates as

$$w = Dz/Dt = (\partial z / \partial t)_{\xi} + \mathbf{u} \cdot \nabla_{\xi} z + \dot{\xi} \partial z / \partial \xi. \quad (\text{P2.13})$$

- (c) Use the above expressions to verify (2.143), the expression for the material derivative in ξ coordinates.

- 2.23 Begin with the mass conservation in height coordinates, namely $D\rho/Dt + \rho \nabla \cdot \mathbf{v} = 0$. Transform this into pressure coordinates using the chain rule (or otherwise) and derive the mass conservation equation in the form $\nabla_p \cdot \mathbf{u} + \partial \omega / \partial p = 0$.

- 2.24 ♦ Starting with the primitive equations in pressure coordinates, derive the form of the primitive equations of motion in sigma-pressure coordinates. In particular, show that the prognostic equation for surface pressure is,

$$\frac{\partial p_s}{\partial t} + \nabla \cdot (p_s \mathbf{u}) + p_s \frac{\partial \sigma}{\partial \sigma} = 0 \quad (\text{P2.14})$$

and that hydrostatic balance may be written $\partial \Phi / \partial \sigma = -RT/\sigma$.

- 2.25 Starting with the primitive equations in pressure coordinates, derive the form of the primitive equations of motion in log-pressure coordinates in which $Z = -H \ln(p/p_r)$ is the vertical coordinate. Here, H is a reference height (e.g., a scale height RT_r/g where T_r is a typical or an average temperature) and p_r is a reference pressure (e.g., 1000 mb). In particular, show that if the ‘vertical velocity’ is $W = DZ/Dt$ then $W = -H\omega/p$ and that

$$\frac{\partial \omega}{\partial p} = -\frac{\partial}{\partial p} \left(\frac{pW}{H} \right) = \frac{\partial W}{\partial Z} - \frac{W}{H}. \quad (\text{P2.15})$$

and obtain the mass conservation equation (2.158b). Show that this can be written in the form

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{1}{\rho_s} \frac{\partial}{\partial Z} (\rho_s W) = 0, \quad (\text{P2.16})$$

where $\rho_s = \rho_r \exp(-Z/H)$.

- 2.26 (a) Prove that the argument of the square root in (2.260) is always positive.

Solution: The largest value or the argument occurs when $m = 0$ and $k^2 = 1/(4H^2)$. The argument is then $1 - 4H^2N^2/c_s^2$. But $c_s^2 = \gamma RT_0 = \gamma gH$ and $N^2 = g\kappa/H$ so that $4N^2H^2/c_s^2 = 4\kappa/\gamma \approx 0.8$.

- (b) ♦ This argument seems to depend on the parameters in the ideal gas equation of state. Is it more general than this? Is a natural system possible for which the argument is negative, and if so what physical interpretation could one ascribe to the situation?

- 2.27 Consider a wind stress imposed by a mesoscale cyclonic storm (in the atmosphere) given by

$$\boldsymbol{\tau} = -Ae^{-(r/\lambda)^2} (y\mathbf{i} - x\mathbf{j}) \quad (\text{P2.17})$$

where $r^2 = x^2 + y^2$, and A and λ are constants. Also assume constant Coriolis gradient $\beta = \partial f/\partial y$ and constant ocean depth H . In the ocean, find (a) the Ekman transport, (b) the vertical velocity $w_E(x, y, z)$ below the Ekman layer, (c) the northward velocity $v(x, y, z)$ below the Ekman layer and (d) indicate how you would find the westward velocity $u(x, y, z)$ below the Ekman layer.

- 2.28 ♦ In an atmospheric Ekman layer on the f -plane let us write the momentum equation as

$$\mathbf{f} \times \mathbf{u} = -\nabla \phi + \frac{1}{\rho_a} \frac{\partial \boldsymbol{\tau}}{\partial z}, \quad (\text{P2.18})$$

where $\boldsymbol{\tau} = A\rho_a \partial \mathbf{u} / \partial z$ and A is a constant eddy viscosity coefficient. An *independent* formula for the stress at the ground is $\boldsymbol{\tau} = C\rho_a \mathbf{u}$, where C is a constant. Let us take $\rho_a = 1$, and assume that in the free atmosphere the wind is geostrophic and zonal, with $\mathbf{u}_g = U\mathbf{i}$.

- (a) Find an expression for the wind vector at the ground. Discuss the limits $C = 0$ and $C = \infty$. Show that when $C = 0$ the frictionally-induced vertical velocity at the top of the Ekman layer is zero.
- (b) Find the vertically integrated horizontal mass flux caused by the boundary layer.
- (c) When the stress on the atmosphere is $\boldsymbol{\tau}$, the stress on the ocean beneath is also $\boldsymbol{\tau}$. Why? Show how this consistent with Newton’s third law.
- (d) Determine the direction and strength of the surface current, and the mass flux in the oceanic Ekman layer, in terms of the geostrophic wind in the atmosphere, the oceanic Ekman depth and the ratio ρ_a/ρ_o , where ρ_o is the density of the seawater. Include a figure showing the directions of the various winds and currents. How does the boundary-layer mass flux in the ocean compare to that in the atmosphere? (Assume, as needed, that the stress in the ocean may be parameterized with an eddy viscosity.)

Partial solution for (a): A useful trick in Ekman layer problems is to write the velocity as a complex number, $\hat{u} = u + iv$ and $\hat{u}_g = u_g + iv_g$. The Ekman layer equation, (2.297a), may then be written as

$$A \frac{\partial^2 \hat{U}}{\partial z^2} = if \hat{U}, \quad (\text{P2.19})$$

where $\hat{U} = \hat{u} - \hat{u}_g$. The solution to this is

$$\hat{u} - \hat{u}_g = [\hat{u}(0) - \hat{u}_g] \exp\left[-\frac{(1+i)z}{d}\right], \quad (\text{P2.20})$$

where $d = \sqrt{2A/f}$ and the boundary condition of finiteness at infinity eliminates the exponentially growing solution. The boundary condition at $z = 0$ is $\partial \hat{u} / \partial z = (C/A) \hat{u}$; applying this gives $[\hat{u}(0) - \hat{u}_g] \exp(i\pi/4) = -Cd\hat{u}(0)/(\sqrt{2}A)$, from which we obtain $\hat{u}(0)$, and the rest of the solution follows. We may also obtain a solution using the same method that was used to obtain (2.301).

2.29 The logarithmic boundary layer

Close to ground rotational effects are unimportant and small-scale turbulence generates a *mixed layer*. In this layer, assume that the stress is constant and that it can be parameterized by an eddy diffusivity the size of which is proportional to the distance from the surface. Show that the velocity then varies logarithmically with height.

Solution: Write the stress as $\tau = \rho_0 u^{*2}$ where the constant u^* is called the friction velocity. Using the eddy diffusivity hypothesis this stress is given by

$$\tau = \rho_0 u^{*2} = \rho_0 A \frac{\partial u}{\partial z} \quad \text{where} \quad A = u^* kz, \quad (\text{P2.21})$$

where k is von Karman's ('universal') constant (approximately equal to 0.4). From (P2.21) we have $\partial u / \partial z = u^* / (kz)$ which integrates to give $u = (u^* / k) \ln(z/z_0)$. The parameter z_0 is known as the roughness length.

Truth is much too complicated to allow anything but approximations.
John von Neumann (1903–1957).

CHAPTER THREE

Shallow Water Systems and Isentropic Coordinates

CONVENTIONALLY, ‘THE’ SHALLOW WATER EQUATIONS describe a thin layer of constant density fluid in hydrostatic balance, rotating or not, bounded from below by a rigid surface and from above by a free surface, above which we suppose is another fluid of negligible inertia. Such a configuration can be generalized to multiple layers of immiscible fluids lying one on top of another, forming a ‘stacked shallow water’ system, and this class of systems is the main subject of this chapter.

The single-layer model is one of the simplest useful models in geophysical fluid dynamics, because it allows for a consideration of the effects of rotation in a simple framework without the complicating effects of stratification. By adding layers we can subsequently study the effects of stratification, and the model with just two layers is not only a simple model of a stratified fluid, it is a surprisingly good model of many phenomena in the ocean and atmosphere. Indeed, the models are more than just pedagogical tools — we will find that there is a close physical and mathematical analogy between the shallow water equations and a description of the continuously stratified ocean or atmosphere written in isopycnal or isentropic coordinates, with a meaning beyond a coincidental similarity in the equations. We begin with the single-layer case.

3.1 DYNAMICS OF A SINGLE, SHALLOW LAYER

Shallow water dynamics apply, by definition, to a fluid layer of constant density in which the horizontal scale of the flow is much greater than the layer depth. The fluid motion is fully determined by the momentum and mass continuity equations, and because of the assumed

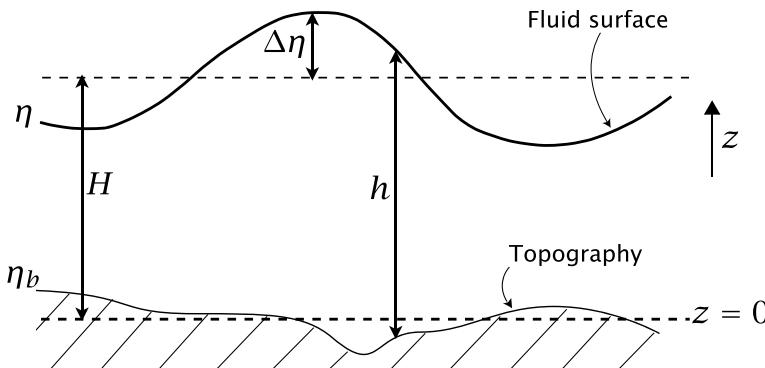


Fig. 3.1 A shallow water system. h is the thickness of a water column, H its mean thickness, η the height of the free surface and η_b is the height of the lower, rigid, surface, above some arbitrary origin, typically chosen such that the average of η_b is zero. $\Delta\eta$ is the deviation free surface height, so we have $\eta = \eta_b + h = H + \Delta\eta$.

small aspect ratio the hydrostatic approximation is well satisfied, and we invoke this from the outset. Consider, then, fluid in a container above which is another fluid of negligible density (and therefore negligible inertia) relative to the fluid of interest, as illustrated in Fig. 3.1. As usual, our notation is that $\mathbf{v} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$ is the three-dimensional velocity and $\mathbf{u} = u\mathbf{i} + v\mathbf{j}$ is the horizontal velocity. $h(x, y)$ is the thickness of the liquid column, H is its mean height, and η is the height of the free surface. In a flat-bottomed container $\eta = h$, whereas in general $h = \eta - \eta_b$, where η_b is the height of the floor of the container.

3.1.1 Momentum equations

The vertical momentum equation is just the hydrostatic equation,

$$\frac{\partial p}{\partial z} = -\rho g, \quad (3.1)$$

and, because density is assumed constant, we may integrate this to

$$p(x, y, z) = -\rho g z + p_o. \quad (3.2)$$

At the top of the fluid, $z = \eta$, the pressure is determined by the weight of the overlying fluid and this is assumed to be negligible. Thus, $p = 0$ at $z = \eta$, giving

$$p(x, y, z) = \rho g(\eta(x, y) - z). \quad (3.3)$$

The consequence of this is that the horizontal gradient of pressure is independent of height. That is

$$\nabla_z p = \rho g \nabla_z \eta, \quad (3.4)$$

where

$$\nabla_z = \mathbf{i} \frac{\partial}{\partial x} + \mathbf{j} \frac{\partial}{\partial y} \quad (3.5)$$

is the gradient operator at constant z . (In the rest of this chapter we will drop the subscript z unless that causes ambiguity. The three-dimensional gradient operator will be denoted by ∇_3 . We will also mostly use Cartesian coordinates, but the shallow water equations may certainly be applied over a spherical planet — indeed, ‘Laplace’s tidal equations’ are essentially the shallow water equations on a sphere.) The horizontal momentum equations therefore become

$$\frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho} \nabla p = -g \nabla \eta. \quad (3.6)$$

The right-hand side of this equation is independent of the vertical coordinate z . Thus, if the flow is initially independent of z , it must stay so. (This z independence is unrelated to that arising from the rapid rotation necessary for the Taylor–Proudman effect.) The velocities u and v are functions of x, y and t only, and the horizontal momentum equation is therefore

$$\frac{D\mathbf{u}}{Dt} = \frac{\partial \mathbf{u}}{\partial t} + u \frac{\partial \mathbf{u}}{\partial x} + v \frac{\partial \mathbf{u}}{\partial y} = -g \nabla \eta. \quad (3.7)$$

That the horizontal velocity is independent of z is a consequence of the hydrostatic equation, which ensures that the horizontal pressure gradient is independent of height. (Another starting point would be to take this independence of the horizontal motion with height as the *definition* of shallow water flow. In real physical situations such independence does not hold exactly — for example, friction at the bottom may induce a vertical dependence of the flow in a boundary layer.) In the presence of rotation, (3.7) easily generalizes to

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -g \nabla \eta,$$

(3.8)

where $\mathbf{f} = f\mathbf{k}$. Just as with the primitive equations, f may be constant or may vary with latitude, so that on a spherical planet $f = 2\Omega \sin \vartheta$ and on the β -plane $f = f_0 + \beta y$.

3.1.2 Mass continuity equation

From first principles

The mass contained in a fluid column of height h and cross-sectional area A is given by $\int_A \rho h \, dA$ (see Fig. 3.2). If there is a net flux of fluid across the column boundary (by advection) then this must be balanced by a net increase in the mass in A , and therefore a net increase in the height of the water column. The mass convergence into the column is given by

$$F_m = \text{mass flux in} = - \int_S \rho \mathbf{u} \cdot d\mathbf{S}, \quad (3.9)$$

where S is the area of the vertical boundary of the column. The surface area of the column is composed of elements of area $h\mathbf{n} \cdot d\mathbf{l}$, where $d\mathbf{l}$ is a line element circumscribing the column and \mathbf{n} is a unit vector perpendicular to the boundary, pointing outwards. Thus (3.9) becomes

$$F_m = - \oint \rho h \mathbf{u} \cdot \mathbf{n} \, dl. \quad (3.10)$$

Using the divergence theorem in two dimensions, (3.10) simplifies to

$$F_m = - \int_A \nabla \cdot (\rho \mathbf{u} h) \, dA, \quad (3.11)$$

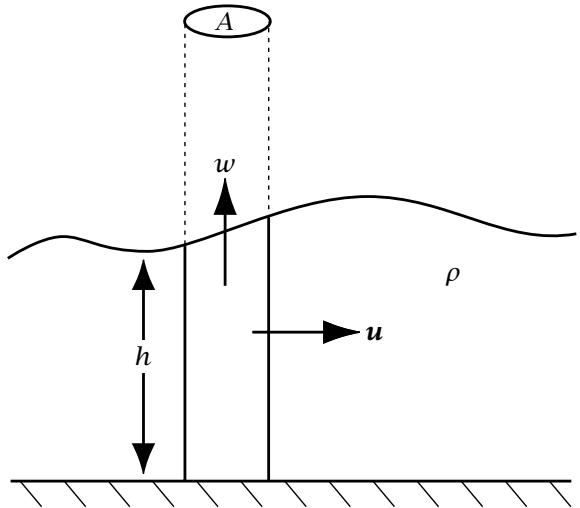


Fig. 3.2 The mass budget for a column of area A in a shallow water system. The fluid leaving the column is $\oint \rho \mathbf{u} \cdot \mathbf{n} dl$ where \mathbf{n} is the unit vector normal to the boundary of the fluid column. There is a non-zero vertical velocity at the top of the column if the mass convergence into the column is non-zero.

where the integral is over the cross-sectional area of the fluid column (looking down from above). This is balanced by the local increase in height of the water column, given by

$$F_m = \frac{d}{dt} \int \rho dV = \frac{d}{dt} \int_A \rho h dA = \int_A \rho \frac{\partial h}{\partial t} dA. \quad (3.12)$$

Because ρ is constant, the balance between (3.11) and (3.12) leads to

$$\int_A \left[\frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{u}h) \right] dA = 0, \quad (3.13)$$

and because the area is arbitrary the integrand itself must vanish, whence,

$$\boxed{\frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{u}h) = 0}, \quad (3.14)$$

or equivalently

$$\boxed{\frac{Dh}{Dt} + h \nabla \cdot \mathbf{u} = 0}. \quad (3.15)$$

This derivation holds whether or not the lower surface is flat. If it is, then $h = \eta$, and if not $h = \eta - \eta_b$. Equations (3.8) and (3.14) or (3.15) form a complete set, summarized in the shaded box on the next page.

From the 3D mass conservation equation

Since the fluid is incompressible, the three-dimensional mass continuity equation is just $\nabla \cdot \mathbf{v} = 0$. Writing this out in component form

$$\frac{\partial w}{\partial z} = - \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = - \nabla \cdot \mathbf{u}. \quad (3.16)$$

The Shallow Water Equations

For a single-layer fluid, and including the Coriolis term, the inviscid shallow water equations are

$$\text{momentum: } \frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -g \nabla \eta. \quad (\text{SW.1})$$

$$\text{mass continuity: } \frac{Dh}{Dt} + h \nabla \cdot \mathbf{u} = 0 \quad \text{or} \quad \frac{\partial h}{\partial t} + \nabla \cdot (h\mathbf{u}) = 0, \quad (\text{SW.2})$$

where \mathbf{u} is the horizontal velocity, h is the total fluid thickness, η is the height of the upper free surface and η_b is the height of the lower surface (the bottom topography). Thus, $h(x, y, t) = \eta(x, y, t) - \eta_b(x, y)$. The material derivative is

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla = \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y}, \quad (\text{SW.3})$$

with the rightmost expression holding in Cartesian coordinates.

Integrate this from the bottom of the fluid ($z = \eta_b$) to the top ($z = \eta$), noting that the right-hand side is independent of z , to give

$$w(\eta) - w(\eta_b) = -h \nabla \cdot \mathbf{u}. \quad (3.17)$$

At the top the vertical velocity is the material derivative of the position of a particular fluid element. But the position of the fluid at the top is just η , and therefore (see Fig. 3.2)

$$w(\eta) = \frac{D\eta}{Dt}. \quad (3.18a)$$

At the bottom of the fluid we have similarly

$$w(\eta_b) = \frac{D\eta_b}{Dt}, \quad (3.18b)$$

where, apart from earthquakes and the like, $\partial\eta_b/\partial t = 0$. Using (3.18a,b), (3.17) becomes

$$\frac{D}{Dt}(\eta - \eta_b) + h \nabla \cdot \mathbf{u} = 0 \quad (3.19)$$

or, as in (3.15),

$$\frac{Dh}{Dt} + h \nabla \cdot \mathbf{u} = 0. \quad (3.20)$$

3.1.3 A rigid lid

The case where the *upper* surface is held flat by the imposition of a rigid lid is sometimes of interest. The ocean suggests one such example, since the bathymetry at the bottom of the ocean provides much larger variations in fluid thickness than do the small variations in the

height of the ocean surface. Suppose then that the upper surface is at a constant height H then, from (3.14) with $\partial h/\partial t = 0$, the mass conservation equation becomes

$$\nabla_h \cdot (\mathbf{u}h_b) = 0, \quad (3.21)$$

where $h_b = H - \eta_b$. Note that (3.21) allows us to define an incompressible *mass-transport velocity*, $\mathbf{U} \equiv h_b \mathbf{u}$.

Although the upper surface is flat, the pressure there is no longer constant because a force must be provided by the rigid lid to keep the surface flat. The horizontal momentum equation is

$$\frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho} \nabla p_{lid}, \quad (3.22)$$

where p_{lid} is the pressure at the lid, and the complete equations of motion are then (3.21) and (3.22).¹ If the lower surface is flat, the two-dimensional flow itself is divergence-free, and the equations reduce to the two-dimensional incompressible Euler equations.

3.1.4 Stretching and the vertical velocity

Because the horizontal velocity is depth independent, the vertical velocity plays no role in advection. However, w is certainly not zero for then the free surface would be unable to move up or down, but because of the vertical independence of the horizontal flow w does have a simple vertical structure; to determine this we write the mass conservation equation as

$$\frac{\partial w}{\partial z} = -\nabla \cdot \mathbf{u} \quad (3.23)$$

and integrate upwards from the bottom to give

$$w = w_b - (\nabla \cdot \mathbf{u})(z - \eta_b). \quad (3.24)$$

Thus, the vertical velocity is a linear function of height. Equation (3.24) can be written as

$$\frac{Dz}{Dt} = \frac{D\eta_b}{Dt} - (\nabla \cdot \mathbf{u})(z - \eta_b), \quad (3.25)$$

and at the upper surface $w = D\eta/Dt$ so that here we have

$$\frac{D\eta}{Dt} = \frac{D\eta_b}{Dt} - (\nabla \cdot \mathbf{u})(\eta - \eta_b). \quad (3.26)$$

Eliminating the divergence term from the last two equations gives

$$\frac{D}{Dt}(z - \eta_b) = \frac{z - \eta_b}{\eta - \eta_b} \frac{D}{Dt}(\eta - \eta_b), \quad (3.27)$$

which in turn gives

$$\frac{D}{Dt} \left(\frac{z - \eta_b}{\eta - \eta_b} \right) = \frac{D}{Dt} \left(\frac{z - \eta_b}{h} \right) = 0. \quad (3.28)$$

This means that the ratio of the height of a fluid parcel above the floor to the total depth of the column is fixed; that is, the fluid stretches uniformly in a column, and this is a kinematic property of the shallow water system.

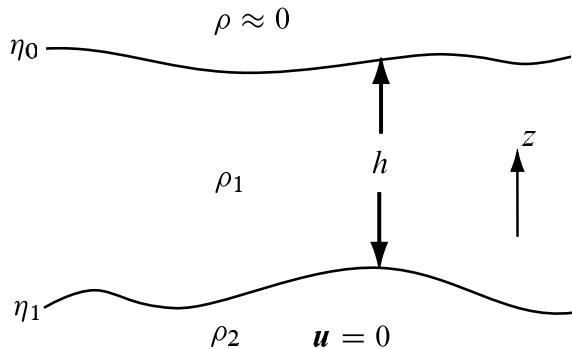


Fig. 3.3 The reduced gravity shallow water system. An active layer lies over a deep, more dense, quiescent layer. In a common variation the upper surface is held flat by a rigid lid, and $\eta_0 = 0$.

3.1.5 Analogy with compressible flow

The shallow water equations (3.8) and (3.14) are analogous to the compressible gas dynamic equations in two dimensions, namely

$$\frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho} \nabla p \quad (3.29)$$

and

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\mathbf{u} \rho) = 0, \quad (3.30)$$

along with an equation of state which we take to be $p = f(\rho)$. The mass conservation equations (3.14) and (3.30) are identical, with the replacement $\rho \leftrightarrow h$. If $p = C\rho^\gamma$, then (3.29) becomes

$$\frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho} \frac{dp}{d\rho} \nabla \rho = -C\gamma \rho^{\gamma-2} \nabla \rho. \quad (3.31)$$

If $\gamma = 2$ then the momentum equations (3.8) and (3.31) become equivalent, with $\rho \leftrightarrow h$ and $C\gamma \leftrightarrow g$. In an ideal gas $\gamma = c_p/c_v$ and values typically are in fact less than 2 (in air $\gamma \approx 7/5$); however, if the equations are linearized, then the analogy is exact for all values of γ , for then (3.31) becomes $\partial \mathbf{v}'/\partial t = -\rho_0^{-1} c_s^2 \nabla \rho'$ where $c_s^2 = dp/d\rho$, and the linearized shallow water momentum equation is $\partial \mathbf{u}'/\partial t = -H^{-1}(gH) \nabla h'$, so that $\rho_0 \leftrightarrow H$ and $c_s^2 \leftrightarrow gH$. The sound waves of a compressible fluid are then analogous to shallow water waves, which are considered in section 3.7.

3.2 REDUCED GRAVITY EQUATIONS

Consider now a single shallow moving layer of fluid on top of a deep, quiescent fluid layer (Fig. 3.3), and beneath a fluid of negligible inertia. This configuration is often used as a model of the upper ocean: the upper layer represents flow in perhaps the upper few hundred metres of the ocean, the lower layer being the near-stagnant abyss. If we turn the model upside-down we have a perhaps slightly less realistic model of the atmosphere: the lower layer represents motion in the troposphere above which lies an inactive stratosphere. The equations of motion are virtually the same in both cases.

3.2.1 Pressure gradient in the active layer

We will derive the equations for the oceanic case (active layer on top) in two cases, which differ slightly in the assumption made about the upper surface.

I Free upper surface

The pressure in the upper layer is given by integrating the hydrostatic equation down from the upper surface. Thus, at a height z in the upper layer

$$p_1(z) = g\rho_1(\eta_0 - z), \quad (3.32)$$

where η_0 is the height of the upper surface. Hence, everywhere in the upper layer,

$$\frac{1}{\rho_1} \nabla p_1 = -g \nabla \eta_0, \quad (3.33)$$

and the momentum equation is

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -g \nabla \eta_0. \quad (3.34)$$

In the lower layer the pressure is also given by the weight of the fluid above it. Thus, at some level z in the lower layer,

$$p_2(z) = \rho_1 g(\eta_0 - \eta_1) + \rho_2 g(\eta_1 - z). \quad (3.35)$$

But if this layer is motionless the horizontal pressure gradient in it is zero and therefore

$$\rho_1 g \eta_0 = -\rho_1 g' \eta_1 + \text{constant}, \quad (3.36)$$

where $g' = g(\rho_2 - \rho_1)/\rho_1$ is the *reduced gravity*. The momentum equation becomes

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = g' \nabla \eta_1. \quad (3.37)$$

The equations are completed by the usual mass conservation equation,

$$\frac{Dh}{Dt} + h \nabla \cdot \mathbf{u} = 0, \quad (3.38)$$

where $h = \eta_0 - \eta_1$. Because $g \gg g'$, (3.36) shows that surface displacements are *much smaller* than the displacements at the interior interface. We see this in the real ocean where the mean interior isopycnal displacements may be several tens of metres but variations in the mean height of ocean surface are of the order of centimetres.

II The rigid lid approximation

The smallness of the upper surface displacement suggests that we will make little error if we impose a *rigid lid* at the top of the fluid. Displacements are no longer allowed, but the lid will in general impart a pressure force to the fluid. Suppose that this is $P(x, y, t)$, then the horizontal pressure gradient in the upper layer is simply

$$\nabla p_1 = \nabla P. \quad (3.39)$$

The pressure in the lower layer is again given by hydrostasy, and is

$$p_2 = -\rho_1 g \eta_1 + \rho_2 g(\eta_1 - z) + P = \rho_1 g h - \rho_2 g(h + z) + P, \quad (3.40)$$

so that

$$\nabla p_2 = -g(\rho_2 - \rho_1) \nabla h + \nabla P. \quad (3.41)$$

Then if $\nabla p_2 = 0$ we have

$$g(\rho_2 - \rho_1) \nabla h = \nabla P, \quad (3.42)$$

and the momentum equation for the upper layer is just

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -g'_1 \nabla h. \quad (3.43)$$

where $g' = g(\rho_2 - \rho_1)/\rho_1$. These equations differ from the usual shallow water equations only in the use of a reduced gravity g' in place of g itself. It is the *density difference* between the two layers that is important. Similarly, if we take a shallow water system, with the moving layer on the bottom, and we suppose that overlying it is a stationary fluid of finite density, then we would easily find that the fluid equations for the moving layer are the same as if the fluid on top had zero inertia, except that g would be replaced by an appropriate reduced gravity (problem 3.1).

3.3 MULTI-LAYER SHALLOW WATER EQUATIONS

We now consider the dynamics of multiple layers of fluid stacked on top of each other. This is a crude representation of continuous stratification, but it turns out to be a powerful model of many geophysically interesting phenomena as well as being physically realizable in the laboratory. The pressure is continuous across the interface, but the density jumps discontinuously and this allows the horizontal velocity to have a corresponding discontinuity. The set up is illustrated in Fig. 3.4.

In each layer pressure is given by the hydrostatic approximation, and so anywhere in the interior we can find the pressure by integrating down from the top. Thus, at a height z in the first layer we have

$$p_1 = \rho_1 g(\eta_0 - z), \quad (3.44)$$

and in the second layer,

$$p_2 = \rho_1 g(\eta_0 - \eta_1) + \rho_2 g(\eta_1 - z) = \rho_1 g \eta_0 + \rho_1 g'_1 \eta_1 - \rho_2 g z, \quad (3.45)$$

where $g'_1 = g(\rho_2 - \rho_1)/\rho_1$, and so on. The term involving z is irrelevant for the dynamics, because only the horizontal derivative enters the equation of motion. Omitting this term, for the n th layer the dynamical pressure is given by the sum from the top down:

$$p_n = \rho_1 \sum_{i=0}^{n-1} g'_i \eta_i, \quad (3.46)$$

where $g'_i = g(\rho_{i+1} - \rho_i)/\rho_1$ (but $g_0 = g$). The interface displacements may be expressed in terms of the layer thicknesses by summing from the bottom up:

$$\eta_n = \eta_b + \sum_{i=n+1}^{i=N} h_i. \quad (3.47)$$

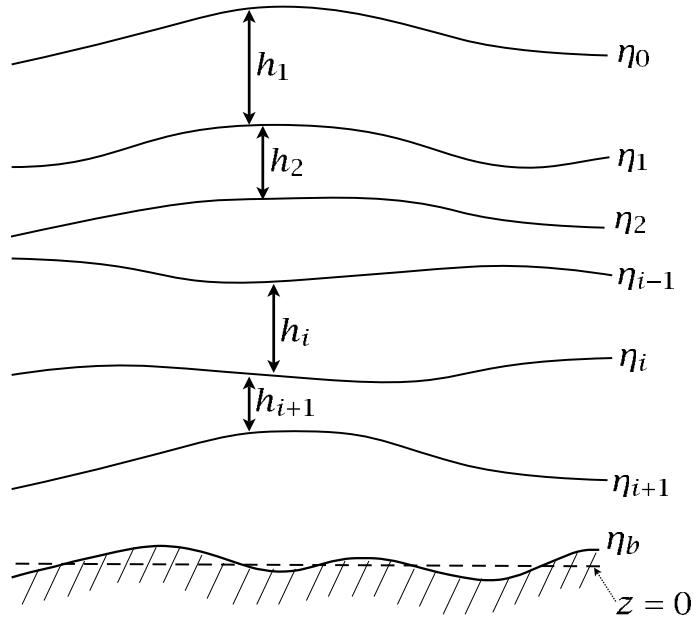


Fig. 3.4 The multi-layer shallow water system. The layers are numbered from the top down. The coordinates of the interfaces are denoted by η , and the layer thicknesses by h , so that $h_i = \eta_i - \eta_{i-1}$.

The momentum equation for each layer may then be written, in general,

$$\frac{D\mathbf{u}_n}{Dt} + \mathbf{f} \times \mathbf{u}_n = -\frac{1}{\rho_n} \nabla p_n, \quad (3.48)$$

where the pressure is given by (3.46) and in terms of the layer depths using (3.48). If we make the Boussinesq approximation then ρ_n on the right-hand side of (3.48) is replaced by ρ_1 .

Finally, the mass conservation equation for each layer has the same form as the single-layer case, and is

$$\frac{Dh_n}{Dt} + h_n \nabla \cdot \mathbf{u}_n = 0. \quad (3.49)$$

The two- and three-layer cases

The two-layer model (Fig. 3.5) is the simplest model to capture the effects of stratification. Evaluating the pressures using (3.46) and (3.47) we find:

$$p_1 = \rho_1 g \eta_0 = \rho_1 g (h_1 + h_2 + \eta_b) \quad (3.50a)$$

$$p_2 = \rho_1 [g \eta_0 + g'_1 \eta_1] = \rho_1 [g (h_1 + h_2 + \eta_b) + g'_1 (h_2 + \eta_b)]. \quad (3.50b)$$

The momentum equations for the two layers are then

$$\frac{D\mathbf{u}_1}{Dt} + \mathbf{f} \times \mathbf{u}_1 = -g \nabla \eta_0 = -g \nabla (h_1 + h_2 + \eta_b). \quad (3.51a)$$

and in the bottom layer

$$\begin{aligned} \frac{D\mathbf{u}_2}{Dt} + \mathbf{f} \times \mathbf{u}_2 &= -\frac{\rho_1}{\rho_2} (g \nabla \eta_0 + g'_1 \nabla \eta_1) \\ &= -\frac{\rho_1}{\rho_2} [g \nabla (\eta_b + h_1 + h_2) + g'_1 \nabla (h_2 + \eta_b)]. \end{aligned} \quad (3.51b)$$

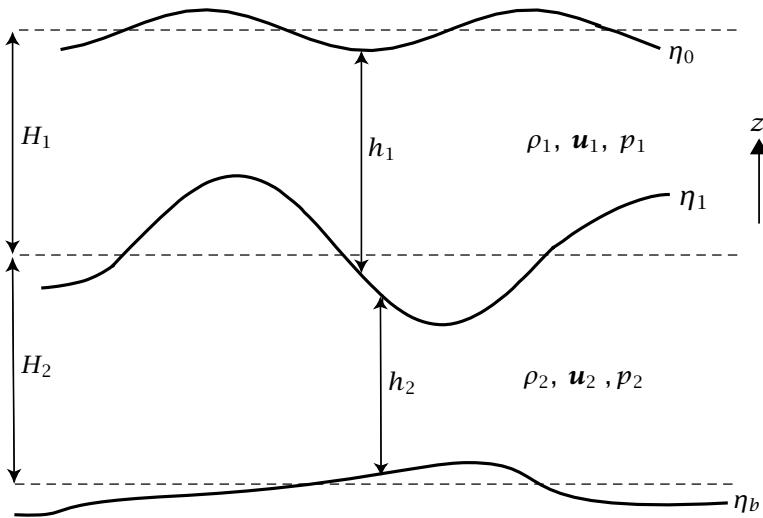


Fig. 3.5 The two-layer shallow water system. A fluid of density ρ_1 lies over a denser fluid of density ρ_2 . In the reduced gravity case the lower layer may be arbitrarily thick and is assumed stationary and so has no horizontal pressure gradient. In the 'rigid-lid' approximation the top surface displacement is neglected, but there is then a non-zero pressure gradient induced by the lid.

In the Boussinesq approximation ρ_1/ρ_2 is replaced by unity.

In a three-layer model the dynamical pressures are found to be

$$p_1 = \rho_1 gh \quad (3.52a)$$

$$p_2 = \rho_1 [gh + g'_1(h_2 + h_3 + \eta_b)] \quad (3.52b)$$

$$p_3 = \rho_1 [gh + g'_1(h_2 + h_3 + \eta_b) + g'_2(h_3 + \eta_b)], \quad (3.52c)$$

where $h = \eta_0 = \eta_b + h_1 + h_2 + h_3$ and $g'_2 = g(\rho_3 - \rho_2)/\rho_1$. More layers can obviously be added in a systematic fashion.

3.3.1 Reduced-gravity multi-layer equation

As with a single active layer, we may envision multiple layers of fluid overlying a deeper stationary layer. This is a useful model of the stratified upper ocean overlying a nearly stationary and nearly unstratified abyss. Indeed we use such a model to study the 'ventilated thermocline' in chapter 16 and a detailed treatment may be found there. If we suppose there is a lid at the top, then the model is almost the same as that of the previous section. However, now the horizontal pressure gradient in the lowest model layer is zero, and so we may obtain the pressures in all the active layers by integrating the hydrostatic equation upwards from this layer. Suppose we have N moving layers, then the reader may verify that the dynamic pressure in the n th layer is given by

$$p_n = - \sum_{i=n}^{i=N} \rho_1 g'_i \eta_i, \quad (3.53)$$

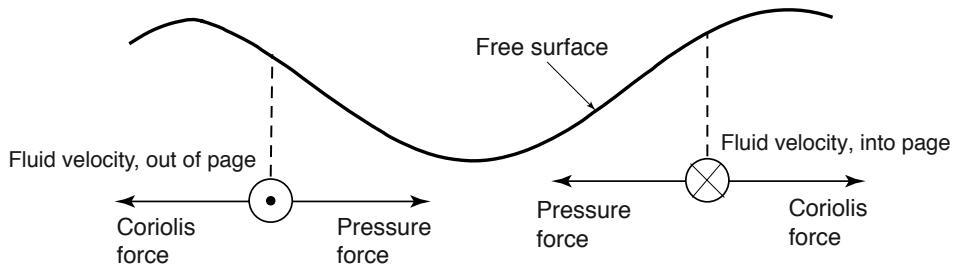


Fig. 3.6 Geostrophic flow in a shallow water system, with a positive value of the Coriolis parameter f , as in the Northern Hemisphere. The pressure force is directed down the gradient of the height field, and this can be balanced by the Coriolis force if the fluid velocity is at right angles to it. If f were negative, the geostrophic flow would be reversed.

where as before $g'_i = g(\rho_{i+1} - \rho_i)/\rho_1$. If we have a lid at the top, and take $\eta_0 = 0$, then the interface displacements are related to the layer thicknesses by

$$\eta_n = - \sum_{i=1}^{i=n} h_i. \quad (3.54)$$

From these expressions the momentum equation in each layer is easily constructed.

3.4 GEOSTROPHIC BALANCE AND THERMAL WIND

Geostrophic balance occurs in the shallow water equations, just as in the continuously stratified equations, when the Rossby number U/fL is small and the Coriolis term dominates the advective terms in the momentum equation. In the single-layer shallow water equations the geostrophic flow is:

$$\mathbf{f} \times \mathbf{u}_g = -g \nabla \eta. \quad (3.55)$$

Thus, the geostrophic velocity is proportional to the slope of the surface, as sketched in Fig. 3.6. (For the rest of this section, we will drop the subscript g , and take all velocities to be geostrophic.)

In both the single-layer and multi-layer cases, the slope of an interfacial surface is directly related to the difference in pressure gradient on either side and so, by geostrophic balance, to the shear of the flow. This is the shallow water analogue of the thermal wind relation. To obtain an expression for this, consider the interface, η , between two layers labelled 1 and 2. The pressure in two layers is given by the hydrostatic relation and so,

$$p_1 = A(x, y) - \rho_1 g z \quad (\text{at some } z \text{ in layer 1}) \quad (3.56a)$$

$$\begin{aligned} p_2 &= A(x, y) - \rho_1 g \eta + \rho_2 g(\eta - z) \\ &= A(x, y) + \rho_1 g'_1 \eta - \rho_2 g z \quad (\text{at some } z \text{ in layer 2}) \end{aligned} \quad (3.56b)$$

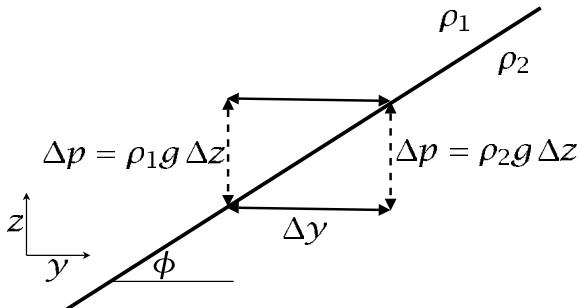


Fig. 3.7 Margules' relation: using hydrostasy, the difference in the horizontal pressure gradient between the upper and the lower layer is given by $-g'\rho_1 s$, where $s = \tan \phi = \Delta z / \Delta y$ is the interface slope and $g' = (\rho_2 - \rho_1) / \rho_1$. Geostrophic balance then gives $f(u_1 - u_2) = g's$, which is a special case of (3.60).

where $A(x, y)$ is a function of integration. Thus we find

$$\frac{1}{\rho_1} \nabla(p_1 - p_2) = -g'_1 \nabla \eta. \quad (3.57)$$

If the flow is geostrophically balanced and Boussinesq then, in each layer, the velocity obeys

$$f \mathbf{u}_i = \frac{1}{\rho_1} \mathbf{k} \times \nabla p_i. \quad (3.58)$$

Using (3.57) then gives

$$f(\mathbf{u}_1 - \mathbf{u}_2) = -\mathbf{k} \times g'_1 \nabla \eta, \quad (3.59)$$

or in general

$$f(\mathbf{u}_n - \mathbf{u}_{n+1}) = -\mathbf{k} \times g'_n \nabla \eta. \quad (3.60)$$

This is the thermal wind equation for the shallow water system. It applies at any interface, and it implies *the shear is proportional to the interface slope*, a result known as the 'Margules relation' (Fig. 3.7).²

Suppose that we represent the atmosphere by two layers of fluid; a meridionally decreasing temperature may then be represented by an interface that slopes upwards toward the pole. Then, in either hemisphere, we have

$$u_1 - u_2 = \frac{g_1}{f} \frac{\partial \eta}{\partial y} > 0, \quad (3.61)$$

and the temperature gradient is associated with a positive shear (see problem 3.2).

3.5 FORM DRAG

When the interface between two layers varies with position — that is, when it is wavy — the layers exert a pressure force on each other. Similarly, if the bottom of the fluid is not flat then the topography and the bottom layer will in general exert forces on each other. This kind of force is known as *form drag*, and it is an important means whereby momentum can be added to or extracted from a flow.³ Consider a layer confined between two interfaces, $\eta_1(x, y)$ and $\eta_2(x, y)$. Then over some zonal interval L the average zonal pressure force on that fluid layer is given by

$$F_p = -\frac{1}{L} \int_{x_1}^{x_2} \int_{\eta_2}^{\eta_1} \frac{\partial p}{\partial x} dx dz. \quad (3.62)$$

Integrating by parts first in z and then in x , and noting that by hydrostasy $\partial p / \partial z$ does not depend on horizontal position within the layer, we obtain

$$F_p = -\frac{1}{L} \int_{x_1}^{x_2} \left[\frac{\partial p}{\partial x} z \right]_{\eta_2}^{\eta_1} dx = -\overline{\eta_1 \frac{\partial p_1}{\partial x}} + \overline{\eta_2 \frac{\partial p_2}{\partial x}} = +\overline{p_1 \frac{\partial \eta_1}{\partial x}} - \overline{p_2 \frac{\partial \eta_2}{\partial x}}, \quad (3.63)$$

where p_1 is the pressure at η_1 , and similarly for p_2 , and to obtain the second line we suppose that the integral is around a closed path, such as a circle of latitude, and the average is denoted with an overbar. These terms represent the transfer of momentum from one layer to the next, and at a particular interface, i , we may define the form drag, τ_i , by

$$\tau_i \equiv \overline{p_i \frac{\partial \eta_i}{\partial x}} = -\overline{\eta_i \frac{\partial p_i}{\partial x}}. \quad (3.64)$$

The form drag is a stress and as the layer depth shrinks to zero its vertical derivative, $\partial \tau / \partial z$, is the force (per unit volume) on the fluid. Form drag is a particularly important means for the vertical transfer of momentum and its ultimate removal in an eddying fluid, and is one of the main mechanisms whereby the wind stress at the top of the ocean is communicated to the ocean bottom. At the fluid bottom the form drag is $\overline{p \eta_{bx}}$, where η_b is the bottom topography, and this is proportional to the momentum exchange with the solid Earth. This is a significant mechanism for the ultimate removal of momentum in the ocean, especially in the Antarctic Circumpolar Current where it is likely to be much larger than bottom (or Ekman) drag arising from small-scale turbulence and friction. In the two-layer, flat-bottomed case the only form drag occurring is that at the interface, and the momentum transfer between the layers is just $p_1 \partial \eta_1 / \partial x$ or $-\eta_1 \partial p / \partial x$; then, the force on each layer due to the other is equal and opposite, as we would expect from momentum conservation. (Form drag is discussed more in an oceanographic context in sections 14.6.3 and 16.6.2.)

For flows in geostrophic balance, the form drag is related to the meridional heat flux. The pressure gradient and velocity are related by $\rho f v' = \partial p' / \partial x$ and the interfacial displacement is proportional to the temperature perturbation, b' [in fact one may show that $\eta' \approx -b' / (\partial \bar{b} / \partial z)$]. Thus $-\eta' \partial p'_\eta / \partial x \propto \bar{v}' b'$, a correspondence that will re-occur when we consider the *Eliassen-Palm* flux in chapter 7.

3.6 CONSERVATION PROPERTIES OF SHALLOW WATER SYSTEMS

There are two common types of conservation property in fluids: (i) material invariants and (ii) integral invariants. Material invariance occurs when a property (ϕ say) is conserved on each fluid element, and so obeys the equation $D\phi/Dt = 0$. An integral invariant is one that is conserved following an integration over some, usually closed, volume; energy is an example.

3.6.1 Potential vorticity: a material invariant

The vorticity of a fluid (considered at greater length in chapter 4), denoted ω , is defined to be the curl of the velocity field. Let us also define the shallow water vorticity, ω^* , as the curl of the horizontal velocity. We therefore have:

$$\omega \equiv \nabla \times \mathbf{v}, \quad \omega^* \equiv \nabla \times \mathbf{u}. \quad (3.65)$$

Because $\partial u / \partial z = \partial v / \partial z = 0$, only the vertical component of \mathbf{w}^* is non-zero and

$$\mathbf{w}^* = \mathbf{k} \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right) \equiv \mathbf{k} \zeta. \quad (3.66)$$

Using the vector identity

$$(\mathbf{u} \cdot \nabla) \mathbf{u} = \frac{1}{2} \nabla (\mathbf{u} \cdot \mathbf{u}) - \mathbf{u} \times (\nabla \times \mathbf{u}), \quad (3.67)$$

we write the momentum equation, (3.8), as

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{w}^* \times \mathbf{u} = -\nabla \left(g\eta + \frac{1}{2} \mathbf{u}^2 \right). \quad (3.68)$$

To obtain an evolution equation for the vorticity we take the curl of (3.68), and make use of the vector identity

$$\begin{aligned} \nabla \times (\mathbf{w}^* \times \mathbf{u}) &= (\mathbf{u} \cdot \nabla) \mathbf{w}^* - (\mathbf{w}^* \cdot \nabla) \mathbf{u} + \mathbf{w}^* \nabla \cdot \mathbf{u} - \mathbf{u} \nabla \cdot \mathbf{w}^* \\ &= (\mathbf{u} \cdot \nabla) \mathbf{w}^* + \mathbf{w}^* \nabla \cdot \mathbf{u}, \end{aligned} \quad (3.69)$$

using the fact that $\nabla \cdot \mathbf{w}^*$ is the divergence of a curl and therefore zero, and $(\mathbf{w}^* \cdot \nabla) \mathbf{u} = 0$ because \mathbf{w}^* is perpendicular to the surface in which \mathbf{u} varies. Taking the curl of (3.68) gives

$$\frac{\partial \zeta}{\partial t} + (\mathbf{u} \cdot \nabla) \zeta = -\zeta \nabla \cdot \mathbf{u}, \quad (3.70)$$

where $\zeta = \mathbf{k} \cdot \mathbf{w}^*$. Now, the mass conservation equation may be written as

$$-\zeta \nabla \cdot \mathbf{u} = \frac{\zeta}{h} \frac{Dh}{Dt}, \quad (3.71)$$

and using this (3.70) becomes

$$\frac{D\zeta}{Dt} = \frac{\zeta}{h} \frac{Dh}{Dt}, \quad (3.72)$$

which simplifies to

$$\frac{D}{Dt} \left(\frac{\zeta}{h} \right) = 0. \quad (3.73)$$

The important quantity ζ/h , often denoted by Q , is known as the *potential vorticity*, and (3.73) is the potential vorticity equation. We re-derive this conservation law in a more general way in section 4.6

Because Q is conserved on parcels, then so is any function of Q ; that is, $F(Q)$ is a material invariant, where F is any function. To see this algebraically, multiply (3.73) by $F'(Q)$, the derivative of F with respect to Q , giving

$$F'(Q) \frac{DQ}{Dt} = \frac{D}{Dt} F(Q) = 0. \quad (3.74)$$

Since F is arbitrary there are an infinite number of material invariants corresponding to different choices of F .

Effects of rotation

In a rotating frame of reference, the shallow water momentum equation is

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -g \nabla \eta, \quad (3.75)$$

where (as before) $\mathbf{f} = f\mathbf{k}$. This may be written in vector invariant form as

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{w}^* + \mathbf{f}) \times \mathbf{u} = -\nabla \left(g\eta + \frac{1}{2}\mathbf{u}^2 \right), \quad (3.76)$$

and taking the curl of this gives the vorticity equation

$$\frac{\partial \zeta}{\partial t} + (\mathbf{u} \cdot \nabla)(\zeta + f) = -(f + \zeta) \nabla \cdot \mathbf{u}. \quad (3.77)$$

This is the same as the shallow water vorticity equation in a non-rotating frame, save that ζ is replaced by $\zeta + f$, the reason for this being that f is the vorticity that the fluid has by virtue of the background rotation. Thus, (3.77) is simply the equation of motion for the total or absolute vorticity, $\mathbf{w}_a = \mathbf{w}^* + \mathbf{f} = (\zeta + f)\mathbf{k}$.

The potential vorticity equation in the rotating case follows, much as in the non-rotating case, by combining (3.77) with the mass conservation equation, giving

$$\frac{D}{Dt} \left(\frac{\zeta + f}{h} \right) = 0. \quad (3.78)$$

That is, $Q \equiv (\zeta + f)/h$, the potential vorticity in a rotating shallow system, is a material invariant.

Vorticity and circulation

Although vorticity itself is not a material invariant, its integral over a horizontal material area is invariant. To demonstrate this in the non-rotating case, consider the integral

$$C = \int_A \zeta \, dA = \int_A Qh \, dA, \quad (3.79)$$

over a surface A , the cross-sectional area of a column of height h (as in Fig. 3.2). Taking the material derivative of this gives

$$\frac{DC}{Dt} = \int_A \frac{DQ}{Dt} h \, dA + \int_A Q \frac{D}{Dt}(h \, dA). \quad (3.80)$$

The first term is zero, by (3.72); the second term is just the derivative of the volume of a column of fluid and it too is zero, by mass conservation. Thus,

$$\frac{DC}{Dt} = \frac{D}{Dt} \int_A \zeta \, dA = 0. \quad (3.81)$$

Thus, the integral of the vorticity over a some cross-sectional area of the fluid is unchanging, although both the vorticity and area of the fluid may individually change. Using Stokes' theorem, it may be written as

$$\frac{DC}{Dt} = \frac{D}{Dt} \oint \mathbf{u} \cdot d\mathbf{l}, \quad (3.82)$$

where the line integral is around the boundary of A . This is an example of Kelvin's circulation theorem, which we shall meet again in a more general form in chapter 4, where we also consider the rotating case.

A slight generalization of (3.81) is possible. Consider the integral $I = \int F(Q)h \, dA$ where again F is any differentiable function of its argument. It is clear that

$$\frac{D}{Dt} \int_A F(Q)h \, dA = 0. \quad (3.83)$$

If the area of integration in (3.68) or (3.83) is the whole domain (enclosed by frictionless walls, for example) then it is clear that the integral of $hF(Q)$ is a constant, including as a special case the integral of ζ .

3.6.2 Energy conservation: an integral invariant

Since we have made various simplifications in deriving the shallow water system, it is not self-evident that energy should be conserved, or indeed what form the energy takes. The kinetic energy density (KE), that is the kinetic energy per unit area, is $\rho_0 h \mathbf{u}^2 / 2$. The potential energy density of the fluid is

$$PE = \int_0^h \rho_0 g z \, dz = \frac{1}{2} \rho_0 g h^2. \quad (3.84)$$

The factor ρ_0 appears in both kinetic and potential energies and, because it is a constant, we will omit it. For algebraic simplicity we also assume the bottom is flat, at $z = 0$.

Using the mass conservation equation (3.15) we obtain an equation for the evolution of potential energy density, namely

$$\frac{D}{Dt} \frac{gh^2}{2} + gh^2 \nabla \cdot \mathbf{u} = 0 \quad (3.85a)$$

or

$$\frac{\partial}{\partial t} \frac{gh^2}{2} + \nabla \cdot \left(\mathbf{u} \frac{gh^2}{2} \right) + \frac{gh^2}{2} \nabla \cdot \mathbf{u} = 0. \quad (3.85b)$$

From the momentum and mass continuity equations we obtain an equation for the evolution of kinetic energy density, namely

$$\frac{D}{Dt} \frac{h \mathbf{u}^2}{2} + \frac{\mathbf{u}^2 h}{2} \nabla \cdot \mathbf{u} = -g \mathbf{u} \cdot \nabla \frac{h^2}{2} \quad (3.86a)$$

or

$$\frac{\partial}{\partial t} \frac{h \mathbf{u}^2}{2} + \nabla \cdot \left(\mathbf{u} \frac{h \mathbf{u}^2}{2} \right) + g \mathbf{u} \cdot \nabla \frac{h^2}{2} = 0. \quad (3.86b)$$

Adding (3.85b) and (3.86b) we obtain

$$\frac{\partial}{\partial t} \frac{1}{2} (h \mathbf{u}^2 + gh^2) + \nabla \cdot \left[\frac{1}{2} \mathbf{u} (gh^2 + h \mathbf{u}^2 + gh^2) \right] = 0, \quad (3.87)$$

or

$$\frac{\partial E}{\partial t} + \nabla \cdot \mathbf{F} = 0, \quad (3.88)$$

where $E = KE + PE = (h\mathbf{u}^2 + gh^2)/2$ is the density of the total energy and $\mathbf{F} = \mathbf{u}(h\mathbf{u}^2 + gh^2 + gh^2)/2$ is the energy flux. If the fluid is confined to a domain bounded by rigid walls, on which the normal component of velocity vanishes, then on integrating (3.87) over that area and using Gauss's theorem, the total energy is seen to be conserved; that is

$$\frac{d\hat{E}}{dt} = \frac{1}{2} \frac{d}{dt} \int_A (h\mathbf{u}^2 + gh^2) dA = 0. \quad (3.89)$$

Such an energy principle also holds in the case with bottom topography. Note that, as we found in the case for a compressible fluid in chapter 2, the energy flux in (3.88) is not just the energy density multiplied by the velocity; it contains an additional term $g\mathbf{u}h^2/2$, and this represents the energy transfer occurring when the fluid does work against the pressure force (see problem 3.3).

3.7 SHALLOW WATER WAVES

Let us now look at the gravity waves that occur in shallow water. To isolate the essence of the phenomena, we will consider waves in a single fluid layer, with a flat bottom and a free upper surface, in which gravity provides the sole restoring force.

3.7.1 Non-rotating shallow water waves

Given a flat bottom the fluid thickness is equal to the free surface displacement (Fig. 3.1), and we let

$$h(x, y, t) = H + h'(x, y, t) = H + \eta'(x, y, t), \quad (3.90a)$$

$$\mathbf{u}(x, y, t) = \mathbf{u}'(x, y, t). \quad (3.90b)$$

The mass conservation equation, (3.15), then becomes

$$\frac{\partial \eta'}{\partial t} + (H + \eta') \nabla \cdot \mathbf{u}' + \mathbf{u}' \cdot \nabla \eta' = 0, \quad (3.91)$$

and neglecting squares of small quantities this yields the linear equation

$$\frac{\partial \eta'}{\partial t} + H \nabla \cdot \mathbf{u}' = 0. \quad (3.92)$$

Similarly, linearizing the momentum equation, (3.8) with $f = 0$, yields

$$\frac{\partial \mathbf{u}'}{\partial t} = -g \nabla \eta'. \quad (3.93)$$

Eliminating velocity by differentiating (3.92) with respect to time and taking the divergence of (3.93) leads to

$$\frac{\partial^2 \eta'}{\partial t^2} - gH \nabla^2 \eta' = 0, \quad (3.94)$$

which may be recognized as a wave equation. We can find the dispersion relationship for this by substituting the trial solution

$$\eta' = \text{Re } \tilde{\eta} e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}, \quad (3.95)$$

where $\tilde{\eta}$ is a complex constant, $\mathbf{k} = ik + jl$ is the horizontal wavenumber and Re indicates that the real part of the solution should be taken. If, for simplicity, we restrict attention to the one-dimensional problem, with no variation in the y -direction, then substituting into (3.94) leads to the dispersion relationship

$$\omega = \pm ck, \quad (3.96)$$

where $c = \sqrt{gH}$; that is, the wave speed is proportional to the square root of the mean fluid depth and is independent of the wavenumber — the waves are dispersionless. The general solution is a superposition of all such waves, with the amplitudes of each wave (or Fourier component) being determined by the Fourier decomposition of the initial conditions.

Because the waves are dispersionless, the general solution can be written as

$$\eta'(x, t) = \frac{1}{2} [F(x - ct) + F(x + ct)], \quad (3.97)$$

where $F(x)$ is the height field at $t = 0$. From this, it is easy to see that the shape of an initial disturbance is preserved as it propagates both to the right and to the left at speed c , (see also problem 3.7).

3.7.2 Rotating shallow water (Poincaré) waves

We now consider the effects of rotation on shallow water waves. Linearizing the rotating, flat-bottomed f -plane shallow water equations [i.e., (SW.1) and (SW.2) on page 127] about a state of rest we obtain

$$\frac{\partial u'}{\partial t} - f_0 v' = -g \frac{\partial \eta'}{\partial x}, \quad \frac{\partial v'}{\partial t} + f_0 u' = -g \frac{\partial \eta'}{\partial y}, \quad \frac{\partial \eta'}{\partial t} + H \left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} \right) = 0. \quad (3.98a,b,c)$$

We non-dimensionalize these equations by writing

$$(x, y) = L(\hat{x}, \hat{y}), \quad (u', v') = U(\hat{u}, \hat{v}), \quad t = \frac{L}{U} \hat{t}, \quad f_0 = \frac{\hat{f}_0}{T}, \quad \eta' = H \hat{\eta}, \quad (3.99)$$

and (3.98) becomes

$$\frac{\partial \hat{u}}{\partial \hat{t}} - \hat{f}_0 \hat{v} = -\hat{c}^2 \frac{\partial \hat{\eta}}{\partial \hat{x}}, \quad \frac{\partial \hat{v}}{\partial \hat{t}} + \hat{f}_0 \hat{u} = -\hat{c}^2 \frac{\partial \hat{\eta}}{\partial \hat{y}}, \quad \frac{\partial \hat{\eta}}{\partial \hat{t}} + \left(\frac{\partial \hat{u}}{\partial \hat{x}} + \frac{\partial \hat{v}}{\partial \hat{y}} \right) = 0. \quad (3.100a,b,c)$$

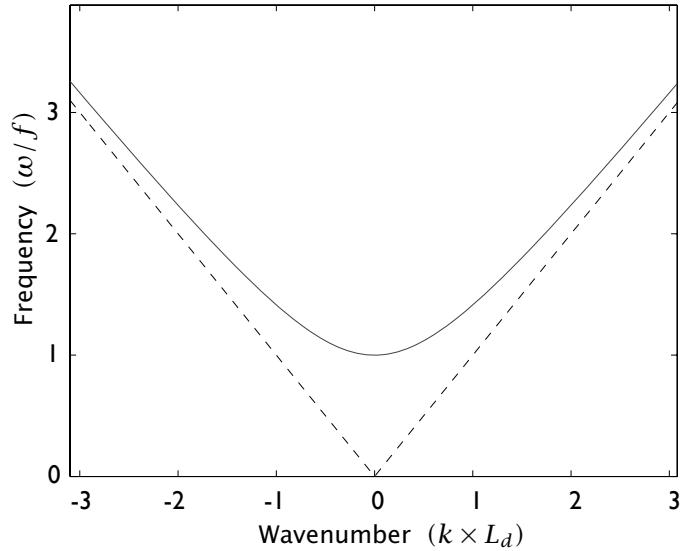
where $\hat{c} = \sqrt{gH}/U$ is the non-dimensional speed of non-rotating shallow water waves. (It is also the inverse of the Froude number U/\sqrt{gH} .) To obtain a dispersion relationship we let

$$(\hat{u}, \hat{v}, \hat{\eta}) = (\tilde{u}, \tilde{v}, \tilde{\eta}) e^{i(\hat{\mathbf{k}} \cdot \hat{\mathbf{x}} - \hat{\omega} \hat{t})}, \quad (3.101)$$

where $\hat{\mathbf{k}} = \hat{k}\mathbf{i} + \hat{l}\mathbf{j}$ and $\hat{\omega}$ is the non-dimensional frequency, and substitute into (3.100), giving

$$\begin{pmatrix} -i\hat{\omega} & -\hat{f}_0 & i\hat{c}^2\hat{k} \\ \hat{f}_0 & -i\hat{\omega} & i\hat{c}^2\hat{l} \\ i\hat{k} & i\hat{l} & -i\hat{\omega} \end{pmatrix} \begin{pmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{\eta} \end{pmatrix} = 0. \quad (3.102)$$

Fig. 3.8 Dispersion relation for Poincaré waves (solid) and non-rotating shallow water waves (dashed). Frequency is scaled by the Coriolis frequency f , and wavenumber by the inverse deformation radius \sqrt{gH}/f . For small wavenumbers the frequency is approximately f ; for high wavenumbers it asymptotes to that of non-rotating waves.



This homogeneous equation has non-trivial solutions only if the determinant of the matrix vanishes. This condition gives

$$\widehat{\omega}(\widehat{\omega}^2 - \widehat{f}_0^2 - \widehat{c}^2 \widehat{K}^2) = 0. \quad (3.103)$$

where $\widehat{K}^2 = \widehat{k}^2 + \widehat{l}^2$. There are two classes of solution to (3.103). The first is simply $\widehat{\omega} = 0$, i.e., time-independent flow corresponding to geostrophic balance in (3.98). (Because geostrophic balance gives a divergence-free velocity field for a constant Coriolis parameter the equations are satisfied by a time-independent solution.) The second set of solutions satisfies the dispersion relation

$$\widehat{\omega}^2 = \widehat{f}_0^2 + \widehat{c}^2(\widehat{k}^2 + \widehat{l}^2), \quad (3.104)$$

which in dimensional form is

$$\omega^2 = f_0^2 + gH(k^2 + l^2) \quad . \quad (3.105)$$

The corresponding waves are known as *Poincaré* waves,⁴ and the dispersion relationship is illustrated in Fig. 3.8. Note that the frequency is always greater than the Coriolis frequency f_0 . There are two interesting limits.

(i) *The short wave limit.* If

$$K^2 \gg \frac{f_0^2}{gH}, \quad (3.106)$$

where $K^2 = k^2 + l^2$, then the dispersion relationship reduces to that of the non-rotating case (3.96). This condition is equivalent to requiring that the wavelength be much shorter than the *deformation radius*, $L_d \equiv \sqrt{gH}/f$. Specifically, if $l = 0$ and $\lambda = 2\pi/k$ is the wavelength, the condition is

$$\lambda^2 \ll L_d^2 (2\pi)^2. \quad (3.107)$$

The numerical factor of $(2\pi)^2$ is more than an order of magnitude, so care must be taken when deciding if the condition is satisfied in particular cases. Furthermore, the wavelength must still be longer than the depth of the fluid, otherwise the shallow water condition is not met.

(ii) *The long wave limit.* If

$$K^2 \ll \frac{f_0^2}{gH}, \quad (3.108)$$

that is if the wavelength is much longer than the deformation radius L_d , then the dispersion relationship is

$$\omega = f_0. \quad (3.109)$$

These are known as *inertial oscillations*. The equations of motion giving rise to them are

$$\frac{\partial u'}{\partial t} - f_0 v' = 0, \quad \frac{\partial v'}{\partial t} + f_0 u' = 0, \quad (3.110)$$

which are equivalent to material equations for free particles in a rotating frame, unconstrained by pressure forces, namely

$$\frac{d^2 x}{dt^2} - f_0 v = 0, \quad \frac{d^2 y}{dt^2} + f_0 u = 0. \quad (3.111)$$

See also problem 3.9.

3.7.3 Kelvin waves

The Kelvin wave is a particular type of gravity wave that exists in the presence of both rotation and a lateral boundary. Suppose there is a solid boundary at $y = 0$; clearly harmonic solutions in the y -direction are not allowable, as these would not satisfy the condition of no normal flow at the boundary. Do any wave-like solutions exist? The affirmative answer to this question was provided by Kelvin and the associated waves are now eponymously known as *Kelvin waves*.⁵ We begin with the linearized shallow water equations, namely

$$\frac{\partial u'}{\partial t} - f_0 v' = -g \frac{\partial \eta'}{\partial x}, \quad \frac{\partial v'}{\partial t} + f_0 u' = -g \frac{\partial \eta'}{\partial y}, \quad \frac{\partial \eta'}{\partial t} + H \left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} \right) = 0. \quad (3.112a,b,c)$$

The fact that $v' = 0$ at $y = 0$ suggests that we look for a solution with $v' = 0$ everywhere, whence these equations become

$$\frac{\partial u'}{\partial t} = -g \frac{\partial \eta'}{\partial x}, \quad f_0 u' = -g \frac{\partial \eta'}{\partial y}, \quad \frac{\partial \eta'}{\partial t} + H \frac{\partial u'}{\partial x} = 0. \quad (3.113a,b,c)$$

Equations (3.113a) and (3.113c) lead to the standard wave equation

$$\frac{\partial^2 u'}{\partial t^2} = c^2 \frac{\partial^2 u'}{\partial x^2}, \quad (3.114)$$

where $c = \sqrt{gH}$, the usual wave speed of shallow water waves. The solution of (3.114) is

$$u' = F_1(x + ct, y) + F_2(x - ct, y), \quad (3.115)$$

with corresponding surface displacement

$$\eta' = \sqrt{H/g} [-F_1(x + ct, y) + F_2(x - ct, y)]. \quad (3.116)$$

The solution represents the superposition of two waves, one (F_1) travelling in the negative x -direction, and the other in the positive x -direction. To obtain the y dependence of these functions we use (3.113b) which gives

$$\frac{\partial F_1}{\partial y} = \frac{f_0}{\sqrt{gH}} F_1, \quad \frac{\partial F_2}{\partial y} = -\frac{f_0}{\sqrt{gH}} F_2, \quad (3.117)$$

with solutions

$$F_1 = F(x + ct) e^{y/L_d} \quad F_2 = G(x - ct) e^{-y/L_d}, \quad (3.118)$$

where $L_d = \sqrt{gH}/f_0$ is the radius of deformation. The solution F_1 grows exponentially away from the wall, and so fails to satisfy the condition of boundedness at infinity. It thus must be eliminated, leaving the general solution

$$u' = e^{-y/L_d} G(x - ct), \quad v' = 0,$$

$$\eta' = \sqrt{H/g} e^{-y/L_d} G(x - ct)$$

(3.119)

These are Kelvin waves, and they decay exponentially away from the boundary. If f_0 is positive, as in the Northern Hemisphere, the boundary is to the right of an observer moving with the wave. Given a constant Coriolis parameter, we could equally well have obtained a solution on a meridional wall, in which case we would find that the wave again moves such that the wall is to the right of the wave direction. (This is obvious once it is realized that f -plane dynamics are isotropic in x and y .) Thus, in the Northern Hemisphere the wave moves anticlockwise round a basin, and conversely in the Southern Hemisphere, and in both hemispheres the direction is cyclonic.

3.8 GEOSTROPHIC ADJUSTMENT

We noted in chapter 2 that the large-scale, extratropical circulation of the atmosphere is in near-geostrophic balance. Why is this? Why should the Rossby number be small? Arguably, the magnitude of the velocity in the atmosphere and ocean is ultimately given by the strength of the forcing, and so ultimately by the differential heating between pole and equator (although even this argument is not satisfactory, since the forcing mainly determines the energy throughput, not directly the energy itself, and the forcing is itself dependent on the atmosphere's response). But even supposing that the velocity magnitudes are given, there is no a priori guarantee that the forcing or the dynamics will produce length scales that are such that the Rossby number is small. However, there is in fact a powerful and ubiquitous process whereby a fluid in an initially unbalanced state naturally evolves toward a state of geostrophic balance, namely *geostrophic adjustment*. This process occurs quite generally in rotating fluids, whether stratified or not. To pose the problem in a simple form we will consider the free evolution of a single shallow layer of fluid whose initial state is manifestly unbalanced, and we will suppose that surface displacements are small so that the

evolution of the system is described by the linearized shallow equations of motion. These are

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{f} \times \mathbf{u} = -g \nabla \eta, \quad \frac{\partial \eta}{\partial t} + H \nabla \cdot \mathbf{u} = 0, \quad (3.120a,b)$$

where η is the free surface displacement and H is the mean fluid depth, and we omit the primes on the linearized variables.

3.8.1 Non-rotating flow

We consider first the non-rotating problem set, with little loss of generality, in one dimension. We suppose that initially the fluid is at rest but with a simple discontinuity in the height field so that

$$\eta(x, t = 0) = \begin{cases} +\eta_0 & x < 0 \\ -\eta_0 & x > 0 \end{cases} \quad (3.121)$$

and $u(x, t = 0) = 0$ everywhere. We can realize these initial conditions physically by separating two fluid masses of different depths by a thin dividing wall, and then quickly removing the wall. What is the subsequent evolution of the fluid? The general solution to the linear problem is given by (3.97) where the functional form is determined by the initial conditions so that here

$$F(x) = \eta(x, t = 0) = -\eta_0 \operatorname{sgn}(x). \quad (3.122)$$

Equation (3.97) states that this initial pattern is propagated to the right and to the left. That is, two discontinuities in fluid height move to the right and left at a speed $c = \sqrt{gH}$. Specifically, the solution is

$$\eta(x, t) = -\frac{1}{2}\eta_0[\operatorname{sgn}(x + ct) + \operatorname{sgn}(x - ct)]. \quad (3.123)$$

The initial conditions may be much more complex than a simple front, but, because the waves are dispersionless, the solution is still simply a sum of the translation of those initial conditions to the right and to the left at speed c . The velocity field in this class of problem is obtained from

$$\frac{\partial \mathbf{u}}{\partial t} = -g \frac{\partial \eta}{\partial x}, \quad (3.124)$$

which gives, using (3.97),

$$u = -\frac{g}{2c}[F(x + ct) - F(x - ct)]. \quad (3.125)$$

Consider the case with initial conditions given by (3.121). At a given location, away from the initial disturbance, the fluid remains at rest and undisturbed until the front arrives. After the front has passed, the fluid surface is again undisturbed and the velocity is uniform and non zero. Specifically:

$$\eta = \begin{cases} -\eta_0 \operatorname{sgn}(x) & \\ 0 & \end{cases} \quad u = \begin{cases} 0 & |x| > ct \\ (\eta_0 g/c) & |x| < ct. \end{cases} \quad (3.126)$$

The solution with ‘top-hat’ initial conditions in the height field, and zero initial velocity, is a superposition two discontinuities similar to (3.126) and is illustrated in Fig. 3.9. Two fronts propagate in either direction from each discontinuity and, in this case, the final velocity, as well as the fluid displacement, is zero after all the fronts have passed. That is, the disturbance is radiated completely away.

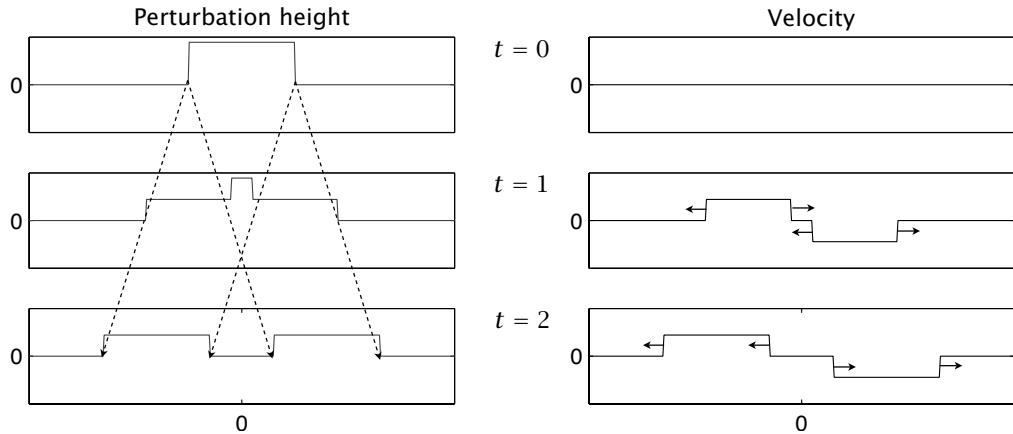


Fig. 3.9 The time development of an initial ‘top hat’ height disturbance, with zero initial velocity, in non-rotating flow. Fronts propagate in both directions, and the velocity is non-zero between fronts, but ultimately the velocity and height disturbance are radiated away to infinity.

3.8.2 Rotating flow

Rotation makes a profound difference to the adjustment problem of the shallow water system, because a steady, adjusted, solution can exist with non-zero gradients in the height field — the associated pressure gradients being balanced by the Coriolis force — and potential vorticity conservation provides a powerful constraint on the fluid evolution.⁶ In a rotating shallow fluid that conservation is represented by

$$\frac{\partial Q}{\partial t} + \mathbf{u} \cdot \nabla Q = 0, \quad (3.127)$$

where $Q = (\zeta + f)/h$. In the linear case with constant Coriolis parameter, (3.127) becomes

$$\frac{\partial q}{\partial t} = 0, \quad q = \left(\zeta - f_0 \frac{\eta}{H} \right). \quad (3.128)$$

This equation may be obtained either from the linearized velocity and mass conservation equations, (3.120), or from (3.127) directly. In the latter case, we write

$$Q = \frac{\zeta + f_0}{H + \eta} \approx \frac{1}{H} (\zeta + f_0) \left(1 - \frac{\eta}{H} \right) \approx \frac{1}{H} \left(f_0 + \zeta - f_0 \frac{\eta}{H} \right) = \frac{f_0}{H} + \frac{q}{H} \quad (3.129)$$

having used $f_0 \gg |\zeta|$ and $H \gg |\eta|$. The term f_0/H is a constant and so dynamically unimportant, as is the H^{-1} factor multiplying q . Further, the advective term $\mathbf{u} \cdot \nabla Q$ becomes $\mathbf{u} \cdot \nabla q$, and this is second order in perturbed quantities and so is neglected. Thus, making these approximations, (3.127) reduces to (3.128). The potential vorticity field is therefore fixed in space! Of course, this was also true in the non-rotating case where the fluid is initially at rest. Then $q = \zeta = 0$ and the fluid remains irrotational throughout the subsequent evolution of the flow. However, this is rather a weak constraint on the subsequent evolution of the fluid; it does nothing, for example, to prevent the conversion

of all the potential energy to kinetic energy. In the rotating case the potential vorticity is non-zero, and potential vorticity conservation and geostrophic balance are all we need to infer the final steady state, assuming it exists, without solving for the details of the flow evolution, as we now see.

With an initial condition for the height field given by (3.121), the initial potential vorticity is given by

$$q(x, y) = \begin{cases} -f_0 \eta_0 / H & x < 0 \\ f_0 \eta_0 / H & x > 0, \end{cases} \quad (3.130)$$

and this remains unchanged throughout the adjustment process. The final steady state is then the solution of the equations

$$\zeta - f_0 \frac{\eta}{H} = q(x, y), \quad f_0 u = -g \frac{\partial \eta}{\partial y}, \quad f_0 v = g \frac{\partial \eta}{\partial x}, \quad (3.131a,b,c)$$

where $\zeta = \partial v / \partial x - \partial u / \partial y$. Because the Coriolis parameter is constant, the velocity field is horizontally non-divergent and we may define a streamfunction $\psi = g\eta / f_0$. Equations (3.131) then reduce to

$$\left(\nabla^2 - \frac{1}{L_d^2} \right) \psi = q(x, y), \quad (3.132)$$

where $L_d = \sqrt{gH} / f_0$ is known as the *Rossby radius of deformation* or often just the ‘deformation radius’ or the ‘Rossby radius’. It is a naturally occurring length scale in problems involving both rotation and gravity, and arises in a slightly different form in stratified fluids.

The initial conditions (3.130) admit of a nice analytic solution, for the flow will remain uniform in y , and (3.132) reduces to

$$\frac{\partial^2 \psi}{\partial x^2} - \frac{1}{L_d^2} \psi = \frac{f_0 \eta_0}{H} \operatorname{sgn}(x). \quad (3.133)$$

We solve this separately for $x > 0$ and $x < 0$ and then match the solutions and their first derivatives at $x = 0$, also imposing the condition that the streamfunction decays to zero as $x \rightarrow \pm\infty$. The solution is

$$\psi = \begin{cases} -(g\eta_0 / f_0)(1 - e^{-x/L_d}) & x > 0 \\ +(g\eta_0 / f_0)(1 - e^{x/L_d}) & x < 0. \end{cases} \quad (3.134)$$

The velocity field associated with this is obtained from (3.131b,c), and is

$$u = 0, \quad v = -\frac{g\eta_0}{f_0 L_d} e^{-|x|/L_d}. \quad (3.135)$$

The velocity is perpendicular to the slope of the free surface, and a jet forms along the initial discontinuity, as illustrated in Fig. 3.10.

The important point of this problem is that the variations in the height and field are not radiated away to infinity, as in the non-rotating problem. Rather, potential vorticity conservation constrains the influence of the adjustment to within a deformation radius (we see now why this name is appropriate) of the initial disturbance. This property is a general one in geostrophic adjustment — it also arises if the initial condition consists of a velocity jump, as considered in problem 3.12.

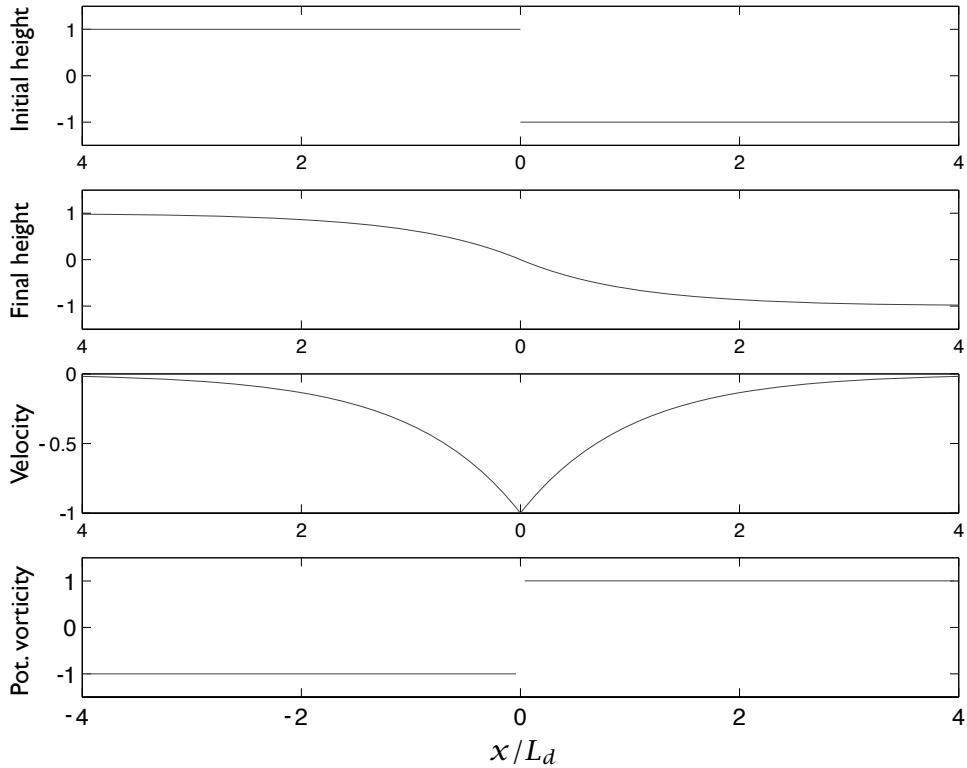


Fig. 3.10 Solutions of a linear geostrophic adjustment problem. Top panel: the initial height field, given by (3.121) with $\eta_0 = 1$. Second panel: equilibrium (final) height field, η given by (3.134) and $\eta = f_0 \psi / g$. Third panel: equilibrium geostrophic velocity (normal to the gradient of height field), given by (3.135). Bottom panel: potential vorticity, given by (3.130), and this does not evolve. The distance, x is non-dimensionalized by the deformation radius $L_d = \sqrt{gH}/f_0$, and the velocity by $\eta_0(g/f_0 L_d)$. Changes to the initial state occur only within $\mathcal{O}(L_d)$ of the initial discontinuity; and as $x \rightarrow \pm\infty$ the initial state is unaltered.

3.8.3 * Energetics of adjustment

How much of the initial potential energy of the flow is lost to infinity by gravity wave radiation, and how much is converted to kinetic energy? The linear equations (3.120) lead to

$$\frac{1}{2} \frac{\partial}{\partial t} (H\mathbf{u}^2 + g\eta^2) + gH \nabla \cdot (\mathbf{u}\eta) = 0, \quad (3.136)$$

so that energy conservation holds in the form

$$E = \frac{1}{2} \int (H\mathbf{u}^2 + g\eta^2) \, dx, \quad \frac{dE}{dt} = 0, \quad (3.137)$$

provided the integral of the divergence term vanishes, as it normally will in a closed domain. The fluid has a non-zero potential energy, $(1/2) \int_{-\infty}^{\infty} g\eta^2 \, dx$, if there are variations in fluid

height, and with the initial conditions (3.121) the initial potential energy is

$$PE_I = \int_0^\infty g\eta_0^2 dx. \quad (3.138)$$

This is nominally infinite if the fluid has no boundaries, and the initial potential energy density is $g\eta_0^2/2$ everywhere.

In the non-rotating case, and with initial conditions (3.121), after the front has passed, the potential energy density is zero and the kinetic energy density is $Hu^2/2 = g\eta_0^2/2$, using (3.126) and $c^2 = gH$. Thus, all the potential energy is locally converted to kinetic energy as the front passes, and eventually the kinetic energy is distributed uniformly along the line. In the case illustrated in Fig. 3.9, the potential energy and kinetic energy are both radiated away from the initial disturbance. (Note that although we can superpose the solutions from different initial conditions, we cannot superpose their potential and kinetic energies.) The general point is that the evolution of the disturbance is not confined to its initial location.

In contrast, in the rotating case the conversion from potential to kinetic energy is *largely confined to within a deformation radius of the initial disturbance*, and at locations far from the initial disturbance the initial state is essentially unaltered. The conservation of potential vorticity has prevented the complete conversion of potential energy to kinetic energy, a result that is not sensitive to the precise form of the initial conditions (see also problem 3.10).

In fact, in the rotating case, some of the initial potential energy is converted to kinetic energy, some remains as potential energy and some is lost to infinity; let us calculate these amounts. The final potential energy, after adjustment, is, using (3.134),

$$PE_F = \frac{1}{2}g\eta_0^2 \left[\int_0^\infty (1 - e^{-x/L_d})^2 dx + \int_{-\infty}^0 (1 - e^{x/L_d})^2 dx \right]. \quad (3.139)$$

This is nominally infinite, but the change in potential energy is finite and is given by

$$PE_I - PE_F = g\eta_0^2 \int_0^\infty (2e^{-x/L_d} - e^{-2x/L_d}) dx = \frac{3}{2}g\eta_0^2 L_d. \quad (3.140)$$

The initial kinetic energy is zero, because the fluid is at rest, and its final value is, using (3.135),

$$KE_F = \frac{1}{2}H \int \mathbf{u}^2 dx = H \left(\frac{g\eta_0}{fL_d} \right)^2 \int_0^\infty e^{-2x/L_d} dx = \frac{g\eta_0^2 L_d}{2}. \quad (3.141)$$

Thus one-third of the difference between the initial and final potential energies is converted to kinetic energy, and this is trapped within a distance of the order of a deformation radius of the disturbance; the remainder, an amount $gL_d\eta_0^2$ is radiated away and lost to infinity. In any finite region surrounding the initial discontinuity the final energy is less than the initial energy.

3.8.4 * General initial conditions

Because of the linearity of the (linear) adjustment problem a spectral viewpoint is useful, in which the fields are represented as the sum or integral of *non-interacting* Fourier modes. For

example, suppose that the height field of the initial disturbance is a two-dimensional field given by

$$\eta(0) = \iint \tilde{\eta}_{k,l}(0) e^{i(kx+ly)} dk dl \quad (3.142)$$

where the Fourier coefficients $\tilde{\eta}_{k,l}(0)$ are given, and the initial velocity field is zero. Then the initial (and final) potential vorticity field is given by

$$q = -\frac{f_0}{H} \iint \tilde{\eta}_{k,l}(0) e^{i(kx+ly)} dk dl. \quad (3.143)$$

To obtain an expression for the final height and velocity fields, we express the potential vorticity field as

$$q = \int \tilde{q}_{k,l} dk dl. \quad (3.144)$$

The potential vorticity field does not evolve, and it is related to the initial height field by

$$\tilde{q}_{k,l} = -\frac{f_0}{H} \eta_{k,l}(0). \quad (3.145)$$

In the final, geostrophically balanced state, the potential vorticity is related to the height field by

$$q = \frac{g}{f_0} \nabla^2 \eta - \frac{f_0}{H} \eta \quad \text{and} \quad \tilde{q}_{k,l} = \left(-\frac{g}{f_0} K^2 - \frac{f_0}{H} \right) \tilde{\eta}_{k,l}, \quad (3.146a,b)$$

where $K^2 = k^2 + l^2$. Using (3.145) and (3.146), the Fourier components of the final height field satisfy

$$\left(-\frac{g}{f_0} K^2 - \frac{f_0}{H} \right) \tilde{\eta}_{k,l} = -\frac{f_0}{H} \tilde{\eta}_{k,l}(0) \quad (3.147)$$

or

$$\tilde{\eta}_{k,l} = \frac{\tilde{\eta}_{k,l}(0)}{K^2 L_d^2 + 1}. \quad (3.148)$$

In physical space the final height field is just the spectral integral of this, namely

$$\eta = \iint \tilde{\eta}_{k,l} e^{i(kx+ly)} dk dl = \iint \frac{\tilde{\eta}_{k,l}(0) e^{i(kx+ly)}}{K^2 L_d^2 + 1} dk dl. \quad (3.149)$$

We see that at large scales ($K^2 L_d^2 \ll 1$) $\eta_{k,l}$ is almost unchanged from its initial state; the velocity field, which is then determined by geostrophic balance, thus adjusts to the pre-existing height field. At large scales most of the energy in geostrophically balanced flow is potential energy; thus, it is energetically easier for the velocity to change to come into balance with the height field than vice versa. At small scales, however, the final height field has much less variability than it did initially.

Conversely, at small scales the height field adjusts to the velocity field. To see this, let us suppose that the initial conditions contain vorticity but have zero height displacement. Specifically, if the initial vorticity is $\nabla^2 \psi(0)$, where $\psi(0)$ is the initial streamfunction, then it is straightforward to show that the final streamfunction is given by

$$\psi = \iint \tilde{\psi}_{k,l} e^{i(kx+ly)} dk dl = \iint \frac{K^2 L_d^2 \tilde{\psi}_{k,l}(0) e^{i(kx+ly)}}{K^2 L_d^2 + 1} dk dl. \quad (3.150)$$

The final height field is then obtained from this, via geostrophic balance, by $\eta = (f_0/g)\psi$. Evidently, for small scales ($K^2 L_d^2 \gg 1$) the streamfunction, and hence the vertical component of the velocity field, are almost unaltered from their initial values. On the other hand, at large scales the final streamfunction has much less variability than it does initially, and so the height field will be largely governed by whatever variation it (and not the velocity field) had initially. In general, the final state is a superposition of the states given by (3.149) and (3.150). The divergent component of the initial velocity field does not affect the final state because it has no potential vorticity, and so all of the associated energy is eventually lost to infinity.

Finally, we remark that just as in the problem with a discontinuous initial height profile, the change in total energy during adjustment is negative — this can be seen from the form of the integrals above, although we leave the specifics as a problem to the reader. That is, some of the initial potential and kinetic energy is lost to infinity, but some is trapped by the potential vorticity constraint.

3.8.5 A variational perspective

In the non-rotating problem, all of the initial potential energy is eventually radiated away to infinity. In the rotating problem, the final state contains both potential and kinetic energy. Why is the energy not all radiated away to infinity? It is because potential vorticity conservation on parcels prevents all of the energy being dispersed. This suggests that it may be informative to think of the geostrophic adjustment problem as a *variational problem*: we seek to minimize the energy consistent with the conservation of potential vorticity. We stay in the linear approximation in which, because the advection of potential vorticity is neglected, potential vorticity remains constant at each point.

The energy of the flow is given by the sum of potential and kinetic energies, namely

$$\text{energy} = \int H(\mathbf{u}^2 + g\eta^2) \, dA, \quad (3.151)$$

(where $dA \equiv dx \, dy$) and the potential vorticity field is

$$q = \zeta - f_0 \frac{\eta}{H} = (v_x - u_y) - f_0 \frac{\eta}{H}, \quad (3.152)$$

where the subscripts x, y denote derivatives. The problem is then to extremize the energy subject to potential vorticity conservation. This is a constrained problem in the calculus of variations, sometimes called an *isoperimetric* problem because of its origins in maximizing the area of a surface for a given perimeter.⁷ The mathematical problem is to extremize the integral

$$I = \int \left\{ H(u^2 + v^2) + g\eta^2 + \lambda(x, y)[(v_x - u_y) - f_0\eta/H] \right\} \, dA, \quad (3.153)$$

where $\lambda(x, y)$ is a Lagrange multiplier, undetermined at this stage. It is a function of space: if it were a constant, the integral would merely extremize energy subject to a given integral of potential vorticity, and rearrangements of potential vorticity (which here we wish to disallow) would leave the integral unaltered.

As there are three independent variables there are three Euler-Lagrange equations that

must be solved in order to minimize I . These are

$$\begin{aligned}\frac{\partial L}{\partial \eta} - \frac{\partial}{\partial x} \frac{\partial L}{\partial \eta_x} - \frac{\partial}{\partial y} \frac{\partial L}{\partial \eta_y} &= 0, \\ \frac{\partial L}{\partial u} - \frac{\partial}{\partial x} \frac{\partial L}{\partial u_x} - \frac{\partial}{\partial y} \frac{\partial L}{\partial u_y} &= 0, \quad \frac{\partial L}{\partial v} - \frac{\partial}{\partial x} \frac{\partial L}{\partial v_x} - \frac{\partial}{\partial y} \frac{\partial L}{\partial v_y} = 0,\end{aligned}\quad (3.154)$$

where L is the integrand on the right-hand side of (3.153). Substituting the expression for L into (3.154) gives, after a little algebra,

$$2g\eta - \lambda f_0 = 0, \quad 2u + \frac{\partial \lambda}{\partial y} = 0, \quad 2v - \frac{\partial \lambda}{\partial x} = 0, \quad (3.155)$$

and then eliminating λ gives the simple relationships

$$u = -\frac{g}{f_0} \frac{\partial \eta}{\partial y}, \quad v = \frac{g}{f_0} \frac{\partial \eta}{\partial x}, \quad (3.156)$$

which are the equations of geostrophic balance. Thus, in the linear approximation, *geostrophic balance is the minimum energy state for a given field of potential vorticity*.

3.9 ISENTROPIC COORDINATES

We now return to the continuously stratified primitive equations, and consider the use of potential density as a vertical coordinate. In practice this means using potential temperature in the atmosphere and (for simple equations of state) buoyancy in the ocean; such coordinate systems are generically called *isentropic coordinates*, and sometimes *isopycnal coordinates* if density is used. This may seem an odd thing to do but for adiabatic flow the resulting equations of motion have an attractive form that aids the interpretation of large-scale flow. The thermodynamic equation becomes a statement for the conservation of the mass of fluid with a given value of potential density and, because the flow of both the atmosphere and the ocean is largely along isentropic surfaces, the momentum and vorticity equations have a quasi-two-dimensional form.

The particular choice of vertical coordinate is determined by the form of the thermodynamic equation in the equation-set at hand; thus, if the thermodynamic equation is $D\theta/Dt = \dot{\theta}$, we transform the equations from (x, y, z) coordinates to (x, y, θ) coordinates. The material derivative in this coordinate system is

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u \left(\frac{\partial}{\partial x} \right)_\theta + v \left(\frac{\partial}{\partial y} \right)_\theta + \frac{D\theta}{Dt} \frac{\partial}{\partial \theta} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_\theta + \dot{\theta} \frac{\partial}{\partial \theta}, \quad (3.157)$$

where the last term on the right-hand side is zero for adiabatic flow.

3.9.1 A hydrostatic Boussinesq fluid

In the simple Boussinesq equations (see the table on page 72) the buoyancy is the relevant thermodynamic variable. With hydrostatic balance the horizontal and vertical momentum equations are, in height coordinates,

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -\nabla \phi, \quad b = \frac{\partial \phi}{\partial z}, \quad (3.158)$$

where b is the buoyancy, the variable analogous to the potential temperature θ of an ideal gas. The thermodynamic equation is

$$\frac{Db}{Dt} = \dot{b}, \quad (3.159)$$

and because $b = -g\delta\rho/\rho_0$, isentropic coordinates are the same as isopycnal coordinates.

Using (2.142) the horizontal pressure gradient may be transformed to isentropic coordinates

$$\left(\frac{\partial\phi}{\partial x}\right)_z = \left(\frac{\partial\phi}{\partial x}\right)_b - \left(\frac{\partial z}{\partial x}\right)_b \frac{\partial\phi}{\partial z} = \left(\frac{\partial\phi}{\partial x}\right)_b - b \left(\frac{\partial z}{\partial x}\right)_b = \left(\frac{\partial M}{\partial x}\right)_b, \quad (3.160)$$

where

$$M \equiv \phi - zb. \quad (3.161)$$

Thus, the horizontal momentum equation becomes

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -\nabla_b M. \quad (3.162)$$

where the material derivative is given by (3.157), with b replacing θ . Using (3.161) the hydrostatic equation becomes

$$\frac{\partial M}{\partial b} = -z. \quad (3.163)$$

The mass continuity equation may be derived by noting that for a Boussinesq fluid the mass element may be written as

$$\delta m = \rho_0 \frac{\partial z}{\partial b} \delta b \delta x \delta y. \quad (3.164)$$

The mass continuity equation, $D\delta m/Dt = 0$, becomes

$$\frac{D}{Dt} \frac{\partial z}{\partial b} + \frac{\partial z}{\partial b} \nabla_3 \cdot \mathbf{v} = 0, \quad (3.165)$$

where $\nabla_3 \cdot \mathbf{v} = \nabla_b \cdot \mathbf{u} + \partial \dot{b} / \partial b$ is the three-dimensional derivative of the velocity in isentropic coordinates. Equation (3.165) may thus be written

$$\frac{D\sigma}{Dt} + \sigma \nabla_b \cdot \mathbf{u} = -\sigma \frac{\partial \dot{b}}{\partial b}, \quad (3.166)$$

where $\sigma \equiv \partial z / \partial b$ is a measure of the thickness between two isentropic surfaces and the material derivative is given by (3.157) with θ replaced by b . Equations (3.162), (3.163) and (3.166) comprise a closed set, with dependent variables \mathbf{u} , M and z in the space of independent variables x , y and b .

3.9.2 A hydrostatic ideal gas

Deriving the equations of motion for this system requires a little more work than in the Boussinesq case but the idea is the same. For an ideal gas in hydrostatic balance we have, using (1.112),

$$\frac{\delta\theta}{\theta} = \frac{\delta T}{T} - \kappa \frac{\delta p}{p} = \frac{\delta T}{T} + \frac{\delta\Phi}{c_p T} = \frac{1}{c_p T} \delta M, \quad (3.167)$$

where $M \equiv c_p T + \Phi$ is the ‘Montgomery potential’, equal to the dry static energy. (We use some of the same symbols as in the Boussinesq case to facilitate comparison, but their meanings are slightly different.) From this

$$\frac{\partial M}{\partial \theta} = \Pi, \quad (3.168)$$

where $\Pi \equiv c_p T / \theta = c_p (p / p_R)^{R/c_p}$ is the ‘Exner function’. Equation (3.168) represents the hydrostatic relation in isentropic coordinates. Note also that $M = \theta \Pi + \Phi$.

To obtain an appropriate form for the horizontal pressure gradient force first note that, in the usual height coordinates, it is given by

$$\frac{1}{\rho} \nabla_z p = \theta \nabla_z \Pi, \quad (3.169)$$

where $\Pi = c_p T / \theta$. Using (2.142) gives

$$\theta \nabla_z \Pi = \theta \nabla_\theta \Pi - \frac{\theta}{g} \frac{\partial \Pi}{\partial z} \nabla_\theta \Phi. \quad (3.170)$$

Then, using the definition of Π and the hydrostatic approximation to help evaluate the vertical derivative, we obtain

$$\frac{1}{\rho} \nabla_z p = c_p \nabla_\theta T + \nabla_\theta \Phi = \nabla_\theta M. \quad (3.171)$$

Thus, the horizontal momentum equation is

$$\frac{D\mathbf{v}}{Dt} + \mathbf{f} \times \mathbf{u} = -\nabla_\theta M. \quad (3.172)$$

Much as in the Boussinesq case, the mass continuity equation may be derived by noting that the mass element may be written as

$$\delta m = -\frac{1}{g} \frac{\partial p}{\partial \theta} \delta \theta \delta x \delta y. \quad (3.173)$$

The mass continuity equation, $D\delta m/Dt = 0$, becomes

$$\frac{D}{Dt} \frac{\partial p}{\partial \theta} + \frac{\partial p}{\partial \theta} \nabla_3 \cdot \mathbf{v} = 0 \quad (3.174)$$

or

$$\frac{D\sigma}{Dt} + \sigma \nabla_\theta \cdot \mathbf{u} = -\sigma \frac{\partial \dot{\theta}}{\partial \theta}, \quad (3.175)$$

where now $\sigma \equiv \partial p / \partial \theta$ is a measure of the (pressure) thickness between two isentropic surfaces. Equations (3.168), (3.172) and (3.175) form a closed set, analogous to (3.163), (3.162) and (3.166).

3.9.3 * Analogy to shallow water equations

The equations of motion in isentropic coordinates have an obvious analogy with the shallow water equations, and we may think of the shallow water equations as a finite-difference

representation of the primitive equations written in isentropic coordinates, or think of the latter as the continuous limit of the shallow water equations as the number of layers increases. For example, consider a two-isentropic-level representation of (3.168), (3.172) and (3.175), in which the lower boundary is an isentrope. A natural finite differencing gives

$$-M_1 = \Pi_0 \Delta \theta_0 \quad (3.176a)$$

$$M_1 - M_2 = \Pi_1 \Delta \theta_1, \quad (3.176b)$$

where the $\Delta \theta$ s are constants, and the momentum equations for each layer become

$$\frac{D\mathbf{u}_1}{Dt} + \mathbf{f} \times \mathbf{u}_1 = -\Delta \theta_0 \nabla \Pi_0 \quad (3.177a)$$

$$\frac{D\mathbf{u}_2}{Dt} + \mathbf{f} \times \mathbf{u}_2 = -\Delta \theta_0 \nabla \Pi_0 - \Delta \theta_1 \nabla \Pi_1. \quad (3.177b)$$

Together with the mass continuity equation for each level these are just like the two-layer shallow water equations (3.51). This means that results that one might easily derive for the shallow water equations will often have a continuous analog.

3.10 AVAILABLE POTENTIAL ENERGY

We now revisit the issue of the internal and potential energy in stratified flow, motivated by the following remarks. In adiabatic, inviscid flow the total amount of energy is conserved, and there are conversions between internal energy, potential energy and kinetic energy. In an ideal gas the potential energy and the internal energy of a column extending throughout the atmosphere are in a constant ratio to each other — their sum is called the total potential energy. In a simple Boussinesq fluid, energetic conversions involve only the potential and kinetic energy, and not the internal energy. Yet, plainly, in neither a Boussinesq fluid nor an ideal gas can *all* the total potential energy in a fluid be converted to kinetic energy, for then all of the fluid would be adjacent to the ground and the fluid would have no thickness, which intuitively seems impossible. Given a state of the atmosphere or ocean, how much of its total potential energy is available for conversion to kinetic energy? In particular, because total energy is conserved only in adiabatic flow, we may usefully ask: how much potential energy is available for conversion to kinetic energy under an adiabatic rearrangement of fluid parcels?

Suppose that at any given time the flow is stably stratified, but that the isentropes (or more generally the surfaces of constant potential density) are sloping, as in Fig. 3.11. The potential energy of the system would be reduced if the isentropes were flattened, for then heavier fluid would be moved to lower altitudes, with lighter fluid replacing it at higher altitudes. In an adiabatic rearrangement the amount of fluid between the isentropes would remain constant, and a state with flat isentropes (meaning parallel to the geopotential surfaces) evidently constitutes a state of minimum total potential energy. The difference between the total potential energy of the fluid and the total potential energy after an adiabatic rearrangement to a state in which the isentropic surfaces are flat is called the *available potential energy*, or APE.⁸

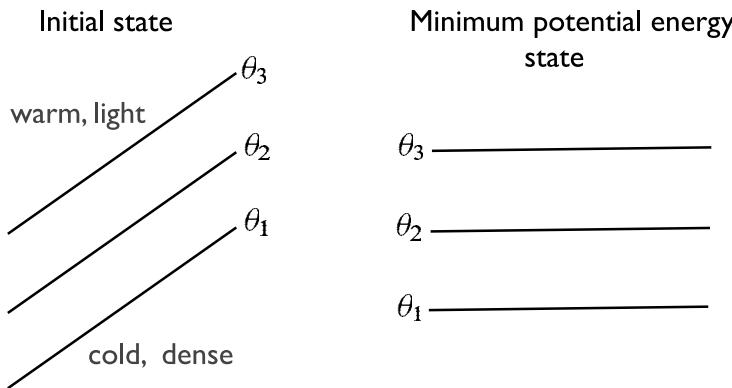


Fig. 3.11 If a stably stratified initial state with sloping isentropes (left) is adiabatically rearranged then the state of minimum potential energy has flat isentropes, as on the right, but the amount of fluid contained between each isentropic surface is unchanged. The difference between the potential energies of the two states is the *available potential energy*.

3.10.1 A Boussinesq fluid

The potential energy of a column of a Boussinesq fluid of unit area is given by

$$P = - \int_0^H b z \, dz = - \int_0^H \frac{b}{2} \, dz^2. \quad (3.178)$$

and the potential energy of the entire fluid is given by the horizontal integral of this. The minimum potential energy of the fluid arises after an adiabatic rearrangement in which the isopycnals are flattened, and the resulting buoyancy is only a function of z . The available potential energy is then the difference between the energy of the initial state and of this minimum state, and to obtain an approximate expression for this we first integrate (3.178) by parts to give

$$P = \frac{1}{2} \int_0^{b_m} z^2 \, db - \left[\frac{b z^2}{2} \right]_0^H = \frac{1}{2} \int_0^{b_m} z^2 \, db - \frac{b_m H^2}{2}, \quad (3.179)$$

where b_m is the maximum value of b in the domain, and we may formally take the upper boundary to have this value of b without affecting the final result. The minimum potential energy state arises when z is a function only of b , $z = Z(b)$ say. Because mass is conserved in the rearrangement, Z is equal to the horizontally averaged value of z on a given isopycnal surface, \bar{z} , and the surfaces \bar{z} and \bar{b} thus define each other completely. The average available potential energy, per unit area, is then given by

$$APE = \frac{1}{2} \int_0^{b_m} (\bar{z}^2 - \bar{z}'^2) \, db = \frac{1}{2} \int_0^{b_m} \bar{z}'^2 \, db, \quad (3.180)$$

where $z = \bar{z} + z'$; that is, z' is the height variation of an isopycnal surface, and the last term on the right-hand side of (3.179) has cancelled with an identical term in the expression for the potential energy of the re-arranged state. The available potential energy is thus

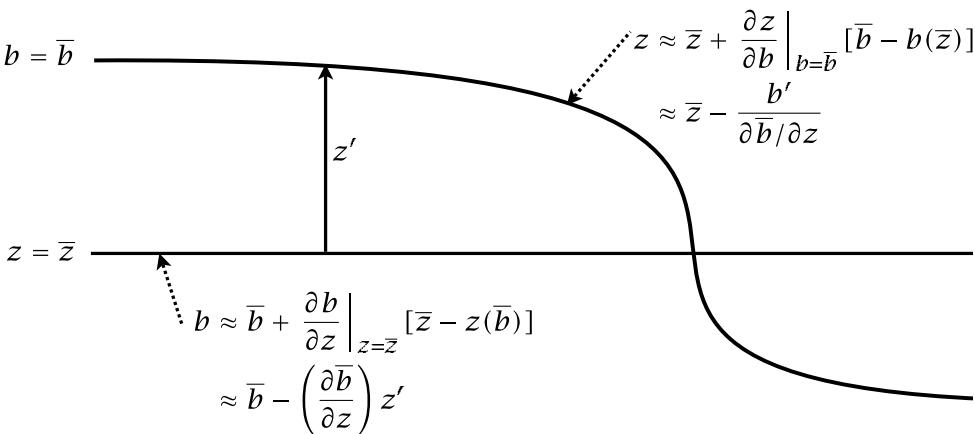


Fig. 3.12 An isopycnal surface, $b = \bar{b}$, and the constant height surface, $z = \bar{z}$. \bar{z} is the height of the isopycnal surface after a rearrangement to a minimum potential energy state, equal to the average height of the isopycnal surface. The values of z on the isopycnal surface, and of b on the constant height surface, can be obtained by the Taylor expansions shown. For an ideal gas in pressure coordinates, replace z by p and b by θ .

proportional to the integral of the variance of the altitude of such a surface, and it is a positive-definite quantity. To obtain an expression in z -coordinates, we express the height variations on an isopycnal surface in terms of buoyancy variations on a surface of constant height by Taylor-expanding the height about its value on the isopycnal surface. Referring to Fig. 3.12 this gives

$$z(\bar{b}) = \bar{z} + \frac{\partial z}{\partial b} \Big|_{b=\bar{b}} [\bar{b} - b(\bar{z})] = \bar{z} - \frac{\partial z}{\partial b} \Big|_{b=\bar{b}} b', \quad (3.181)$$

where $b' = b(\bar{z}) - \bar{b}$ is corresponding buoyancy perturbation on the \bar{z} surface and \bar{b} is the average value of b on the \bar{z} surface. Furthermore, $\partial z / \partial b|_{z=\bar{z}} \approx \partial \bar{z} / \partial b \approx (\partial \bar{b} / \partial z)^{-1}$, and (3.181) thus becomes

$$z' = z(\bar{b}) - \bar{z} \approx -b' \left(\frac{\partial \bar{z}}{\partial b} \right) \approx -\frac{b'}{(\partial \bar{b} / \partial z)}, \quad (3.182)$$

where $z' = z(b) - \bar{z}$ is the height perturbation of the isopycnal surface, from its average value. Using (3.182) in (3.180) we obtain an expression for the APE per unit area, to wit

$$APE \approx \frac{1}{2} \int_0^H \frac{\bar{b}'^2}{\partial \bar{b} / \partial z} dz.$$

(3.183)

The total APE of the fluid is the horizontal integral of the above, and so is proportional to the variance of the buoyancy on a height surface. We emphasize that APE is not defined for a single column of fluid, for it depends on the variations of buoyancy over a horizontal surface. Note too that the derivation neglects the effects of topography; this, and the use of a basic-state stratification, effectively restrict the use of (3.183) to a single ocean basin, and even for that the approximations used limit the accuracy of the expressions.

3.10.2 An ideal gas

The expression for the APE for an ideal gas is obtained, *mutatis mutandis*, in the same way as for a Boussinesq fluid and the trusting reader may skip directly to (3.191). The internal energy of an ideal gas column of unit area is given by

$$I = \int_0^{\infty} c_v T \rho \, dz = \int_0^{p_s} \frac{c_v}{g} T \, dp, \quad (3.184)$$

where p_s is the surface pressure, and the corresponding potential energy is given by

$$P = \int_0^{\infty} \rho g z \, dz = \int_0^{p_s} z \, dp = \int_0^{\infty} p \, dz = \int_0^{p_s} \frac{R}{g} T \, dp. \quad (3.185)$$

In (3.184) we use hydrostasy, and in (3.185) the equalities make successive use of hydrostasy, an integration by parts, hydrostasy and the ideal gas relation. Thus, the total potential energy (TPE) is given by

$$TPE \equiv I + P = \frac{c_p}{g} \int_0^{p_s} T \, dp. \quad (3.186)$$

Using the ideal gas equation of state we can write this as

$$TPE = \frac{c_p}{g} \int_0^{p_s} \left(\frac{p}{p_s} \right)^\kappa \theta \, dp = \frac{c_p p_s}{g(1 + \kappa)} \int_0^{\infty} \left(\frac{p}{p_s} \right)^{\kappa+1} \, d\theta, \quad (3.187)$$

after an integration by parts. (We omit a term proportional to $p_s \theta_s$ that arises in the integration by parts, because it cancels in a similar fashion to the boundary term in the Boussinesq derivation; or take $\theta_s = 0$.) The total potential energy of the entire fluid is equal to a horizontal integral of (3.187). The minimum total potential energy arises when the pressure in (3.187) is a function only of θ , $p = P(\theta)$, where by conservation of mass P is the average value of the original pressure on the isentropic surface, $P = \bar{p}$. The average available potential energy per unit area is then given by the difference between the initial state and this minimum, namely

$$APE = \frac{c_p p_s}{g(1 + \kappa)} \int_0^{\infty} \left[\overline{\left(\frac{p}{p_s} \right)^{\kappa+1}} - \left(\frac{\bar{p}}{p_s} \right)^{\kappa+1} \right] \, d\theta, \quad (3.188)$$

which is a positive-definite quantity. A useful approximation to this is obtained by expressing the right-hand side in terms of the variance of the potential temperature on a pressure surface. We first use the binomial expansion to expand $p^{\kappa+1} = (\bar{p} + p')^{\kappa+1}$. Neglecting third- and higher-order terms (3.188) becomes

$$APE = \frac{R \bar{p}_s}{2g} \int_0^{\infty} \left(\frac{\bar{p}}{\bar{p}_s} \right)^{\kappa+1} \overline{\left(\frac{p'}{\bar{p}} \right)^2} \, d\theta. \quad (3.189)$$

The variable $p' = p(\theta) - \bar{p}$ is a pressure perturbation on an isentropic surface, and is related to the potential temperature perturbation on an isobaric surface by [cf. (3.182)]

$$p' \approx -\theta' \frac{\partial \bar{p}}{\partial \theta} \approx -\frac{\theta'}{\partial \bar{\theta} / \partial p}, \quad (3.190)$$

where $\theta' = \theta(p) - \theta(\bar{p})$ is the potential temperature perturbation on the \bar{p} surface. Using (3.190) in (3.189) we finally obtain

$$APE = \frac{R\bar{p}_s^{-\kappa}}{2} \int_0^{p_s} p^{\kappa-1} \left(-g \frac{\partial \bar{\theta}}{\partial p} \right)^{-1} \bar{\theta}^2 dp. \quad (3.191)$$

The APE is thus proportional to the variance of the potential temperature on the pressure surface or, from (3.189), proportional to the variance of the pressure on an isentropic surface.

3.10.3 Use, interpretation, and the atmosphere and ocean

The potential energy of a fluid is reduced when the dynamics acts to flatten the isentropes. Consider, for example, the Earth's atmosphere, with isentropes sloping upwards toward the pole (Fig. 3.11 with the pole on the right). Flattening these isentropes amounts to a sinking of dense air and a rising of light air, and this reduction of potential energy leads to a corresponding production of kinetic energy. Thus, if the dynamics is such as to reduce the temperature gradient between equator and pole by flattening the isentropes then APE is converted to KE by that process. A statistically steady state is achieved because the heating from the Sun continually acts to restore the horizontal temperature gradient between equator and pole, thus replenishing the pool of APE, and to this extent the large-scale atmospheric circulation acts like a heat engine.

It is a useful exercise to calculate the total potential energy, the available potential energy and the kinetic energy of atmosphere and the ocean, (problem 3.16). One finds

$$TPE \gg APE > KE \quad (3.192)$$

with, very approximately, $TPE \sim 100 APE$ and $APE \sim 10 KE$. The first inequality should not surprise us (as it was this that led us to define APE in the first instance), but the second inequality is not obvious (and in fact the ratio is larger in the ocean). It is related to the fact that the instabilities of the atmosphere and ocean occur at a scale smaller than the size of the domain, and are unable to release all the potential energy that might be available. Understanding this more fully is the topic of chapters 6 and 9.

Notes

- 1 The algorithm to solve these equations numerically differs from that of the free-surface shallow water equations because the mass conservation equation can no longer be stepped forward in time. Rather, an elliptic equation for p_{lid} must be derived by eliminating time derivatives between (3.22) using (3.21), and this is then solved at each timestep.
- 2 After Margules (1903). Margules sought to relate the energy of fronts to their slope. In this same paper the notion of available potential energy arose.
- 3 The expression 'form drag' is also commonly used in aerodynamics, and the two usages are related. In aerodynamics, form drag is the force due to the pressure difference between the front and the rear of an object, or any other 'form', moving through a fluid. Aerodynamic

form drag may include frictional effects between the wind and the surface itself, but these are omitted in most geophysical uses.

- 4 (Jules) Henri Poincaré (1854–1912) was a prodigious French mathematician, physicist and philosopher, regarded as one of the greatest mathematicians living at the turn of the twentieth century. He is remembered for his original work in algebra and topology, and in dynamical systems and celestial mechanics, obtaining many results in what would be called nonlinear dynamics and chaos when these fields re-emerged some 60 years later — the notion of ‘sensitive dependence on initial conditions’, for example, is present in his work. He also obtained a number of the results of special relativity independently of Einstein, and worked on the theory of rotating fluids — hence the Poincaré waves of this chapter. He wrote extensively and successfully for the general public on the meaning, importance and philosophy of science. Among other things he discussed whether scientific knowledge was an arbitrary convention, a notion that remains discussed and controversial to this day. (His answer: ‘convention’, in part, yes; ‘arbitrary’, no.) He was a proponent of the role of intuition in mathematical and scientific progress, and did not believe that mathematics could ever be wholly reduced to pure logic.
- 5 Thomson (1869). W. Thompson later became Lord Kelvin. See Gill (1982) or Philander (1990) for more discussion.
- 6 As considered by Rossby (1938).
- 7 An introduction to variational problems may be found in Weinstock (1952) and a number of other textbooks. Applications to many traditional problems in mechanics are discussed by Lanczos (1970).
- 8 Margules (1903) introduced the concept of potential energy that is available for conversion to kinetic energy, Lorenz (1955) clarified its meaning and derived useful, approximate formulae for its computation. Shepherd (1993) showed that the APE is just the non-kinetic part of the pseudoenergy, an interpretation that naturally leads to a number of extensions of the concept. There are a host of other papers on the subject, including that of Huang (1998) who looked at some of the limitations of the approximate expressions in an oceanic context.

Further reading

Gill, A. E., 1982. *Atmosphere–Ocean Dynamics*.

This remains a good reference for geostrophic adjustment and gravity waves. The time-dependent geostrophic adjustment problem is discussed in section 7.3.

Problems

- 3.1 Derive the appropriate shallow water equations for a single moving layer of fluid of density ρ_1 above a rigid floor, and where above the moving fluid is a stationary fluid of density ρ_0 , where $\rho_0 < \rho_1$. Show that as $(\rho_0/\rho_1) \rightarrow 0$ the usual shallow water equations emerge.
- 3.2 (a) Model the atmosphere as two immiscible, ‘shallow water’ fluids of different density stacked one above the other. Using reasonable values for any required physical parameters, estimate the displacement of the interfacial surface associated with a pole-equator temperature gradient of 60 K.
(b) Similarly estimate an interfacial displacement in the ocean associated with a temperature gradient of 20 K over a distance of 4000 km. (This is a crude representation of the main oceanic thermocline.)
- 3.3 ♦ For a shallow water fluid the energy equation, (3.88), has the form $\partial E/\partial t + \nabla \cdot [\mathbf{v}(E + gh^2/2)] = 0$. But for a compressible fluid, the corresponding energy equation, (1.187), has

the form $\partial E/\partial t + \nabla \cdot [\mathbf{v}(E + p)] = 0$. In a shallow water fluid, $p = gh^2/2$ at a point so these equations are superficially different. Explain this and reconcile the two forms. (Hint: What is the average pressure in a fluid column?)

- 3.4 ♦ Can the shallow water equations for an incompressible fluid be derived by way of an asymptotic expansion in the aspect ratio? If so, do it. That is, without assuming hydrostatics ab initio, expand the Euler equations with a free surface in a small parameter equal to the ratio of the depth of the fluid to the horizontal scale of the motion, and so obtain the shallow water equations.
- 3.5 The inviscid shallow water equations, whether rotating or not, can support gravity waves of arbitrarily short wavelengths. For sufficiently high wavenumber, the wavelength will be shorter than the depth of the fluid. Is this consistent with an asymptotic nature of the shallow water equations? Discuss.

- 3.6 Show that the vertical velocity within a shallow water system is given by

$$w = \frac{z - \eta_b}{h} \frac{Dh}{Dt} + \frac{D\eta_b}{Dt}. \quad (\text{P3.1})$$

Interpret this result, showing that it gives sensible answers at the top and bottom of the fluid layer.

- 3.7 What is the appropriate generalization of (3.97) to two dimensions? Suppose that at time $t = 0$ the height field is given by a Gaussian distribution $h' = A e^{-r^2/\sigma^2}$, where $r^2 = x^2 + y^2$. What is the subsequent evolution of this, in the linear approximation? Show that the distribution remains Gaussian, and that its width increases at speed \sqrt{gH} , where H is the mean depth of the fluid.
- 3.8 In an adiabatic shallow water fluid in a rotating reference frame show that the potential vorticity conservation law is

$$\frac{D}{Dt} \frac{\zeta + f}{\eta - h_b} = 0 \quad (\text{P3.2})$$

where η is the height of the free surface and h_b is the height of the bottom topography, both referenced to the same flat surface.

- (a) A cylindrical column of air at 30° latitude with radius 100 km expands horizontally to twice its original radius. If the air is initially at rest, what is the mean tangential velocity at the perimeter after the expansion?
- (b) An air column at 60°N with zero relative vorticity ($\zeta = 0$) stretches from the surface to the tropopause, which we assume is a rigid lid, at 10 km. The air column moves zonally on to a plateau 2.5 km high. What is its relative vorticity? Suppose it then moves southwards to 30°N . What is its vorticity? (Assume that the density is constant.)
- 3.9 ♦ In the long-wave limit of Poincaré waves, fluid parcels behave as free-agents; that is, like free solid particles moving in a rotating frame unencumbered by pressure forces. Why then, is their frequency given by $\omega = f = 2\Omega$ where Ω is the rotation rate of the coordinate system, and not by Ω itself? Do particles that are stationary or move in a straight line in the inertial frame of reference satisfy the dispersion relationship for Poincaré waves in this limit? Explain. [See also Durran (1993), Egger (1999) and Phillips (2000).]
- 3.10 Linearize the f -plane shallow water system about a state of rest. Suppose that there is an initial disturbance given in the general form

$$\eta = \iint \tilde{\eta}_{k,l} e^{i(kx+ly)} dk dl, \quad (\text{P3.3})$$

where η is the deviation surface height and the Fourier coefficients $\tilde{\eta}_{k,l}$ are given, and that the initial velocity is zero.

- (a) Obtain the geopotential field at the completion of geostrophic adjustment, and show that the deformation scale is a natural length scale in the problem.
 (b) Show that the change in total energy during the adjustment is always less than or equal to zero. Neglect any initial divergence.
 N.B. Because the problem is linear, the Fourier modes do not interact.

3.11 One way of assimilating observations into numerical models is via a method known as ‘nudging’ or ‘robust diagnostics’. In this method, a relaxation term is added to the right-hand side of an evolution equation, relaxing the field back to observations. Thus, for example, $D\phi/Dt = \text{other terms} + \lambda(\phi_{\text{obs}} - \phi)$. If ϕ is potential temperature, is this method likely to work better in the ocean or in the atmosphere, as regards reproducing the large scale mean potential temperature and velocity fields? What steps might be taken to improve the method?

3.12 *Geostrophic adjustment of a velocity jump*

Consider the evolution of the linearized f -plane shallow water equations in an infinite domain. Suppose that initially the fluid surface is flat, the zonal velocity is zero and the meridional velocity is given by

$$v(x) = v_0 \operatorname{sgn}(x) \quad (\text{P3.4})$$

(a) Find the equilibrium height and velocity fields at $t = \infty$.

(b) What are the initial and final kinetic and potential energies?

Partial Solution:

The potential vorticity is $q = \zeta - f_0 \eta / H$, so that the initial and final state are

$$q = 2v_0 \delta(x). \quad (\text{P3.5})$$

(Why?) The final state streamfunction is thus given by $(\partial^2/\partial x^2 - L_d^{-2})\psi = q$, with solution $\psi = \psi_0 \exp(x/L_d)$ and $\psi = \psi_0 \exp(-x/L_d)$ for $x < 0$ and $x > 0$, where $\psi_0 = L_d v_0$ (why?), and $\eta = f_0 \psi / g$. The energy is $E = \int (Hv^2 + g\eta^2) / 2 \, dx$. The initial KE is infinite, the initial PE is zero, and the final state has $PE = KE = gL_d \eta_0^2 / 4$; that is, the energy is equipartitioned between kinetic and potential energies.

- 3.13 In the shallow water equations show that, if the flow is approximately geostrophically balanced, the energy at large scales is predominantly potential energy and that energy at small scales is predominantly kinetic energy. Define precisely what ‘large scale’ and ‘small scale’ mean in this context.
- 3.14 In the shallow water geostrophic adjustment problem, show that at large scales the velocity adjusts to the height field, and that at small scales the height field adjusts to the velocity field.
- 3.15 ♦ Consider the problem of minimizing the full energy [i.e., $\int (hu^2) + g\eta^2 \, dx$], given the potential vorticity field $q(x, y) = (\zeta + f)/h$. Show that the balance relations analogous to (3.8.5) are $uh = -\partial(Bq^{-1})/\partial y$ and $vh = \partial(Bq^{-1})/\partial x$ where B is the Bernoulli function $B = gn + \mathbf{u}^2/2$. Show that steady flow does not necessarily satisfy these equations. Discuss.
- 3.16 Using realistic values for temperature, velocity, etc., calculate *approximate* values for the total potential energy, the available potential energy and the kinetic energy, of either a hemisphere in the atmosphere or an ocean basin.

All real fluid motions are rotational.
Clifford Truesdell, *The Kinematics of Vorticity*, 1954.

CHAPTER FOUR

Vorticity and Potential Vorticity

VORTICITY AND POTENTIAL VORTICITY both play a central role in geophysical fluid dynamics — indeed, we shall find that the large-scale circulations of both the ocean and atmosphere are in large part governed by the evolution of the potential vorticity. In this chapter we define and discuss these quantities and deduce some of their dynamical properties. Along the way we will come across *Kelvin's circulation theorem*, one of the most fundamental conservation laws in all of fluid mechanics, and we will find that the conservation of potential vorticity is intimately tied to this.

4.1 VORTICITY AND CIRCULATION

4.1.1 Preliminaries

Vorticity, ω , is defined to be the curl of velocity and so is given by

$$\omega \equiv \nabla \times \mathbf{v}. \quad (4.1)$$

Circulation, C , is defined to be the integral of velocity around a closed fluid loop and so is given by

$$C \equiv \oint \mathbf{v} \cdot d\mathbf{r} = \int_S \omega \cdot dS, \quad (4.2)$$

where the second expression uses Stokes' theorem and S is any surface bounded by the loop. The circulation around the path is equal to the integral of the normal component of vorticity over *any* surface bounded by that path. The circulation is not a field like vorticity and velocity; rather, we think of the circulation around a particular material line of finite length, and so its value generally depends on the path chosen. If δS is an infinitesimal

surface element whose normal points in the direction of the unit vector $\hat{\mathbf{n}}$, then

$$\hat{\mathbf{n}} \cdot (\nabla \times \mathbf{v}) = \frac{1}{\delta S} \oint_{\delta r} \mathbf{v} \cdot d\mathbf{r}, \quad (4.3)$$

where the line integral is around the infinitesimal area. Thus at a point the component of vorticity in the direction of \mathbf{n} is proportional to the circulation around the surrounding infinitesimal fluid element, divided by the elemental area bounded by the path of the integral. A heuristic test for the presence of vorticity is to imagine a small paddle wheel in the flow; the paddle wheel acts as a ‘circulation-meter’, and rotates if the vorticity is non-zero. Vorticity might seem to be similar to angular momentum, in that it is a measure of spin. However, unlike angular momentum, *the value of vorticity at a point does not depend on the particular choice of an axis of rotation*; indeed, the definition of vorticity makes no reference at all to an axis of rotation or to a coordinate system. Rather, vorticity is a measure of the *local* spin of a fluid element.

4.1.2 Simple axisymmetric examples

Consider axisymmetric motion in two dimensions, so that the flow is confined to a plane. We use cylindrical coordinates (r, ϕ, z) , where z is the direction perpendicular to the plane, with velocity components (u^r, u^ϕ, u^z) . For axisymmetric flow $u^z = u^r = 0$ but $u^\phi \neq 0$.

Rigid body motion

For a body in rigid body rotation, the velocity distribution is given by

$$u^\phi = \Omega r, \quad (4.4)$$

where Ω is the angular velocity of the fluid and r is the distance from the axis of rotation. Associated with this rotation is a vorticity given by

$$\boldsymbol{\omega} = \nabla \times \mathbf{v} = \omega^z \mathbf{k}, \quad (4.5)$$

where

$$\omega^z = \frac{1}{r} \frac{\partial}{\partial r} (r u^\phi) = \frac{1}{r} \frac{\partial}{\partial r} (r^2 \Omega) = 2\Omega. \quad (4.6)$$

The vorticity of a fluid in solid body rotation is thus twice the angular velocity of the fluid about the axis of rotation, and is pointed in a direction orthogonal to the plane of rotation.

The ‘ vr ’ vortex

This vortex is so-called because the tangential velocity (historically denoted by ‘ v ’ in this context) is such that the product vr is constant. In our notation we would have

$$u^\phi = \frac{K}{r}, \quad (4.7)$$

where K is a constant determining the vortex strength. Evaluating the z -component of vorticity gives

$$\omega^z = \frac{1}{r} \frac{\partial}{\partial r} (r u^\phi) = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{K}{r} \right) = 0, \quad (4.8)$$

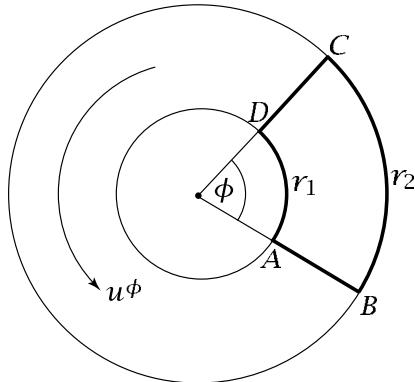


Fig. 4.1 Evaluation of circulation in the axi-symmetric vr vortex. The circulation around the path $A-B-C-D$ is zero. This result does not depend on the radii r_1 or r_2 or the angle ϕ , and the circulation around any infinitesimal path not enclosing the origin is zero. Thus the vorticity is zero everywhere except at the origin.

except where $r = 0$, at which the expression is singular and the vorticity is infinite. Our paddle wheel rotates when placed at the vortex center, but, less obviously, does not if placed elsewhere.

The circulation around a circle that encloses the origin is given by

$$C = \oint \frac{K}{r} r d\phi = 2\pi K. \quad (4.9)$$

This does not depend on the radius, and so it is true even as the radius tends to zero. Since the vorticity is the circulation divided by the area, the vorticity at the origin must be infinite. Consider now an integration path that does *not* enclose the origin, for example the contour $A-B-C-D-A$ in Fig. 4.1. Over the segments $A-B$ and $C-D$ the velocity is orthogonal to the contour, and so the contribution is zero. Over $B-C$ and $D-A$ we have

$$C_{BC} = \frac{K}{r_2} \phi r_2 = K\phi, \quad C_{DA} = -\frac{K}{r_1} \phi r_1 = -K\phi. \quad (4.10)$$

Adding these two expressions we see that the net circulation around the contour C_{ABCD} is zero. If we shrink the integration path to an infinitesimal size then, within the path, by Stokes' theorem, the vorticity is zero. We can of course place the path anywhere we wish, except surrounding the origin, and obtain this result. Thus the vorticity is everywhere zero, except at the origin.

4.2 THE VORTICITY EQUATION

Using the vector identity $\mathbf{v} \times (\nabla \times \mathbf{v}) = \nabla(\mathbf{v} \cdot \mathbf{v})/2 - (\mathbf{v} \cdot \nabla)\mathbf{v}$, we write the momentum equation as

$$\frac{\partial \mathbf{v}}{\partial t} + \boldsymbol{\omega} \times \mathbf{v} = -\frac{1}{\rho} \nabla p - \frac{1}{2} \nabla \mathbf{v}^2 + \mathbf{F}, \quad (4.11)$$

where \mathbf{F} represents viscous and body forces. Taking the curl of (4.11) gives the vorticity equation

$$\frac{\partial \boldsymbol{\omega}}{\partial t} + \nabla \times (\boldsymbol{\omega} \times \mathbf{v}) = \frac{1}{\rho^2} (\nabla p \times \nabla p) + \nabla \times \mathbf{F}. \quad (4.12)$$

Now, the vector identity

$$\nabla \times (\mathbf{a} \times \mathbf{b}) = (\mathbf{b} \cdot \nabla) \mathbf{a} - (\mathbf{a} \cdot \nabla) \mathbf{b} + \mathbf{a} \nabla \cdot \mathbf{b} - \mathbf{b} \nabla \cdot \mathbf{a}, \quad (4.13)$$

implies that the second term on the left-hand side of (4.12) may be written as

$$\nabla \times (\boldsymbol{\omega} \times \mathbf{v}) = (\mathbf{v} \cdot \nabla) \boldsymbol{\omega} - (\boldsymbol{\omega} \cdot \nabla) \mathbf{v} + \boldsymbol{\omega} \nabla \cdot \mathbf{v} - \mathbf{v} \nabla \cdot \boldsymbol{\omega}. \quad (4.14)$$

Because vorticity is the curl of velocity its divergence vanishes and so (4.12) becomes

$$\frac{\partial \boldsymbol{\omega}}{\partial t} + (\mathbf{v} \cdot \nabla) \boldsymbol{\omega} = (\boldsymbol{\omega} \cdot \nabla) \mathbf{v} - \boldsymbol{\omega} \nabla \cdot \mathbf{v} + \frac{1}{\rho^2} (\nabla \rho \times \nabla p) + \nabla \times \mathbf{F}. \quad (4.15)$$

The divergence term may be eliminated with the aid of the mass-conservation equation to give

$$\frac{D\tilde{\boldsymbol{\omega}}}{Dt} = (\tilde{\boldsymbol{\omega}} \cdot \nabla) \mathbf{v} + \frac{1}{\rho^3} (\nabla \rho \times \nabla p) + \frac{1}{\rho} \nabla \times \mathbf{F}, \quad (4.16)$$

where $\tilde{\boldsymbol{\omega}} \equiv \boldsymbol{\omega}/\rho$. We will set $\mathbf{F} = 0$ in most of what follows.

The third term on the right-hand side of (4.15), as well as the second term on the right-hand side of (4.16), is variously called the *baroclinic* term, the *non-homentropic* term, or the *solenoidal* term. The solenoidal vector, \mathbf{S}_o , is defined by

$$\mathbf{S}_o \equiv \frac{1}{\rho^2} \nabla \rho \times \nabla p = -\nabla \alpha \times \nabla p. \quad (4.17)$$

A solenoid is a tube directed perpendicular to both $\nabla \alpha$ and ∇p , with elements of length proportional to $\nabla p \times \nabla \alpha$. If the isolines of p and α are parallel to each other, then solenoids do not exist. This occurs when the density is a function only of pressure for then

$$\nabla \rho \times \nabla p = \nabla \rho \times \nabla \rho \frac{dp}{d\rho} = 0. \quad (4.18)$$

The solenoidal vector may also be written

$$\mathbf{S}_o = -\nabla \eta \times \nabla T. \quad (4.19)$$

This follows most easily by first writing the momentum equation in the form $\partial \mathbf{v}/\partial t + \boldsymbol{\omega} \times \mathbf{v} = T \nabla \eta - \nabla B$, and taking its curl (see problem 2.3). Evidently the solenoidal term vanishes if: (i) isolines of pressure and density are parallel; (ii) isolines of temperature and entropy are parallel; (ii) density, entropy, temperature or pressure are constant. A *barotropic* fluid has by definition $p = p(\rho)$ and therefore no solenoids. A *baroclinic* fluid is one for which ∇p is not parallel to $\nabla \rho$. From (4.16) we see that the baroclinic term must be balanced by terms involving velocity or its tendency and therefore, in general, *a baroclinic fluid is a moving fluid*, even in the presence of viscosity.

For a barotropic fluid the vorticity equation takes the simple form,

$$\frac{D\tilde{\boldsymbol{\omega}}}{Dt} = (\tilde{\boldsymbol{\omega}} \cdot \nabla) \mathbf{v}. \quad (4.20)$$

If the fluid is also incompressible, meaning that $\nabla \cdot \mathbf{v} = 0$, then we have the even simpler form,

$$\frac{D\boldsymbol{\omega}}{Dt} = (\boldsymbol{\omega} \cdot \nabla) \mathbf{v}. \quad (4.21)$$

When expanded into their components, the terms on the right-hand side of (4.20) or (4.21) are conventionally divided into 'stretching' and 'tipping' (or 'tilting') terms, and we return to that in section 4.3.1.

An integral conservation property

Consider a single Cartesian component in (4.15). Then, using superscripts to denote components,

$$\begin{aligned}\frac{\partial \omega^x}{\partial t} &= -\mathbf{v} \cdot \nabla \omega^x - \omega^x \nabla \cdot \mathbf{v} + (\mathbf{w} \cdot \nabla) v^x + S_o^x \\ &= -\nabla \cdot (\mathbf{v} \omega^x) + \nabla \cdot (\mathbf{w} v^x) + S_o^x,\end{aligned}\quad (4.22)$$

where S_o^x is the (x -component of) the solenoidal term. Eq. (4.22) may be written as

$$\frac{\partial \omega^x}{\partial t} + \nabla \cdot (\mathbf{v} \omega^x - \mathbf{w} v^x) = S_o^x, \quad (4.23)$$

and this implies the Cartesian tensor form of the vorticity equation, namely

$$\frac{\partial \omega_i}{\partial t} + \frac{\partial}{\partial x_j} (v_j \omega_i - v_i \omega_j) = S_{oi}, \quad (4.24)$$

with summation over repeated indices. The tendency of the components of vorticity are thus given by the solenoidal term plus the divergence of a vector field, and if the former vanishes the volume integrated vorticity can only be altered by boundary effects. However, in both the atmosphere and the ocean the solenoidal term is important, although we will see in section 4.5 that a useful conservation law for a scalar quantity can still be obtained.

4.2.1 Two-dimensional flow

In two-dimensional flow the fluid is confined to a surface, and independent of the third dimension normal to that surface. Let us initially stay in a Cartesian geometry, and then two-dimensional flow is flow on a plane; the velocity normal to the plane and the rate of change of any quantity normal to that plane, are zero. Let the normal direction be the z -direction; the fluid velocity in the plane, \mathbf{u} , is $\mathbf{u} = u\mathbf{i} + v\mathbf{j}$, and the velocity normal to the plane, w , is zero. Only one component of vorticity is non-zero and this is given by

$$\boldsymbol{\omega} = \mathbf{k} \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right). \quad (4.25)$$

That is, in two-dimensional flow the vorticity is perpendicular to the velocity. We let $\zeta \equiv \omega^z = \boldsymbol{\omega} \cdot \mathbf{k}$. Both the stretching and tilting terms vanish in two-dimensional flow, and the two-dimensional vorticity equation becomes, for incompressible flow,

$$\frac{D\zeta}{Dt} = 0, \quad (4.26)$$

where $D\zeta/Dt = \partial\zeta/\partial t + \mathbf{u} \cdot \nabla \zeta$. That is, in two-dimensional flow vorticity is conserved following the fluid elements; each material parcel of fluid keeps its value of vorticity even as it is being advected around. Furthermore, specification of the vorticity completely determines the flow field. To see this, we use the incompressibility condition to define a streamfunction ψ such that

$$u = -\frac{\partial \psi}{\partial y}, \quad v = \frac{\partial \psi}{\partial x}, \quad \zeta = \nabla^2 \psi. \quad (4.27a,b,c)$$

Given the vorticity, the Poisson equation (4.27c) can be solved for the streamfunction and the velocity fields obtained through (4.27a,b), and this process is called ‘inverting the vorticity’.

Numerical integration of (4.26) is then a process of timestepping plus inversion. The vorticity equation may then be written as an advection equation for vorticity,

$$\frac{\partial \zeta}{\partial t} + \mathbf{u} \cdot \nabla \zeta = 0, \quad (4.28)$$

in conjunction with (4.27). The vorticity is stepped forward one timestep using a finite-difference representation of (4.28), and the vorticity inverted to obtain a velocity using (4.27). (The notion that complete or nearly complete information about the flow may be obtained by inverting one field plays an important role in geophysical fluid dynamics, as we will see later on.)

Two-dimensional flow is not restricted to a Cartesian plane, and we may certainly envision two-dimensional flow on the surface of a sphere; in this case the velocity normal to the spherical surface (the ‘vertical velocity’) vanishes, and the equations are naturally expressed in spherical coordinates. Nevertheless, vorticity (absolute vorticity if the sphere is rotating) is still conserved on parcels as they move over the spherical surface.

4.3 VORTICITY AND CIRCULATION THEOREMS

4.3.1 The ‘frozen-in’ property of vorticity

Let us first consider some simple topological properties of the vorticity field and its evolution. We define a *vortex line* to be a line drawn through the fluid which is everywhere in the direction of the local vorticity. This definition is analogous to that of a streamline, which is everywhere in the direction of the local velocity. A *vortex tube* is formed by the collection of vortex lines passing through a closed curve (Fig. 4.2). A *material line* is just a line that connects material fluid elements. Suppose we draw a vortex line through the fluid; such a line obviously connects fluid elements and therefore defines a coincident material line. As the fluid moves the material line deforms, and the vortex line also evolves in a manner determined by the equations of motion. A remarkable property of vorticity is that, for an unforced and inviscid barotropic fluid, the flow evolution is such that a vortex line remains coincident with the material line that it was initially associated with. Put another way, a vortex line always contains the same material elements — the vorticity is ‘frozen’ or ‘glued’ to the material fluid.¹

To prove this we consider how an infinitesimal material line element $\delta \mathbf{l}$ evolves, $\delta \mathbf{l}$ being the infinitesimal material element connecting \mathbf{l} with $\mathbf{l} + \delta \mathbf{l}$. The rate of change of $\delta \mathbf{l}$ following the flow is given by

$$\frac{D\delta \mathbf{l}}{Dt} = \frac{1}{\delta t} [\delta \mathbf{l}(t + \delta t) - \delta \mathbf{l}(t)], \quad (4.29)$$

which follows from the definition of the material derivative in the limit $\delta t \rightarrow 0$. From the Taylor expansion of $\delta \mathbf{l}(t)$ and the definition of velocity it is also apparent that

$$\delta \mathbf{l}(t + \delta t) = \mathbf{l}(t) + \delta \mathbf{l}(t) + (\mathbf{v} + \delta \mathbf{v})\delta t - (\mathbf{l}(t) + \mathbf{v}\delta t) = \delta \mathbf{l} + \delta \mathbf{v} \delta t, \quad (4.30)$$

as illustrated in Fig. 4.3. Substituting (4.30) into (4.29) gives $D\delta \mathbf{l}/Dt = \delta \mathbf{v}$, as expected, and

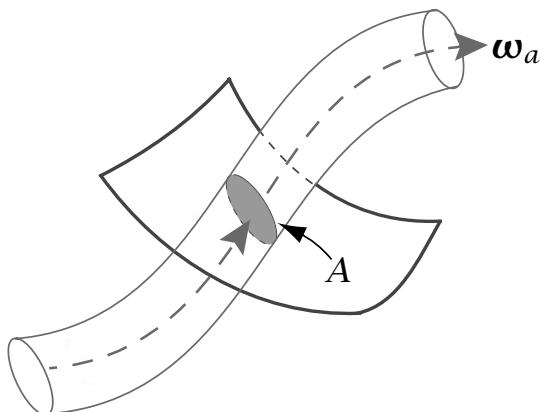


Fig. 4.2 A vortex tube passing through a material sheet. The circulation is the integral of the velocity around the boundary of A , and is equal to the integral of the normal component of vorticity over A .

because $\delta\mathbf{v} = (\delta\mathbf{l} \cdot \nabla)\mathbf{v}$ we obtain

$$\frac{D\delta\mathbf{l}}{Dt} = (\delta\mathbf{l} \cdot \nabla)\mathbf{v}. \quad (4.31)$$

Comparing this with (4.16), we see that vorticity evolves in the same way as a line element. To see what this means, at some initial time we can define an infinitesimal material line element parallel to the vorticity at that location, that is,

$$\delta\mathbf{l}(\mathbf{x}, t = 0) = A\boldsymbol{\omega}(\mathbf{x}, t = 0), \quad (4.32)$$

where A is a constant. Then, for all subsequent times the magnitude of the vorticity of that fluid element, even as it moves to a new location \mathbf{x}' , remains proportional to the length of the fluid element at that point and is oriented in the same way; that is $\boldsymbol{\omega}(\mathbf{x}', t) = A^{-1}\delta\mathbf{l}(\mathbf{x}', t)$.

To verify this result in a different way note that a vortex line element is determined by the condition $\delta\mathbf{l} = A\boldsymbol{\omega}$ which implies $\boldsymbol{\omega} \times \delta\mathbf{l} = 0$. Now, for any line element we have that

$$\frac{D}{Dt}(\boldsymbol{\omega} \times \delta\mathbf{l}) = \frac{D\boldsymbol{\omega}}{Dt} \times \delta\mathbf{l} - \frac{D\delta\mathbf{l}}{Dt} \times \boldsymbol{\omega}. \quad (4.33)$$

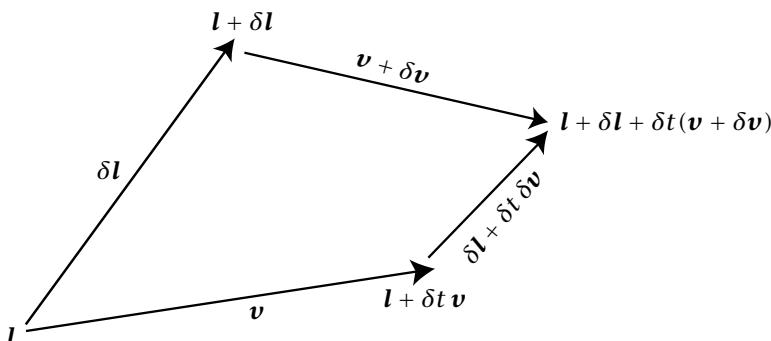


Fig. 4.3 Evolution of an infinitesimal material line $\delta\mathbf{l}$ from time t to time $t + \delta t$. It can be seen from the diagram that $D\delta\mathbf{l}/Dt = \delta\mathbf{v}$.

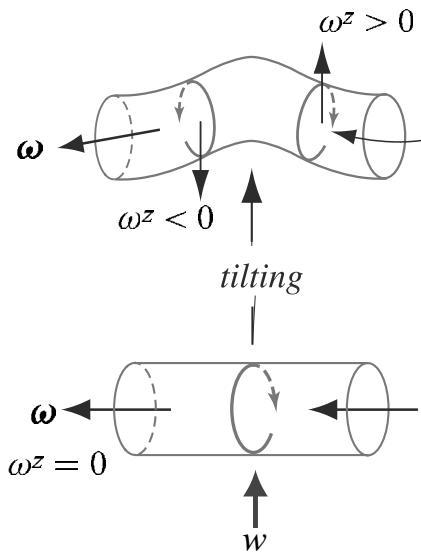


Fig. 4.4 The tilting of vorticity. Suppose that the vorticity, ω is initially directed horizontally, as in the lower figure, so that ω^z , its vertical component, is zero. The material lines, and therefore also the vortex lines, are tilted by the positive vertical velocity W , thus creating a vertically oriented vorticity. This mechanism is important in creating vertical vorticity in the atmospheric boundary layer (and, one may show, the β -effect in large-scale flow).

We also have that

$$\frac{D\delta l}{Dt} = \delta v = \delta l \cdot \nabla v \quad (4.34a)$$

and

$$\frac{D\omega}{Dt} = \omega \cdot \nabla v. \quad (4.34b)$$

If the line element is initially a vortex line element then, at $t = 0$, $\delta l = A\omega$ and, using (4.34), the right-hand side of (4.33) vanishes. Thus, the *tendency* of $\omega \times \delta l$ is zero, and the vortex line continues to be a material line.

Stretching and tilting

The terms on the right-hand side of (4.20) or (4.21) may be interpreted in terms of 'stretching' and 'tipping' (or 'tilting'). Consider a single Cartesian component of (4.21),

$$\frac{D\omega^x}{Dt} = \omega^x \frac{\partial u}{\partial x} + \omega^y \frac{\partial u}{\partial y} + \omega^z \frac{\partial u}{\partial z}. \quad (4.35)$$

The second and third terms of this are the tilting or tipping terms because they involve changes in the orientation of the vorticity vector. They tell us that vorticity in the x -direction may be generated from vorticity in the y - and z -directions if the advection acts to tilt the material lines. Because vorticity is tied to these lines, vorticity oriented in one direction becomes oriented in another, as in Fig. 4.4.

The first term on the right-hand side of (4.35) is the stretching term, and it acts to intensify the x -component of vorticity if the velocity is increasing in the x -direction — that is, if the material lines are being stretched (Fig. 4.5). This effect arises because a vortex line is tied to a material line, and therefore vorticity is amplified in proportion to the stretching of the material line aligned with it. This effect is important in tornadoes, to give one example. If the fluid is incompressible, stretching of a fluid mass in one direction must be accompanied

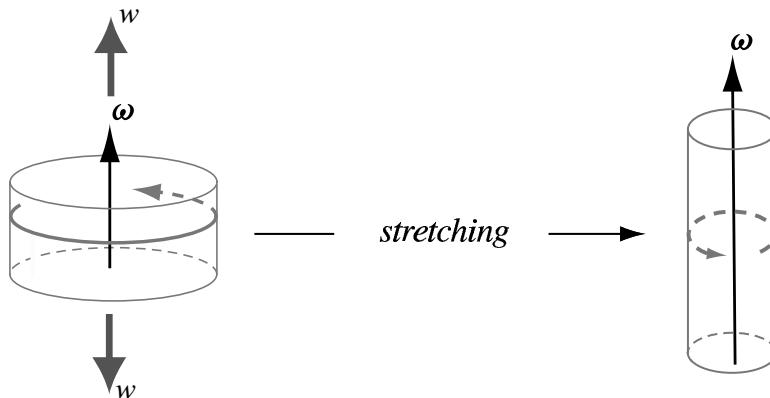


Fig. 4.5 Stretching of material lines distorts the cylinder of fluid as shown. Vorticity is tied to material lines, and so is amplified in the direction of the stretching. However, because the volume of fluid is conserved, the end surfaces shrink, the material lines through the cylinder ends converge and the integral of vorticity over a material surface (the circulation) remains constant, as discussed in section 4.3.2.

by convergence in another, and this leads to the conservation of circulation, as we now discuss.

4.3.2 Kelvin's circulation theorem

Kelvin's circulation theorem states that under certain circumstances the circulation around a material fluid parcel is conserved; that is, the circulation is conserved 'following the flow'.² The primary restrictions are that body forces are conservative (i.e., they are representable as potential forces, and therefore that the flow be inviscid), and that the fluid is barotropic [i.e., $p = p(\rho)$]. Of these, the latter is more restrictive for geophysical fluids. The circulation in the theorem is defined with respect to an inertial frame of reference; specifically, the velocity in (4.39) is the velocity relative to an inertial frame. To prove the theorem, we begin with the inviscid momentum equation,

$$\frac{D\mathbf{v}}{Dt} = -\frac{1}{\rho} \nabla p - \nabla \Phi, \quad (4.36)$$

where $\nabla \Phi$ represents the conservative body forces on the system. Applying the material derivative to the circulation, (4.2), gives

$$\begin{aligned} \frac{DC}{Dt} &= \frac{D}{Dt} \oint \mathbf{v} \cdot d\mathbf{r} = \oint \left(\frac{D\mathbf{v}}{Dt} \cdot d\mathbf{r} + \mathbf{v} \cdot d\mathbf{v} \right) \\ &= \oint \left[\left(-\frac{1}{\rho} \nabla p - \nabla \Phi \right) \cdot d\mathbf{r} + \mathbf{v} \cdot d\mathbf{v} \right] \\ &= \oint -\frac{1}{\rho} \nabla p \cdot d\mathbf{r} \end{aligned} \quad (4.37)$$

using (4.36) and $D(d\mathbf{r})/Dt = d\mathbf{v}$, where $d\mathbf{r}$ is the line element and with the line integration being over a closed, material, circuit. The second and third terms on the second line vanish

separately, because they are exact differentials integrated around a closed loop. The term on the last line vanishes if the density is constant or, more generally, if the pressure is a function of density alone, in which case ∇p is parallel to $\nabla \rho$. To see this, note that

$$\oint \frac{1}{\rho} \nabla p \cdot d\mathbf{r} = \int_S \nabla \times \left(\frac{\nabla p}{\rho} \right) \cdot d\mathbf{S} = \int_S \frac{-\nabla \rho \times \nabla p}{\rho^2} \cdot d\mathbf{S}, \quad (4.38)$$

using Stokes' theorem where S is any surface bounded by the path of the line integral. The integral evidently vanishes identically if p is a function of ρ alone. The right-most expression above is the integral of the solenoidal vector, and if it is zero (4.37) becomes

$$\frac{D}{Dt} \oint \mathbf{v} \cdot d\mathbf{r} = 0.$$

(4.39)

This is Kelvin's circulation theorem. In words, *the circulation around a material loop is invariant for a barotropic fluid that is subject only to conservative forces*. Using Stokes' theorem, the circulation theorem may also be written as

$$\frac{D}{Dt} \int_S \boldsymbol{\omega} \cdot d\mathbf{S} = 0.$$

(4.40)

That is, the area integral of the normal component of vorticity across any material surface is constant, under the same conditions. This form is both natural and useful, and it arises because of the way vorticity is tied to material fluid elements.

Stretching and circulation

Let us informally consider how vortex stretching and mass conservation work together to give the circulation theorem. Let the fluid be incompressible so that the volume of a fluid mass is constant, and consider a surface normal to a vortex tube, as in Fig. 4.5). Let the volume of a small material box around the surface be δV , the length of the material lines be δl and the surface area be δA . Then

$$\delta V = \delta l \delta A. \quad (4.41)$$

Because of the frozen-in property, the vorticity passing through the surface is proportional to the length of the material lines. That is, $\omega \propto \delta l$, and

$$\delta V \propto \omega \delta A. \quad (4.42)$$

The right-hand side is just the circulation around the surface. Now, if the corresponding material tube is stretched δl increases, but the volume, δV , remains constant by mass conservation. Thus, the circulation given by the right-hand side of (4.42) also remains constant. In other words, because of the frozen-in property vorticity is amplified by the stretching, but the vortex lines get closer together in such a way that the product $\omega \delta A$ remains constant and circulation is conserved.

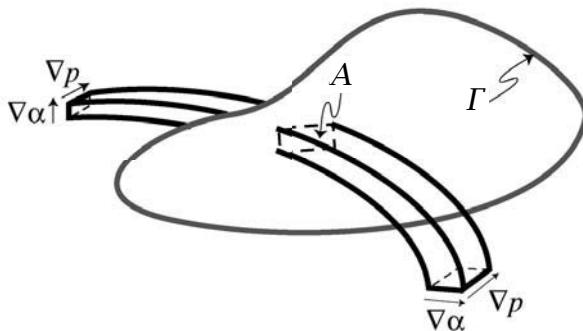


Fig. 4.6 Solenoids and the circulation theorem. Solenoids are tubes perpendicular to both $\nabla\alpha$ and ∇p , and they have a non-zero cross-sectional area if isolines of α and p do not coincide. The rate of change of circulation over a material surface is given by the sum of all the solenoidal areas crossing the surface. If $\nabla\alpha \times \nabla p = 0$ there are no solenoids.

4.3.3 Baroclinic flow and the solenoidal term

In baroclinic flow, the circulation is not generally conserved, and from (4.37) we have

$$\frac{DC}{Dt} = - \oint \frac{\nabla p}{\rho} \cdot d\mathbf{r} = - \oint \frac{dp}{\rho}, \quad (4.43)$$

and this is called the baroclinic circulation theorem.³ Noting the fundamental thermodynamic relation $T d\eta = dI + p d\alpha$, where $\alpha = \rho^{-1}$, we have

$$\alpha dp = d(p\alpha) - T d\eta + dI, \quad (4.44)$$

so that the solenoidal term on the right-hand side of (4.43) may be written as

$$S_o \equiv - \oint \alpha dp = \oint T d\eta = - \oint \eta dT = - R \oint T d \log p, \quad (4.45)$$

where the last equality holds only for an ideal gas. Using Stokes' theorem, S_o can also be written as

$$S_o = - \int_S \nabla\alpha \times \nabla p \cdot d\mathbf{S} = - \int_S \left(\frac{\partial \alpha}{\partial T} \right)_p \nabla T \times \nabla p \cdot d\mathbf{S} = \int_S \nabla T \times \nabla\eta \cdot d\mathbf{S}. \quad (4.46)$$

The rate of change of the circulation across a surface depends on the existence of this solenoidal term (Fig. 4.6 and, for an example, problem 4.8). However, even if the solenoidal vector is in general non-zero, circulation *is* conserved if the material path is in a surface of constant entropy, η , and if $D\eta/Dt = 0$. In this case the solenoidal term vanishes and, because $D\eta/Dt = 0$, entropy remains constant on that same material loop as it evolves. This result gives rise to the conservation of potential vorticity, discussed in section 4.5

4.3.4 Circulation in a rotating frame

The absolute and relative velocities are related by $\mathbf{v}_a = \mathbf{v}_r + \boldsymbol{\Omega} \times \mathbf{r}$ so that in a rotating frame the rate of change of circulation is given by

$$\frac{D}{Dt} \oint (\mathbf{v}_r + \boldsymbol{\Omega} \times \mathbf{r}) \cdot d\mathbf{r} = \oint \left[\left(\frac{D\mathbf{v}_r}{Dt} + \boldsymbol{\Omega} \times \mathbf{v}_r \right) \cdot d\mathbf{r} + (\mathbf{v}_r + \boldsymbol{\Omega} \times \mathbf{r}) \cdot d\mathbf{v}_r \right]. \quad (4.47)$$

But $\oint \mathbf{v}_r \cdot d\mathbf{v}_r = 0$ and, integrating by parts,

$$\begin{aligned} \oint (\boldsymbol{\Omega} \times \mathbf{r}) \cdot d\mathbf{v}_r &= \oint \left\{ d[(\boldsymbol{\Omega} \times \mathbf{r}) \cdot \mathbf{v}_r] - (\boldsymbol{\Omega} \times d\mathbf{r}) \cdot \mathbf{v}_r \right\} \\ &= \oint \left\{ d[(\boldsymbol{\Omega} \times \mathbf{r}) \cdot \mathbf{v}_r] + (\boldsymbol{\Omega} \times \mathbf{v}_r) \cdot d\mathbf{r} \right\}. \end{aligned} \quad (4.48)$$

The first term on the right-hand side is zero and so (4.47) becomes

$$\frac{D}{Dt} \oint (\mathbf{v}_r + \boldsymbol{\Omega} \times \mathbf{r}) \cdot d\mathbf{r} = \oint \left(\frac{D\mathbf{v}_r}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{v}_r \right) \cdot d\mathbf{r} = - \oint \frac{dp}{\rho}, \quad (4.49)$$

where the second equality uses the momentum equation. The term on the last line vanishes if the fluid is barotropic, and if so the circulation theorem is, unsurprisingly,

$$\frac{D}{Dt} \oint (\mathbf{v}_r + \boldsymbol{\Omega} \times \mathbf{r}) \cdot d\mathbf{r} = 0, \quad \text{or} \quad \frac{D}{Dt} \int_S (\boldsymbol{\omega}_r + 2\boldsymbol{\Omega}) \cdot d\mathbf{S} = 0, \quad (4.50a,b)$$

where the second equation uses Stokes' theorem and we have used $\nabla \times (\boldsymbol{\Omega} \times \mathbf{r}) = 2\boldsymbol{\Omega}$, and where $\boldsymbol{\omega}_r = \nabla \times \mathbf{v}_r$ is the *relative vorticity*.⁴

4.3.5 The circulation theorem for hydrostatic flow

Kelvin's circulation theorem holds for hydrostatic flow, with a slightly different form. For simplicity we restrict attention to the *f*-plane, and start with the hydrostatic momentum equations,

$$\frac{D\mathbf{u}_r}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{u}_r = -\frac{1}{\rho} \nabla_z p, \quad 0 = -\frac{1}{\rho} \frac{\partial p}{\partial z} - \nabla \Phi, \quad (4.51a,b)$$

where $\Phi = gz$ is the gravitational potential and $\boldsymbol{\Omega} = \Omega \mathbf{k}$. The advecting field is three-dimensional, and in particular we still have $D\delta\mathbf{r}/Dt = \delta\mathbf{v} = (\delta\mathbf{r} \cdot \nabla)\mathbf{v}$. Thus, using (4.51) we have

$$\begin{aligned} \frac{D}{Dt} \oint (\mathbf{u}_r + \boldsymbol{\Omega} \times \mathbf{r}) \cdot d\mathbf{r} &= \oint \left[\left(\frac{D\mathbf{u}_r}{Dt} + \boldsymbol{\Omega} \times \mathbf{v}_r \right) \cdot d\mathbf{r} + (\mathbf{u}_r + \boldsymbol{\Omega} \times \mathbf{r}) \cdot d\mathbf{v}_r \right] \\ &= \oint \left(\frac{D\mathbf{u}_r}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{u}_r \right) \cdot d\mathbf{r} \\ &= \oint \left(-\frac{1}{\rho} \nabla p - \nabla \Phi \right) \cdot d\mathbf{r}, \end{aligned} \quad (4.52)$$

as with (4.49), having used $\boldsymbol{\Omega} \times \mathbf{v}_r = \boldsymbol{\Omega} \times \mathbf{u}_r$, and where the gradient operator ∇ is three-dimensional. The last term on the right-hand side vanishes because it is the integral of the gradient of a potential around a closed path. The first term vanishes if the fluid is barotropic, so that the circulation theorem is

$$\frac{D}{Dt} \oint (\mathbf{u}_r + \boldsymbol{\Omega} \times \mathbf{r}) \cdot d\mathbf{r} = 0. \quad (4.53)$$

Using Stokes' theorem we have the equivalent form

$$\frac{D}{Dt} \int_S (\boldsymbol{\omega}_{hy} + 2\boldsymbol{\Omega}) \cdot d\mathbf{S} = 0, \quad (4.54)$$

where the subscript 'hy' denotes hydrostatic and, in Cartesian coordinates,

$$\boldsymbol{\omega}_{hy} = \nabla \times \mathbf{u}_r = -\mathbf{i} \frac{\partial v_r}{\partial z} + \mathbf{j} \frac{\partial u_r}{\partial z} + \mathbf{k} \left(\frac{\partial v_r}{\partial x} - \frac{\partial u_r}{\partial y} \right). \quad (4.55)$$

4.4 VORTICITY EQUATION IN A ROTATING FRAME

Perhaps the easiest way to derive the vorticity equation appropriate for a rotating reference frame is to begin with the momentum equation in the form

$$\frac{\partial \mathbf{v}_r}{\partial t} + (2\boldsymbol{\Omega} + \boldsymbol{\omega}_r) \times \mathbf{v}_r = -\frac{1}{\rho} \nabla p + \nabla \left(\Phi - \frac{1}{2} \mathbf{v}_r^2 \right), \quad (4.56)$$

where the potential Φ contains the gravitational and centrifugal forces. Take the curl of this and use the identity (4.13), which here implies

$$\nabla \times [(2\boldsymbol{\Omega} + \boldsymbol{\omega}_r) \times \mathbf{v}_r] = (2\boldsymbol{\Omega} + \boldsymbol{\omega}_r) \nabla \cdot \mathbf{v}_r + (\mathbf{v}_r \cdot \nabla) (2\boldsymbol{\Omega} + \boldsymbol{\omega}_r) - [(2\boldsymbol{\Omega} + \boldsymbol{\omega}_r) \cdot \nabla] \mathbf{v}_r, \quad (4.57)$$

(noting that $\nabla \cdot (2\boldsymbol{\Omega} + \boldsymbol{\omega}) = 0$), to give the vorticity equation

$$\frac{D\boldsymbol{\omega}_r}{Dt} = [(2\boldsymbol{\Omega} + \boldsymbol{\omega}_r) \cdot \nabla] \mathbf{v} - (2\boldsymbol{\Omega} + \boldsymbol{\omega}_r) \nabla \cdot \mathbf{v}_r + \frac{1}{\rho^2} (\nabla \rho \times \nabla p). \quad (4.58)$$

Note that because $\boldsymbol{\Omega}$ is a constant, $D\boldsymbol{\omega}_r/Dt = D\boldsymbol{\omega}_a/Dt$ where $\boldsymbol{\omega}_a = 2\boldsymbol{\Omega} + \boldsymbol{\omega}_r$ is the absolute vorticity. The only difference between the vorticity equation in the rotating and inertial frames of reference is in the presence of the solid-body vorticity $2\boldsymbol{\Omega}$ on the right-hand side. The second term on the right-hand side may be folded in to the material derivative using mass continuity, and after a little manipulation (4.58) becomes

$$\frac{D}{Dt} \left(\frac{\boldsymbol{\omega}_a}{\rho} \right) = \frac{1}{\rho} (2\boldsymbol{\Omega} + \boldsymbol{\omega}_r) \cdot \nabla \mathbf{v}_r + \frac{1}{\rho^3} (\nabla \rho \times \nabla p). \quad (4.59)$$

However, note that it is the absolute vorticity, $\boldsymbol{\omega}_a$, that now appears on the left-hand side. If ρ is constant, $\boldsymbol{\omega}_a$ may be replaced by $\boldsymbol{\omega}_r$.

4.4.1 The circulation theorem and the beta effect

What are the implications of the circulation theorem on a rotating, spherical planet? Let us define relative circulation over some material loop as

$$C_r \equiv \oint \mathbf{v}_r \cdot d\mathbf{r}. \quad (4.60)$$

Because $\mathbf{v}_r = \mathbf{v}_a - 2\boldsymbol{\Omega} \times \mathbf{r}$ (where \mathbf{r} is the distance from the center of the earth), we have

$$C_r = C_a - \int 2\boldsymbol{\Omega} \cdot d\mathbf{S} = C_a - 2\boldsymbol{\Omega} A_{\perp}, \quad (4.61)$$

where C_a is the total or absolute circulation and A_{\perp} is the area enclosed by the projection of the material circuit on to the plane normal to the rotation vector; that is, on to the equatorial plane (Fig. 4.7). If the solenoidal term is zero then the circulation theorem, (4.50), may be written as

$$\frac{D}{Dt} (C_r + 2\boldsymbol{\Omega} A_{\perp}) = 0. \quad (4.62)$$

Thus, the relative circulation around a circuit changes if the orientation of the plane changes; that is, if the area of its projection on to the equatorial plane changes. In large scale dynamics the most common cause of this is when a fluid parcel changes its latitude. For example,

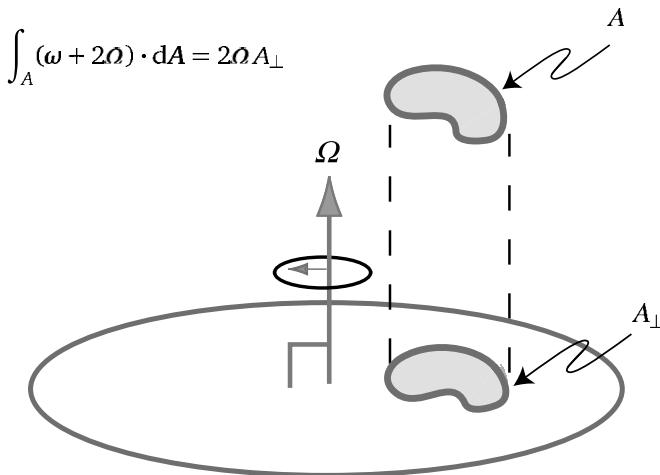


Fig. 4.7 The projection of a material circuit on to the equatorial plane. If a fluid element moves poleward, keeping its orientation to the local vertical fixed (e.g., it stays horizontal) then the area of its projection on to the equatorial plane increases. If its total (absolute) circulation is to be maintained, then the vertical component of the relative vorticity must diminish; that is, $\int_A (\omega + 2\Omega) \cdot dA = \int_A (\zeta + f) dA = \text{constant}$. Thus, the β term in $D(\zeta + f)/Dt = D\zeta/Dt + \beta v = 0$ ultimately arises from the *tilting* of a parcel relative to the axis of rotation as it moves meridionally.

consider the flow of a two-dimensional, infinitesimal, horizontal (i.e., tangent to the the radial vector), constant-density fluid parcel at a latitude ϑ with area A , so that the projection of its area on to the equatorial plane is $A_{\perp} = A \sin \vartheta$ and $C_r = \zeta_r A$. If the fluid surface moves, but remains horizontal, its area is preserved (because it is incompressible) and directly from (4.62) its relative vorticity changes as

$$\frac{D\zeta_r}{Dt} = -\frac{2\Omega}{A} \frac{DA_{\perp}}{Dt} = -2\Omega \frac{D}{Dt} \sin \vartheta = -v_r \frac{2\Omega \cos \vartheta}{a} = -\beta v_r, \quad (4.63)$$

where

$$\beta \equiv \frac{df}{dy} = \frac{2\Omega}{a} \cos \vartheta. \quad (4.64)$$

The means by which the vertical component of the relative vorticity of a parcel changes by virtue of its latitudinal displacement is known as the *beta effect*, or the β -effect. It is a manifestation of the tilting term in the vorticity equation, and it is often the most important means by which relative vorticity does change in large-scale flow. The β -effect arises in the full vorticity equation, as we now see.

4.4.2 The vertical component of the vorticity equation

In large-scale dynamics, the most important, although not the largest, component of the vorticity is often the vertical one, because this contains much of the information about the horizontal flow. We can obtain an explicit expression for its evolution by taking the

vertical component of (4.58), although care must be taken because the unit vectors (\mathbf{i} , \mathbf{j} , \mathbf{k}) are functions of position (see section 2.2).

An alternative derivation begins with the horizontal momentum equations

$$\frac{\partial u}{\partial t} - v(\zeta + f) + w\frac{\partial u}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial x} - \frac{1}{2}\frac{\partial}{\partial x}(u^2 + v^2) + F^x \quad (4.65a)$$

$$\frac{\partial v}{\partial t} + u(\zeta + f) + w\frac{\partial v}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial y} - \frac{1}{2}\frac{\partial}{\partial y}(u^2 + v^2) + F^y, \quad (4.65b)$$

where in this section we again drop the subscript r on variables measured in the rotating frame. Cross-differentiating gives, after a little algebra,

$$\begin{aligned} \frac{D}{Dt}(\zeta + f) = & -(\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \left(\frac{\partial u}{\partial z}\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\frac{\partial w}{\partial x}\right) \\ & + \frac{1}{\rho^2}\left(\frac{\partial \rho}{\partial x}\frac{\partial p}{\partial y} - \frac{\partial \rho}{\partial y}\frac{\partial p}{\partial x}\right) + \left(\frac{\partial F^y}{\partial x} - \frac{\partial F^x}{\partial y}\right). \end{aligned} \quad (4.66)$$

We interpret the various terms as follows.

$D\zeta/Dt = \partial\zeta/\partial t + \mathbf{v} \cdot \nabla\zeta$. The material derivative of the vertical component of the vorticity.

$Df/Dt = v\partial f/\partial y = v\beta$. The β -effect. The vorticity is affected by the meridional motion of the fluid, so that, apart from the terms on the right-hand side, $(\zeta + f)$ is conserved on parcels. Because the Coriolis parameter changes with latitude this is like saying that the system has differential rotation. This effect is precisely that due to the change in orientation of fluid surfaces with latitude, as given above in section 4.4.1 and Fig. 4.7.

$-(\zeta + f)(\partial u/\partial x + \partial v/\partial y)$. The divergence term, which gives rise to vortex stretching.

In an incompressible fluid this may be written $(\zeta + f)\partial w/\partial z$, so that vorticity is amplified if the vertical velocity increases with height, so stretching the material lines and the vorticity.

$(\partial u/\partial z)(\partial w/\partial y) - (\partial v/\partial z)(\partial w/\partial x)$. The tilting term, whereby a vertical component of vorticity may be generated by a vertical velocity acting on a horizontal vorticity. See Fig. 4.4.

$\rho^{-2}[(\partial \rho/\partial x)(\partial p/\partial y) - (\partial \rho/\partial y)(\partial p/\partial x)] = \rho^{-2}J(\rho, p)$. The solenoidal term, also called the non-homentropic or baroclinic term, arising when isosurfaces of pressure and density are not parallel.

$(\partial F^y/\partial x - \partial F^x/\partial y)$. The forcing and friction term. If the only contribution is from molecular viscosity then this term is $\nu\nabla^2\zeta$.

Two-dimensional and shallow water vorticity equations

In an inviscid two-dimensional incompressible flow, all of the terms on the right-hand side of (4.66) vanish and we have the simple equation

$$\frac{D(\zeta + f)}{Dt} = 0, \quad (4.67)$$

implying that the absolute vorticity, $\zeta_a \equiv \zeta + f$, is materially conserved. If f is a constant, then (4.67) reduces to (4.28), and background rotation plays no role. If f varies linearly with y , so that $f = f_0 + \beta y$, then (4.67) becomes

$$\frac{\partial \zeta}{\partial t} + \mathbf{u} \cdot \nabla \zeta + \beta v = 0, \quad (4.68)$$

which is known as the two-dimensional β -plane vorticity equation.

For inviscid shallow water flow, we can show that (see chapter 3)

$$\frac{D(\zeta + f)}{Dt} = -(\zeta + f) \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right). \quad (4.69)$$

In this equation the vanishing of the tilting term is perhaps the only aspect which is perhaps not immediately apparent, but this succumbs to a little thought.

4.5 POTENTIAL VORTICITY CONSERVATION

Too much of a good thing is wonderful.

Mae West (1892–1990).

Although Kelvin's circulation theorem is a general statement about vorticity conservation, in its original form it is not always a practically useful statement for two reasons. First, it is not a statement about a *field*, such as vorticity itself. Second, it is not satisfied for baroclinic flow, such as is found in the atmosphere and ocean. (Of course non-conservative forces such as viscosity also lead to circulation non-conservation, but this applies to virtually all conservation laws and does not diminish them.) It turns out that it is possible to derive a beautiful conservation law that overcomes both of these failings and one, furthermore, that is extraordinarily useful in geophysical fluid dynamics. This is the conservation of *potential vorticity* (PV) introduced first by Rossby and then in a more general form by Ertel.⁵ The idea is that we can use a scalar field that is being advected by the flow to keep track of, or to take care of, the evolution of fluid elements. For a baroclinic fluid this scalar field must be chosen in a special way (it must be a function of the density and pressure alone), but there is no restriction to a barotropic fluid. Then using the scalar evolution equation in conjunction with the vorticity equation gives us a scalar conservation equation. In the next few subsections we derive the equation for potential vorticity conservation in a number of superficially different ways.

4.5.1 PV conservation from the circulation theorem

Barotropic fluids

Let us begin with the simple case of a barotropic fluid. For an infinitesimal volume we write Kelvin's theorem as

$$\frac{D}{Dt} [(\boldsymbol{\omega}_a \cdot \mathbf{n}) \delta A] = 0, \quad (4.70)$$

where \mathbf{n} is a unit vector normal to an infinitesimal surface δA . Now consider a volume bounded by two isosurfaces of values χ and $\chi + \delta\chi$, where χ is any materially conserved tracer, thus satisfying $D\chi/Dt = 0$, so that δA initially lies in an isosurface of χ (see Fig. 4.8).

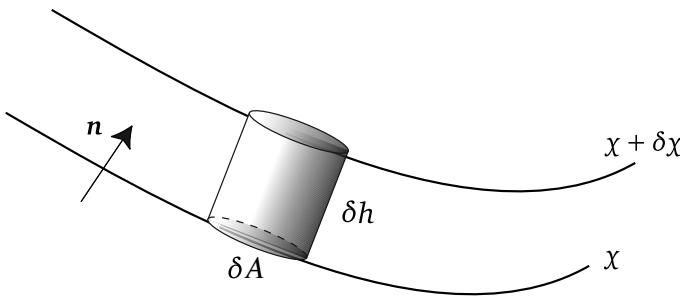


Fig. 4.8 An infinitesimal fluid element, bounded by two isosurfaces of the conserved tracer χ . As $D\chi/Dt = 0$, then $D\delta\chi/Dt = 0$.

Since $\mathbf{n} = \nabla\chi/|\nabla\chi|$ and the infinitesimal volume $\delta V = \delta h \delta A$, where δh is the separation between the two surfaces, we have

$$\mathbf{w}_a \cdot \mathbf{n} \delta A = \mathbf{w}_a \cdot \frac{\nabla\chi}{|\nabla\chi|} \frac{\delta V}{\delta h}. \quad (4.71)$$

Now, the separation between the two surfaces, δh may be obtained from

$$\delta\chi = \delta\mathbf{x} \cdot \nabla\chi = \delta h |\nabla\chi|, \quad (4.72)$$

and using this in (4.70) we obtain

$$\frac{D}{Dt} \left[\frac{(\mathbf{w}_a \cdot \nabla\chi) \delta V}{\delta\chi} \right] = 0. \quad (4.73)$$

Now, as χ is conserved on material elements, then so is $\delta\chi$, and it may be taken out of the differentiation. The mass of the volume element $\rho \delta V$ is also conserved, so that (4.73) becomes

$$\frac{\rho \delta V}{\delta\chi} \frac{D}{Dt} \left(\frac{\mathbf{w}_a \cdot \nabla\chi}{\rho} \right) = 0 \quad (4.74)$$

or

$$\frac{D}{Dt} (\widetilde{\mathbf{w}}_a \cdot \nabla\chi) = 0, \quad (4.75)$$

where $\widetilde{\mathbf{w}}_a = \mathbf{w}_a/\rho$. Equation (4.75) is a statement of potential vorticity conservation for a barotropic fluid. The field χ may be chosen arbitrarily, provided that it is materially conserved.

The general case

For a baroclinic fluid the above derivation fails simply because the statement of the conservation of circulation, (4.70) is not, in general, true: there are solenoidal terms on the right-hand side and from (4.43) and (4.45) we have

$$\frac{D}{Dt} [(\mathbf{w}_a \cdot \mathbf{n}) \delta A] = \mathbf{S}_o \cdot \mathbf{n} \delta A, \quad \mathbf{S}_o = -\nabla\alpha \times \nabla p = -\nabla\eta \times \nabla T. \quad (4.76a,b)$$

However, the right-hand side of (4.76a) may be annihilated by choosing the circuit around which we evaluate the circulation to be such that the solenoidal term is identically zero.

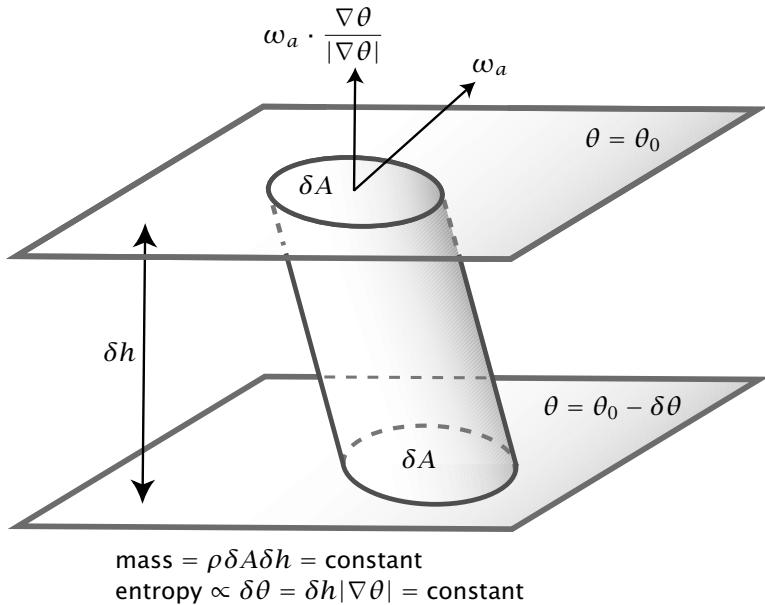


Fig. 4.9 Geometry of potential vorticity conservation. The circulation equation is $D[(\omega_a \cdot \mathbf{n})\delta A]/Dt = S_o \cdot \mathbf{n} \delta A$, where $S_o \propto \nabla \theta \times \nabla T$. We choose $\mathbf{n} = \nabla \theta / |\nabla \theta|$, where θ is materially conserved, to annihilate the solenoidal term on the right-hand side, and we note that $\delta A = \delta V / \delta h$, where δV is the volume of the cylinder, and the height of the column is $\delta h = \delta \theta / |\nabla \theta|$. The circulation is $C \equiv \omega_a \cdot \mathbf{n} \delta A = \omega_a \cdot (\nabla \theta / |\nabla \theta|)(\delta V / \delta h) = [\rho^{-1} \omega_a \cdot \nabla \theta](\delta M / \delta \theta)$, where $\delta M = \rho \delta V$ is the mass of the cylinder. As δM and $\delta \theta$ are materially conserved, so is the potential vorticity $\rho^{-1} \omega_a \cdot \nabla \theta$.

Given the form of S_o , this occurs if the values of any of p, ρ, η, T are constant on that circuit; that is, if $\chi = p, \rho, \eta$ or T . But the derivation also demands that χ be a materially conserved quantity, which usually restricts the choice of χ to be η (or potential temperature), or to be ρ itself if the thermodynamic equation is $D\rho/Dt = 0$. Thus, the conservation of potential vorticity for inviscid, adiabatic flow is

$$\frac{D}{Dt} (\widetilde{\omega}_a \cdot \nabla \theta) = 0, \quad (4.77)$$

where $D\theta/Dt = 0$. For diabatic flow source terms appear on the right-hand side, and we derive these later on. A summary of this derivation is provided by Fig. 4.9.

4.5.2 PV conservation from the frozen-in property

In this section we show that potential vorticity conservation is a consequence of the frozen-in property of vorticity. This is not surprising, because the circulation theorem itself has a similar origin. Thus, this derivation is not independent of the derivation in the previous section, just a re-expression of it. We first consider the case in which the solenoidal term vanishes from the outset.

Barotropic fluids

If χ is a materially conserved tracer then the difference in χ between two infinitesimally close fluid elements is also conserved and

$$\frac{D}{Dt}(\chi_1 - \chi_2) = \frac{D\delta\chi}{Dt} = 0. \quad (4.78)$$

But $\delta\chi = \nabla\chi \cdot \delta\mathbf{l}$, where $\delta\mathbf{l}$ is the infinitesimal vector connecting the two fluid elements. Thus

$$\frac{D}{Dt}(\nabla\chi \cdot \delta\mathbf{l}) = 0. \quad (4.79)$$

However, as the line element and the vorticity (divided by density) obey the same equation, we can replace the line element by vorticity (divided by density) in (4.79) to obtain again

$$\frac{D}{Dt}\left(\frac{\nabla\chi \cdot \mathbf{\omega}_a}{\rho}\right) = 0. \quad (4.80)$$

That is, the potential vorticity, $Q = (\widetilde{\mathbf{\omega}}_a \cdot \nabla\chi)$ is a material invariant, where χ is any scalar quantity that satisfies $D\chi/Dt = 0$.

Baroclinic fluids

In baroclinic fluids we cannot casually substitute the vorticity for that of a line element in (4.79) because of the presence of the solenoidal term, and in any case a little more detail would not be amiss. From (4.79) we obtain

$$\delta\mathbf{l} \cdot \frac{D\nabla\chi}{Dt} + \nabla\chi \cdot \frac{D\delta\mathbf{l}}{Dt} = 0 \quad (4.81)$$

or, using (4.31),

$$\delta\mathbf{l} \cdot \frac{D\nabla\chi}{Dt} + \nabla\chi \cdot [(\delta\mathbf{l} \cdot \nabla)\mathbf{v}] = 0. \quad (4.82)$$

Now, let us choose $\delta\mathbf{l}$ to correspond to a vortex line, so that at the initial time $\delta\mathbf{l} = \epsilon\widetilde{\mathbf{\omega}}_a$. (Note that in this case the association of $\delta\mathbf{l}$ with a vortex line can only be made instantaneously, and we cannot set $D\delta\mathbf{l}/Dt \propto D\mathbf{\omega}_a/Dt$.) Then,

$$\widetilde{\mathbf{\omega}}_a \cdot \frac{D\nabla\chi}{Dt} + \nabla\chi \cdot [(\mathbf{\omega}_a \cdot \nabla)\mathbf{v}] = 0, \quad (4.83)$$

or, using the vorticity equation (4.16),

$$\widetilde{\mathbf{\omega}}_a \cdot \frac{D\nabla\chi}{Dt} + \nabla\chi \cdot \left(\frac{D\widetilde{\mathbf{\omega}}_a}{Dt} - \frac{1}{\rho^3}\nabla\rho \times \nabla p\right) = 0. \quad (4.84)$$

This may be written as

$$\frac{D}{Dt}\widetilde{\mathbf{\omega}}_a \cdot \nabla\chi = \frac{1}{\rho^3}\nabla\chi \cdot (\nabla\rho \times \nabla p). \quad (4.85)$$

The term on the right-hand side is, in general, non-zero for an arbitrary choice of scalar, but it will evidently vanish if ∇p , $\nabla\rho$ and $\nabla\chi$ are coplanar. If χ is any function of p and ρ this will be satisfied, but χ must also be a materially conserved scalar. If, as for an ideal gas, $\rho = \rho(\eta, p)$ (or $\eta = \eta(p, \rho)$) where η is the entropy (which is materially conserved), and if χ

is a function of entropy η alone, then χ satisfies both conditions. Explicitly, the solenoidal term vanishes because

$$\nabla \chi \cdot (\nabla \rho \times \nabla p) = \frac{d\chi}{d\eta} \nabla \eta \cdot \left[\left(\frac{\partial \rho}{\partial p} \nabla p + \frac{\partial \rho}{\partial \eta} \nabla \eta \right) \times \nabla p \right] = 0. \quad (4.86)$$

Thus, provided χ satisfies the two conditions

$$\frac{D\chi}{Dt} = 0 \quad \text{and} \quad \chi = \chi(p, \rho), \quad (4.87)$$

then (4.85) becomes

$$\frac{D}{Dt} \left(\frac{\boldsymbol{\omega}_a \cdot \nabla \chi}{\rho} \right) = 0. \quad (4.88)$$

The natural choice for χ is potential temperature, whence

$$\frac{D}{Dt} \left(\frac{\boldsymbol{\omega}_a \cdot \nabla \theta}{\rho} \right) = 0. \quad (4.89)$$

The presence of a density term in the denominator is not necessary for incompressible flows (i.e., if $\nabla \cdot \mathbf{v} = 0$).

4.5.3 PV conservation: an algebraic derivation

Finally, we give an algebraic derivation of potential vorticity conservation. We will take the opportunity to include frictional and diabatic processes, although these may also be included in the derivations above.⁶ We begin with the frictional vorticity equation in the form

$$\frac{D\tilde{\boldsymbol{\omega}}_a}{Dt} = (\tilde{\boldsymbol{\omega}}_a \cdot \nabla) \mathbf{v} + \frac{1}{\rho^3} (\nabla \rho \times \nabla p) + \frac{1}{\rho} (\nabla \times \mathbf{F}). \quad (4.90)$$

where \mathbf{F} represents any non-conservative force term on the right-hand side of the momentum equation (i.e., $D\mathbf{v}/Dt = -\rho^{-1} \nabla p + \mathbf{F}$). We have also the equation for our materially conserved scalar χ ,

$$\frac{D\chi}{Dt} = \dot{\chi}, \quad (4.91)$$

where $\dot{\chi}$ represents any sources and sinks of χ . Now

$$(\tilde{\boldsymbol{\omega}}_a \cdot \nabla) \frac{D\chi}{Dt} = \tilde{\boldsymbol{\omega}}_a \cdot \frac{D\nabla\chi}{Dt} + [(\tilde{\boldsymbol{\omega}}_a \cdot \nabla) \mathbf{v}] \cdot \nabla \chi, \quad (4.92)$$

which may be obtained just by expanding the left-hand side. Thus, using (4.91),

$$\tilde{\boldsymbol{\omega}}_a \cdot \frac{D\nabla\chi}{Dt} = (\tilde{\boldsymbol{\omega}}_a \cdot \nabla) \dot{\chi} - [(\tilde{\boldsymbol{\omega}}_a \cdot \nabla) \mathbf{v}] \cdot \nabla \chi. \quad (4.93)$$

Now take the dot product of (4.90) with $\nabla \chi$:

$$\nabla \chi \cdot \frac{D\tilde{\boldsymbol{\omega}}_a}{Dt} = \nabla \chi \cdot [(\tilde{\boldsymbol{\omega}}_a \cdot \nabla) \mathbf{v}] + \nabla \chi \cdot \left[\frac{1}{\rho^3} (\nabla \rho \times \nabla p) \right] + \nabla \chi \cdot \left[\frac{1}{\rho} (\nabla \times \mathbf{F}) \right]. \quad (4.94)$$

The sum of the last two equations yields

$$\frac{D}{Dt} (\tilde{\omega}_a \cdot \nabla \chi) = \tilde{\omega}_a \cdot \nabla \dot{\chi} + \nabla \chi \cdot \left[\frac{1}{\rho^3} (\nabla \rho \times \nabla p) \right] + \frac{\nabla \chi}{\rho} \cdot (\nabla \times \mathbf{F}). \quad (4.95)$$

This equation reprises (4.85), but with the addition of frictional and diabatic terms. As before, the solenoidal term is annihilated if we choose $\chi = \theta(p, \rho)$, so giving the evolution equation for potential vorticity in the presence of forcing and diabatic terms, namely

$$\frac{D}{Dt} (\tilde{\omega}_a \cdot \nabla \theta) = \tilde{\omega}_a \cdot \nabla \dot{\theta} + \frac{\nabla \theta}{\rho} \cdot (\nabla \times \mathbf{F}). \quad (4.96)$$

4.5.4 Effects of salinity and moisture

For seawater the equation of state may be written as

$$\theta = \theta(\rho, p, S), \quad (4.97)$$

where θ is the potential temperature and S is the salinity. In the absence of diabatic terms (which in general include saline diffusion) the potential temperature is a materially conserved quantity. However, because of the presence of salinity, θ cannot be used to annihilate the solenoidal term; that is

$$\nabla \theta \cdot (\nabla \rho \times \nabla p) = \left(\frac{\partial \theta}{\partial S} \right)_{p, \rho} \nabla S \cdot (\nabla \rho \times \nabla p) \neq 0. \quad (4.98)$$

Strictly speaking then, *there is no potential vorticity conservation principle for seawater*. However, such a blunt statement rather overemphasizes the importance of salinity in the ocean, and the non-conservation of potential vorticity because of this effect is rather small.

In a moist atmosphere in which condensational heating occurs there is no 'moist potential vorticity' that is generally materially conserved.⁷ We may choose to define a moist PV (Q_e say) based on moist equivalent potential temperature but it does not always obey $DQ_e/Dt = 0$.

4.5.5 Effects of rotation, and summary remarks

In a rotating frame the potential vorticity conservation equation is obtained simply by replacing ω_a by $\omega + 2\Omega$, where Ω is the rotation rate of the rotating frame. The operator D/Dt is reference-frame invariant, and so may be evaluated using the usual formulae with velocities measured in the rotating frame.

In the above derivations, we have generally referred to the quantity $\omega_a \cdot \nabla \theta / \rho$ as potential vorticity; however, it is clear that this form is not unique. If θ is a materially conserved variable, then so is $g(\theta)$ where g is any function, so that $\omega_a \cdot \nabla g(\theta) / \rho$ is also a potential vorticity. When such a non-standard definition is used we qualify the expression potential vorticity with some adjective, whereas the unqualified phrase 'potential vorticity', or the phrase 'Ertel potential vorticity', will refer to the standard expression $\omega_a \cdot \nabla \theta / \rho$.

The conservation of potential vorticity has profound consequences in fluid dynamics, especially in a rotating, stratified fluid. The non-conservative terms are often small, and large-scale flow in both the ocean and the atmosphere is characterized by conservation

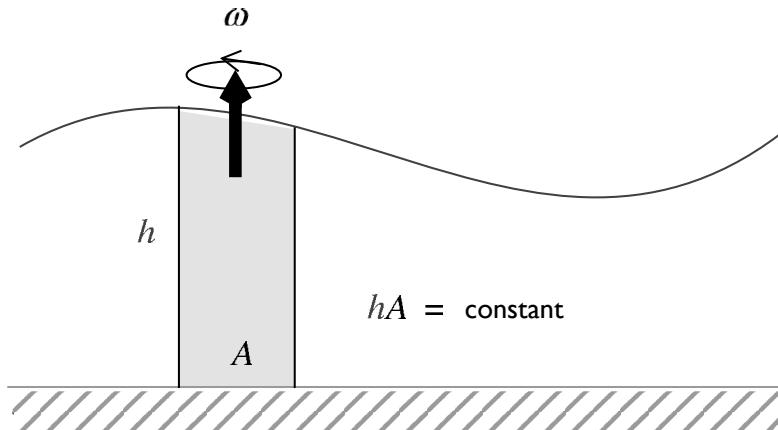


Fig. 4.10 The volume of a column of fluid, hA , is conserved in the shallow water system. Furthermore, the vorticity is tied to material lines so that ζA is also a material invariant, where $\zeta = \omega \cdot \mathbf{k}$ is the vertical component of the vorticity. From this, ζ/h must be materially conserved; that is, $D(\zeta/h)/Dt = 0$, which is the conservation of potential vorticity in a shallow water system. In a rotating system this generalizes to $D[(\zeta + f)/h]/Dt = 0$.

of potential vorticity. Such conservation is a very powerful constraint on the flow, and indeed it turns out that potential vorticity is usually a more useful quantity for baroclinic, or non-homentropic, fluids than for barotropic fluids, because the required use of a special conserved scalar imparts additional information. A large fraction of the remainder of this book explores, in one way or another, the consequences of potential vorticity conservation.

4.6 * POTENTIAL VORTICITY IN THE SHALLOW WATER SYSTEM

In chapter 3 we derived potential vorticity conservation by direct manipulation of the shallow water equations. We now show that shallow water potential vorticity is also derivable from the conservation of circulation. Specifically, we will begin with the three-dimensional form of Kelvin's theorem, and then make the small aspect ratio assumption (which is the key assumption underlying shallow water dynamics), and thereby recover shallow water potential vorticity conservation. In the following two subsections we give two variants of such a derivation (see also Fig. 4.10).

4.6.1 Using Kelvin's theorem

We begin with

$$\frac{D}{Dt} (\omega_3 \cdot \delta S) = 0, \quad (4.99)$$

where ω_3 is the curl of the three-dimensional velocity and $\delta S = \mathbf{n} \delta S$ is an arbitrary infinitesimal vector surface element, with \mathbf{n} being a unit vector pointing in the direction normal to the surface. If we separate the vorticity and the surface element into vertical and

horizontal components we can write (4.99) as

$$\frac{D}{Dt} [(\zeta + f)\delta A + \mathbf{\omega}_h \cdot \delta \mathbf{S}_h] = 0, \quad (4.100)$$

where $\mathbf{\omega}_h$ and $\delta \mathbf{S}_h$ are the horizontally directed components of the vorticity and the surface element, and $\delta A = \mathbf{k} \cdot \delta \mathbf{S}$ is the area of a horizontal cross-section of a fluid column. In Cartesian form the horizontal component of the vorticity is

$$\mathbf{\omega}_h = \mathbf{i} \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \right) - \mathbf{j} \left(\frac{\partial w}{\partial x} - \frac{\partial v}{\partial z} \right) = \mathbf{i} \frac{\partial w}{\partial y} - \mathbf{j} \frac{\partial w}{\partial x}, \quad (4.101)$$

where vertical derivatives of the horizontal velocity are zero by virtue of the nature of the shallow water system. Now, the vertical velocity in the shallow water system is smaller than the horizontal velocity by the order of the aspect ratio — the ratio of the fluid depth to the horizontal scale of the motion. Furthermore, the size of the horizontally directed surface element is also smaller than the vertically-directed component by the aspect ratio; that is,

$$|\mathbf{\omega}_h| \sim \alpha |\zeta| \quad \text{and} \quad |\delta \mathbf{S}_h| \sim \alpha |\delta A|, \quad (4.102)$$

where $\alpha = H/L$ is the aspect ratio. Thus $\mathbf{\omega}_h \cdot \delta \mathbf{S}_h$ is smaller than the term $\zeta \delta A$ by the aspect number squared, and in the small aspect ratio approximation should be neglected. Kelvin's circulation theorem, (4.100), becomes

$$\frac{D}{Dt} [(\zeta + f)\delta A] = 0 \quad \text{or} \quad \frac{D}{Dt} \left[\frac{(\zeta + f)}{h} h \delta A \right] = 0, \quad (4.103a,b)$$

where h is the depth of the fluid column. But $h \delta A$ is the volume of the fluid column, and this is constant. Thus, (4.103b) gives, as in (3.78),

$$\frac{D}{Dt} \left(\frac{\zeta + f}{h} \right) = 0, \quad (4.104)$$

where, because horizontal velocities are independent of the vertical coordinate, the advection is purely horizontal.

4.6.2 Using an appropriate scalar field

In a constant density fluid we can write potential vorticity conservation as

$$\frac{D}{Dt} (\mathbf{\omega}_3 \cdot \nabla \chi) = 0, \quad (4.105)$$

where χ is any materially conserved scalar [cf. (4.75) or (4.80)]. In the flat-bottomed shallow water system, a useful choice of scalar is the ratio z/h , where h is the local thickness of the fluid column because, from (3.28),

$$\frac{D}{Dt} \left(\frac{z}{h} \right) = 0, \quad (4.106)$$

if (for simplicity) the fluid is flat-bottomed. With this choice of scalar, potential vorticity conservation becomes

$$\frac{D}{Dt} \left[\mathbf{\omega} \cdot \nabla \left(\frac{z}{h} \right) \right] = 0, \quad (4.107)$$

where ω and D/Dt are fully three-dimensional. Expanding the dot product gives

$$\frac{D}{Dt} \left(\frac{\zeta + f}{h} - \frac{z}{h^2} \omega_h \cdot \nabla_z h \right) = 0. \quad (4.108)$$

For an order-unity Rossby number, the ratio of the size of the two terms in this equation is

$$\frac{|\zeta|}{|(z/h)\omega_h \cdot \nabla_z h|} \sim \frac{[U/L]}{[WH/L^2]} = \frac{UL}{WH} = \alpha^{-2} \gg 1. \quad (4.109)$$

Thus, the second term in (4.108) is smaller than the first by the aspect-ratio squared and, upon its neglect, (4.104) is recovered.

4.7 POTENTIAL VORTICITY IN APPROXIMATE, STRATIFIED MODELS

If approximate models of stratified flow (Boussinesq, hydrostatic and so on) are to be useful then they should conserve an appropriate form of potential vorticity, and we consider a few such cases here.

4.7.1 The Boussinesq equations

A Boussinesq fluid is incompressible (that is, the volume of a fluid element is conserved, and $\nabla \cdot \mathbf{v} = 0$) and the equation for vorticity itself is isomorphic to that for a line element. However, the Boussinesq equations are not barotropic — $\nabla \rho$ is not parallel to ∇p — and although the pressure gradient term $\nabla \phi$ disappears on taking its curl (or equivalently disappears on integration around a closed path) the buoyancy term $b\mathbf{k}$ does not, and it is this that prevents Kelvin's circulation theorem from holding. Specifically, the evolution of circulation in the Boussinesq equations obeys

$$\frac{D}{Dt} [(\omega_a \cdot \mathbf{n}) \delta A] = (\nabla \times b\mathbf{k}) \cdot \mathbf{n} \delta A, \quad (4.110)$$

where here, as in (4.70), \mathbf{n} is a unit vector orthogonal to an infinitesimal surface element of area δA . The right-hand side is annihilated if we choose \mathbf{n} to be parallel to ∇b , because $\nabla b \cdot \nabla \times (b\mathbf{k}) = 0$. In the simple Boussinesq equations the thermodynamic equation is

$$\frac{Db}{Dt} = 0, \quad (4.111)$$

and potential vorticity conservation is therefore (with $\omega_a = \omega + 2\Omega$)

$$\frac{DQ}{Dt} = 0, \quad Q = (\omega + 2\Omega) \cdot \nabla b. \quad (4.112a,b)$$

Expanding (4.112b) in Cartesian coordinates with $2\Omega = f\mathbf{k}$ we obtain:

$$Q = (v_x - u_y)b_z + (w_y - v_z)b_x + (u_z - w_x)b_y + fb_z. \quad (4.113)$$

In the general Boussinesq equations b itself is not materially conserved. We cannot expect to obtain a conservation law if salinity is present, but if the equation of state and the thermodynamic equation are:

$$b = b(\theta, z), \quad \frac{D\theta}{Dt} = 0, \quad (4.114)$$

then potential vorticity conservation follows, because taking \mathbf{n} to be parallel to $\nabla\theta$ will cause the right-hand side of (4.110) to vanish; that is,

$$\nabla\theta \cdot \nabla \times (b\mathbf{k}) = \left(\frac{\partial\theta}{\partial z} \nabla z + \frac{\partial\theta}{\partial b} \nabla b \right) \cdot \nabla \times (b\mathbf{k}) = 0. \quad (4.115)$$

The materially conserved potential vorticity is thus

$$Q = \boldsymbol{\omega}_a \cdot \nabla\theta. \quad (4.116)$$

Note that if the equation of state is $b = b(\theta, \phi)$, where ϕ is the pressure, then potential vorticity is not conserved because then, in general, $\nabla\phi \cdot \nabla \times (b\mathbf{k}) \neq 0$.

4.7.2 The hydrostatic equations

Making the hydrostatic approximation has no effect on whether or not the circulation theorem is satisfied. Thus, in a baroclinic hydrostatic fluid we have

$$\frac{D}{Dt} \int (\boldsymbol{\omega}_{hy} + 2\boldsymbol{\Omega}) \cdot d\mathbf{S} = - \int \nabla\alpha \times \nabla p \cdot d\mathbf{S}. \quad (4.117)$$

where, from (4.55) $\boldsymbol{\omega}_{hy} = \nabla \times \mathbf{u} = -i\nu_z + j u_z + \mathbf{k}(\nu_x - u_y)$, but the gradient operator and material derivative are fully three-dimensional. Derivation of potential vorticity conservation then proceeds, as in section 4.5.1, by choosing the circuit over which the circulation is calculated to be such that the right-hand side vanishes; that is, to be such that the solenoidal term is annihilated. Precisely as before, this occurs if the circuit is barotropic and without further ado we write

$$\frac{DQ_{hy}}{Dt} = \frac{D}{Dt} \left[\frac{(\boldsymbol{\omega}_{hy} + 2\boldsymbol{\Omega}) \cdot \nabla\theta}{\rho} \right] = 0. \quad (4.118)$$

Expanding this gives in Cartesian coordinates

$$Q_{hy} = \frac{1}{\rho} \left[(\nu_x - u_y)\theta_z - \nu_z\theta_x + u_z\theta_y + 2\Omega\theta_z \right]. \quad (4.119)$$

In spherical coordinates the hydrostatic approximation is usually accompanied by the traditional approximation and the expanded expression for a conserved potential vorticity is more complicated. It can still be derived from Kelvin's theorem, but this is left as an exercise for the reader (problem 4.5).

4.7.3 Potential vorticity on isentropic surfaces

If we begin with the primitive equations in isentropic coordinates then potential vorticity conservation follows quite simply. Cross-differentiating the horizontal momentum equations (3.162) gives the vorticity equation [cf. (3.70)]

$$\frac{D}{Dt}(\zeta + f) + (\zeta + f)\nabla_\theta \cdot \mathbf{u} = 0, \quad (4.120)$$

where $D/Dt = \partial/\partial t + \mathbf{u} \cdot \nabla_\theta$. The thermodynamic equation is

$$\frac{D\sigma}{Dt} + \sigma \nabla \cdot \mathbf{u} = 0, \quad (4.121)$$

where $\sigma = \partial z / \partial b$ (Boussinesq) or $\partial p / \partial \theta$ (ideal gas) is the thickness of an isopycnal layer. Eliminating the divergence between (4.120) and (4.121) gives

$$\frac{D}{Dt} \left(\frac{\zeta + f}{\sigma} \right) = 0. \quad (4.122)$$

The derivation, and the result, are precisely the same as with the shallow water equations (sections 3.6.1 and 4.6).

A connection between isentropic and height coordinates

The hydrostatic potential vorticity written in height coordinates may be transformed into a form that reveals its intimate connection with isentropic surfaces. Let us make the Boussinesq approximation for which the potential vorticity is

$$Q_{hy} = (v_x - u_y)b_z - v_z b_x + u_z b_y, \quad (4.123)$$

where b is the buoyancy. We can write this as

$$Q_{hy} = b_z \left[\left(v_x - v_z \frac{b_x}{b_z} \right) - \left(u_y - u_z \frac{b_y}{b_z} \right) \right]. \quad (4.124)$$

But the terms in the inner brackets are just the horizontal velocity derivatives at constant b . To see this, note that

$$\left(\frac{\partial v}{\partial x} \right)_b = \left(\frac{\partial v}{\partial x} \right)_z + \frac{\partial v}{\partial z} \left(\frac{\partial z}{\partial x} \right)_b = \left(\frac{\partial v}{\partial x} \right)_z - \frac{\partial v}{\partial z} \left(\frac{\partial b}{\partial x} \right)_z / \frac{\partial b}{\partial z}, \quad (4.125)$$

with a similar expression for $(\partial u / \partial y)_b$. (These relationships follow from standard rules of partial differentiation. Derivatives with respect to z are taken at constant x and y .) Thus, we obtain

$$Q_{hy} = \frac{\partial b}{\partial z} \left[\left(\frac{\partial v}{\partial x} \right)_b - \left(\frac{\partial u}{\partial y} \right)_b \right] = \frac{\partial b}{\partial z} \zeta_b. \quad (4.126)$$

Thus, potential vorticity is simply the horizontal vorticity evaluated on a surface of constant buoyancy, multiplied by the vertical derivative of buoyancy. An analogous derivation, with a similar result, proceeds for the ideal gas equations, with potential temperature replacing buoyancy.

4.8 * THE IMPERMEABILITY OF ISENTROPIES TO POTENTIAL VORTICITY

A kinematical result is a result valid forever.

Clifford Truesdell, *The Kinematics of Vorticity*, 1954.

An interesting property of isentropic surfaces is that they are ‘impermeable’ to potential vorticity, meaning that the mass integral of potential vorticity ($\int Q \rho dV$) over a volume bounded by an isentropic surface remains constant, even in the presence of diabatic sources, provided the surfaces do not intersect a non-isentropic surface such as the ground.⁸ This may seem surprising, especially because unlike most conservation laws the result does not require adiabatic flow, and for that reason it leads to interesting interpretations of a number of phenomena. However, at the same time impermeability is a consequence of the definition

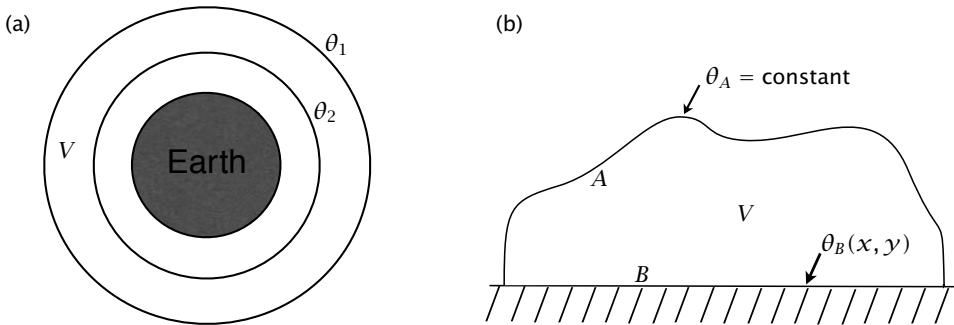


Fig. 4.11 (a) Two isentropic surfaces that do not intersect the ground. The integral of PV concentration over the volume between them, V , is zero, even if there is heating and the contours move. (b) An isentropic surface, A , intersects the ground, B , thus enclosing a volume V . The rate of change of PV concentration over the volume is given by an integral over B .

of potential vorticity rather than the equations of motion, and in that sense is a kinematic property.

To derive the result we define $s \equiv \rho Q = \nabla \cdot (\theta \mathbf{w}_a)$, where \mathbf{w}_a is the absolute vorticity, and integrate over some volume V to give

$$I = \int_V s \, dV = \int_V \nabla \cdot (\theta \mathbf{w}_a) \, dV = \int_S \theta \mathbf{w}_a \cdot d\mathbf{S}, \quad (4.127)$$

using the divergence theorem, where S is the surface surrounding the volume V . If this is an isentropic surface then we have

$$I = \theta \int_S \mathbf{w}_a \cdot d\mathbf{S} = \theta \int_V \nabla \cdot \mathbf{w}_a \, dV = 0, \quad (4.128)$$

again using the divergence theorem. That is, over a volume wholly enclosed by a single isentropic surface the integral of s vanishes. If the volume is bounded by more than one isentropic surface neither of which intersect the surface, for example by concentric spheres of different radii as in Fig. 4.11(a), the result still holds. The quantity s is called 'potential vorticity concentration', or 'PV concentration'. The integral of s over a volume is akin to the total amount of a conserved material property, such as salt content, and so may be called 'PV substance'. That is, the PV concentration is the amount of potential vorticity substance per unit volume (following the meaning for concentration introduced in section 1.2.3 on page 11) and

$$\text{PV substance} = \int s \, dV = \int \rho Q \, dV. \quad (4.129)$$

Suppose now that the fluid volume is enclosed by an isentrope that intersects the ground, as in Fig. 4.11(b). Let A denote the isentropic surface, B denote the ground, θ_A the constant value of θ on the isentrope, and $\theta_B(x, y, t)$ the non-constant value of θ on the ground. The

integral of s over the volume is then

$$\begin{aligned} I &= \int_V \nabla \cdot (\theta \mathbf{w}_a) \, dV = \theta_A \int_A \mathbf{w}_a \cdot d\mathbf{S} + \int_B \theta_B \mathbf{w}_a \cdot d\mathbf{S} \\ &= \theta_A \int_{A+B} \mathbf{w}_a \cdot d\mathbf{S} + \int_B (\theta_B - \theta_A) \mathbf{w}_a \cdot d\mathbf{S} \\ &= \int_B (\theta_B - \theta_A) \mathbf{w}_a \cdot d\mathbf{S}. \end{aligned} \quad (4.130)$$

The first term on the second line vanishes after using the divergence theorem. Thus, the value of I , and hence its rate of change, is a function *only of an integral over the surface B* , and the PV flux there must be calculated using the full equations of motion. However, we do not need to be concerned with any flux of PV concentration through the isentropic surface; put another way, the PV substance in a volume can change only when isentropes enclosing the volume intersect a boundary such as the Earth's surface.

4.8.1 Interpretation and application

Motion of the isentropic surface

How can the above results hold in the presence of heating? The isentropic surfaces must move in such a way that the total amount of PV concentration contained between them nevertheless stays fixed, and we now demonstrate this explicitly. The potential vorticity equation may be written

$$\frac{\partial Q}{\partial t} + \mathbf{v} \cdot \nabla Q = S_Q, \quad (4.131)$$

where, from (4.96), $S_Q = (\mathbf{w}_a/\rho) \cdot \nabla \dot{\theta} + \nabla \theta \cdot (\nabla \times \mathbf{F})/\rho$. Using mass continuity this may be written as

$$\frac{\partial s}{\partial t} + \nabla \cdot \mathbf{J} = 0, \quad (4.132)$$

where $\mathbf{J} \equiv \rho \mathbf{v} Q + \mathbf{N}$ and $\nabla \cdot \mathbf{N} = -\rho S_Q$. Written in this way, the quantity \mathbf{J}/s is a notional velocity, \mathbf{v}_Q say, and s satisfies

$$\frac{\partial s}{\partial t} + \nabla \cdot (\mathbf{v}_Q s) = 0. \quad (4.133)$$

That is, s evolves as if it were being fluxed by the velocity \mathbf{v}_Q . The concentration of a chemical tracer χ (i.e., χ is the amount of tracer per unit volume) obeys a similar equation, to wit

$$\frac{\partial \chi}{\partial t} + \nabla \cdot (\mathbf{v} \chi) = 0. \quad (4.134)$$

However, whereas (4.134) implies that $D(\chi/\rho)/Dt = 0$, (4.133) does not imply that $\partial Q/\partial t + \mathbf{v}_Q \cdot \nabla Q = 0$ because $\partial \rho/\partial t + \nabla \cdot (\rho \mathbf{v}_Q) \neq 0$.

Now, the impermeability result tells us that there can be no notional velocity across an isentropic surface. How can this be satisfied by the equations of motion? We write the right-hand side of (4.131) as

$$\rho S_Q = \nabla \cdot (\dot{\theta} \mathbf{w}_a + \theta \nabla \times \mathbf{F}) = \nabla \cdot (\dot{\theta} \mathbf{w}_a + \mathbf{F} \times \nabla \theta). \quad (4.135)$$

Thus, $\mathbf{N} = -\dot{\theta} \mathbf{w}_a - \mathbf{F} \times \nabla \theta$ and we may write the \mathbf{J} vector as

$$\mathbf{J} = \rho \mathbf{v} Q - \dot{\theta} \mathbf{w}_a - \mathbf{F} \times \nabla \theta = \rho Q (\mathbf{v}_\perp + \mathbf{v}_\parallel) - \dot{\theta} \mathbf{w}_\parallel - \mathbf{F} \times \nabla \theta, \quad (4.136)$$

where, making use of the thermodynamic equation,

$$\mathbf{v}_{\parallel} = \mathbf{v} - \frac{\mathbf{v} \cdot \nabla \theta}{|\nabla \theta|^2} \nabla \theta, \quad \mathbf{v}_{\perp} = -\frac{\partial \theta / \partial t}{|\nabla \theta|^2} \nabla \theta, \quad (4.137a)$$

$$\boldsymbol{\omega}_{\parallel} = \boldsymbol{\omega}_a - \frac{\boldsymbol{\omega}_a \cdot \nabla \theta}{|\nabla \theta|^2} \nabla \theta = \boldsymbol{\omega}_a - \frac{Q\rho}{|\nabla \theta|^2} \nabla \theta. \quad (4.137b)$$

The subscripts ‘ \perp ’ and ‘ \parallel ’ denote components perpendicular and parallel to the local isentropic surface, and \mathbf{v}_{\perp} is the velocity of the isentropic surface normal to itself. Equation (4.136) may be verified by using (4.137) and $D\theta/Dt = \dot{\theta}$.

The ‘parallel’ terms in (4.137) are all vectors parallel to the local isentropic surface, and therefore do not lead to any flux of PV concentration across that surface. Furthermore, the term $\rho Q \mathbf{v}_{\perp}$ is ρQ multiplied by the normal velocity of the surface. That is to say, the notional velocity associated with the flux normal to the isentropic surface is equal to the normal velocity of the isentropic surface itself, and so it too provides no flux of PV concentration across that surface (even though there may well be a mass flux across the surface). Put simply, the isentropic surface always moves in such a way as to ensure that there is no flux of PV concentration across it. In our proof of the impermeability result in the previous subsection we used the fact that the potential vorticity multiplied by the density is the divergence of a vector. In the demonstration above we used the fact that the terms *forcing* potential vorticity are the divergence of a vector.

* Dynamical choices of PV flux and a connection to Bernoulli's theorem

If we add a non-divergent vector to the flux, \mathbf{J} , then it has no effect on the evolution of s . This gauge invariance means that the notional velocity, $v_Q = \mathbf{J}/(\rho Q)$ is similarly non-unique, although it does not mean that there are not dynamical choices for it that are more appropriate in given circumstances. To explore this, let us obtain a general expression for \mathbf{J} by starting with the definition of s , so that

$$\begin{aligned} \frac{\partial s}{\partial t} &= \nabla \theta \cdot \frac{\partial \boldsymbol{\omega}_a}{\partial t} + \boldsymbol{\omega}_a \cdot \nabla \frac{\partial \theta}{\partial t} \\ &= \nabla \theta \cdot \nabla \times \frac{\partial \mathbf{v}}{\partial t} + \nabla \cdot \left(\boldsymbol{\omega}_a \frac{\partial \theta}{\partial t} \right) = -\nabla \cdot \mathbf{J}', \end{aligned} \quad (4.138)$$

where

$$\mathbf{J}' = \nabla \theta \times \frac{\partial \mathbf{v}}{\partial t} - \frac{\partial \theta}{\partial t} \boldsymbol{\omega}_a + \nabla \phi \times \nabla \chi. \quad (4.139)$$

The last term in this expression is an arbitrary divergence-free vector. If we choose $\phi = \theta$ and $\chi = B$, where B is the Bernoulli function given by $B = I + \mathbf{v}^2/2 + p/\rho$ where I is the internal energy per unit mass, then

$$\mathbf{J}' = \nabla \theta \times \left(\nabla B + \frac{\partial \mathbf{v}}{\partial t} \right) - \boldsymbol{\omega}_a (\dot{\theta} - \mathbf{v} \cdot \nabla \theta), \quad (4.140)$$

having used the thermodynamic equation $D\theta/Dt = \dot{\theta}$. Now, the momentum equation may be written, without approximation, in the form (see problems 2.2 and 2.3)

$$\frac{\partial \mathbf{v}}{\partial t} = -\boldsymbol{\omega}_a \times \mathbf{v} + T \nabla \eta + \mathbf{F} - \nabla B, \quad (4.141)$$

where η is the specific entropy ($d\eta = c_p d \ln \theta$). Using (4.140) and (4.141) gives

$$\mathbf{J}' = \rho Q \mathbf{v} - \dot{\theta} \mathbf{w}_a + \nabla \theta \times \mathbf{F}. \quad (4.142)$$

which is the same as (4.136). Furthermore, using (4.139) for steady flow,

$$\mathbf{J} = \nabla \theta \times \nabla B. \quad (4.143)$$

That is, the flux of potential vorticity (in this gauge) is aligned with the intersection of θ - and B -surfaces. For steady *inviscid and adiabatic* flow the Bernoulli function is constant along streamlines; that is, surfaces of constant Bernoulli function are aligned with streamlines, and, because θ is materially conserved, streamlines are formed at intersecting θ - and B -surfaces, as in (1.190). In the presence of forcing, this property is replaced by (4.143), and the flux of PV concentration is along such intersections.

This choice of gauge leading to (4.142) is physical in that it reduces to the true advective flux $\mathbf{v} \rho Q$ for unforced, adiabatic flow, but it is not a unique choice, nor is it mandated by the dynamics. Choosing $\chi = 0$ leads to the flux

$$\mathbf{J}_1 = \rho Q \mathbf{v} - \dot{\theta} \mathbf{w}_a + \nabla \theta \times (\mathbf{F} - \nabla B), \quad (4.144)$$

and using (4.139) this vanishes for steady flow — a potentially useful property.

Summary remarks

The impermeability result is kinematic, but it nevertheless provides an interesting point of view and useful diagnostic tool.⁹ In summary, we will just make the following remarks.

- ★ There can be no net transport of potential vorticity across an isentropic surface, and the total amount of potential vorticity in a volume wholly enclosed by isentropic surfaces is zero. Thus, and with hindsight trivially, the amount of potential vorticity contained between two isentropes isolated from the Earth's surface in the Northern Hemisphere is the negative of the corresponding amount in the Southern Hemisphere.
- ★ Potential vorticity flux lines (i.e., lines everywhere parallel to \mathbf{J}) can either close in on themselves or begin and end at boundaries (e.g., the ground or the ocean surface). However, \mathbf{J} may change its character. Thus, for example, at the base of the oceanic mixed layer \mathbf{J} may change from being a diabatic flux above to an adiabatic advective flux below. There may be a similar change in character at the atmospheric tropopause.
- ★ The flux vector \mathbf{J} is defined only to within the curl of a vector. Thus the vector $\mathbf{J}' = \mathbf{J} + \nabla \times \mathbf{A}$, where \mathbf{A} is arbitrary, is as valid as is \mathbf{J} in the above derivations.

Notes

- 1 The frozen-in property — that vortex lines are material lines — was derived by Helmholtz (1858) and is sometimes called Helmholtz's theorem.
- 2 The theorem originates with Thomson (1869), who later became later Lord Kelvin.
- 3 Silberstein (1896) proved that 'the necessary and sufficient condition for the generation of vortical flow... influenced only by conservative forces ... is that the surface of constant

pressure and surface of constant density... intersect', as we derived in section 4.2, and this leads to (4.43). Bjerknes (1898a,b) explicitly put this into the form of a circulation theorem and applied it to problems of meteorological and oceanographic importance (see Thorpe *et al.* 2003), and the theorem is sometimes called the Bjerknes theorem or the Bjerknes–Silberstein theorem.

Vilhelm Bjerknes (1862–1951) was a physicist and hydrodynamicist who in 1917 moved to the University of Bergen as founding head of the Bergen Geophysical Institute. Here he did what was probably his most influential work in meteorology, setting up and contributing to the 'Bergen School of Meteorology'. Among other things he and his colleagues were the first to consider, as a practical proposition, the use of numerical methods — initial data in conjunction with the fluid equations of motion — to forecast the state of the atmosphere, based on earlier work describing how that task might be done (Bjerknes 1904). Inaccurate initial velocity fields compounded with the shear complexity of the effort ultimately defeated them, but the effort was continued (also unsuccessfully) by L. F. Richardson (Richardson 1922), before J. Charney, R. Fjørtoft and J. Von Neumann eventually made what may be regarded as the first successful numerical forecast (Charney *et al.* 1950). Their success can be attributed to the used of a simplified, filtered, set of equations and the use of an electronic computer.

Vilhelm's son, Jacob Bjerknes (1897–1975) was a leading player in the Bergen school. He was responsible for the now-famous frontal model of cyclones (Bjerknes 1919), and was one of the first to seriously discuss the role of cyclones in the general circulation of the atmosphere. In collaboration with Halvor Solberg and Tor Bergeron the frontal model lead to a prescient picture of the lifecycle of extratropical cyclones (see chapter 9), in which a wave grows initially on the polar front (akin to baroclinic instability with the meridional temperature gradient compressed to a front, but baroclinic instability theory was not yet developed), develops into a mature cyclone, occludes and decays. In 1939 Bjerknes moved to the U.S. and, largely because of World War II, stayed, joining UCLA and heading its Department of Meteorology after its formation in 1945. He developed an interest in air-sea interactions, and notably proposed the essential mechanism governing El Niño, a feedback between sea-surface temperatures and the strength of the trade winds (Bjerknes 1969). [See also Friedman (1989), Cressman (1996), Shapiro & Grønås (1999), and the memoir by Arnt Eliassen at <http://www.nap.edu/readingroom/books/biomems/jbjerknes.html>.]

- 4 The result (4.50) was given by Poincaré (1893) (section 158), although it is sometimes attributed to Bjerknes (1902).
- 5 The first derivation of the PV conservation law was given for the hydrostatic shallow water equations by Rossby (1936), with a generalization to the stratified case, via the use of isentropic coordinates, in Rossby (1938) and Rossby (1940). In the 1936 paper Rossby noted [his eq. (75)] that a fluid column satisfies $f + \zeta = cD$, where c is a constant and D is the thickness of a fluid column; equivalently, $(f + \zeta)/D$ is a material invariant. The expression 'potential vorticity' was introduced in Rossby (1940), as follows: '*This quantity, which may be called the potential vorticity, represents the vorticity the air column would have if it were brought, isopycnally or isentropically, to a standard latitude (f_0) and stretched or shrunk vertically to a standard depth D_0 or weight Δ_0 .*' (Rossby's italics.) That is,

$$\text{potential vorticity} = \zeta_0 = \left(\frac{\zeta + f}{D} \right) D_0 - f_0, \quad (4.145)$$

which follows from his eq. (11), and this is the sense he uses it in that paper. However, potential vorticity has come to mean the quantity $(\zeta + f)/D$, which of course does not have the dimensions of vorticity. We use it in this latter, now conventional, sense throughout

this book. Ironically, quasi-geostrophic potential vorticity as usually defined does have the dimensions of vorticity.

The expression for potential vorticity in a non-hydrostatic, continuously stratified fluid was given by Ertel (1942a), and its relationship to circulation was given by Ertel (1942b). It is now commonly known as the *Ertel potential vorticity*. Interestingly, in Rossby (1940) we find the Fermat-like comment ‘It is possible to derive corresponding results for an atmosphere in which the potential temperature varies continuously with elevation. . . . The generalized treatment will be presented in another place.’ Given his prior (1938) derivation of the stratified quantity in isentropic coordinates, he must not have regarded his own derivations as very general. Opinions differ as to whether Rossby’s and Ertel’s derivations were independent, and Cressman (1996) remarks that the origin of the concept of potential vorticity is a ‘delicate one that has aroused some passion in private correspondences’. In fact, Ertel visited MIT in autumn 1937 and presumably talked to Rossby and became aware of his work. It seems almost certain that Ertel knew of Rossby’s shallow water and isentropic theorems, but it is also clear that Ertel subsequently provided a significant generalization, most likely independently. Rossby and Ertel apparently remained on good terms, but further collaboration was stymied by World War II. They later published a pair of short joint papers, one in German and the other in English, describing related conservation theorems (Ertel & Rossby 1949a,b). English translations of a number of Ertel’s papers are to be found in Schubert *et al.* (2004).

- 6 Truesdell (1951, 1954) and Obukhov (1962) were early explorers of the consequences of heating and friction on potential vorticity.
- 7 Schubert *et al.* (2001) provide more discussion of this topic. They derive a ‘moist PV’ that, although not materially conserved, is an extension of the dry Ertel PV to moist atmospheres and has an impermeability result.
- 8 Haynes & McIntyre (1987, 1990). See also Danielsen (1990), Schär (1993) [who obtained the result (4.143)], Bretherton & Schär (1993) and Davies-Jones (2003).
- 9 See, for example, McIntyre & Norton (1990) and Marshall & Nurser (1992). The latter use J vectors to study the creation and transport of potential vorticity in the oceanic thermocline.

Further reading

Truesdell, C., 1954. *The Kinematics of Vorticity*.

Written in Truesdell’s inimitable style, this book discusses many aspects of vorticity and has many historical references (as well as a generous definition of what constitutes a ‘kinematic’ result).

Batchelor, G. K., 1967. *An Introduction to Fluid Dynamics*.

Contains an extensive discussion of vorticity and vortices.

Dutton, J. A., 1986. *The Ceaseless Wind: An Introduction to the Theory of Atmospheric Motion*.

Contains derivations of the main vorticity theorems as well as more general material on the dynamics of atmospheric flows.

Salmon, R., 1998. *Lecture on Geophysical Fluid Dynamics*.

Chapter 4 contains a brief discussion of potential vorticity, and chapter 7 a longer discussion of Hamiltonian fluid dynamics, in which the particle relabeling symmetry that gives rise to potential vorticity conservation is discussed.

Problems

- 4.1 For the vr vortex, choose a contour of arbitrary shape (e.g., a square) with segments neither parallel nor perpendicular to the radius, and not enclosing the origin. Show explicitly that the circulation around the contour is zero.

- 4.2 Show that the viscous term in the momentum equation vanishes for a fluid in solid-body rotation.
- 4.3 ♦ *Vortex stretching and viscosity*
Suppose there is an incompressible swirling flow given in cylindrical coordinates (r, ϕ, z) :

$$\mathbf{v} = (v_r, v_\phi, v_z) = \left(-\frac{1}{2} \alpha r, v_\phi, \alpha z \right). \quad (\text{P4.1})$$

Show that this satisfies the mass conservation equation. Also, show that vorticity is only non-zero in the vertical direction. Show that the vertical component of the vorticity equation contains only the stretching term and that in a steady state it is

$$-\frac{1}{2} \alpha r \frac{\partial \zeta}{\partial r} = \zeta \alpha + \nu \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \zeta}{\partial r} \right). \quad (\text{P4.2})$$

Show that this may be integrated to $\zeta = \zeta_0 \exp(-\alpha r^2/4\nu)$. Thus deduce that there is a rotational core of thickness $r_o = 2(\nu/\alpha)^{1/2}$, and that the radial velocity field is given by

$$v_\phi(r) = -\frac{1}{r} \frac{2\nu}{\alpha} \zeta_0 \exp \left[-\frac{\alpha r^2}{4\nu} \right] + \frac{A}{r} \quad (\text{P4.3})$$

where $A = 2\nu\zeta_0/\alpha$. What is the swirling velocity field? (Adapted from Batchelor 1967.)

- 4.4 Beginning with the three-dimensional vorticity equation in a rotating frame of reference [e.g., (4.58) or (4.59)], or otherwise, obtain an expression for the evolution of the vertical and radial coordinate of relative vorticity in Cartesian and spherical coordinates, respectively. Discuss the differences (if any) between the resulting equations and (4.66). Show carefully how the β -term arises, and in particular that it may be interpreted as arising from tilting term in the vorticity equation.
- 4.5 ♦ Making use of Kelvin's circulation theorem obtain an expression for the potential vorticity that is conserved following the flow (for an adiabatic and unforced fluid) and that is appropriate for the hydrostatic primitive equations on a spherical planet. Express this in terms of the components of the spherical coordinate system.
- 4.6 In pressure coordinates for hydrostatic flow on the f -plane, the horizontal momentum equation takes the form

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -\nabla \phi \quad (\text{P4.4})$$

On taking the curl of this, there appears to be no baroclinic term. Show that Kelvin's circulation theorem is nevertheless not in general satisfied, even for unforced, adiabatic flow. By appropriately choosing a path on which to evaluate the circulation, obtain an expression for potential vorticity, in this coordinate system, that is conserved following the flow. *Hint:* look at the hydrostatic equation.

- 4.7 ♦ *Shallow water hurricanes*

Consider the shallow water equations on an f -plane, flat bottom, forced by an imposed steady mass source, and damped by a linear drag:

$$\frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{u}h) = Q(x, y) \quad (\text{P4.5})$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + \mathbf{f}_0 \times \mathbf{u} = -g \nabla h - \alpha \mathbf{u}, \quad (\text{P4.6})$$

where α is a positive constant. The term Q represents fluid being added to or removed from the layer by heating.

- (a) If we consider the fluid layer to be near the surface of the atmosphere, cumulus convection leads to a mass sink with $Q < 0$. Linearize the equations about a state of rest and solve them to obtain steady solutions for the radial and tangential velocities, and the height field, for the mass sink

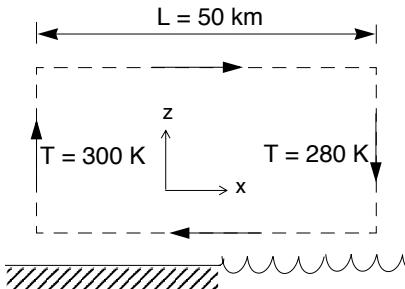
$$Q = -Q_0, \quad r < L; \quad Q = 0 \quad r > L \quad (\text{P4.7})$$

with $Q_0 > 0$. Here $r = \sqrt{x^2 + y^2}$ is radius from the origin. (You may wish to work in plane polar coordinates).

- (b) Obtain a scaling estimate for the value of Q_0 at which the linearization will become invalid.
 (c) Now consider the *upper* troposphere in a hurricane, where convection leads to an effective mass source so Q is *positive* in some region and zero outside that. Without linearizing the momentum equation, show that in the region of the mass source, in the limit as $\alpha \rightarrow 0$, the flow must obtain a state of zero absolute vorticity, $\zeta + f = 0$. You may use the spatial form of the mass source from the previous part of the problem as necessary, although only its sign and not its spatial distribution matters. Explain why this flow violates quasi-geostrophic scaling. (Contributed by Adam Sobel.)

4.8 Solenoids and sea breezes

A land-sea temperature contrast of 20 K forces a sea breeze in the surface 'mixed layer' (potential temperature nearly uniform with height), as illustrated schematically. The layer extends through the lowest 10% of the mass of the atmosphere.



- (a) In the absence of dissipation and diffusion, at what rate does the circulation change on the material circuit indicated? You may assume the horizontal flow is isobaric, and express your answer in m s^{-1} per hour.
 (b) Suppose the sea breeze is equilibrated by a nonlinear surface drag of the form

$$\frac{dV}{dt} = -\frac{V^2}{L_F} \quad (\text{P4.8})$$

with $L_F = (3 \text{ m s}^{-1})(3600 \text{ s})$. What is the steady speed of the horizontal wind in the case $L = 50 \text{ km}$?

- (c) Suppose that the width of the circulation is determined by a horizontal thermal diffusion of the form

$$\frac{D\theta}{Dt} = \kappa_H \frac{\partial^2 \theta}{\partial x^2} \quad (\text{P4.9})$$

Provide an estimate of κ_H that is consistent with $L = 50 \text{ km}$. Comment on whether you think the extent of real sea breezes is really determined this way.

This extreme generality whereby the equations of motion apply to the entire spectrum of possible motions — to sound waves as well as cyclone waves — constitutes a serious defect of the equations from the meteorological point of view.

Jule Charney, *On the scale of atmospheric motions*, 1948.

CHAPTER FIVE

Simplified Equations for the Ocean and Atmosphere

LARGE-SCALE FLOW IN THE OCEAN AND THE ATMOSPHERE is characterized by an approximate balance in the vertical direction between the pressure gradient and gravity (hydrostatic balance), and in the horizontal direction between the pressure gradient and the Coriolis force (geostrophic balance). In this chapter we exploit these balances to simplify the Navier-Stokes equations and thereby obtain various sets of simplified 'geostrophic equations'. Depending on the precise nature of the assumptions we make, we are led to the *quasi-geostrophic* (QG) system for horizontal scales similar to that on which most synoptic activity takes place and, for very large-scale motion, to the *planetary-geostrophic* (PG) set of equations. By eliminating unwanted or unimportant modes of motion, in particular sound waves and gravity waves, and by building in the important balances between flow fields, these filtered equation sets allow the investigator to better focus on a particular class of phenomena and to potentially achieve a deeper understanding than might otherwise be possible.¹

Simplifying the equations in this way relies first on scaling the equations. The idea is that we *choose* the scales we wish to describe, typically either on some a priori basis or by using observations as a guide. We then attempt to derive a set of equations that is simpler than the original set but that consistently describes motion of the chosen scale. An asymptotic method is one approach to this, for it systematically tells us which terms we can drop and which we should keep. The combined approach — scaling plus asymptotics — has proven enormously useful, but it is useful to always remember two things: (i) that scaling is a choice; (ii) that the approach does not explain the existence of particular scales of motion, it

just describes the motion that might occur on such scales. We have already employed this general approach in deriving the hydrostatic primitive equations, but now we go further.

5.1 GEOSTROPHIC SCALING

I have no satisfaction in formulas unless I feel their numerical magnitude.

William Thomson, Lord Kelvin (1824–1907).

5.1.1 Scaling in the shallow water equations

Postponing the complications that come with stratification, we begin with the shallow water equations. With the odd exception, we will denote the scales of variables by capital letters; thus, if L is a typical length scale of the motion we wish to describe, and U is a typical velocity scale, and assuming the scales are horizontally isotropic, we write

$$\begin{aligned} (x, y) &\sim L & \text{or} & (x, y) = \mathcal{O}(L) \\ (u, v) &\sim U & \text{or} & (u, v) = \mathcal{O}(U). \end{aligned} \quad (5.1)$$

and similarly for other variables. We may then non-dimensionalize the variables by writing

$$(x, y) = L(\hat{x}, \hat{y}), \quad (u, v) = U(\hat{u}, \hat{v}), \quad (5.2)$$

where the hatted variables are non-dimensional and, by supposition, are $\mathcal{O}(1)$. The various terms in the momentum equation then scale as:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + \mathbf{f} \times \mathbf{u} = -g \nabla \eta, \quad (5.3a)$$

$$\frac{U}{T} \quad \frac{U^2}{L} \quad fU \quad g \frac{\mathcal{H}}{L}, \quad (5.3b)$$

where the ∇ operator acts in the x, y plane and \mathcal{H} is the amplitude of the variations in the surface displacement. (We use η to denote the height of the free surface above some arbitrary reference level, as in Fig. 3.1. Thus, $\eta = H + \Delta\eta$, where $\Delta\eta$ denotes the variation of η about its mean position.)

The ratio of the advective term to the rotational term in the momentum equation (5.3) is $(U^2/L)/(fU) = U/fL$; this is the Rossby number, first encountered in chapter 2.² Using values typical of the large-scale circulation (e.g., from Table 2.1) we find that $Ro \approx 0.1$ for the atmosphere and $Ro \approx 0.01$ for the ocean: small in both cases. If we are interested in motion that has the advective time scale $T = L/U$ then we scale time by L/U so that

$$t = \frac{L}{U} \hat{t}, \quad (5.4)$$

and the local time derivative and the advective term then both scale as U^2/L , and both are smaller than the rotation term by a factor of the order of the Rossby number. Then, either the Coriolis term is the dominant term in the equation, in which case we have a state of no motion with $-f v = 0$, or else the Coriolis force is balanced by the pressure force, and the dominant balance is

$$-f v = -g \frac{\partial \eta}{\partial x}, \quad (5.5)$$

namely *geostrophic balance*, as encountered in chapter 2. If we make this non-trivial choice, then the equation informs us that variations in η scale according to

$$\Delta\eta \sim \mathcal{H} = \frac{fUL}{g}. \quad (5.6)$$

We can also write \mathcal{H} as

$$\mathcal{H} = Ro \frac{f^2 L^2}{g} = Ro H \frac{L^2}{L_d^2}, \quad (5.7)$$

where $L_d = \sqrt{gH}/f$ is the deformation radius and H is the mean depth of the fluid. The variations in fluid height thus scale as

$$\frac{\Delta\eta}{H} \sim Ro \frac{L^2}{L_d^2}, \quad (5.8)$$

and the height of the fluid may be written as

$$\eta = H \left(1 + Ro \frac{L^2}{L_d^2} \hat{\eta} \right) \quad \text{and} \quad \Delta\eta = Ro \frac{L^2}{L_d^2} H \hat{\eta}, \quad (5.9)$$

where $\hat{\eta}$ is the $\mathcal{O}(1)$ non-dimensional value of the surface height deviation.

Non-dimensional momentum equation

If we use (5.9) to scale height variations, (5.2) to scale lengths and velocities, and (5.4) to scale time, then the momentum equation (5.3) becomes

$$Ro \left[\frac{\partial \hat{\mathbf{u}}}{\partial \hat{t}} + (\hat{\mathbf{u}} \cdot \nabla) \hat{\mathbf{u}} \right] + \hat{\mathbf{f}} \times \hat{\mathbf{u}} = -\nabla \hat{\eta}, \quad (5.10)$$

where $\hat{\mathbf{f}} = \mathbf{k} \hat{f} = \mathbf{k} f / f_0$, where f_0 is a representative value of the Coriolis parameter. (If f is a constant, then $\hat{f} = 1$, but it is useful to keep it in the equations to indicate the presence of the Coriolis parameter. Also, where the operator ∇ operates on a non-dimensional variable then the differentials are taken with respect to the non-dimensional variables \hat{x}, \hat{y} .) All the variables in (5.10) will be assumed to be of order unity, and the Rossby number multiplying the local time derivative and the advective terms indicates the smallness of those terms. By construction, the dominant balance in this equation is the geostrophic balance between the last two terms.

Non-dimensional mass continuity (height) equation

The (dimensional) mass continuity equation can be written as

$$\frac{1}{H} \frac{D\eta}{Dt} + \left(1 + \frac{\Delta\eta}{H} \right) \nabla \cdot \mathbf{u} = 0, \quad (5.11)$$

Using (5.2), (5.4) and (5.9) this equation may be written

$$Ro \left(\frac{L}{L_d} \right)^2 \frac{D\hat{\eta}}{D\hat{t}} + \left[1 + Ro \left(\frac{L}{L_d} \right)^2 \hat{\eta} \right] \nabla \cdot \hat{\mathbf{u}} = 0. \quad (5.12)$$

Equations (5.10) and (5.12) are the non-dimensional versions of the full shallow water equations of motion. Evidently, some terms in the equations of motion are small and may be eliminated with little loss of accuracy, and the way this is done will depend on the size of the second non-dimensional parameter, $(L/L_d)^2$. We explore this in sections 5.2 and 5.3.

Froude and Burger numbers

The Froude number may be generally defined as the ratio of a fluid particle speed to a wave speed. In a shallow water system this gives

$$Fr \equiv \frac{U}{\sqrt{gH}} = \frac{U}{f_0 L_d} = Ro \frac{L}{L_d}. \quad (5.13)$$

The Burger number³ is a useful measure of the scale of motion of the fluid, relative to the deformation radius, and may be defined by

$$Bu \equiv \left(\frac{L_d}{L} \right)^2 = \frac{gH}{f_0^2 L^2} = \left(\frac{Ro}{Fr} \right)^2. \quad (5.14)$$

It is also useful to define the parameter $F \equiv Bu^{-1}$, which is like the square of a Froude number but uses the rotational speed fL instead of U in the numerator.

5.1.2 Geostrophic scaling in the stratified equations

We now apply the same scaling ideas, *mutatis mutandis*, to the stratified primitive equations. We use the hydrostatic anelastic equations, which we write as

$$\frac{D\mathbf{u}}{Dt} + \mathbf{f} \times \mathbf{u} = -\nabla_z \phi, \quad (5.15a)$$

$$\frac{\partial \phi}{\partial z} = b, \quad (5.15b)$$

$$\frac{Db}{Dt} = 0, \quad (5.15c)$$

$$\nabla \cdot (\tilde{\rho} \mathbf{v}) = 0. \quad (5.15d)$$

where b is the buoyancy and $\tilde{\rho}$ is a reference density profile. Anticipating that the average stratification may not scale in the same way as the deviation from it, let us separate out the contribution of the advection of a reference stratification in (5.15c) by writing

$$b = \tilde{b}(z) + b'(x, y, z, t). \quad (5.16)$$

Then the thermodynamic equation becomes

$$\frac{Db'}{Dt} + N^2 w = 0, \quad (5.17)$$

where $N^2 \equiv \partial \tilde{b} / \partial z$ (and the advective derivative is still three-dimensional). We then let $\phi = \tilde{\phi}(z) + \phi'$, where $\tilde{\phi}$ is hydrostatically balanced by \tilde{b} , and the hydrostatic equation becomes

$$\frac{\partial \phi'}{\partial z} = b'. \quad (5.18)$$

Equations (5.17) and (5.18) replace (5.15c) and (5.15b), and ϕ' is used in (5.15a).

Non-dimensional equations

We scale the basic variables by supposing that

$$(x, y) \sim L, \quad (u, v) \sim U, \quad t \sim \frac{L}{U}, \quad z \sim H, \quad f \sim f_0, \quad (5.19)$$

where the scaling variables (capitalized, except for f_0) are chosen to be such that the non-dimensional variables have magnitudes of the order of unity. We presume that the scales chosen are such that the Rossby number is small; that is $Ro = U/(f_0 L) \ll 1$. In the momentum equation the pressure term then balances the Coriolis force,

$$|\mathbf{f} \times \mathbf{u}| \sim |\nabla \phi'| \quad (5.20)$$

and so the pressure scales as

$$\phi' \sim \Phi = f_0 U L. \quad (5.21)$$

Using the hydrostatic relation, (5.21) implies that the buoyancy scales as

$$b' \sim B = \frac{f_0 U L}{H}, \quad (5.22)$$

and from this we obtain

$$\frac{(\partial b' / \partial z)}{N^2} \sim Ro \frac{L^2}{L_d^2}, \quad (5.23)$$

where $L_d = NH/f_0$ is the deformation radius in the continuously stratified fluid, analogous to the quantity \sqrt{gH}/f_0 in the shallow water system, and we use the same symbol, L_d , for both. In the continuously stratified system, *if the scale of motion is the same as or smaller than the deformation radius, and the Rossby number is small, then the variations in stratification are small*. The choice of scale is the key difference between the planetary-geostrophic and quasi-geostrophic equations.

Finally, we will non-dimensionalize the vertical velocity by using the mass conservation equation,

$$\frac{1}{\tilde{\rho}} \frac{\partial \tilde{\rho} w}{\partial z} = - \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right), \quad (5.24)$$

and we suppose that this implies

$$w \sim W = \frac{UH}{L}. \quad (5.25)$$

This is a naïve scaling for rotating flow: if the Coriolis parameter is nearly constant the geostrophic velocity is nearly horizontally non-divergent and the right-hand side of (5.24) is small, and $W \ll UH/L$. We might then estimate w by cross-differentiating geostrophic balance to obtain the linear geostrophic vorticity equation and corresponding scaling:

$$\beta v \approx f \frac{\partial w}{\partial z}, \quad w \sim W = \frac{\beta UH}{f_0}. \quad (5.26a,b)$$

However, rather than using (5.26b) from the outset, we will use (5.25) and let the asymptotics guide us to a proper scaling in the fullness of time. Note that if variations in the Coriolis parameter are large and $\beta \sim f_0/L$, then (5.26b) is the same as (5.25).

Non-dimensional Primitive Equations

Horizontal momentum: $Ro \frac{D\hat{\mathbf{u}}}{Dt} + \hat{\mathbf{f}} \times \hat{\mathbf{u}} = -\nabla \hat{\phi} \quad (\text{NDPE.1})$

Hydrostatic: $\frac{\partial \hat{\phi}}{\partial \hat{z}} = \hat{b} \quad (\text{NDPE.2})$

Mass continuity: $\left(\frac{\partial \hat{u}}{\partial \hat{x}} + \frac{\partial \hat{v}}{\partial \hat{y}} + \frac{1}{\tilde{\rho}} \frac{\partial \tilde{\rho} \hat{w}}{\partial \hat{z}} \right) = 0 \quad (\text{NDPE.3})$

Thermodynamic: $Ro \frac{D\hat{b}}{Dt} + \left(\frac{L_d}{L} \right)^2 \hat{w} = 0 \quad (\text{NDPE.4})$

These equations are written for the anelastic equations. The Boussinesq equations result if we take $\tilde{\rho} = 1$. The equations in pressure coordinates have a very similar form to the Boussinesq equations, but with a slight difference in hydrostatic equation.

Given the scalings above [using (5.25) for w] we non-dimensionalize by setting

$$\begin{aligned} (\hat{x}, \hat{y}) &= L^{-1}(x, y), & \hat{z} &= H^{-1}z, & (\hat{u}, \hat{v}) &= U^{-1}(u, v), & \hat{t} &= \frac{U}{L}t, \\ \hat{w} &= \frac{L}{UH}w, & \hat{f} &= f_0^{-1}f, & \hat{\phi} &= \frac{\phi'}{f_0UL}, & \hat{b} &= \frac{H}{f_0UL}b', \end{aligned} \quad (5.27)$$

where the hatted variables are non-dimensional. The horizontal momentum and hydrostatic equations then become

$$Ro \frac{D\hat{\mathbf{u}}}{Dt} + \hat{\mathbf{f}} \times \hat{\mathbf{u}} = -\nabla \hat{\phi}, \quad (5.28)$$

and

$$\frac{\partial \hat{\phi}}{\partial \hat{z}} = \hat{b}. \quad (5.29)$$

The non-dimensional mass conservation equation is simply

$$\frac{1}{\tilde{\rho}} \nabla \cdot (\tilde{\rho} \hat{\mathbf{v}}) = \left(\frac{\partial \hat{u}}{\partial \hat{x}} + \frac{\partial \hat{v}}{\partial \hat{y}} + \frac{1}{\tilde{\rho}} \frac{\partial \tilde{\rho} \hat{w}}{\partial \hat{z}} \right) = 0, \quad (5.30)$$

and the non-dimensional thermodynamic equation is

$$\frac{f_0UL}{H} \frac{U}{L} \frac{D\hat{b}}{Dt} + N^2 \frac{HU}{L} \hat{w} = 0, \quad (5.31)$$

or

$$Ro \frac{D\hat{b}}{Dt} + \left(\frac{L_d}{L} \right)^2 \hat{w} = 0. \quad (5.32)$$

The non-dimensional primitive equations are summarized in the box above.

5.2 THE PLANETARY-GEOSTROPHIC EQUATIONS

We now use the low Rossby number scalings above to derive equation sets that are simpler than the original, ‘primitive’, ones. The planetary-geostrophic equations are probably the simplest such set of equations, and we derive these equations first for the shallow water equations, and then for the stratified primitive equations.

5.2.1 Using the shallow water equations

Informal derivation

The advection and time derivative terms in the momentum equation (5.10) are order Rossby number smaller than the Coriolis and pressure terms (the term in square brackets is multiplied by Ro), and therefore let us neglect them. The momentum equation straightforwardly becomes

$$\hat{\mathbf{f}} \times \hat{\mathbf{u}} = -\nabla \hat{\eta}. \quad (5.33)$$

The mass conservation equation (5.12), contains two non-dimensional parameters, $Ro = U/(f_0 L)$ (the Rossby number), and $F = (L/L_d)^2$ (the ratio of the length scale of the motion to the deformation scale; $F = Bu^{-1}$) and we must make a choice as to the relationship between these two numbers. We will choose

$$FRo = \mathcal{O}(1), \quad (5.34)$$

which implies

$$L^2 \gg L_d^2 \quad \text{or equivalently} \quad F \gg 1, \quad Bu \ll 1. \quad (5.35)$$

That is to say, we suppose that the scales of motion are much larger than the deformation scale. Given this choice, all the terms in the mass conservation equation, (5.12), are of roughly the same size, and we retain them all. Thus, the shallow water planetary geostrophic equations are the full mass continuity equation along with geostrophic balance and a geometric relationship between the height field and the fluid thickness, and in dimensional form these are:

$$\frac{Dh}{Dt} + h \nabla \cdot \mathbf{u} = 0$$

$$\mathbf{f} \times \mathbf{u} = -g \nabla \eta, \quad \eta = h + \eta_b$$

(5.36a,b)

We emphasize that *the planetary-geostrophic equations are only valid for scales of motion much larger than the deformation radius*. The height variations are then as large as the mean height field itself; that is, using (5.8), $\Delta\eta/H = \mathcal{O}(1)$.

Formal derivation

We make the following assumptions.

- (i) The Rossby number is small. $Ro = U/f_0 L \ll 1$.
- (ii) The scale of the motion is significantly larger than the deformation scale. That is, (5.34) holds or

$$F = Bu^{-1} = \left(\frac{L}{L_d} \right)^2 \gg 1 \quad (5.37)$$

and in particular

$$FRo = \mathcal{O}(1). \quad (5.38)$$

(iii) Time scales advectively, so that $T = L/U$.

The idea is now to expand the non-dimensional variables velocity and height fields in an asymptotic series with the Rossby number as the small parameter, substitute into the equations of motion and derive a simpler set of equations. It is a nearly trivial exercise in this instance, and so it illustrates the methodology well. The expansions are

$$\hat{\mathbf{u}} = \hat{\mathbf{u}}_0 + Ro \hat{\mathbf{u}}_1 + Ro^2 \hat{\mathbf{u}}_2 + \dots \quad (5.39a)$$

and

$$\hat{\eta} = \hat{\eta}_0 + Ro \hat{\eta}_1 + Ro^2 \hat{\eta}_2 + \dots \quad (5.39b)$$

Then substituting (5.39a) and (5.39b) into the momentum equation gives

$$Ro \left[\frac{\partial \hat{\mathbf{u}}_0}{\partial \hat{t}} + \hat{\mathbf{u}}_0 \cdot \nabla \hat{\mathbf{u}}_0 + \hat{\mathbf{f}} \times \hat{\mathbf{u}}_1 \right] + \hat{\mathbf{f}} \times \hat{\mathbf{u}}_0 = -\nabla \hat{\eta}_0 - Ro [\nabla \hat{\eta}_1] + \mathcal{O}(Ro^2) \quad (5.40)$$

The Rossby number is an asymptotic ordering parameter; thus, the sum of all the terms at any particular order in Rossby number must vanish. At lowest order we obtain the simple expression

$$\hat{\mathbf{f}} \times \hat{\mathbf{u}}_0 = -\nabla \hat{\eta}_0. \quad (5.41)$$

Note that although f_0 is a representative value of f , we have made no assumptions about the constancy of f . In particular, f is allowed to vary by an order one amount, provided that it does not become so small that the Rossby number $U/(f_0 L)$ is not small.

The appropriate height (mass conservation) equation is similarly obtained by substituting (5.39a) and (5.39b) into the shallow water mass conservation equation. Because $FRo = \mathcal{O}(1)$ at lowest order we simply retain all the terms in the equation to give

$$FRo \left[\frac{\partial \hat{\eta}_0}{\partial \hat{t}} + \hat{\mathbf{u}}_0 \cdot \nabla \hat{\eta}_0 \right] + [1 + FRo \hat{\eta}] \nabla \cdot \hat{\mathbf{u}}_0 = 0. \quad (5.42)$$

Equations (5.41) and (5.42) are a closed set, and constitute the non-dimensional planetary-geostrophic equations. The dimensional forms of these equations are just (5.36).

Variation of the Coriolis parameter

Suppose then that f is a constant (f_0). Then, from the curl of (5.41), $\nabla \cdot \mathbf{u}_0 = 0$. This means that we can define a streamfunction for the flow and, from geostrophic balance, the height field is just that streamfunction. That is, in dimensional form,

$$\psi = \frac{g}{f_0} \eta, \quad \mathbf{u} = -\mathbf{k} \times \nabla \psi, \quad (5.43)$$

and (5.42) becomes, in dimensional form,

$$\frac{\partial \eta}{\partial t} + \mathbf{u} \cdot \nabla \eta = 0, \quad \text{or} \quad \frac{\partial \eta}{\partial t} + J(\psi, \eta) = 0, \quad (5.44)$$

where $J(a, b) \equiv a_x b_y - a_y b_x$. But since $\eta \propto \psi$ the advective term is proportional to $J(\psi, \psi)$, which is zero. Thus, the flow does not evolve at this order. The planetary-geostrophic equations are *uninteresting* if the scale of the motion is such that the Coriolis parameter is not variable. On Earth, the scale of motion on which this parameter regime

exists is rather limited, since the planetary-geostrophic equations require that the scale of motion also be larger than the deformation radius. In the Earth's atmosphere, any scale that is larger than the deformation radius will be such that the Coriolis parameter varies significantly over it, and we do not encounter this parameter regime. On the other hand, in the Earth's ocean the deformation radius is relatively small and there exists a small parameter regime that has scales larger than the deformation radius but smaller than that on which the Coriolis parameter varies.⁴

Potential vorticity

The shallow water PG equations may be written as an evolution equation for an approximated potential vorticity. A little manipulation reveals that (5.36) are equivalent to:

$$\boxed{\frac{DQ}{Dt} = 0} \quad (5.45)$$

$$Q = \frac{f}{h}, \quad \mathbf{f} \times \mathbf{u} = -g \nabla \eta, \quad \eta = h + \eta_b$$

Thus, potential vorticity is a material invariant in the approximate equation set, just as it is in the full equations. The other variables — the free surface height and the velocity — are diagnosed from it, a process known as *potential vorticity inversion*. In the planetary geostrophic approximation, the inversion proceeds using the approximate form f/h rather than the full potential vorticity, $(f + \zeta)/h$. (Strictly speaking, we do not approximate potential vorticity, because this is the evolving variable. Rather, we approximate the inversion relations from which we derive the height and velocity fields.) The simplest way of all to derive the shallow water PG equations is to *begin* with the conservation of potential vorticity, and to note that at small Rossby number the expression $(\zeta + f)/h$ may be approximated by f/h . Then, noting in addition that the flow is geostrophic, (5.45) immediately emerges. Every approximate set of equations that we derive in this chapter may be expressed as the evolution of potential vorticity, with the other fields being obtained diagnostically from it.

5.2.2 Planetary-geostrophic equations for stratified flow

To explore the stratified system we will use the inviscid and adiabatic Boussinesq equations of motion with the hydrostatic approximation. The derivation carries through easily enough using the anelastic or pressure-coordinate equations, but as the PG equations have more oceanographic than atmospheric importance, using the incompressible equations is quite appropriate.

Simplifying the equations

The non-dimensional equations we begin with are (5.28)–(5.32). As in the shallow water case we expand these in a series in the Rossby number, so that:

$$\hat{u} = \hat{u}_0 + Ro \hat{u}_1 + Ro^2 \hat{u}_2 + \dots, \quad \hat{b} = \hat{b}_0 + Ro \hat{b}_1 + Ro^2 \hat{b}_2 + \dots, \quad (5.46)$$

and similarly for \hat{v} , \hat{w} and $\hat{\phi}$. Substituting into the non-dimensional equations of motion (on page 202) and equating powers of Ro gives the lowest-order momentum, hydrostatic,

and mass conservation equations:

$$\hat{\mathbf{f}} \times \hat{\mathbf{u}}_0 = -\nabla \hat{\phi}_0, \quad \frac{\partial \hat{\phi}_0}{\partial \hat{z}} = \hat{b}_0, \quad \nabla \cdot \hat{\mathbf{v}}_0 = 0. \quad (5.47a,b,c)$$

If we also assume that $L_d/L = \mathcal{O}(1)$, then the thermodynamic equation (5.32) becomes

$$\left(\frac{L_d}{L}\right)^2 \hat{w}_0 = 0. \quad (5.48)$$

Of course we have neglected any diabatic terms in this equation, which would in general provide a non-zero right-hand side. Nevertheless, this is not a useful equation, because the set of the equations we have derived, (5.47) and (5.48), can no longer evolve: all the time derivatives have been scaled away! Thus, although instructive, these equations are not very useful. If instead we assume that the scale of motion is much larger than the deformation scale then the other terms in the thermodynamic equation will become equally important. Thus, we suppose that $L_d^2 \ll L^2$ or, more formally, that $L^2 = \mathcal{O}(Ro^{-1})L_d^2$, and then all the terms in the thermodynamic equation are retained. A closed set of equations is then given by (5.47) and the thermodynamic equation (5.32).

Dimensional equations

Restoring the dimensions, dropping the asymptotic subscripts, and allowing for the possibility of a source term, denoted by $S[b']$, in the thermodynamic equation, the *planetary-geostrophic* equations of motion are:

$$\frac{Db'}{Dt} + wN^2 = S[b']$$

$$\mathbf{f} \times \mathbf{u} = -\nabla \phi', \quad \frac{\partial \phi'}{\partial z} = b', \quad \nabla \cdot \mathbf{v} = 0$$

(5.49)

The thermodynamic equation may also be written simply as

$$\frac{Db}{Dt} = \dot{b}, \quad (5.50)$$

where b now represents the total stratification. The relevant pressure, ϕ , is then the pressure that is in hydrostatic balance with b , so that geostrophic and hydrostatic balance are most usefully written as

$$\mathbf{f} \times \mathbf{u} = -\nabla \phi, \quad \frac{\partial \phi}{\partial z} = b. \quad (5.51a,b)$$

Potential vorticity

Manipulation of (5.49) reveals that we can equivalently write the equations as an evolution equation for potential vorticity. Thus, the evolution equations may be written as

$$\frac{DQ}{Dt} = \dot{Q}$$

$$Q = f \frac{\partial b}{\partial z},$$

(5.52)

where $\dot{Q} = f \partial \dot{b} / \partial z$, and the inversion — i.e., the diagnosis of velocity, pressure and buoyancy — is carried out using the hydrostatic, geostrophic and mass conservation equations.

Applicability to the ocean and atmosphere

In the atmosphere a typical deformation radius NH/f is about 1 000 km. The constraint that the scale of motion be much larger than the deformation radius is thus quite hard to satisfy, since one quickly runs out of room on a planet whose equator-to-pole distance is 10 000 km. Thus, only the largest planetary waves can satisfy the planetary-geostrophic scaling in the atmosphere and we should then also write the equations in spherical coordinates. In the ocean the deformation radius is about 100 km, so there is lots of room for the planetary-geostrophic equations to hold, and indeed much of the theory of the large-scale structure of the ocean involves the planetary-geostrophic equations.

5.3 THE SHALLOW WATER QUASI-GEOSTROPHIC EQUATIONS

We now derive a set of geostrophic equations that is valid (unlike the PG equations) when the horizontal scale of motion is similar to that of the deformation radius. These equations are called the *quasi-geostrophic* equations, and are perhaps the most widely used set of equations for theoretical studies of the atmosphere and ocean. The specific assumptions we make are as follows.

- (i) The Rossby number is small, so that the flow is in near-geostrophic balance.
- (ii) The scale of the motion is not significantly larger than the deformation scale. Specifically, we shall require that

$$Ro \left(\frac{L}{L_d} \right)^2 = \mathcal{O}(Ro). \quad (5.53)$$

For the shallow water equations, this assumption implies, using (5.9), that the variations in fluid depth are small compared to its total depth. For the continuously stratified system it implies, using (5.23), that the variations in stratification are small compared to the background stratification.

- (iii) Variations in the Coriolis parameter are small; that is, $|\beta L| \ll |f_0|$ where L is the length scale of the motion.
- (iv) Time scales advectively; that is, the scaling for time is given by $T = L/U$.

The second and third of these differ from the planetary-geostrophic counterparts: we make the second assumption because we wish to explore a different parameter regime, and we then find that the third assumption is necessary to avoid a rather trivial state [i.e., a leading order balance of $\beta v = 0$, see the discussion surrounding (5.77)]. All of the assumptions are the same whether we consider the shallow water equations or a continuously stratified flow, and in this section we consider the former.

5.3.1 Single-layer shallow water quasi-geostrophic equations

The algorithm is, again, to expand the variables $\hat{u}, \hat{v}, \hat{\eta}$ in an asymptotic series with the Rossby number as the small parameter, substitute into the equations of motion, and derive a simpler set of equations. Thus we let

$$\hat{u} = \hat{u}_o + Ro \hat{u}_1 + Ro^2 \hat{u}_2 + \dots, \quad \hat{v} = \hat{v}_o + Ro \hat{v}_1 + Ro^2 \hat{v}_2 + \dots \quad (5.54a)$$

$$\hat{\eta} = \hat{\eta}_0 + Ro \hat{\eta}_1 + Ro^2 \hat{\eta}_2 \dots . \quad (5.54b)$$

We will recognize the smallness of β compared to f_0/L by letting $\beta = \hat{\beta}U/L^2$, where $\hat{\beta}$ is assumed to be a parameter of order unity. Then the expression $f = f_0 + \beta y$ becomes

$$\hat{f} = f/f_0 = \hat{f}_0 + Ro\hat{\beta}\hat{y}. \quad (5.55)$$

where \hat{f}_0 is the non-dimensional value of f_0 ; its value is unity, but it is helpful to denote it explicitly. Substitute (5.54) into the non-dimensional momentum equation (5.10), and equate powers of Ro . At lowest order we obtain

$$\hat{f}_0 \hat{u}_0 = -\frac{\partial \hat{\eta}_0}{\partial \hat{y}}, \quad \hat{f}_0 \hat{v}_0 = \frac{\partial \hat{\eta}_0}{\partial \hat{x}}. \quad (5.56)$$

Cross-differentiating gives

$$\nabla \cdot \hat{\mathbf{u}}_0 = 0, \quad (5.57)$$

where, when ∇ operates on a non-dimensional variable, the derivatives are taken with respect to the non-dimensional variables \hat{x} and \hat{y} . From (5.57) we see that the velocity field is divergence-free, and that this arises from the momentum equation rather than the mass conservation equation.

The mass conservation equation is also, at lowest order, $\nabla \cdot \hat{\mathbf{u}}_0 = 0$, and at next order we have

$$F \frac{\partial \hat{\eta}_0}{\partial \hat{t}} + F \hat{\mathbf{u}}_0 \cdot \nabla \hat{\eta}_0 + \nabla \cdot \hat{\mathbf{u}}_1 = 0 \quad (5.58)$$

This equation is not closed, because the evolution of the zeroth-order term involves evaluation of a first-order quantity. For closure, we go to the next order in the momentum equation,

$$\frac{\partial \hat{\mathbf{u}}_0}{\partial \hat{t}} + (\hat{\mathbf{u}}_0 \cdot \nabla) \hat{\mathbf{u}}_0 + \hat{\beta} \hat{y} \mathbf{k} \times \hat{\mathbf{u}}_0 + \hat{f}_0 \mathbf{k} \times \hat{\mathbf{u}}_1 = -\nabla \hat{\eta}_1, \quad (5.59)$$

and take its curl to give the vorticity equation:

$$\frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{\mathbf{u}}_0 \cdot \nabla) (\hat{\zeta}_0 + \hat{\beta} \hat{y}) = -\hat{f}_0 \nabla \cdot \hat{\mathbf{u}}_1. \quad (5.60)$$

The term on the right-hand side is the *vortex stretching* term. Only vortex stretching by the background or planetary vorticity is present, because the vortex stretching by the relative vorticity is smaller by a factor of the Rossby number. Equation (5.60) is also not closed; however, we may use (5.58) to eliminate the divergence term to give

$$\frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{\mathbf{u}}_0 \cdot \nabla) (\hat{\zeta}_0 + \hat{\beta} \hat{y}) = \hat{f}_0 \left(F \frac{\partial \hat{\eta}_0}{\partial \hat{t}} + F \hat{\mathbf{u}}_0 \cdot \nabla \hat{\eta}_0 \right), \quad (5.61)$$

or

$$\frac{\partial}{\partial \hat{x}} (\hat{\zeta}_0 - \hat{f}_0 F \hat{\eta}_0) + (\hat{\mathbf{u}}_0 \cdot \nabla) (\hat{\zeta}_0 + \hat{\beta} \hat{y} - F \hat{f}_0 \hat{\eta}_0) = 0. \quad (5.62)$$

The final step is to note that the lowest-order vorticity and height fields are related through geostrophic balance, so that using (5.56) we can write

$$\hat{u}_0 = -\frac{\partial \hat{\psi}_0}{\partial \hat{y}}, \quad \hat{v}_0 = \frac{\partial \hat{\psi}_0}{\partial \hat{x}}, \quad \hat{\zeta}_0 = \nabla^2 \hat{\psi}_0, \quad (5.63)$$

where $\hat{\psi}_0 = \hat{\eta}_0 / \hat{f}_0$ is the streamfunction. Equation (5.62) can thus be written as

$$\frac{\partial}{\partial \hat{t}} (\nabla^2 \hat{\psi}_0 - \hat{f}_0^2 F \hat{\psi}_0) + (\hat{\mathbf{u}}_0 \cdot \nabla) (\hat{\zeta}_0 + \hat{\beta} \hat{\gamma} - \hat{f}_0^2 F \hat{\psi}_0) = 0 \quad (5.64)$$

or

$$\frac{D_0}{D \hat{t}} (\nabla^2 \hat{\psi}_0 + \hat{\beta} \hat{\gamma} - \hat{f}_0^2 F \hat{\psi}_0) = 0, \quad (5.65)$$

where the subscript '0' on the material derivative indicates that the lowest order velocity, the geostrophic velocity, is the advecting velocity. Restoring the dimensions, (5.65) becomes

$$\frac{D}{Dt} \left(\nabla^2 \psi + \beta \gamma - \frac{1}{L_d^2} \psi \right) = 0, \quad (5.66)$$

where $\psi = (g/f_0)\eta$, $L_d^2 = gH/f_0^2$, and the advective derivative is

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u_g \frac{\partial}{\partial x} + v_g \frac{\partial}{\partial y} = \frac{\partial}{\partial t} - \frac{\partial \psi}{\partial y} \frac{\partial}{\partial x} + \frac{\partial \psi}{\partial x} \frac{\partial}{\partial y} = \frac{\partial}{\partial t} + J(\psi, \cdot). \quad (5.67)$$

Another form of (5.66) is

$$\frac{D}{Dt} \left(\zeta + \beta \gamma - \frac{f_0}{H} \eta \right) = 0, \quad (5.68)$$

with $\zeta = (g/f_0)\nabla^2 \eta$. Equations (5.66) and (5.68) are forms of the shallow water quasi-geostrophic potential vorticity equation. The quantity

$$q \equiv \zeta + \beta \gamma - \frac{f_0}{H} \eta = \nabla^2 \psi + \beta \gamma - \frac{1}{L_d^2} \psi \quad (5.69)$$

is the *shallow water quasi-geostrophic potential vorticity*.

Connection to shallow water potential vorticity

The quantity q given by (5.69) is an approximation (except for dynamically unimportant constant additive and multiplicative factors) to the shallow water potential vorticity. To see the truth of this statement, begin with the expression for the shallow water potential vorticity,

$$Q = \frac{f + \zeta}{h}. \quad (5.70)$$

Now let $h = H(1 + \eta'/H)$, where η' is the perturbation of the free-surface height, and assume that η'/H is small to obtain

$$Q = \frac{f + \zeta}{H(1 + \eta'/H)} \approx \frac{1}{H} (f + \zeta) \left(1 - \frac{\eta'}{H} \right) \approx \frac{1}{H} \left(f_0 + \beta \gamma + \zeta - f_0 \frac{\eta'}{H} \right). \quad (5.71)$$

Because f_0/H is a constant it has no effect in the evolution equation, and the quantity given by

$$q = \beta \gamma + \zeta - f_0 \frac{\eta'}{H} \quad (5.72)$$

is materially conserved. Using geostrophic balance we have $\zeta = \nabla^2 \psi$ and $\eta' = f_0 \psi / g$ so that (5.72) is identical to (5.69). [Note that only the variation in η is important in (5.68) or (5.69).]

The approximations needed to go from (5.70) to (5.72) are the same as those used in our earlier, more long-winded, derivation of the quasi-geostrophic equations. That is, we assumed that f itself is nearly constant, and that f_0 is much larger than ζ , equivalent to a low Rossby number assumption. It was also necessary to assume that $H \gg \eta'$ to enable the expansion of the height field which, using assumption (ii) on page 207, is equivalent to requiring that the scale of motion not be significantly larger than the deformation scale. The derivation is completed by noting that the advection of the potential vorticity should be by the geostrophic velocity alone, and we recover (5.66) or (5.68).

Two interesting limits

There are two interesting limits to the quasi-geostrophic potential vorticity equation which, taking $\beta = 0$ for simplicity, are as follows.

- (i) *Motion on scales much smaller than the deformation radius.* That is, $L \ll L_d$ and thus $Bu \gg 1$ or $F \ll 1$. Then (5.66) becomes

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) = 0, \quad (5.73)$$

where $\zeta = \nabla^2 \psi$ and $J(\psi, \zeta) = \psi_x \zeta_y - \psi_y \zeta_x$. Thus, the motion obeys the two-dimensional vorticity equation. Physically, on small length scales the deviations in the height field are very small and may be neglected.

- (ii) *Motion on scales much larger than the deformation radius.* Although scales are not allowed to become so large that $Ro(L/L_d)^2$ is of order unity, we may, a posteriori, still have $L \gg L_d$, whence the potential vorticity equation, (5.66), becomes

$$\frac{\partial \psi}{\partial t} + J(\psi, \psi) = 0 \quad \text{or} \quad \frac{\partial \eta}{\partial t} + J(\psi, \eta) = 0, \quad (5.74)$$

because $\psi = g\eta/f_0$. The Jacobian term evidently vanishes. Thus, one is left with a trivial equation that implies there is no advective evolution of the height field. There is nothing wrong with our reasoning; the mathematics has indeed pointed out a limit interesting in its uninterestingness. From a physical point of view, however, such a lack of motion is likely to be rare, because on such large scales the Coriolis parameter varies considerably, and we are led to the planetary-geostrophic equations.

In practice, often the most severe restriction of quasi-geostrophy is that variations in layer thickness are small: what does this have to do with geostrophy? If we scale η assuming geostrophic balance then $\eta \sim fUL/g$ and $\eta/H \sim Ro(L/L_d)^2$. Thus, if Ro is to remain small, η/H can only be of order one if $(L/L_d)^2 \gg 1$. That is, the height variations must occur on a large scale, or we are led to a scaling inconsistency. Put another way, *if there are order-one height variations over a length scale of less than or of the order of the deformation scale, the Rossby number will not be small.* Large height variations are allowed if the scale of motion is large, but this contingency is described by the planetary-geostrophic equations.

Another flow regime

Although perhaps of little terrestrial interest, we can imagine a regime in which the Coriolis parameter varies fully, but the scale of motion remains no larger than the deformation radius. This parameter regime is not quasi-geostrophic, but it gives an interesting result. Because $\eta'/H \sim Ro(L/L_d)^2$ deviations of the height field are at least of order Rossby number smaller than the reference height and $|\eta'| \ll H$. The dominant balance in the height equation is then

$$H \nabla \cdot \mathbf{u} = 0, \quad (5.75)$$

presuming that time still scales advectively. This zero horizontal divergence must remain consistent with geostrophic balance

$$\mathbf{f} \times \mathbf{u} = -g \nabla \eta, \quad (5.76)$$

where now f is a fully variable Coriolis parameter. Taking the curl of (that is, cross-differentiating) (5.76) gives

$$\beta v + f \nabla \cdot \mathbf{u} = 0, \quad (5.77)$$

whence, using (5.75), $v = 0$, and the flow is purely zonal. Although not at all useful as an evolution equation, this illustrates the constraining effect that differential rotation has on meridional velocity. This effect may be the cause of the banded, highly zonal flow on some of the giant planets, and we will revisit this issue in our discussion of geostrophic turbulence.

5.3.2 Two-layer and multi-layer quasi-geostrophic systems

Just as for the one-layer case, the multi-layer shallow water equations simplify to a corresponding quasi-geostrophic system in appropriate circumstances. The assumptions are virtually the same as before, although we assume that the variation in the thickness of *each* layer is small compared to its mean thickness. The basic fluid system for a two-layer case is sketched in Fig. 5.1 (and see also Fig. 3.5), and for the multi-layer case in Fig. 5.2.

Let us proceed directly from the potential vorticity equation for each layer. We will also stay in dimensional variables, foregoing a strict asymptotic approach for the sake of informality and insight, and use the Boussinesq approximation. For each layer the potential vorticity equation is just

$$\frac{DQ_i}{Dt} = 0, \quad Q_i = \frac{\zeta_i + f}{h_i}. \quad (5.78)$$

Let $h_i = H_i + h'_i$ where $|h'_i| \ll H_i$. The potential vorticity then becomes

$$Q_i \approx \frac{1}{H_i} (\zeta_i + f) \left(1 - \frac{h'_i}{H_i} \right) \quad - \text{variations in layer thickness are small} \quad (5.79a)$$

$$\approx \frac{1}{H_i} \left(f + \zeta_i - f \frac{h'_i}{H_i} \right) \quad - \text{the Rossby number is small} \quad (5.79b)$$

$$\approx \frac{1}{H_i} \left(f + \zeta_i - f_0 \frac{h'_i}{H_i} \right) \quad - \text{variations in Coriolis parameter are small.} \quad (5.79c)$$

Now, because Q appears in the equations only as an advected quantity, it is only the *variations* in the Coriolis parameter that are important in the first term on the right-hand

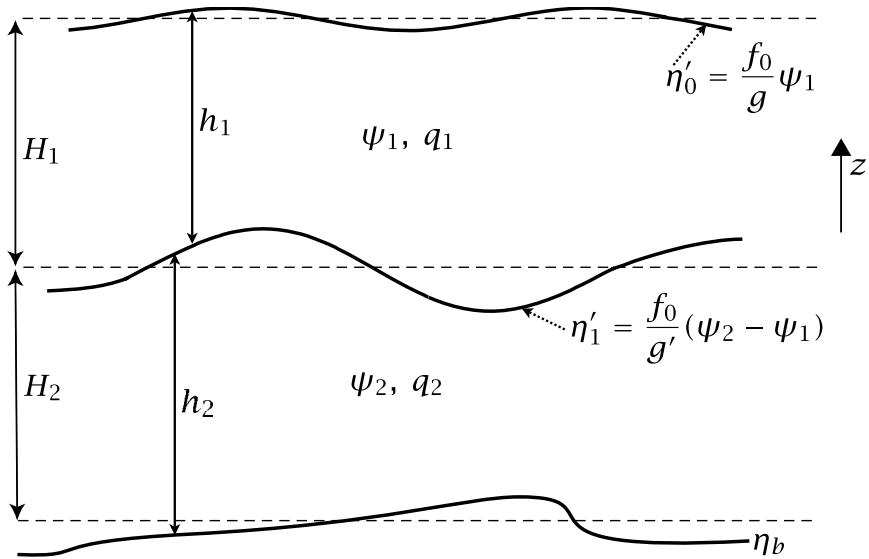


Fig. 5.1 A quasi-geostrophic fluid system consisting of two immiscible fluids of different density. The quantities η' are the interface displacements from the resting basic state, denoted with dashed lines, with η_b being the bottom topography.

side of (5.79c), and given this all three terms are of the same approximate magnitude. Then, because mean layer thicknesses are constant, we can define the quasi-geostrophic potential vorticity in each layer by

$$q_i = \left(\beta y + \zeta_i - f_0 \frac{h'_i}{H_i} \right), \quad (5.80)$$

and this will evolve according to $Dq_i/Dt = 0$, where the advective derivative is by the geostrophic wind. As in the one-layer case, the quasi-geostrophic potential vorticity has different dimensions from the full shallow water potential vorticity.

Two-layer model

To obtain a closed set of equations we must obtain an advecting field from the potential vorticity. We use geostrophic balance to do this, and neglecting the advective derivative in (3.51) gives

$$f_0 \times \mathbf{u}_1 = -g \nabla \eta_0 = -g \nabla (h'_1 + h'_2 + \eta_b), \quad (5.81a)$$

$$f_0 \times \mathbf{u}_2 = -g \nabla \eta_0 - g' \nabla \eta_1 = -g \nabla (h'_1 + h'_2 + \eta_b) - g' \nabla (h'_2 + \eta_b), \quad (5.81b)$$

where $g' = (\rho_2 - \rho_1)/\rho_1$ and η_b is the height of any bottom topography, and, because variations in the Coriolis parameter are presumptively small, we use a constant value of f (i.e., f_0) on the left-hand side. For each layer there is therefore a streamfunction, given by

$$\psi_1 = \frac{g}{f_0} (h'_1 + h'_2 + \eta_b), \quad \psi_2 = \frac{g}{f_0} (h'_1 + h'_2 + \eta_b) + \frac{g'}{f_0} (h'_2 + \eta_b), \quad (5.82a,b)$$

and these two equations may be manipulated to give

$$h'_1 = \frac{f_0}{g'}(\psi_1 - \psi_2) + \frac{f_0}{g}\psi_1, \quad h'_2 = \frac{f_0}{g'}(\psi_2 - \psi_1) - \eta_b. \quad (5.83a,b)$$

We note as an aside that the interface displacements are given by

$$\eta'_0 = \frac{f_0}{g}\psi_1, \quad \eta'_1 = \frac{f_0}{g'}(\psi_2 - \psi_1). \quad (5.84a,b)$$

Using (5.80) and (5.83) the quasi-geostrophic potential vorticity for each layer becomes

$$q_1 = \beta y + \nabla^2 \psi_1 + \frac{f_0^2}{g'H_1}(\psi_2 - \psi_1) - \frac{f_0^2}{gH_1}\psi_1$$

$$q_2 = \beta y + \nabla^2 \psi_2 + \frac{f_0^2}{g'H_2}(\psi_1 - \psi_2) + f_0 \frac{\eta_b}{H_2}$$

(5.85a,b)

In the rigid-lid approximation the last term in (5.85a) is neglected. The potential vorticity in each layer is advected by the geostrophic velocity, so that the evolution equation for each layer is just

$$\frac{\partial q_i}{\partial t} + J(\psi_i, q_i) = 0, \quad i = 1, 2. \quad (5.86)$$

* Multi-layer model

A multi-layer quasi-geostrophic model may be constructed by a straightforward extension of the above two-layer procedure (see Fig. 5.2). The quasi-geostrophic potential vorticity for each layer is still given by (5.80). The pressure field in each layer can be expressed in terms of the thickness of each layer using (3.46) and (3.47) on page 131, and by geostrophic balance the pressure is proportional to the streamfunction, ψ_i , for each layer. Carrying out these steps we obtain, after a little algebra, the following expression for the quasi-geostrophic potential vorticity of an interior layer, in the Boussinesq approximation:

$$q_i = \beta y + \nabla^2 \psi_i + \frac{f_0^2}{H_i} \left(\frac{\psi_{i-1} - \psi_i}{g'_{i-1}} - \frac{\psi_i - \psi_{i+1}}{g'_i} \right),$$

(5.87)

and for the top and bottom layers,

$$q_1 = \beta y + \nabla^2 \psi_1 + \frac{f_0^2}{H_1} \left(\frac{\psi_2 - \psi_1}{g'_1} \right) - \frac{f_0^2}{gH_1}\psi_1, \quad (5.88a)$$

$$q_N = \beta y + \nabla^2 \psi_N + \frac{f_0^2}{H_N} \left(\frac{\psi_{N-1} - \psi_N}{g'_{N-1}} \right) + \frac{f_0}{H_N}\eta_b. \quad (5.88b)$$

In these equations H_i is the basic-state thickness of the i th layer, and $g'_i = g(\rho_{i+1} - \rho_i)/\rho_1$. In each layer the evolution equation is (5.86), now for $i = 1 \dots N$. The displacements of each interface are given, similarly to (5.84), by

$$\eta'_0 = \frac{f_0}{g}\psi_1, \quad \eta'_i = \frac{f_0}{g'_i}(\psi_{i+1} - \psi_i). \quad (5.89a,b)$$

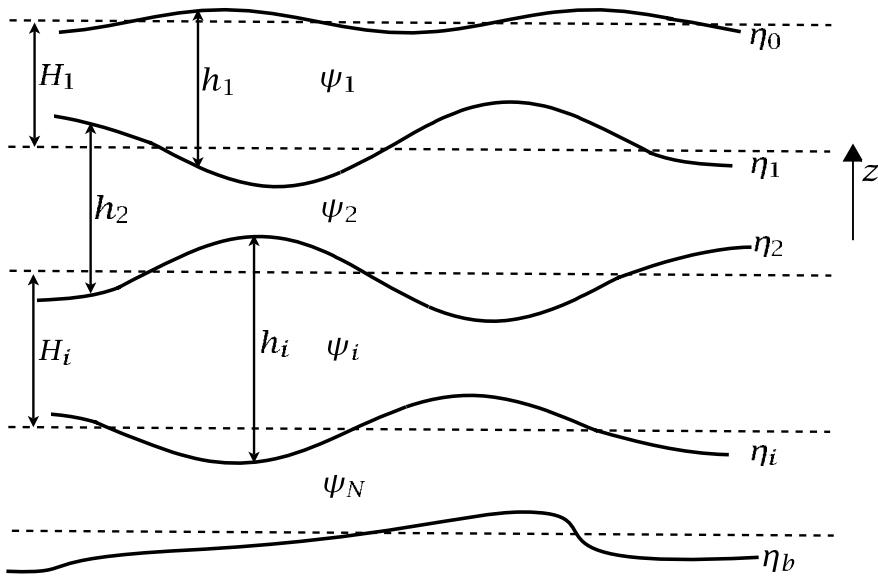


Fig. 5.2 A multi-layer quasi-geostrophic fluid system. Layers are numbered from the top down, i denotes a general interior layer and N denotes the bottom layer.

5.3.3 † Non-asymptotic and intermediate models

The form of the derivation of the previous section suggests that we might be able to improve on the accuracy and the range of applicability of the quasi-geostrophic equations, whilst still filtering gravity waves. For example, a seemingly improved set of geostrophic evolution equations might be

$$\frac{\partial q_i}{\partial t} + \mathbf{u}_i \cdot \nabla q_i = 0, \quad (5.90)$$

with

$$q_i = \frac{f + \zeta_i}{h_i}, \quad \zeta_i = \frac{\partial v_i}{\partial x} - \frac{\partial u_i}{\partial y}, \quad (5.91a,b)$$

and with the velocities given by geostrophic balance, and therefore a function of the layer depths. Thus, the vorticity, height and velocity fields may all be inverted from potential vorticity. Note that the inversion does not involve the linearization of potential vorticity about a resting state [compare (5.91a) with (5.80)], and we might also choose to keep the full variation of the Coriolis parameter in (5.81). Thus, the model consisting of (5.90) and (5.91) contains both the planetary geostrophic and quasi-geostrophic equations. However, the informality of the derivation hides the fact that this is not an asymptotically consistent set of equations: it mixes asymptotic orders in the same equation, and good conservation properties are not assured. The set above does not, in fact, exactly conserve energy. Models that are either more accurate or more general than the quasi-geostrophic or planetary-geostrophic equations yet that still filter gravity waves are called ‘intermediate models’.⁵

A model that is derived asymptotically will, in general, maintain the conservation properties of the original set. To see this, albeit in a rather abstract way, suppose that the original

equations (e.g., the primitive equations) may be written in non-dimensional form, as

$$\frac{\partial \phi}{\partial t} = F(\phi, \epsilon) \quad (5.92)$$

where ϕ is a set of variables, F is some operator and ϵ is a small parameter, such as the Rossby number. Suppose also that this set of equations has various invariants (such as energy and potential vorticity) that hold for any value of ϵ . The asymptotically derived lowest-order model (such as quasi-geostrophy) is simply a version of this equation set valid in the limit $\epsilon = 0$, and therefore it will preserve the invariants of the original set. These invariants may seem to have a different form in the simplified set: for example, in deriving the hydrostatic primitive equations from the Navier-Stokes equations the small parameter is the aspect ratio, and this multiplies the vertical velocity. Thus, in the limit of zero aspect ratio, and therefore in the primitive equations, the kinetic energy component of the energy invariant has contributions only from the horizontal velocity. In other cases, some invariants may be reduced to trivialities in the simplified set. On the other hand, there is nothing to preclude new invariants emerging that hold only in the limit $\epsilon = 0$, and enstrophy (considered later in this chapter) is one example.

5.4 THE CONTINUOUSLY STRATIFIED QUASI-GEOSTROPHIC SYSTEM

We now consider the quasi-geostrophic equations for the continuously stratified hydrostatic system. The primitive equations of motion are given by (5.15), and we extract the mean stratification so that the thermodynamic equation is given by (5.17). We also stay on the β -plane for simplicity. Readers who wish for a briefer, more informal derivation may peruse the box on page 222; however, it is important to realize that there is a systematic asymptotic derivation of the quasi-geostrophic equations, for it is this that ensures that the resulting equations have good conservation properties, as explained in section 5.3.3.

5.4.1 Scaling and assumptions

The scaling assumptions we make are just those we made for the shallow water system on page 207, with a deformation radius now given by $L_d = NH/f_0$. The non-dimensionalization and scaling are initially precisely that of section 5.1.2, and we obtain the following non-dimensional equations:

$$\text{horizontal momentum:} \quad Ro \frac{D\hat{\mathbf{u}}}{Dt} + \hat{\mathbf{f}} \times \hat{\mathbf{u}} = -\nabla_z \hat{\phi}, \quad (5.93)$$

$$\text{hydrostatic:} \quad \frac{\partial \hat{\phi}}{\partial \hat{z}} = \hat{b}, \quad (5.94)$$

$$\text{mass continuity:} \quad \frac{\partial \hat{u}}{\partial \hat{x}} + \frac{\partial \hat{v}}{\partial \hat{y}} + \frac{1}{\hat{\rho}} \frac{\partial \hat{\rho} \hat{w}}{\partial \hat{z}} = 0, \quad (5.95)$$

$$\text{thermodynamic:} \quad Ro \frac{D\hat{b}}{Dt} + \left(\frac{L_d}{L} \right)^2 \hat{w} = 0. \quad (5.96)$$

In Cartesian coordinates we may express the Coriolis parameter as

$$\mathbf{f} = f_0 + \beta y \mathbf{k}, \quad (5.97)$$

where $f_0 = f_0 \mathbf{k}$. The variation of the Coriolis parameter is assumed to be small (this is a key difference between the quasi-geostrophic system and the planetary-geostrophic system), and in particular we shall assume that βy is approximately the size of the relative vorticity, and so is much smaller than f_0 itself.⁶ Thus,

$$\beta y \sim \frac{U}{L}, \quad \beta \sim \frac{U}{L^2}, \quad (5.98)$$

and so we define an $\mathcal{O}(1)$ non-dimensional beta parameter by

$$\hat{\beta} = \frac{\beta L^2}{U} = \frac{\beta L}{Ro f_0}. \quad (5.99)$$

From this it follows that if $f = f_0 + \beta y$, the corresponding non-dimensional version is

$$\hat{f} = \hat{f}_0 + Ro \hat{\beta} \hat{y}. \quad (5.100)$$

where $\hat{f} = f/f_0$ and $\hat{f}_0 = f_0/f_0 = 1$.

5.4.2 Asymptotics

We now expand the non-dimensional dependent variables in an asymptotic series in Rossby number, and write

$$\hat{\mathbf{u}} = \hat{\mathbf{u}}_0 + Ro \hat{\mathbf{u}}_1 + \dots, \quad \hat{\phi} = \hat{\phi}_0 + Ro \hat{\phi}_1 + \dots, \quad \hat{b} = \hat{b}_0 + Ro \hat{b}_1 + \dots. \quad (5.101)$$

Substituting these into the equations of motion, the lowest-order momentum equation is simply geostrophic balance,

$$\hat{f}_0 \times \hat{\mathbf{u}}_0 = -\nabla \hat{\phi}_0 \quad (5.102)$$

with a *constant* value of the Coriolis parameter. (For the rest of this chapter we drop the subscript z from the ∇ operator.) From (5.102) it is evident that

$$\nabla \cdot \hat{\mathbf{u}}_0 = 0. \quad (5.103)$$

Thus, the horizontal flow is, to leading order, non-divergent; this is a consequence of geostrophic balance, and is *not* a mass conservation equation. Using (5.103) in the mass conservation equation, (5.95), gives

$$\frac{\partial}{\partial z} (\tilde{\rho} \hat{w}_0) = 0, \quad (5.104)$$

which implies that if w_0 is zero somewhere (e.g., at a solid surface) then w_0 is zero everywhere (essentially the Taylor-Proudman effect). A physical way of saying this is that the scaling estimate $W = UH/L$ is an overestimate of the size of the vertical velocity, because even though $\partial w/\partial z \approx -\nabla \cdot \mathbf{u}$, the horizontal divergence of the geostrophic flow is small if f is nearly constant and $|\nabla \cdot \mathbf{u}| \ll U/L$. We might have anticipated this from the outset, and scaled w differently, perhaps using the geostrophic vorticity balance estimate, $w \sim \beta UH/f_0 = Ro UH/L$ as the scaling factor for w , but there is no a priori guarantee that this would be correct.

At next order the momentum equation is

$$\frac{D_0 \hat{\mathbf{u}}_0}{Dt} + \hat{\beta} \hat{\mathbf{y}} \mathbf{k} \times \hat{\mathbf{u}}_0 + \hat{\mathbf{f}} \times \hat{\mathbf{u}}_1 = -\nabla \hat{\phi}_1, \quad (5.105)$$

where $D_0/Dt = \partial/\partial \hat{t} + (\hat{\mathbf{u}}_0 \cdot \nabla)$, and the next order mass conservation equation is

$$\nabla_z \cdot (\tilde{\rho} \hat{\mathbf{u}}_1) + \frac{\partial}{\partial z} (\tilde{\rho} \hat{w}_1) = 0. \quad (5.106)$$

From (5.96), the lowest-order thermodynamic equation is just

$$\left(\frac{L_d}{L} \right)^2 \hat{w}_0 = 0 \quad (5.107)$$

provided that, as we have assumed, the scales of motion are not sufficiently large that $Ro(L/L_d)^2 = \mathcal{O}(1)$. (This is a key difference between quasi-geostrophy and planetary geostrophy.) At next order we obtain an evolution equation for the buoyancy, and this is

$$\frac{D_0 \hat{b}_0}{Dt} + \hat{w}_1 \left(\frac{L_d}{L} \right)^2 = 0. \quad (5.108)$$

The potential vorticity equation

To obtain a single evolution equation for lowest-order quantities we eliminate w_1 between the thermodynamic and momentum equations. Cross-differentiating the first-order momentum equation (5.105) gives the vorticity equation,

$$\frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{\mathbf{u}}_0 \cdot \nabla) \hat{\zeta}_0 + \hat{v}_0 \hat{\beta} = -\hat{f}_0 \nabla_z \cdot \hat{\mathbf{u}}_1. \quad (5.109)$$

(In dimensional terms, the divergence on the right-hand side is small, but is multiplied by the large term f_0 , and their product is of the same order as the terms on the left-hand side.) Using the mass conservation equation (5.106), (5.109) becomes

$$\frac{D_0}{Dt} (\zeta_0 + \hat{f}) = \frac{\hat{f}_0}{\tilde{\rho}} \frac{\partial}{\partial z} (w_1 \tilde{\rho}). \quad (5.110)$$

Combining (5.110) and (5.108) gives

$$\frac{D_0}{Dt} (\zeta_0 + \hat{f}) = -\frac{\hat{f}_0}{\tilde{\rho}} \frac{\partial}{\partial z} \left[\frac{D_0}{Dt} (F \tilde{\rho} \hat{b}_0) \right], \quad (5.111)$$

where $F \equiv (L/L_d)^2$. The right-hand side of this equation is

$$\frac{\partial}{\partial z} \left(\frac{D_0 \hat{b}_0}{Dt} \right) = \frac{D_0}{Dt} \left(\frac{\partial \hat{b}_0}{\partial z} \right) + \frac{\partial \hat{\mathbf{u}}_0}{\partial z} \cdot \nabla \hat{b}_0. \quad (5.112)$$

The second term on the right-hand side vanishes identically using the thermal wind equation

$$\mathbf{k} \times \frac{\partial \hat{\mathbf{u}}_0}{\partial z} = -\frac{1}{\hat{f}_0} \nabla \hat{b}_0, \quad (5.113)$$

and so (5.111) becomes

$$\frac{D_0}{D\hat{t}} \left[\hat{\zeta}_0 + \hat{f} + \frac{\hat{f}_0}{\tilde{\rho}} \frac{\partial}{\partial \hat{z}} \left(\tilde{\rho} F \hat{b}_0 \right) \right] = 0, \quad (5.114)$$

or, after using the hydrostatic equation,

$$\frac{D_0}{D\hat{t}} \left[\hat{\zeta}_0 + \hat{f} + \frac{\hat{f}_0}{\tilde{\rho}} \frac{\partial}{\partial \hat{z}} \left(\tilde{\rho} F \frac{\partial \hat{\psi}}{\partial \hat{z}} \right) \right] = 0. \quad (5.115)$$

Since the lowest-order horizontal velocity is divergence-free, we can define a streamfunction $\hat{\psi}$ such that

$$\hat{u}_0 = -\frac{\partial \hat{\psi}}{\partial \hat{y}}, \quad \hat{v}_0 = \frac{\partial \hat{\psi}}{\partial \hat{x}} \quad (5.116)$$

where also, using (5.102), $\phi_0 = \hat{f}_0 \hat{\psi}$. The vorticity is then given by $\hat{\zeta}_0 = \nabla^2 \hat{\psi}$ and (5.115) becomes a single equation in a single unknown, to wit

$$\frac{D_0}{D\hat{t}} \left[\nabla^2 \hat{\psi} + \hat{\beta} \hat{y} + \frac{\hat{f}_0^2}{\tilde{\rho}} \frac{\partial}{\partial \hat{z}} \left(\tilde{\rho} F \frac{\partial \hat{\psi}}{\partial \hat{z}} \right) \right] = 0, \quad (5.117)$$

where the material derivative is evaluated using $\hat{\mathbf{u}}_0 = \mathbf{k} \times \nabla \hat{\psi}$. This is the non-dimensional form of the quasi-geostrophic potential vorticity equation, one of the most important equations in dynamical meteorology and oceanography. In deriving it we have reduced the Navier–Stokes equations, which are six coupled nonlinear partial differential equations in six unknowns (u, v, w, T, p, ρ) to a single (albeit nonlinear) first-order partial differential equation in a single unknown.⁷

Dimensional equations

The dimensional version of the quasi-geostrophic potential vorticity equation may be written as

$$\frac{Dq}{Dt} = 0, \\ q = \nabla^2 \psi + f + \frac{f_0^2}{\tilde{\rho}} \frac{\partial}{\partial z} \left(\frac{\tilde{\rho}}{N^2} \frac{\partial \psi}{\partial z} \right), \quad (5.118a,b)$$

where only the variable part of f (e.g., βy) is relevant in the second term on the right-hand side of the expression for q . The quantity q is known as the *quasi-geostrophic potential vorticity*. It is analogous to the exact (Ertel) potential vorticity (see section 5.5 for more about this), and it is conserved when advected by the *horizontal* geostrophic flow. All the other dynamical variables may be obtained from potential vorticity as follows.

- (i) Streamfunction, using (5.118b).
- (ii) Velocity: $\mathbf{u} = \mathbf{k} \times \nabla \psi$ [$\equiv \nabla^\perp \psi = -\nabla \times (\mathbf{k} \psi)$].
- (iii) Relative vorticity: $\zeta = \nabla^2 \psi$.
- (iv) Perturbation pressure: $\phi = f_0 \psi$.
- (v) Perturbation buoyancy: $b' = f_0 \partial \psi / \partial z$.

The length scale $L_d = NH/f_0$, emerges naturally from the QG dynamics. It is the scale at which buoyancy and relative vorticity effects contribute equally to the potential vorticity, and is called the *deformation radius*; it is analogous to the quantity \sqrt{gH}/f_0 arising in shallow water theory. In the upper ocean, with $N \approx 10^{-2} \text{ s}^{-1}$, $H \approx 10^3 \text{ m}$ and $f_0 \approx 10^{-4} \text{ s}^{-1}$, then $L_d \approx 100 \text{ km}$. At high latitudes the ocean is much less stratified and f is somewhat larger, and the deformation radius may be as little as 30 km (see Fig. 9.11 on page 401, where the deformation radius is defined slightly differently). In the atmosphere, with $N \approx 10^{-2} \text{ s}^{-1}$, $H \approx 10^4 \text{ m}$, then $L_d \approx 1000 \text{ km}$. It is this order of magnitude difference in the deformation scales that accounts for a great deal of the quantitative difference in the dynamics of the ocean and the atmosphere. If we take the limit $L_d \rightarrow \infty$ then the stratified quasi-geostrophic equations reduce to

$$\frac{Dq}{Dt} = 0, \quad q = \nabla^2 \psi + f. \quad (5.119)$$

This is the two-dimensional vorticity equation, identical to (4.67). The high stratification of this limit has suppressed all vertical motion, and variations in the flow become confined to the horizontal plane. Finally, we note that it is typical in quasi-geostrophic applications to omit the prime on the buoyancy perturbations, and write $b = f_0 \partial \psi / \partial z$; however, we will keep the prime in this chapter.

5.4.3 Buoyancy advection at the surface

The solution of the elliptic equation in (5.118) requires vertical boundary conditions on ψ at the ground and at the top of the atmosphere, and these are given by use of the thermodynamic equation. For a flat, slippery, rigid surface the vertical velocity is zero so that the thermodynamic equation may be written as

$$\frac{Db'}{Dt} = 0, \quad b' = f_0 \frac{\partial \psi}{\partial z}. \quad (5.120)$$

We apply this at the ground and at the tropopause, treating the latter as a lid on the lower atmosphere. In the presence of friction and topography the vertical velocity is not zero, but is given by

$$w = r \nabla^2 \psi + \mathbf{u} \cdot \nabla \eta_b \quad (5.121)$$

where the first term represents Ekman friction (with the constant r proportional to the thickness of the Ekman layer) and the second term represents topographic forcing. The boundary condition becomes

$$\frac{\partial}{\partial t} \left(\frac{\partial \psi}{\partial z} \right) + \mathbf{u} \cdot \nabla \left(\frac{\partial \psi}{\partial z} + N^2 \eta_b \right) + N^2 r \nabla^2 \psi = 0, \quad (5.122)$$

where all the fields are evaluated at $z = 0$ and $z = H$, the height of the lid. Thus, the quasi-geostrophic system is characterized by the horizontal advection of potential vorticity in the interior and the advection of buoyancy at the boundary. Instead of a lid at the top, then in a compressible fluid such as the atmosphere we may suppose that all disturbances tend to zero as $z \rightarrow \infty$.

** A potential vorticity sheet at the boundary*

Rather than regarding buoyancy advection as providing the boundary condition, it is sometimes useful to think of there being a very thin sheet of potential vorticity just above the ground and another just below the lid, specifically with a vertical distribution proportional to $\delta(z - \epsilon)$ or $\delta(z - H + \epsilon)$, where ϵ is small. The boundary condition (5.120) or (5.122) can be replaced by this, along with the condition that there are no variations of buoyancy at the boundary and $\partial\psi/\partial z = 0$ at $z = 0$ and $z = H$.⁸

To see this, we first note that the differential of a step function is a delta function. Thus, a discontinuity in $\partial\psi/\partial z$ at a level $z = z_1$ is equivalent to a delta function in potential vorticity there:

$$q(z_1) = \left[\frac{f_0^2}{N^2} \frac{\partial\psi}{\partial z} \right]_{z_1-}^{z_1+} \delta(z - z_1). \quad (5.123)$$

Now, suppose that the lower boundary condition, given by (5.120), has some arbitrary distribution of buoyancy on it. We can replace this condition by the simpler condition $\partial\psi/\partial z = 0$ at $z = 0$, provided we also add to our definition of potential vorticity a term given by (5.123) with $z_1 = \epsilon$. This term is then advected by the horizontal flow, as are the other contributions. A buoyancy source at the boundary must similarly be treated as a sheet of potential vorticity source in the interior. Any flow with buoyancy variations over a horizontal boundary is thus equivalent to a flow with uniform buoyancy at the boundary, but with a spike in potential vorticity adjacent to the boundary. This approach brings notational and conceptual advantages, in that now everything is expressed in terms of potential vorticity and its advection. However, in practice there may be less to be gained, because the boundary terms must still be included in any particular calculation that is to be performed.

5.4.4 Quasi-geostrophy in pressure coordinates

The derivation of the quasi-geostrophic system in pressure coordinates is very similar to that in height coordinates, with the main difference coming at the boundaries, and we give only the results. The starting point is the primitive equations in pressure coordinates, (2.153). In pressure coordinates, the quasi-geostrophic potential vorticity is found to be

$$q = f + \nabla^2\psi + \frac{\partial}{\partial p} \left(\frac{f_0^2}{S^2} \frac{\partial\psi}{\partial p} \right), \quad (5.124)$$

where $\psi = \Phi/f_0$ is the streamfunction and Φ the geopotential, and

$$S^2 \equiv -\frac{R}{p} \left(\frac{p}{p_R} \right)^\kappa \frac{d\tilde{\theta}}{dp} = -\frac{1}{\rho\theta} \frac{d\tilde{\theta}}{dp}, \quad (5.125)$$

where $\tilde{\theta}$ is a reference profile and a function of pressure only. In log-pressure coordinates, with $Z = -H \ln p$, the potential vorticity may be written as

$$q = f + \nabla^2\psi + \frac{1}{\rho_*} \frac{\partial}{\partial Z} \left(\frac{\rho_* f_0^2}{N_Z^2} \frac{\partial\psi}{\partial Z} \right), \quad (5.126)$$

where

$$N_Z^2 = S^2 \left(\frac{p}{H} \right)^2 = -\left(\frac{R}{H} \right) \left(\frac{p}{p_R} \right)^\kappa \frac{d\tilde{\theta}}{dZ} \quad (5.127)$$

is the buoyancy frequency and $\rho_* = \exp(-z/H)$. Temperature and potential temperature are related to the streamfunction by

$$T = -\frac{f_0 p}{R} \frac{\partial \psi}{\partial p} = \frac{H f_0}{R} \frac{\partial \psi}{\partial Z}, \quad (5.128a)$$

$$\theta = -\left(\frac{p_R}{p}\right)^k \left(\frac{f_0 p}{R}\right) \frac{\partial \psi}{\partial p} = \left(\frac{p_R}{p}\right)^k \left(\frac{H f_0}{R}\right) \frac{\partial \psi}{\partial Z}. \quad (5.128b)$$

In pressure or log-pressure coordinates, potential vorticity is advected along isobaric surfaces, analogous to the horizontal advection in height coordinates.

The surface boundary condition again is derived from the thermodynamic equation. In log-pressure coordinates this is

$$\frac{D}{Dt} \left(\frac{\partial \psi}{\partial Z} \right) + \frac{N_Z^2}{f_0} W = 0, \quad (5.129)$$

where $W = DZ/Dt$. This is not the real vertical velocity, w , but it is related to it by

$$w = \frac{f_0}{g} \frac{\partial \psi}{\partial t} + \frac{RT}{gH} W. \quad (5.130)$$

Thus, choosing $H = RT(0)/g$, we have, at $Z = 0$,

$$\frac{\partial}{\partial t} \left(\frac{\partial \psi}{\partial Z} - \frac{N_Z^2}{g} \psi \right) + \mathbf{u} \cdot \nabla \psi = -\frac{N^2}{f_0} w, \quad (5.131)$$

where

$$w = \mathbf{u} \cdot \nabla \eta_b + r \nabla^2 \psi. \quad (5.132)$$

This differs from the expression in height coordinates only by the second term in the local time derivative. In applications where accuracy is not the main issue the simpler boundary condition $D(\partial_Z \psi)/Dt = 0$ is sometimes used.

5.4.5 The two-level quasi-geostrophic system

The quasi-geostrophic system has, in general, continuous variation in the vertical direction (and horizontal, of course). By finite-differencing the continuous equations we can obtain a *multi-level* model, and a crude but important special case of this is the *two-level* model, also known as the Phillips model.⁹ To obtain the equations of motion one way to proceed is to take a crude finite difference of the continuous relation between potential vorticity and streamfunction given in (5.118b). In the Boussinesq case (or in pressure coordinates, with a slight reinterpretation of the meaning of the symbols) the continuous expression for potential vorticity is

$$q = \zeta + f + \frac{\partial}{\partial z} \left(\frac{f_0 b'}{N^2} \right), \quad (5.133)$$

where $b' = f_0 \partial \psi / \partial z$. In the case with a flat bottom and rigid lid at the top (and incorporating topography is an easy extension) the boundary condition of $w = 0$ is satisfied by $D\partial_z \psi / Dt = 0$ at the top and bottom. An obvious finite-differencing of (5.133) in the vertical direction (see Fig. 5.3) then gives

Informal Derivation of Stratified QG Equations

We will use the Boussinesq equations, but similar derivations could be given using the anelastic equations or pressure coordinates. The first ingredient is the vertical component of the vorticity equation, (4.66); in the Boussinesq version (or the pressure coordinate or anelastic versions) there is no baroclinic term and we have

$$\frac{D_3}{Dt}(\zeta + f) = -(\zeta + f) \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + \left(\frac{\partial u}{\partial z} \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \frac{\partial w}{\partial x} \right). \quad (\text{QG.1})$$

We now apply the assumptions on page 207. The advection and the vorticity on the left-hand side are geostrophic, but we keep the horizontal divergence (which is small) on the right-hand side where it is multiplied by the big term f . Furthermore, because f is nearly constant we replace it with f_0 except where it is differentiated. The second term (tilting) on the right-hand side is smaller than the advection terms on the left-hand side by the ratio $[UW/(HL)]/[U^2/L^2] = [W/H]/[U/L] \ll 1$, because w is small ($\partial w/\partial z$ equals the divergence of the ageostrophic velocity). We therefore neglect it, and given all this (QG.1) becomes

$$\frac{D_g}{Dt}(\zeta_g + f) = -f_0 \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = f_0 \frac{\partial w}{\partial z}, \quad (\text{QG.2})$$

where the second equality uses mass continuity and $D_g/Dt = \partial/\partial t + \mathbf{u}_g \cdot \nabla$.

The second ingredient is the three-dimensional thermodynamic equation,

$$\frac{D_3 b}{Dt} = 0. \quad (\text{QG.3})$$

The stratification is assumed to be nearly constant, so we write $b = \tilde{b}(z) + b'(x, y, z, t)$, where \tilde{b} is the basic state buoyancy. Furthermore, because w is small it only advects the basic state, and with $N^2 = \partial \tilde{b} / \partial z$ (QG.3) becomes

$$\frac{D_g b'}{Dt} + w N^2 = 0. \quad (\text{QG.4})$$

Hydrostatic and geostrophic wind balance enable us to write the geostrophic velocity, vorticity, and buoyancy in terms of streamfunction ψ [= $p/(f_0 \rho_0)$]:

$$\mathbf{u}_g = \mathbf{k} \times \nabla \psi, \quad \zeta_g = \nabla^2 \psi, \quad b' = f_0 \partial \psi / \partial z. \quad (\text{QG.5})$$

The quasi-geostrophic potential vorticity equation is obtained by eliminating w between (QG.2) and (QG.4), and this gives

$$\frac{D_g q}{Dt} = 0, \quad q = \zeta_g + f + \frac{\partial}{\partial z} \left(\frac{f_0 b'}{N^2} \right). \quad (\text{QG.6})$$

This equation is the Boussinesq version of (5.118), and using (QG.5) it may be expressed entirely in terms of the streamfunction, with $D_g \cdot /Dt = \partial/\partial t + J(\psi, \cdot)$. The vertical boundary conditions, at $z = 0$ and $z = H$ say, are given by (QG.4) with $w = 0$, with straightforward generalizations if topography or friction are present.

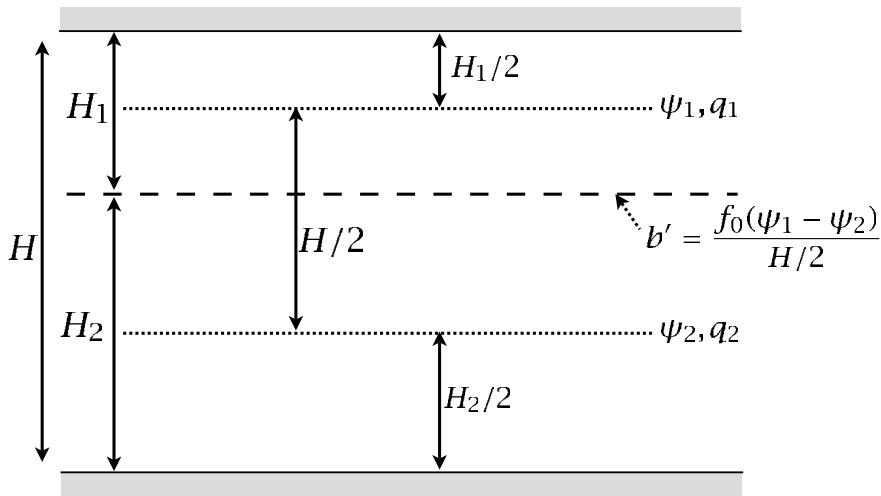


Fig. 5.3 The two-level quasi-geostrophic system with a flat bottom and rigid lid at which $w = 0$.

$$q_1 = \zeta_1 + f + \frac{2f_0^2}{N^2 H_1 H} (\psi_2 - \psi_1), \quad q_2 = \zeta_2 + f + \frac{2f_0^2}{N^2 H_2 H} (\psi_1 - \psi_2). \quad (5.134)$$

In atmospheric problems it is common to choose $H_1 = H_2$, whereas in oceanic problems we might choose to have a thinner upper layer, representing the flow above the main thermocline. Note that the boundary conditions of $w = 0$ at the top and bottom are already taken care of in (5.134): *they are incorporated into the definition of the potential vorticity* — a finite-difference analogue of the delta-function construction of section 5.4.3. At each level the potential vorticity is advected by the streamfunction so that the evolution equation for each level is:

$$\frac{Dq_i}{Dt} = \frac{\partial q_i}{\partial t} + \mathbf{u}_i \cdot \nabla q_i = \frac{\partial q_i}{\partial t} + J(\psi_i, q_i) = 0, \quad i = 1, 2. \quad (5.135)$$

Models with more than two levels can be constructed by extending the finite-differencing procedure in a natural way.

Connection to the layered system

The two-level expressions, (5.134), have an obvious similarity to the *two-layer* expressions, (5.85). Noting that $N^2 = \partial \hat{b} / \partial z$ and that $b = -g\delta\rho/\rho_0$ it is natural to let

$$N^2 = -\frac{g}{\rho_0} \frac{\rho_1 - \rho_2}{H/2} = \frac{g'}{H/2}. \quad (5.136)$$

With this identification we find that (5.134) becomes

$$q_1 = \zeta_1 + f + \frac{f_0^2}{g'H_1} (\psi_2 - \psi_1), \quad q_2 = \zeta_2 + f + \frac{f_0^2}{g'H_2} (\psi_1 - \psi_2). \quad (5.137)$$

These expressions are identical to (5.85) in the flat-bottomed, rigid lid case. Similarly, a multi-layered system with n layers is equivalent to a finite-difference representation with n levels. It should be said, though, that in the pantheon of quasi-geostrophic models the two-level and two-layer models hold distinguished places.

5.5 * QUASI-GEOSTROPHY AND ERTEL POTENTIAL VORTICITY

When using the shallow water equations, quasi-geostrophic theory could be naturally developed beginning with the expression for potential vorticity. Is such an approach possible for the stratified primitive equations? The answer is yes, although the algebra is more complicated, as we will see.

5.5.1 * Using height coordinates

Noting the general expression, (4.119), for potential vorticity in a hydrostatic fluid, the potential vorticity in the Boussinesq hydrostatic equations is given by

$$Q = \left[(v_x - u_y) b_z - v_z b_x + u_z b_y + f b_z \right], \quad (5.138)$$

where the x, y, z subscripts denote derivatives. Without approximation, we write the stratification as $b = \tilde{b}(z) + b'(x, y, z, t)$, and (5.138) becomes

$$Q = [f_0 N^2] + [(\beta y + \zeta) N^2 + f_0 b'_z] + [(\beta y + \zeta) b'_z - (v_z b'_x - u_z b'_y)], \quad (5.139)$$

where, under quasi-geostrophic scaling, the terms in square brackets are in decreasing order of size. Neglecting the third term, and taking the velocity and buoyancy fields to be in geostrophic and thermal wind balance, we can write the potential vorticity as $Q \approx \tilde{Q} + Q'$, where $\tilde{Q} = f_0 N^2$ and

$$Q' = (\beta y + \zeta) N^2 + f_0 b'_z = (\beta y + \nabla^2 \psi) N^2 + f_0^2 \frac{\partial \psi}{\partial z}. \quad (5.140)$$

The potential vorticity evolution equation is then

$$\frac{DQ'}{Dt} + w \frac{\partial \tilde{Q}}{\partial z} = 0. \quad (5.141)$$

The vertical advection is important only in advecting the basic state potential vorticity \tilde{Q} . Thus, after dividing by N^2 , (5.141) becomes

$$\frac{\partial q_*}{\partial t} + \mathbf{u}_g \cdot \nabla q_* + \frac{w}{N^2} \frac{\partial \tilde{Q}}{\partial z} = 0, \quad (5.142)$$

where

$$q_* = (\beta y + \zeta) + \frac{f_0}{N^2} b'_z. \quad (5.143)$$

This is the approximation to the (perturbation) Ertel potential vorticity in the quasi-geostrophic limit. However, it is not the same as the expression for the quasi-geostrophic potential vorticity, (5.118b) and, furthermore, (5.142) involves a vertical advection. (Thus, we might refer to the expression in (5.118) as the ‘quasi-geostrophic pseudopotential vorticity’, but the prefix ‘quasi-geostrophic’ alone normally suffices.) We can derive (5.118) by eliminating w between (5.142) and the quasi-geostrophic thermodynamic equation $\partial b'/\partial t + \mathbf{u}_g \cdot \nabla b' + w \partial \tilde{b}/\partial z = 0$.

5.5.2 Using isentropic coordinates

An illuminating and somewhat simpler path from Ertel potential vorticity to the quasi-geostrophic equations goes by way of isentropic coordinates.¹⁰ We begin with the isentropic expression for the Ertel potential vorticity of an ideal gas,

$$Q = \frac{f + \zeta}{\sigma}, \quad (5.144)$$

where $\sigma = -\partial p / \partial \theta$ is the thickness density (which we will just call the thickness), and in adiabatic flow the potential vorticity is advected along isopycnals. We now employ quasi-geostrophic scaling to derive an approximate equation set from this. First, assume that variations in thickness are small compared with the reference state, so that

$$\sigma = \tilde{\sigma}(\theta) + \sigma', \quad |\sigma'| \ll |\sigma|, \quad (5.145)$$

and similarly for pressure and density. Assuming also that the variations in the Coriolis parameter are small, (5.144) becomes

$$Q \approx \left[\frac{f_0}{\tilde{\sigma}} \right] + \left[\frac{1}{\tilde{\sigma}} (\zeta + \beta y) - \frac{f_0}{\tilde{\sigma}} \frac{\sigma'}{\tilde{\sigma}} \right]. \quad (5.146)$$

We now use geostrophic and hydrostatic balance to express the terms on the right-hand side in terms of a single variable, noting that the first term does not vary along isentropic surfaces. Hydrostatic balance is

$$\frac{\partial M}{\partial \theta} = \Pi, \quad (5.147)$$

where $M = c_p T + gz$ and $\Pi = c_p (p/p_R)^\kappa$. Writing $M = \tilde{M}(\theta) + M'$ and $\Pi = \tilde{\Pi}(\theta) + \Pi'$, where \tilde{M} and $\tilde{\Pi}$ are hydrostatically balanced reference profiles, we obtain

$$\frac{\partial M'}{\partial \theta} = \Pi' \approx \frac{d\tilde{\Pi}}{dp} p' = \frac{1}{\theta \tilde{\rho}} p', \quad (5.148)$$

where the last equality follows using the equation of state for an ideal gas and $\tilde{\rho}$ is a reference profile. The perturbation thickness field may then be written as

$$\sigma' = -\frac{\partial}{\partial \theta} \left(\tilde{\rho} \theta \frac{\partial M'}{\partial \theta} \right). \quad (5.149)$$

Geostrophic balance is $\mathbf{f}_0 \times \mathbf{u} = -\nabla_\theta M'$ where the velocity, and the horizontal derivatives, are along isentropic surfaces. This enables us to define a flow streamfunction by

$$\psi \equiv \frac{M'}{f_0} \quad (5.150)$$

and we can then write all the variables in terms of ψ :

$$\begin{aligned} u &= -\left(\frac{\partial \psi}{\partial y} \right)_\theta, & v &= \left(\frac{\partial \psi}{\partial x} \right)_\theta, \\ \zeta &= \nabla_\theta^2 \psi, & \sigma' &= f_0 \frac{\partial}{\partial \theta} \left(\tilde{\rho} \theta \frac{\partial M'}{\partial \theta} \right). \end{aligned} \quad (5.151)$$

Using (5.146) (5.150) and (5.151), the quasi-geostrophic system in isentropic coordinates may be written

$$\boxed{\frac{Dq}{Dt} = 0 \\ q = f + \nabla_\theta^2 \psi + \frac{f_0^2}{\tilde{\sigma}} \frac{\partial}{\partial \theta} \left(\tilde{\rho} \theta \frac{\partial \psi}{\partial \theta} \right)}, \quad (5.152a,b)$$

where the advection of potential vorticity is by the geostrophically balanced flow, along isentropes. The variable q is an approximation to the second term in square brackets in (5.146), multiplied by $\tilde{\sigma}$.

Projection back to physical-space coordinates

We can recover the height or pressure coordinate quasi-geostrophic systems by projecting (5.152) on to the appropriate coordinate. This is straightforward because, by assumption, the isentropes in a quasi-geostrophic system are nearly flat. Recall that [cf. (2.142)] a transformation between vertical coordinates may be effected by

$$\frac{\partial}{\partial x} \Big|_\theta = \frac{\partial}{\partial x} \Big|_p + \frac{\partial p}{\partial x} \Big|_\theta \frac{\partial}{\partial p}, \quad (5.153)$$

but the second term is $\mathcal{O}(Ro)$ smaller than the first one because, under quasi-geostrophic scaling, isentropic slopes are small. Thus $\nabla_\theta^2 \psi$ in (5.152b) may be replaced by $\nabla_p^2 \psi$ or $\nabla_z^2 \psi$. The vortex stretching term in (5.152) becomes, in pressure coordinates,

$$\frac{f_0^2}{\tilde{\sigma}} \frac{\partial}{\partial \theta} \left(\tilde{\rho} \theta \frac{\partial \psi}{\partial \theta} \right) \approx \frac{f_0^2}{\tilde{\sigma}} \frac{d\tilde{p}}{d\theta} \frac{\partial}{\partial p} \left(\tilde{\rho} \theta \frac{d\tilde{p}}{d\theta} \frac{\partial \psi}{\partial p} \right) = \frac{\partial}{\partial p} \left(\frac{f_0^2}{S^2} \frac{\partial \psi}{\partial p} \right) \quad (5.154)$$

where S^2 is given by (5.125). The expression for the quasi-geostrophic potential vorticity in isentropic coordinates is thus approximately equal to the quasi-geostrophic potential vorticity in pressure coordinates. This near-equality holds because the isentropic expression, (5.152b), does not contain a component proportional to the mean stratification: the second square-bracketed term on the right-hand side (5.146) is the only dynamically relevant one, and its evolution along isentropes is mirrored by the evolution along isobaric surfaces of quasi-geostrophic potential vorticity in pressure coordinates.

5.6 * ENERGETICS OF QUASI-GEOSTROPHY

If the quasi-geostrophic set of equations is to represent a real fluid system in a physically meaningful way, then it should have a consistent set of energetics. In particular, the total energy should be conserved, and there should be analogs of kinetic and potential energy and conversion between the two. We now show that such energetic properties do hold, using the Boussinesq set as an example.

Let us write the governing equations as a potential vorticity equation in the interior,

$$\frac{D}{Dt} \left[\nabla^2 \psi + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial \psi}{\partial z} \right) \right] + \beta \frac{\partial \psi}{\partial x} = 0, \quad 0 < z < 1, \quad (5.155)$$

and buoyancy advection at the boundary,

$$\frac{D}{Dt} \left(\frac{\partial \psi}{\partial z} \right) = 0, \quad z = 0, 1. \quad (5.156)$$

For lateral boundary conditions we may assume that $\psi = \text{constant}$, or impose periodic conditions. If we multiply (5.155) by $-\psi$ and integrate over the domain, using the boundary conditions, we easily find

$$\frac{d\hat{E}}{dt} = 0, \quad \hat{E} = \frac{1}{2} \int_V \left[(\nabla \psi)^2 + \frac{f_0^2}{N^2} \left(\frac{\partial \psi}{\partial z} \right)^2 \right] dV. \quad (5.157a,b)$$

The term involving β makes no direct contribution to the energy budget. Equation (5.157) is the fundamental energy equation for quasi-geostrophic motion, and it states that in the absence of viscous or diabatic terms the total energy is conserved. The two terms in (5.157b) can be identified as the kinetic energy (KE) and available potential energy (APE) of the flow, where

$$KE = \frac{1}{2} \int_V (\nabla \psi)^2 dV, \quad APE = \frac{1}{2} \int_V \frac{f_0^2}{N^2} \left(\frac{\partial \psi}{\partial z} \right)^2 dV. \quad (5.158a,b)$$

The available potential energy may also be written as

$$APE = \frac{1}{2} \int_V \frac{H^2}{L_d^2} \left(\frac{\partial \psi}{\partial z} \right)^2 dV, \quad (5.159)$$

where L_d is the deformation radius NH/f_0 and we may choose H such that $z \sim H$. At some scale L the ratio of the kinetic energy to the potential energy is thus, roughly,

$$\frac{KE}{APE} \sim \frac{L_d^2}{L^2}. \quad (5.160)$$

For scales much larger than L_d the potential energy dominates the kinetic energy, and contrariwise.

5.6.1 Conversion between APE and KE

Let us return to the vorticity and thermodynamic equations,

$$\frac{D\zeta}{Dt} = f \frac{\partial w}{\partial z} \quad (5.161)$$

where $\zeta = \nabla^2 \psi$, and

$$\frac{Db'}{Dt} + N^2 w = 0 \quad (5.162)$$

where $b' = f_0 \partial \psi / \partial z$. From (5.161) we form a kinetic energy equation namely

$$\frac{1}{2} \frac{d}{dt} \int_V (\nabla \psi)^2 dV = - \int_V f_0 \frac{\partial w}{\partial z} \psi dV = \int_V f_0 w \frac{\partial \psi}{\partial z} dV. \quad (5.163)$$

From (5.162) we form a potential energy equation, namely

$$\frac{d}{dt} \frac{1}{2} \int_V \frac{f_0^2}{N^2} \left(\frac{\partial \psi}{\partial z} \right)^2 dV = - \int_V f_0 w \frac{\partial \psi}{\partial z} dV. \quad (5.164)$$

Thus, the *conversion* from APE to KE is represented by

$$\frac{d}{dt} KE = -\frac{d}{dt} APE = \int_v f_0 w \frac{\partial \psi}{\partial z} dV. \quad (5.165)$$

Because the buoyancy is proportional to $\partial \psi / \partial z$, when warm fluid rises there is a correlation between w and $\partial \psi / \partial z$ and APE is converted to KE. Whether such a phenomenon occurs depends of course on the dynamics of the flow; however, such a conversion is, in fact, a common feature of geophysical flows, and in particular of baroclinic instability, as we shall see in chapter 6.

5.6.2 Energetics of two-layer flows

Two-layer or two-level flows are an important special case. For layers of equal thickness let us write the evolution equations as

$$\frac{D}{Dt} \left[\nabla^2 \psi_1 - \frac{1}{2} k_d^2 (\psi_1 - \psi_2) \right] + \beta \frac{\partial \psi_1}{\partial x} = 0 \quad (5.166a)$$

$$\frac{D}{Dt} \left[\nabla^2 \psi_2 + \frac{1}{2} k_d^2 (\psi_1 - \psi_2) \right] + \beta \frac{\partial \psi_2}{\partial x} = 0 \quad (5.166b)$$

where $k_d^2/2 = (2f_0/NH)^2$. On multiplying these two equations by $-\psi_1$ and $-\psi_2$, respectively, and integrating over the horizontal domain, the advective term in the material derivatives and the beta term all vanish, and we obtain

$$\frac{d}{dt} \int_A \left[\frac{1}{2} (\nabla \psi_1)^2 + \frac{1}{2} k_d^2 \psi_1 (\psi_1 - \psi_2) \right] dA = 0, \quad (5.167a)$$

$$\frac{d}{dt} \int_A \left[\frac{1}{2} (\nabla \psi_2)^2 - \frac{1}{2} k_d^2 \psi_2 (\psi_1 - \psi_2) \right] dA = 0. \quad (5.167b)$$

Adding these gives

$$\frac{d}{dt} \int_A \frac{1}{2} \left[(\nabla \psi_1)^2 + (\nabla \psi_2)^2 + k_d^2 (\psi_1 - \psi_2)^2 \right] dA = 0. \quad (5.168)$$

This is the energy conservation statement for the two layer model. The first two terms represent the kinetic energy and the last term represents the available potential energy.

Energy in the baroclinic and barotropic modes

A useful partitioning of the energy is between the energy in the barotropic and baroclinic modes. The barotropic streamfunction, $\bar{\psi}$, is the vertically averaged streamfunction and the baroclinic mode is the difference between the streamfunctions in the two layers. That is, for equal layer thicknesses,

$$\bar{\psi} \equiv \frac{1}{2} (\psi_1 + \psi_2), \quad \tau \equiv \frac{1}{2} (\psi_1 - \psi_2). \quad (5.169)$$

Substituting (5.169) into (5.168) reveals that

$$\frac{d}{dt} \int_A \left[(\nabla \bar{\psi})^2 + (\nabla \tau)^2 + k_d^2 \tau^2 \right] dA = 0. \quad (5.170)$$

The energy density in the barotropic mode is thus just $(\nabla \bar{\psi})^2$, and that in the baroclinic mode is $(\nabla \tau)^2 + k_d^2 \tau^2$. This partitioning will prove particularly useful when we consider baroclinic turbulence in chapter 9.

5.6.3 Enstrophy conservation

Potential vorticity is advected only by the horizontal flow, and thus it is materially conserved on the horizontal surface at every height and

$$\frac{Dq}{Dt} = \frac{\partial q}{\partial t} + \mathbf{u} \cdot \nabla q = 0. \quad (5.171)$$

Furthermore, the advecting flow is divergence-free so that $\mathbf{u} \cdot \nabla q = \nabla \cdot (\mathbf{u}q)$. Thus, on multiplying (5.171) by q and integrating over a horizontal domain, A , using either no-normal flow or periodic boundary conditions, we straightforwardly obtain

$$\frac{d\hat{Z}}{dt} = 0, \quad \hat{Z} = \frac{1}{2} \int_A q^2 dA. \quad (5.172)$$

The quantity \hat{Z} is known as the enstrophy, and this is conserved at each height as well as, naturally, over the entire volume.

The enstrophy is just one of an infinity of invariants in quasi-geostrophic flow. Because the potential vorticity of a fluid element is conserved, *any* function of the potential vorticity must be a material invariant and we can immediately write

$$\frac{D}{Dt} F(q) = 0. \quad (5.173)$$

To verify that this is true, simply note that (5.173) implies that $(dF/dq)Dq/Dt = 0$, which is true by virtue of (5.171). [However, by virtue of the material advection, the function $F(q)$ need not be differentiable in order for (5.173) to hold.] Each of the material invariants corresponding to different choices of $F(q)$ has a corresponding integral invariant; that is,

$$\frac{d}{dt} \int_A F(q) dA = 0. \quad (5.174)$$

The enstrophy invariant corresponds to choosing $F(q) = q^2$; it plays a particularly important role because, like energy, it is a quadratic invariant, and its presence profoundly alters the behaviour of two-dimensional and quasi-geostrophic flow compared to three-dimensional flow (see section 8.3).

5.7 ROSSBY WAVES

In the final topic of this chapter we consider wave motion in a quasi-geostrophic system. (A brief introduction to wave kinematics is given in the appendix to this chapter.) Although we consider the closely related topics of hydrodynamic instability and wave-mean flow interaction in Part II, Rossby waves are such a fundamental part of geophysical fluid dynamics, and intimately tied to quasi-geostrophic dynamics, that they find a natural place in this chapter.

5.7.1 Waves in a single layer

Consider flow of a single homogeneous layer on a flat-bottomed β -plane. The unforced, inviscid equation of motion is

$$\frac{D}{Dt} (\zeta + f - \psi/L_d^2) = 0, \quad (5.175)$$

where $\zeta = \nabla^2 \psi$ is the vorticity and ψ is the streamfunction.

Infinite deformation radius

If the scale of motion is much less than the deformation scale then on the β -plane the equation of motion is governed by

$$\frac{D}{Dt}(\zeta + \beta y) = 0. \quad (5.176)$$

Expanding the material derivative gives

$$\frac{\partial \zeta}{\partial t} + \mathbf{u} \cdot \nabla \zeta + \beta v = 0 \quad \text{or} \quad \frac{\partial \zeta}{\partial t} + J(\psi, \zeta) + \beta \frac{\partial \psi}{\partial x} = 0. \quad (5.177)$$

We now *linearize* this equation; that is, we suppose that the flow consists of a time-independent component (the ‘basic state’) plus a perturbation, with the perturbation being small compared with the mean flow. Such a mean flow must satisfy the time-independent equation of motion, and purely zonal flow will do this. For simplicity we choose a flow with no meridional dependence and let

$$\psi = \Psi + \psi'(x, y, t), \quad (5.178)$$

where $\Psi = -Uy$ and $|\psi'| \ll |\Psi|$. (The symbol U represents the zonal flow of the basic state, not a magnitude for scaling purposes.) Substitute (5.178) into (5.177) and neglect the nonlinear terms involving products of ψ' to give

$$\frac{\partial \zeta'}{\partial t} + J(\Psi, \zeta') + \beta \frac{\partial \psi'}{\partial x} = 0 \quad \text{or} \quad \frac{\partial}{\partial t} \nabla^2 \psi' + U \frac{\partial \nabla^2 \psi'}{\partial x} + \beta \frac{\partial \psi'}{\partial x} = 0. \quad (5.179a,b)$$

Solutions to this equation may be found in the form of a plane wave,

$$\psi' = \text{Re } \tilde{\psi} e^{i(kx + ly - \omega t)}, \quad (5.180)$$

where Re indicates the real part of the function (and this will sometimes be omitted if no ambiguity is so-caused). Solutions of the form (5.180) are valid in a domain with doubly-periodic boundary conditions; solutions in a channel can be obtained using a meridional variation of $\sin ly$, with no essential changes to the dynamics. The amplitude of the oscillation is given by $\tilde{\psi}$ and the phase by $kx + ly - \omega t$, where k and l are the x - and y -wavenumbers and ω is the frequency of the oscillation.

Substituting (5.180) into (5.179) yields

$$[(-\omega + Uk)(K^2) + \beta k] \tilde{\psi} = 0, \quad (5.181)$$

where $K^2 = k^2 + l^2$. For non-trivial solutions this implies

$$\omega = Uk - \frac{\beta k}{K^2}.$$

(5.182)

This is the *dispersion relation* for Rossby waves. The phase speed, c_p , and group velocity, c_g , in the x -direction are

$$c_p^x \equiv \frac{\omega}{k} = U - \frac{\beta}{K^2}, \quad c_g^x \equiv \frac{\partial \omega}{\partial k} = U + \frac{\beta(k^2 - l^2)}{(k^2 + l^2)^2}. \quad (5.183a,b)$$

The velocity U provides a uniform translation, and Doppler shifts the frequency. The phase speed in the absence of a mean flow is *westwards*, with waves of longer wavelengths travelling more quickly, and the eastward current speed required to hold the waves of a particular wavenumber stationary (i.e., $c_p^x = 0$) is $U = \beta/K^2$. We discuss the meaning of the group velocity in the appendix.

Finite deformation radius

For a finite deformation radius the basic state $\Psi = -Uy$ is still a solution of the original equations of motion, but the potential vorticity corresponding to this state is $Q = Uy/L_d^2 + \beta y$ and its gradient is $\nabla Q = (\beta + U/L_d^2)\mathbf{j}$. The linearized equation of motion is thus

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) (\nabla^2 \psi' - \psi'/L_d^2) + (\beta + U/L_d^2) \frac{\partial \psi'}{\partial x} = 0. \quad (5.184)$$

Substituting $\psi' = \tilde{\psi} e^{i(kx+ly-\omega t)}$ we obtain the dispersion relation,

$$\omega = \frac{k(UK^2 - \beta)}{K^2 + 1/L_d^2} = U k - k \frac{\beta + U/L_d^2}{K^2 + 1/L_d^2}.$$

(5.185)

The corresponding x -components of phase speed and group velocity are

$$c_p^x = U - \frac{\beta + U k_d^2}{K^2 + k_d^2} = \frac{U K^2 - \beta}{K^2 + k_d^2}, \quad c_g^x = U + \frac{(\beta + U k_d^2)(k^2 - l^2 - k_d^2)}{(k^2 + l^2 + k_d^2)^2}, \quad (5.186a,b)$$

where $k_d = 1/L_d$. The uniform velocity field now no longer provides just a simple Doppler shift of the frequency, nor a uniform addition to the phase speed. From (5.186a) the waves are stationary when $K^2 = K_s^2 \equiv \beta/U$; that is, the current speed required to hold waves of a particular wavenumber stationary is $U = \beta/K^2$. However, this is *not* simply the magnitude of the phase speed of waves of that wavenumber in the absence of a current — this is given by

$$c_p^x = -\frac{\beta}{K_s^2 + k_d^2} = -\frac{U}{1 + k_d^2/K_s^2}. \quad (5.187)$$

Why is there a difference? It is because the current does not just provide a uniform translation, but, if L_d is non-zero, it also modifies the basic potential vorticity gradient. The basic state height field η_0 is sloping, that is $\eta_0 = -(f_0/g)\Psi y$, and the ambient potential vorticity field increases with y , that is $Q = (\beta + U/L_d^2)y$. Thus, the basic state defines a preferred frame of reference, and the problem is not Galilean invariant.¹¹

The mechanism of Rossby waves

The fundamental mechanism underlying Rossby waves is easily understood. Consider a material line of stationary fluid parcels along a line of constant latitude, and suppose that some disturbance causes their displacement to the line marked $\eta(t = 0)$ in Fig. 5.4. In the displacement, the potential vorticity of the fluid parcels is conserved, and in the simplest case of barotropic flow on the β -plane the potential vorticity is the absolute vorticity, $\beta y + \zeta$. Thus, in either hemisphere, a northward displacement leads to the production of negative relative vorticity and a southward displacement leads to the production of positive relative

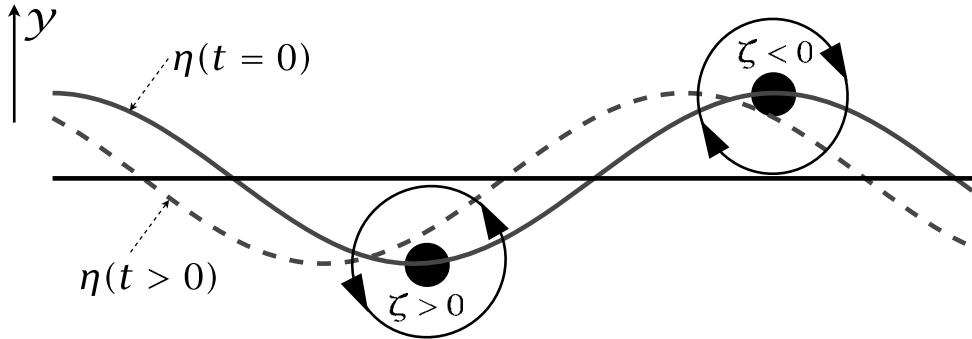


Fig. 5.4 The mechanism of a two-dimensional (x - y) Rossby wave. An initial disturbance displaces a material line at constant latitude (the straight horizontal line) to the solid line marked $\eta(t = 0)$. Conservation of potential vorticity, $\beta y + \zeta$, leads to the production of relative vorticity, as shown for two parcels. The associated velocity field (arrows on the circles) then advects the fluid parcels, and the material line evolves into the dashed line. The phase of the wave has propagated westwards.

vorticity. The relative vorticity gives rise to a velocity field which, in turn, advects the parcels in material line in the manner shown, and the wave propagates westwards.

In more complicated situations, such as flow in two layers, considered below, or in a continuously stratified fluid, the mechanism is essentially the same. A displaced fluid parcel carries with it its potential vorticity and, in the presence of a potential vorticity gradient in the basic state, a potential vorticity anomaly is produced. The potential vorticity anomaly produces a velocity field (an example of potential vorticity inversion) which further displaces the fluid parcels, leading to the formation of a Rossby wave. The vital ingredient is a basic state potential vorticity gradient, such as that provided by the change of the Coriolis parameter with latitude.

5.7.2 Rossby waves in two layers

Now consider the dynamics of the two-layer model, linearized about a state of rest. The two, coupled, linear equations describing the motion in each layer are

$$\frac{\partial}{\partial t} \left[\nabla^2 \psi'_1 + F_1 (\psi'_2 - \psi'_1) \right] + \beta \frac{\partial \psi'_1}{\partial x} = 0, \quad (5.188a)$$

$$\frac{\partial}{\partial t} \left[\nabla^2 \psi'_2 + F_2 (\psi'_1 - \psi'_2) \right] + \beta \frac{\partial \psi'_2}{\partial x} = 0, \quad (5.188b)$$

where $F_1 = f_0^2/g'H_1$ and $F_2 = f_0^2/g'H_2$. By inspection (5.188) may be transformed into two uncoupled equations: the first is obtained by multiplying (5.188a) by F_2 and (5.188b) by F_1 and adding, and the second is the difference of (5.188a) and (5.188b). Then, defining

$$\bar{\psi} = \frac{F_1 \psi'_2 + F_2 \psi'_1}{F_1 + F_2}, \quad \tau = \frac{1}{2} (\psi'_1 - \psi'_2), \quad (5.189a,b)$$

(think ‘ τ for temperature’), (5.188) become

$$\frac{\partial}{\partial t} \nabla^2 \bar{\psi} + \beta \frac{\partial \bar{\psi}}{\partial x} = 0, \quad (5.190a)$$

$$\frac{\partial}{\partial t} \left[(\nabla^2 - k_d^2) \tau \right] + \beta \frac{\partial \tau}{\partial x} = 0, \quad (5.190b)$$

where now $k_d = (F_1 + F_2)^{1/2}$. The internal radius of deformation for this problem is the inverse of this, namely

$$L_d = k_d^{-1} = \frac{1}{f_0} \left(\frac{g' H_1 H_2}{H_1 + H_2} \right)^{1/2}. \quad (5.191)$$

The variables $\bar{\psi}$ and τ are the *normal modes* for the two-layer model, as they oscillate independently of each other. [For the continuous equations the analogous modes are the eigenfunctions of $\partial_z [(f_0^2/N^2) \partial_z \phi] = \lambda^2 \phi$.] The equation for $\bar{\psi}$, the *barotropic mode*, is identical to that of the single-layer, rigid-lid model, namely (5.179) with $U = 0$, and its dispersion relation is just

$$\omega = -\frac{\beta k}{K^2}. \quad (5.192)$$

The barotropic mode corresponds to synchronous, depth-independent, motion in the two layers, with no undulations in the dividing interface.

The displacement of the interface is given by $2f_0\tau/g'$ and so proportional to the amplitude of τ , the *baroclinic mode*. The dispersion relation for the baroclinic mode is

$$\omega = -\frac{\beta k}{K^2 + k_d^2}. \quad (5.193)$$

The mass transport associated with this mode is identically zero, since from (5.189) we have

$$\psi_1 = \bar{\psi} + \frac{2F_1\tau}{F_1 + F_2}, \quad \psi_2 = \bar{\psi} - \frac{2F_2\tau}{F_1 + F_2}, \quad (5.194a,b)$$

and this implies

$$H_1\psi_1 + H_2\psi_2 = (H_1 + H_2)\bar{\psi}. \quad (5.195)$$

The left-hand side is proportional to the total mass transport, which is evidently associated with the barotropic mode.

The dispersion relation and associated group and phase velocities are plotted in Fig. 5.5. The x -component of the phase speed, ω/k , is negative (westwards) for both baroclinic and barotropic Rossby waves. The group velocity of the barotropic waves is always positive (eastwards), but the group velocity of long baroclinic waves may be negative (westwards). For very short waves, $k^2 \gg k_d^2$, the baroclinic and barotropic velocities coincide and their phase and group velocities are equal and opposite. With a deformation radius of 50 km, typical for the mid-latitude ocean, then a non-dimensional frequency of unity in the figure corresponds to a dimensional frequency of $5 \times 10^{-7} \text{ s}^{-1}$ or a period of about 100 days. In an atmosphere with a deformation radius of 1000 km a non-dimensional frequency of unity corresponds to $1 \times 10^{-5} \text{ s}^{-1}$ or a period of about 7 days. Non-dimensional velocities of unity correspond to respective dimensional velocities of about 0.25 m s^{-1} (ocean) and 10 m s^{-1} (atmosphere).

The deformation radius only affects the baroclinic mode. For scales much smaller than the deformation radius, $K^2 \gg k_d^2$, we see from (5.190b) that the baroclinic mode obeys the same equation as the barotropic mode so that

$$\frac{\partial}{\partial t} \nabla^2 \tau + \beta \frac{\partial \tau}{\partial x} = 0. \quad (5.196)$$

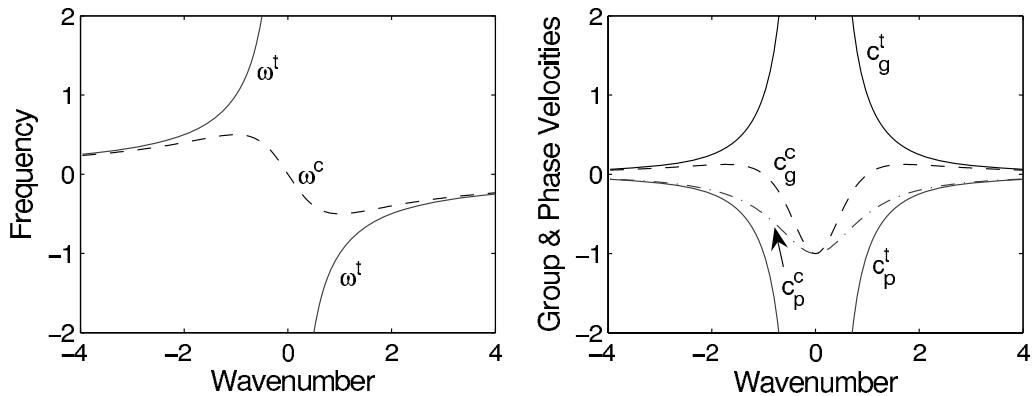


Fig. 5.5 Left: the dispersion relation for barotropic (ω^t , solid line) and baroclinic (ω^c , dashed line) Rossby waves in the two-layer model, calculated using (5.192) and (5.193) with $k^y = 0$, plotted for both positive and negative zonal wavenumbers and frequencies. The wavenumber is non-dimensionalized by k_d , and the frequency is non-dimensionalized by β/k_d . Right: the corresponding zonal group and phase velocities, $c_g = \partial\omega/\partial k^x$ and $c_p = \omega/k^x$, with superscript ‘t’ or ‘c’ for the barotropic or baroclinic mode, respectively. The velocities are non-dimensionalized by β/k_d^2 .

Using this and (5.190a) implies that

$$\frac{\partial}{\partial t} \nabla^2 \psi_i + \beta \frac{\partial \psi_i}{\partial x} = 0, \quad i = 1, 2. \quad (5.197)$$

That is to say, the two layers themselves are uncoupled from each other. At the other extreme, for very long baroclinic waves the relative vorticity is unimportant.

5.8 * ROSSBY WAVES IN STRATIFIED QUASI-GEOSTROPHIC FLOW

5.8.1 Setting up the problem

Let us now consider the dynamics of linear waves in stratified quasi-geostrophic flow on a β -plane, with a resting basic state. (In chapter 13 we explore the role of Rossby waves in a more realistic setting.) The interior flow is governed by the potential vorticity equation, (5.118), and linearizing this about a state of rest gives

$$\frac{\partial}{\partial t} \left[\nabla^2 \psi' + \frac{1}{\tilde{\rho}(z)} \frac{\partial}{\partial z} \left(\tilde{\rho}(z) F(z) \frac{\partial \psi'}{\partial z} \right) \right] + \beta \frac{\partial \psi'}{\partial x} = 0, \quad (5.198)$$

where $\tilde{\rho}$ is the density profile of the basic state and $F(z) = f_0^2/N^2$. (F is the square of the inverse Prandtl ratio, N/f_0 .) In the Boussinesq approximation $\tilde{\rho} = \rho_0$, i.e., a constant. The vertical boundary conditions are determined by the thermodynamic equation, (5.120). If the boundaries are flat, rigid, slippery surfaces then $w = 0$ at the boundaries and if there is no surface buoyancy gradient the linearized thermodynamic equation is

$$\frac{\partial}{\partial t} \left(\frac{\partial \psi'}{\partial z} \right) = 0. \quad (5.199)$$

We apply this at the ground and, with somewhat less justification, at the tropopause — the higher static stability of the stratosphere inhibits vertical motion. If the ground is not flat or if friction provides a vertical velocity by way of an Ekman layer, the boundary condition must be correspondingly modified, but we will stay with the simplest case here and apply (5.199) at $z = 0$ and $z = H$.

5.8.2 Wave motion

As in the single-layer case, we seek solutions of the form

$$\psi' = \operatorname{Re} \tilde{\psi}(z) e^{i(kx+ly-\omega t)}, \quad (5.200)$$

where $\tilde{\psi}(z)$ will determine the vertical structure of the waves. The case of a sphere is more complicated but introduces no truly new physical phenomena.

Substituting (5.200) into (5.198) gives

$$\omega \left[-K^2 \tilde{\psi}(z) + \frac{1}{\tilde{\rho}} \frac{\partial}{\partial z} \left(\tilde{\rho} F(z) \frac{\partial \tilde{\psi}}{\partial z} \right) \right] - \beta k \tilde{\psi}(z) = 0. \quad (5.201)$$

Now, if $\tilde{\psi}$ satisfies

$$\frac{1}{\tilde{\rho}} \frac{\partial}{\partial z} \left(\tilde{\rho} F(z) \frac{\partial \tilde{\psi}}{\partial z} \right) = -\Gamma \tilde{\psi}, \quad (5.202)$$

where Γ is a constant, then the equation of motion becomes

$$-\omega \left[K^2 + \Gamma \right] \tilde{\psi} - \beta k \tilde{\psi} = 0, \quad (5.203)$$

and the dispersion relation follows, namely

$$\omega = -\frac{\beta k}{K^2 + \Gamma}.$$

(5.204)

Equation (5.202) constitutes an eigenvalue problem for the vertical structure; the boundary conditions, derived from (5.199), are $\partial \tilde{\psi} / \partial z = 0$ at $z = 0$ and $z = H$. The resulting eigenvalues, Γ are proportional to the inverse of the squares of the deformation radii for the problem and the eigenfunctions are the vertical structure functions.

A simple example

Consider the case in which $F(z)$ and $\tilde{\rho}$ are constant, and in which the domain is confined between two rigid surfaces at $z = 0$ and $z = H$. Then the eigenvalue problem for the vertical structure is

$$F \frac{\partial^2 \tilde{\psi}}{\partial z^2} = -\Gamma \tilde{\psi} \quad (5.205a)$$

with boundary conditions of

$$\frac{\partial \tilde{\psi}}{\partial z} = 0, \quad \text{at } z = 0, H. \quad (5.205b)$$

There is a sequence of solutions to this, namely

$$\tilde{\psi}_n(z) = \cos(n\pi z/H), \quad n = 1, 2, \dots \quad (5.206)$$

with corresponding eigenvalues

$$\Gamma_n = n^2 \frac{F\pi^2}{H^2} = (n\pi)^2 \left(\frac{f_0}{NH} \right)^2, \quad n = 1, 2, \dots \quad (5.207)$$

Equation (5.207) may be used to define the deformation radii for this problem, namely

$$L_n \equiv \frac{1}{\sqrt{\Gamma_n}} = \frac{NH}{n\pi f_0}. \quad (5.208)$$

The first deformation radius is the same as the expression obtained by dimensional analysis, namely NH/f , except for a factor of π . (Definitions of the deformation radii both with and without the factor of π are common in the literature, and neither is obviously more correct. In the latter case, the first deformation radius in a problem with uniform stratification is given by NH/f , equal to $\pi/\sqrt{\Gamma_1}$.) In addition to these baroclinic modes, the case with $n = 0$, that is with $\tilde{\psi} = 1$, is also a solution of (5.205) for any $F(z)$.

Using (5.204) and (5.207) the dispersion relation becomes

$$\omega = -\frac{\beta k}{K^2 + (n\pi)^2 (f_0/NH)^2}, \quad n = 0, 1, 2, \dots \quad (5.209)$$

and, of course, the horizontal wavenumbers k and l are also quantized in a finite domain. The dynamics of the barotropic mode are independent of height and independent of the stratification of the basic state, and so these Rossby waves are *identical* with the Rossby waves in a homogeneous fluid contained between two flat rigid surfaces. The structure of the baroclinic modes, which in general depends on the structure of the stratification, becomes increasingly complex as the vertical wavenumber n increases. This increasing complexity naturally leads to a certain delicacy, making it rare that they can be unambiguously identified in nature. The eigenproblem for a realistic atmospheric profile is further complicated because of the lack of a rigid lid at the top of the atmosphere.¹²

APPENDIX: WAVE KINEMATICS, GROUP VELOCITY AND PHASE SPEED

This appendix provides a brief and informal look at wave kinematics and the meaning of phase speed and group velocity.¹³

5.A.1 Kinematics and definitions

A wave may be defined as a disturbance that satisfies a dispersion relation. To see what this means, suppose a disturbance, $\psi(\mathbf{x}, t)$ (where ψ might be velocity, streamfunction, pressure, etc), satisfies some equation

$$L(\psi) = 0, \quad (5.210)$$

where L is a linear operator, typically a polynomial in time and space derivatives; an example is $\partial \nabla^2 \psi / \partial t + \beta \partial \psi / \partial x$. If (5.210) has constant coefficients (if β is constant in our example) then solutions may often be found as a superposition of *plane waves*, each of which satisfy

$$\psi = \operatorname{Re} \tilde{\psi} e^{i\theta(x,t)} = \operatorname{Re} \tilde{\psi} e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}. \quad (5.211)$$

where $\tilde{\psi}$ is a constant, θ is the phase, \mathbf{k} is the vector wavenumber (k^x, k^y, k^z) , and ω is the wave frequency. [We also often write the wave vector as $\mathbf{k} = (k, l, m)$.] The frequency and wavevector are related by the *dispersion relation*,

$$\omega = \Omega(\mathbf{k}), \quad (5.212)$$

where $\Omega(\mathbf{k})$ [meaning $\Omega(k, l, m)$] is some function determined by the form of L . Unless it is necessary to explicitly distinguish the function Ω from the frequency ω , we will often write $\omega = \omega(\mathbf{k})$. Two examples of dispersion relations are (2.248) and (5.182).

If the medium in which the waves are propagating is inhomogeneous, then (5.210) will probably not have constant coefficients (for example, β may vary meridionally). Nevertheless, if the medium is slowly varying, wave solutions may often still be found — although we do not prove it here — with the the general form

$$\psi = \operatorname{Re} \tilde{\psi}(\mathbf{x}, t) e^{i\theta(\mathbf{x}, t)}, \quad (5.213)$$

where $\tilde{\psi}(\mathbf{x}, t)$ varies slowly compared to the variation of the phase, θ . The frequency and wavenumber are then defined by

$$\mathbf{k} \equiv \nabla \theta, \quad \omega \equiv -\frac{\partial \theta}{\partial t}, \quad (5.214)$$

which in turn implies the formal relation between \mathbf{k} and ω :

$$\frac{\partial \mathbf{k}}{\partial t} + \nabla \omega = 0. \quad (5.215)$$

Even if the medium is inhomogeneous, we may still have a *local* dispersion relation between frequency and wavevector,

$$\omega = \Omega(\mathbf{k}, \mathbf{x}, t). \quad (5.216)$$

5.A.2 Wave propagation

Phase speed

First, consider the propagation of monochromatic plane waves. Given (5.211) a wave will propagate in the direction of \mathbf{k} (Fig. 5.6). At a given instant and location we can align our coordinate axis along this direction, and we may write $\mathbf{k} \cdot \mathbf{x} = Kx^*$, where x^* increases in the direction of \mathbf{k} and $K^2 = |\mathbf{k}|^2$ is the magnitude of the wavenumber. With this, we can write (5.211) as

$$\psi = \operatorname{Re} \tilde{\psi} e^{i(Kx^* - \omega t)} = \operatorname{Re} \tilde{\psi} e^{iK(x^* - ct)}, \quad (5.217)$$

where $c = \omega/K$. From this equation it is evident that the phase of the wave propagates at the speed c in the direction of \mathbf{k} , and we define the *phase speed* by

$$c_p \equiv \frac{\omega}{K}. \quad (5.218)$$

The wavelength of the wave, λ , is the distance between two wavecrests — that is, the distance between two locations along the line of travel whose phase differs by 2π — and evidently this is given by

$$\lambda = \frac{2\pi}{K}. \quad (5.219)$$

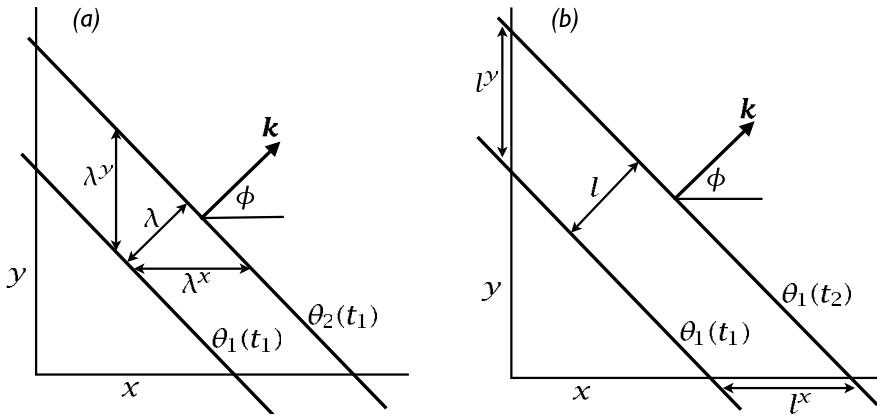


Fig. 5.6 The propagation of a two-dimensional wave. (a) Two lines of constant phase (e.g., two wavecrests) at a time t_1 . The wave is propagating in the direction \mathbf{k} with wavelength λ . (b) The same line of constant phase at two successive times. The phase speed is the speed of advancement of the wavecrest in the direction of travel, and so $c_p = l/(t_2 - t_1)$. The phase speed in the x -direction is the speed of propagation of the wavecrest along the x -axis, and $c_p^x = l^x/(t_2 - t_1) = c_p / \cos \phi$.

In (for simplicity) a two-dimensional wave, and referring to Fig. 5.6(a), the wavelength and wave vectors in the x - and y -directions are given by,

$$\lambda^x = \frac{\lambda}{\cos \phi}, \quad \lambda^y = \frac{\lambda}{\sin \phi}, \quad k^x = K \cos \phi, \quad k^y = K \sin \phi. \quad (5.220)$$

In general, lines of constant phase intersect both the coordinate axes and propagate along them. The speed of propagation along these axes is given by

$$c_p^x = c_p \frac{l^x}{\cos \phi} = c_p \frac{K}{k^x} = c_p \frac{\omega}{k^x}, \quad c_p^y = c_p \frac{l^y}{\sin \phi} = c_p \frac{K}{k^y} = c_p \frac{\omega}{k^y}, \quad (5.221)$$

using (5.220). The speed of phase propagation along any one of the axis is in general *larger* than the phase speed in the primary direction of the wave. The phase speeds are clearly *not* components of a vector: for example, $c_p^x \neq c_p \cos \phi$. To summarize, the phase speed and its components are given by

$$c_p = \frac{\omega}{K}, \quad c_p^x = \frac{\omega}{k^x}, \quad c_p^y = \frac{\omega}{k^y}.$$

(5.222)

Group velocity

Let us consider how the wavevector and frequency might change with position and time. Using (5.216) in (5.215) gives

$$\frac{D_{c_g} \mathbf{k}}{Dt} \equiv \frac{\partial \mathbf{k}}{\partial t} + \mathbf{c}_g \cdot \nabla \mathbf{k} = -\nabla \Omega \quad (5.223)$$

where

$$\mathbf{c}_g \equiv \frac{\partial \Omega}{\partial \mathbf{k}} \equiv \left(\frac{\partial \Omega}{\partial k}, \frac{\partial \Omega}{\partial l}, \frac{\partial \Omega}{\partial m} \right), \quad (5.224)$$

is the *group velocity*, sometimes written as $\mathbf{c}_g = \nabla_{\mathbf{k}}\omega$ or, in subscript notation, as $c_{gi} = \partial\Omega/\partial k_i$. The group velocity is a vector. If the frequency is spatially constant the wavevector is evidently propagated at the group velocity.

The frequency is, in general, a function of space, wavenumber and time, and from the dispersion relation, (5.216), its variation is governed by

$$\frac{\partial \omega}{\partial t} = \frac{\partial \Omega}{\partial t} + \frac{\partial \Omega}{\partial \mathbf{k}} \frac{\partial \mathbf{k}}{\partial t} = \frac{\partial \Omega}{\partial t} - \frac{\partial \Omega}{\partial \mathbf{k}} \nabla \omega \quad (5.225)$$

using (5.215). Thus, using the definition of group velocity, we have

$$\frac{D_{c_g} \omega}{Dt} \equiv \frac{\partial \omega}{\partial t} + \mathbf{c}_g \cdot \nabla \omega = \frac{\partial \Omega}{\partial t}. \quad (5.226)$$

If the dispersion relation is not an function of time, the frequency propagates at the group velocity. We may define a *ray* as the trajectory traced by the group velocity, and if the frequency is not an explicit function of space or time, then both the wavevector and the frequency are constant along a ray.

5.A.3 Meaning of group velocity

In the discussion above we introduced the group velocity in a kinematic and rather formal way. What is the group velocity, physically? Information and energy travel clearly cannot travel at the phase speed, for as the direction of propagation of the phase line tends to a direction parallel to the y -axis, the phase speed in the x -direction tends to infinity! Rather, they travel at the group velocity, and to see this we consider the superposition of plane waves, noting that a monochromatic plane wave already fills space uniformly. We will restrict attention to waves propagating in one direction, but the argument may be extended to two or three dimensions.

Superposition of two waves

Consider the linear superposition of two waves. Limiting attention to the one-dimensional case for simplicity, consider a disturbance represented by

$$\psi = \operatorname{Re} \tilde{\psi} (e^{i(k_1 x - \omega_1 t)} + e^{i(k_2 x - \omega_2 t)}). \quad (5.227)$$

Let us further suppose that the two waves have similar wavenumbers and frequency, and, in particular, that $k_1 = k + \Delta k$ and $k_2 = k - \Delta k$, and $\omega_1 = \omega + \Delta\omega$ and $\omega_2 = \omega - \Delta\omega$. With this, (5.227) becomes

$$\begin{aligned} \psi &= \operatorname{Re} \tilde{\psi} e^{i(kx - \omega t)} [e^{i(\Delta k x - \Delta\omega t)} + e^{-i(\Delta k x - \Delta\omega t)}] \\ &= 2 \operatorname{Re} \tilde{\psi} e^{i(kx - \omega t)} \cos(\Delta k x - \Delta\omega t). \end{aligned} \quad (5.228)$$

The resulting disturbance, illustrated in Fig. 5.7 has two aspects: a rapidly varying component, with wavenumber k and frequency ω , and a more slowly varying envelope, with wavenumber

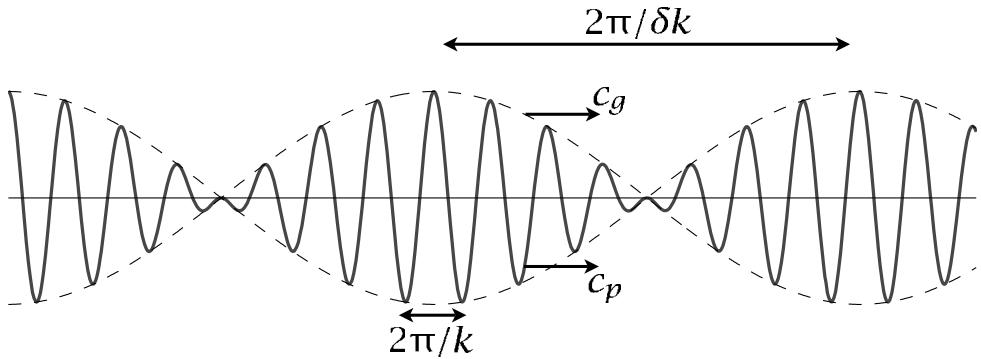


Fig. 5.7 Superposition of two sinusoidal waves with wavenumbers k and $k + \delta k$, producing a wave (solid line) that is modulated by a slowly varying wave envelope or wave packet (dashed line). The envelope moves at the group velocity, $c_g = \partial \omega / \partial k$ and the phase of the wave moves at the group speed $c_p = \omega / k$.

Δk and frequency $\Delta\omega$. The envelope modulates the fast oscillation, and moves with velocity $\Delta\omega / \Delta k$; in the limit $\Delta k \rightarrow 0$ and $\Delta\omega \rightarrow 0$ this is the *group velocity*, $c_g = \partial \omega / \partial k$. Group velocity is equal to the phase speed, ω / k , only when the frequency is a linear function of wavenumber. The energy in the disturbance must move at the group velocity — note that the node of the envelope moves at the speed of the envelope and no energy can cross the node. These concepts generalize to more than one dimension, and if the wavenumber is the three-dimensional vector $\mathbf{k} = (k, l, m)$ then the three-dimensional envelope propagates at the group velocity given by (5.224).

* Superposition of many waves

Now consider a generalization of the above arguments. We will suppose that the disturbance is a *wave packet* of the form

$$\psi = A(x, t) e^{i(kx - \omega t)} \quad (5.229)$$

where $A(x, t)$, like the envelope in Fig. 5.7, modulates the amplitude of the wave on a scale much longer than that of the wavelength $2\pi/k$. We assume that the packet is confined to a finite region of space, and that it contains a superposition of Fourier modes with wavenumbers near to k . That is, at some fixed time, $t = 0$ say,

$$\psi(x) = A(x) e^{ikx}, \quad A(x) = \int_{-\Delta k}^{+\Delta k} \tilde{A}(k') e^{ik'x} dk', \quad (5.230a,b)$$

where Δk is small. Each of the wavenumber components oscillates at a frequency given by the dispersion relation at hand, so that the evolution of the packet is given by

$$A(x, t) = \int_{-\Delta k}^{+\Delta k} \tilde{A}(k') e^{i(k'x - \omega' t)} dk' \approx \int_{-\Delta k}^{+\Delta k} \tilde{A}(k') e^{ik'(x - t\partial\omega/\partial k')} dk', \quad (5.231)$$

where we have used the fact that the wavenumber range is small, so that $\omega' \approx k'(\partial\omega/\partial k')$, where the derivative is evaluated at k itself. Thus, comparing (5.230b) and (5.231), we see that

$$A(x, t) = A(x - c_g t), \quad (5.232)$$

where $c_g = \partial \omega / \partial k$. Thus, the packet moves at the group velocity.

Notes

- 1 The phrase ‘quasi-geostrophic’ seems to have been introduced by Durst & Sutcliffe (1938) and the concept used in Sutcliffe’s development theory of baroclinic systems (Sutcliffe 1939, 1947). The first systematic derivation of the quasi-geostrophic equations based on scaling theory was given by Charney (1948). The planetary-geostrophic equations were used by Robinson & Stommel (1959) and Welander (1959) in studies of the thermocline (and were first known as the ‘thermocline equations’), and were put in the context of other approximate equation sets by Phillips (1963).
- 2 Carl-Gustav Rossby (1898–1957) played a dominant role in the development of dynamical meteorology in the early and middle parts of the twentieth century, and his work permeates all aspects of dynamical meteorology today. Perhaps the most fundamental non-dimensional number in rotating fluid dynamics, the Rossby number, is named for him, as is the perhaps most fundamental wave, the Rossby wave. He also discovered the conservation of potential vorticity (later generalized by Ertel) and contributed important ideas to atmospheric turbulence and the theory of air masses. Swedish born, he studied first with V. Bjerknes before taking a position in Stockholm in 1922 with the Swedish Meteorological Hydrologic Service and receiving a ‘Licentiat’ from the University of Stockholm in 1925. Shortly thereafter he moved to the United States, joining the Government Weather Bureau, a precursor of NOAA’s National Weather Service. In 1928 he moved to MIT, playing an important role in developing the meteorology department there, while still maintaining connections with the Weather Bureau. In 1940 he moved to the University of Chicago, where he similarly helped develop meteorology there. In 1947 he became director of the newly formed Institute of Meteorology in Stockholm, and subsequently divided his time between there and the United States. Thus, as well as his scientific contributions, he played a very influential role in the institutional development of the field.
- 3 Burger (1958).
- 4 This is the so-called ‘frontal geostrophic’ regime (Cushman-Roisin 1994).
- 5 Numerical integrations of the potential vorticity equation using (5.91), and performing the inversion without linearizing potential vorticity, do in fact indicate improved accuracy over either the quasi-geostrophic or planetary-geostrophic equations (Mundt *et al.* 1997). In a similar vein, McIntyre & Norton (2000) show how useful potential vorticity inversion can be, and Allen *et al.* (1990a,b) demonstrate the high accuracy of certain intermediate models. Certainly, asymptotic correctness should not be the only criterion used in constructing a filtered model, because the parameter range in which the model is useful may be too limited. Note that there is a difference between extending the parameter range in which a filtered model is useful, as in the inversion of (5.91), and going to higher asymptotic order accuracy in a given parameter regime, as in Allen (1993) and Warn *et al.* (1995). Using Hamiltonian mechanics it is possible to derive equations that span different asymptotic regimes, and that also have good conservation properties (Salmon 1983, Allen *et al.* 2002).
- 6 There is a difference between the *dynamical* demands of the quasi-geostrophic system in requiring β to be small, and the *geometric* demands of the Cartesian geometry. On Earth the two demands are similar in practice. But without dynamical inconsistency we may imagine a Cartesian system in which $\beta y \sim f$, and indeed this is common in idealized, planetary geostrophic, models of the large-scale ocean circulation.

- 7 The atmospheric and oceanic sciences are sometimes thought of as not being ‘beautiful’ in the same way as some branches of theoretical physics. Yet surely quasi-geostrophic theory, and the quasi-geostrophic potential vorticity equation, are quite beautiful, both for their austerity of description and richness of behaviour.
- 8 Bretherton (1966). Schneider *et al.* (2003) look at the non-QG extension. The equivalence between boundary conditions and delta-function sources is a common feature of elliptic and similar problems, and is analogous to the generation of electromagnetic fields by point charges. It is sometimes exploited in the numerical solution of elliptic equations, both as a simple way to include non-homogeneous boundary conditions and, using the so-called capacitance matrix method, to solve problems in irregular domains (e.g., Hockney 1970).
- 9 Phillips (1954, 1956) used a two-level model for instability studies and to construct a simple general circulation model of the atmosphere.
- 10 Charney & Stern (1962). See also Berrisford *et al.* (1993) and Vallis (1996).
- 11 This non-Doppler effect also arises quite generally, even in models in height coordinates. See White (1977) and problem 5.9.
- 12 See Chapman & Lindzen (1970).
- 13 For a review of waves and group velocity, see Lighthill (1965).

Further reading

Majda, A. J., 2003. *Introduction to PDEs and Waves for the Atmosphere and Ocean*.

Provides a compact, somewhat mathematical introduction to various equation sets and their properties, including quasi-geostrophy.

Problems

- 5.1 In the derivation of the quasi-geostrophic equations, geostrophic balance leads to the lowest-order velocity being divergence-free — that is, $\nabla_z \cdot \mathbf{u}_0 = 0$. It seems that this can also be obtained from the mass conservation equation at lowest order. Is this a coincidence? Suppose that the Coriolis parameter varied, and that the momentum equation yielded $\nabla_z \cdot \mathbf{u}_0 \neq 0$. Would there be an inconsistency?
- 5.2 ♦ In the planetary-geostrophic approximation, obtain an evolution equation and corresponding inversion conditions that conserves potential vorticity and that are accurate to one higher order in Rossby number than the usual shallow water planetary-geostrophic equations.
- 5.3 Consider the flat-bottomed shallow water potential vorticity equation in the form

$$\frac{D}{Dt} \frac{\zeta + f}{h} = 0 \quad (\text{P5.1})$$

- (a) Suppose that deviations of the height field are small compared to the mean height field, and that the Rossby number is small (so $|\zeta| \ll f$). Further consider flow on a β -plane such that $f = f_0 + \beta y$ where $|\beta y| \ll f_0$. Show that the evolution equation becomes

$$\frac{D}{Dt} \left(\zeta + \beta y - \frac{f_0 \eta}{H} \right) = 0 \quad (\text{P5.2})$$

where $h = H + \eta$ and $|\eta| \ll H$. Using geostrophic balance in the form $f_0 u = -g \partial \eta / \partial y$, $f_0 v = g \partial \eta / \partial x$, obtain an expression for ζ in terms of η .

- (b) Linearize (P5.2) about a state of rest, and show that the resulting system supports two-dimensional Rossby waves that are similar to those of the usual two-dimensional barotropic system. Discuss the limits in which the wavelength is much shorter or much longer than the deformation radius.

- (c) Linearize (P5.2) about a *geostrophically balanced state* that is translating uniformly eastwards. Note that this means that:

$$u = U + u' \quad \eta = \eta(y) + \eta',$$

where $\eta(y)$ is in geostrophic balance with U . Obtain an expression for the form of $\eta(y)$.

- (d) Obtain the dispersion relation for Rossby waves in this system. Show that their speed is different from that obtained by adding a constant U to the speed of Rossby waves in part (b), and discuss why this should be so. (That is, why is the problem not *Galilean invariant*?)

5.4 Obtain solutions to the two-layer Rossby wave problem by seeking solutions of the form

$$\psi_1 = \operatorname{Re} \tilde{\psi}_1 e^{i(k_x x + k_y y - \omega t)}, \quad \psi_2 = \operatorname{Re} \tilde{\psi}_2 e^{i(k_x x + k_y y - \omega t)}. \quad (\text{P5.3})$$

Substitute (P5.3) directly into (5.188) to obtain the dispersion relation, and show that the ensuing two roots correspond to the baroclinic and barotropic modes.

- 5.5 ♦ (Not difficult, but messy.) Obtain the vertical normal modes and the dispersion relationship of the two-layer quasi-geostrophic problem with a free surface, for which the equations of motion linearized about a state of rest are

$$\frac{\partial}{\partial t} \left[\nabla^2 \psi_1 + F_1(\psi_2 - \psi_1) \right] + \beta \frac{\partial \psi_1}{\partial x} = 0 \quad (\text{P5.4a})$$

$$\frac{\partial}{\partial t} \left[\nabla^2 \psi_2 + F_2(\psi_1 - \psi_2) - F_{\text{ext}} \psi_1 \right] + \beta \frac{\partial \psi_2}{\partial x} = 0, \quad (\text{P5.4b})$$

where $F_{\text{ext}} = f_0 / (gH_2)$.

- 5.6 Given the baroclinic dispersion relation, $\omega = -\beta k^x / (k^x{}^2 + k_d^2)$, for what value of k_x is the x -component of the group velocity the largest (i.e., the most positive), and what is the corresponding value of the group velocity?
- 5.7 Beginning with the vorticity and thermodynamic equations for a two-layer model, obtain an expression for the conversion between available potential energy and kinetic energy in the two-layer model. Show that these expressions are consistent with the conservation of total energy as expressed by (5.168). Show also that the expression might be considered to be a simple finite-difference approximation to (5.165).
- 5.8 ♦ The vertical normal modes are the eigenfunctions of

$$\frac{1}{\tilde{\rho}} \frac{\partial}{\partial z} \left(\frac{\tilde{\rho}}{N^2} \frac{\partial \Psi}{\partial z} \right) = -\Gamma \Psi \quad (\text{P5.5})$$

along with boundary conditions on $\Psi(z)$. Obtain numerically the vertical normal modes for some or all of the following profiles, or others of your choice.

- (a) $\tilde{\rho} = 1$, $N^2(z) = 1$, $\Psi_z = 0$ at $z = 0, 1$. (This is profile 'b' of Fig. 9.10. An analytic solution is possible.)
- (b) $\tilde{\rho} = 1$ and an $N^2(z)$ profile corresponding to a density profile similar to profile 'a' of Fig. 9.10 (e.g., an exponential), and $\Psi_z = 0$ at $z = 0, 1$.
- (c) An isothermal atmosphere, with $\Psi_z = 0$ at $z = 0, 1$. (Similar to (i), except that $\tilde{\rho}$ varies with height.)
- (d) An isothermal atmosphere, now assuming that $\psi \rightarrow 0$ as $z \rightarrow \infty$.
- (e) A fluid with $N^2 = 1$ for $0 < z < 0.5$ and $N^2 = 4$ for $0.5 < z < 1$, with continuous b , and with $\Psi_z = 0$ at $z = 0, 1$.

In the atmospheric cases it is easiest to do the problem first with $\tilde{\rho} = 1$ (the Boussinesq case) and then extend the problem (and the code) to the compressible case. Then remove the upper boundary to larger and larger values of z .

- 5.9 ♦ Show that the non-Doppler effect arises using geometric height as the vertical coordinate, using the modified quasi-geostrophic set of White (1977). In particular, obtain the dispersion relation for stratified quasi-geostrophic flow with a resting basic state. Then obtain the dispersion relation for the equations linearized about a uniformly translating state, paying attention to the lower boundary condition, and note the conditions under which the waves are stationary. Discuss.
- 5.10 Derive the quasi-geostrophic potential vorticity equation in isopycnal coordinates for a Boussinesq fluid. Show that the isopycnal expression for potential vorticity is approximately equal to the corresponding expression in height coordinates, carefully stating any assumptions that may be necessary to show this.
- 5.11 (a) Obtain the dispersion relationship for Rossby waves in the single-layer quasi-geostrophic potential vorticity equation with linear drag.
(b) Obtain the dispersion relation for Rossby waves in the linearized two-layer potential vorticity equation with linear drag in the lowest layer.
(c) ♦ Obtain the dispersion relation for Rossby waves in the continuously stratified quasi-geostrophic equations, with the effects of linear drag appearing in the thermodynamic equation for the lower boundary condition. That is, the boundary condition at $z = 0$ is $\partial_t(\partial_z \psi) + N^2 w = 0$ where $w = \alpha \zeta$ with α being a constant. You may make the Boussinesq approximation and assume N^2 is constant if you like.

Part II

INSTABILITIES, WAVE-MEAN FLOW INTERACTION AND TURBULENCE

There is a degree of instability which is inconsistent with civilization. But on the whole, the great ages have been unstable ones.

Alfred North Whitehead (1861–1947).

CHAPTER SIX

Barotropic and Baroclinic Instability

WHAT HYDRODYNAMIC STATES OCCUR IN NATURE? If we take as given the applicability of the Navier-Stokes equations, any flow must be a solution of these equations, subject to the relevant initial and boundary conditions. Most of the flows we experience are of course *time-dependent* solutions, not steady solutions. Why should this be? There are, in fact, many steady solutions to the equations of motion — certain purely zonal flows, for example. However, steady solutions do not abound in nature because, in order to persist, they must be stable to those small perturbations that inevitably arise. Indeed, all the steady solutions that are known for the large-scale flow in the Earth's atmosphere and ocean have been found to be unstable. It is such instability that makes the subject an interesting one.

There are a myriad forms of hydrodynamic instability, but our focus in this chapter is on barotropic and baroclinic instability. *Baroclinic instability* (and we will define the term more precisely later on) is an instability that arises in rotating, stratified fluids that are subject to a horizontal temperature gradient. It is the instability that gives rise to the large- and mesoscale motion in the atmosphere and ocean — it produces atmospheric weather systems, for example — and so is, perhaps, the form of hydrodynamic instability that most affects the human condition. *Barotropic instability* is an instability that arises because of the shear in a flow, and may occur in fluids of constant density. It is important to us for two reasons: first, it is important in its own right as an instability mechanisms for jets and vortices, and as an important process in both two- and three-dimensional turbulence; second, many problems in barotropic and baroclinic instability are formally and dynamically similar, so that the solutions and insight we obtain in the often simpler problems in barotropic instability may be useful in the baroclinic problem.

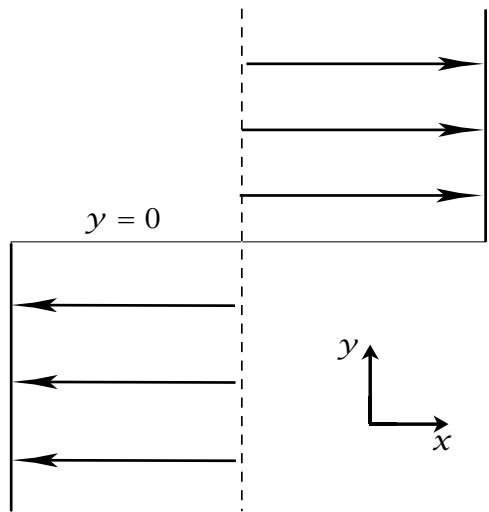


Fig. 6.1 A simple basic state giving rise to shear-flow instability. The velocity profile is discontinuous and the density is uniform.

6.1 KELVIN-HELMHOLTZ INSTABILITY

To introduce the issue, we will first consider, rather informally, perhaps the simplest physically interesting instance of a fluid-dynamical instability — that of a constant-density flow with a shear perpendicular to the fluid's mean velocity, this being an example of a *Kelvin-Helmholtz instability*,¹ and of a barotropic instability. (More generally, Kelvin-Helmholtz instability can involve fluid with varying density.) Let us consider two fluid masses of equal density, with a common interface at $y = 0$, moving with velocities $-U$ and $+U$ in the x -direction, respectively (Fig. 6.1). There is no variation in the basic flow in the z -direction (normal to the page), and we will assume this is also true for the instability (these restrictions are not essential). This flow is clearly a solution of the Euler equations. What happens if the flow is perturbed slightly? If the perturbation is initially small then even if it grows we can, for small times after the onset of instability, neglect the nonlinear interactions in the governing equations because these are the squares of small quantities. The equations determining the evolution of the initial perturbation are then the Euler equations linearized about the steady solution. Thus, denoting perturbation quantities with a prime and basic state variables with capital letters, for $y > 0$ the perturbation satisfies

$$\frac{\partial \mathbf{u}'}{\partial t} + U \frac{\partial \mathbf{u}'}{\partial x} = -\nabla p', \quad \nabla \cdot \mathbf{u}' = 0 \quad (6.1a,b)$$

and a similar equation holds for $y < 0$, but with U replaced by $-U$. Given periodic boundary conditions in the x -direction, we may seek solutions of the form

$$\phi'(x, y, t) = \text{Re} \sum_k \tilde{\phi}_k(y) \exp[ik(x - ct)], \quad (6.2)$$

where ϕ is any field variable (e.g., pressure or velocity), and Re denotes that only the real part should be taken. (Typically we use tildes over variables to denote Fourier-like modes, and we will often omit the marker 'Re'.) Because (6.1a) is linear, the Fourier modes do not interact and we may confine attention to just one. Taking the divergence of (6.1a), the

left-hand side vanishes and the pressure satisfies Laplace's equation

$$\nabla^2 p' = 0. \quad (6.3)$$

This has solutions in the form

$$p' = \begin{cases} \operatorname{Re} \tilde{p}_1 e^{ikx - ky} e^{\sigma t} & y > 0, \\ \operatorname{Re} \tilde{p}_2 e^{ikx + ky} e^{\sigma t} & y < 0, \end{cases} \quad (6.4)$$

where, anticipating the possibility of growing solutions, we have written the time variation in terms of a growth rate, $\sigma = -ikc$. In general σ is complex: if it has a positive real component, the amplitude of the perturbation will grow and there is an instability; if σ has a non-zero imaginary component, then there will be oscillatory motion, and there may of course be both oscillatory motion *and* an instability. To obtain the dispersion relationship, we consider the y -component of (6.1a), namely (for $y > 0$)

$$\frac{\partial v'_1}{\partial t} + U \frac{\partial v'_1}{\partial x} = -\frac{\partial p'_1}{\partial y}. \quad (6.5)$$

Substituting a solution of the form $v'_1 = \tilde{v}_1 \exp(ikx + \sigma t)$ yields, with (6.4),

$$(\sigma + ikU) \tilde{v}_1 = k \tilde{p}_1. \quad (6.6)$$

But the velocity normal to the interface is, at the interface, nothing but the rate of change of the position of the interface itself; that is, at $y = +0$

$$v_1 = \frac{\partial \eta'}{\partial t} + U \frac{\partial \eta'}{\partial x}, \quad (6.7)$$

or

$$\tilde{v}_1 = (\sigma + ikU) \tilde{\eta}. \quad (6.8)$$

where η' is the displacement of the interface from its equilibrium position. Using this in (6.6) gives

$$(\sigma + ikU)^2 \tilde{\eta} = k \tilde{p}_1. \quad (6.9)$$

The above few equations pertain to motion on the $y > 0$ side of the interface. Similar reasoning on the other side gives (at $y = -0$)

$$(\sigma - ikU)^2 \tilde{\eta} = -k \tilde{p}_2. \quad (6.10)$$

But at the interface $p_1 = p_2$ (because pressure must be continuous). The dispersion relationship then emerges from (6.9) and (6.10), giving

$$\sigma^2 = k^2 U^2. \quad (6.11)$$

This equation has two roots, one of which is positive. Thus, the amplitude of the perturbation grows exponentially, like $e^{\sigma t}$, and the flow is *unstable*. The instability itself can be seen in the natural world when billow clouds appear wrapped up into spirals: the clouds are acting as tracers of fluid flow, and are a manifestation of the instability at finite amplitude, as seen later in Fig. 6.6.

6.2 INSTABILITY OF PARALLEL SHEAR FLOW

We now consider a little more systematically the instability of *parallel shear flows*, such as are illustrated in Fig. 6.2.² This is a classic problem in hydrodynamic stability theory, and there are two particular reasons for our own interest.

- (i) The instability is an example of *barotropic instability*, which abounds in the ocean and atmosphere. Loosely, barotropic instability arises when a flow is unstable by virtue of its horizontal shear, with gravitational and buoyancy effects being secondary.
- (ii) The instability is in many ways analogous to *baroclinic instability*, which is the main instability giving rise to weather systems in the atmosphere and similar phenomena in the ocean.

We will restrict attention to two-dimensional, incompressible flow; this illustrates the physical mechanisms in the most transparent way, in part because it allows for the introduction of a streamfunction and the automatic satisfaction of the mass continuity equation. In fact, for parallel two-dimensional shear flows the most unstable disturbances are two-dimensional ones.³

The vorticity equation for incompressible two-dimensional flow is just

$$\frac{D\zeta}{Dt} = 0. \quad (6.12)$$

We suppose the basic state to be a parallel flow in the x -direction that may vary in the y -direction. That is

$$\bar{\mathbf{u}} = U(y)\mathbf{i}. \quad (6.13)$$

The linearized vorticity equation is then

$$\frac{\partial \zeta'}{\partial t} + U \frac{\partial \zeta'}{\partial x} + v' \frac{\partial Z}{\partial y} = 0 \quad (6.14)$$

where $Z = -\partial_y U$. Because the mass continuity equation has the simple form

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} = 0, \quad (6.15)$$

we may introduce a streamfunction ψ such that $u' = -\partial\psi'/\partial y$, $v' = \partial\psi'/\partial x$ and $\zeta' = \nabla^2\psi'$. The linear vorticity equation is then

$$\frac{\partial \nabla^2\psi'}{\partial t} + U \frac{\partial \nabla^2\psi'}{\partial x} + \frac{\partial Z}{\partial y} \frac{\partial \psi'}{\partial x} = 0. \quad (6.16)$$

The coefficients of the x -derivatives are not themselves functions of x ; thus, we may seek solutions that are harmonic functions (sines and cosines) in the x -direction, but the y dependence must remain arbitrary at this stage and we write

$$\psi' = \text{Re } \tilde{\psi}(y) e^{ik(x-ct)}. \quad (6.17)$$

The full solution is a superposition of all wavenumbers, but since the problem is linear the waves do not interact and it suffices to consider them separately. If c is purely real then c is the phase speed of the wave; if c has a positive imaginary component then the wave will grow exponentially and is thus *unstable*.

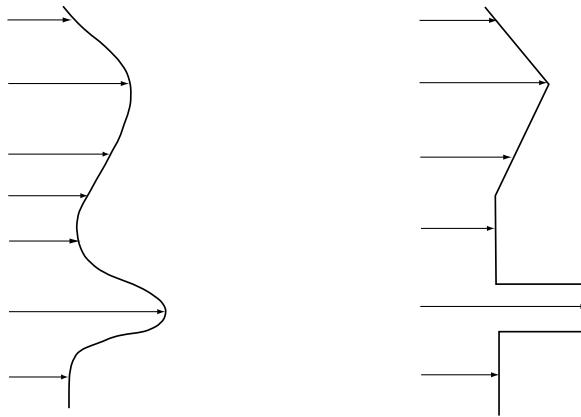


Fig. 6.2 Left: example of a smooth velocity profile, in which both the velocity and the vorticity are continuous. Right: example piecewise continuous profile, in which the velocity and vorticity may have finite discontinuities.

From (6.17) we have

$$u' = \tilde{u}(\gamma) e^{ik(x-ct)} = -\tilde{\psi}_\gamma e^{ik(x-ct)}, \quad (6.18a)$$

$$v' = \tilde{v}(\gamma) e^{ik(x-ct)} = ik\tilde{\psi} e^{ik(x-ct)}, \quad (6.18b)$$

$$\zeta' = \tilde{\zeta}(\gamma) e^{ik(x-ct)} = (-k^2\tilde{\psi} + \tilde{\psi}_{\gamma\gamma}) e^{ik(x-ct)}, \quad (6.18c)$$

where the γ subscript denotes a derivative. Using (6.18) in (6.14) gives

$$(U - c)(\tilde{\psi}_{\gamma\gamma} - k^2\tilde{\psi}) - U_{\gamma\gamma}\tilde{\psi} = 0, \quad (6.19)$$

which is known as *Rayleigh's equation*.⁴ It is the linear vorticity equation for disturbances to parallel shear flow, and in the presence of a β -effect it generalizes slightly to

$$(U - c)(\tilde{\psi}_{\gamma\gamma} - k^2\tilde{\psi}) + (\beta - U_{\gamma\gamma})\tilde{\psi} = 0. \quad (6.20)$$

6.2.1 Piecewise linear flows

Although Rayleigh's equation is linear and has a simple form, it is nevertheless quite difficult to analytically solve for an arbitrary smoothly varying profile. It is simpler to consider *piecewise linear* flows, in which U_γ is a constant over some interval, with U or U_γ changing abruptly to another value at a line of discontinuity, as illustrated in Fig. 6.2. The curvature, $U_{\gamma\gamma}$ is accounted for through the satisfaction of matching conditions, analogous to boundary conditions, at the lines of discontinuity (as in section 6.1), and solutions in each interval are then exponential functions.

Jump or matching conditions

The idea, then, is to solve the linearized vorticity equation separately in the continuous intervals in which vorticity is constant, matching the solution with that in the adjacent regions. The matching conditions arise from two physical conditions:

- (i) That normal stress should be continuous across the interface. For an inviscid fluid this implies that pressure be continuous.
- (ii) That the normal velocity of the fluid on either side of the interface should be consistent with the motion of the interface itself.

Let us consider the implications of these two conditions.

(i) *Continuity of pressure*

The linearized momentum equation in the direction along the interface is:

$$\frac{\partial u'}{\partial t} + U \frac{\partial u'}{\partial x} + v' \frac{\partial U}{\partial y} = -\frac{\partial p'}{\partial x}. \quad (6.21)$$

For normal modes, $u' = -\tilde{\psi}_y e^{ik(x-ct)}$, $v' = ik\tilde{\psi} e^{ik(x-ct)}$ and $p' = \tilde{p} e^{ik(x-ct)}$, (6.21) becomes

$$ik(U - c)\tilde{\psi}_y - ik\tilde{\psi}U_y = -ik\tilde{p}. \quad (6.22)$$

Because pressure is continuous across the interface we have the first *matching* or *jump condition*,

$$\boxed{\Delta[(U - c)\tilde{\psi}_y - \tilde{\psi}U_y] = 0}, \quad (6.23)$$

where the operator Δ denotes the difference in the values of the argument (in square brackets) across the interface. That is, the quantity $(U - c)\tilde{\psi}_y - \tilde{\psi}U_y$ is continuous.

We can obtain this condition directly from Rayleigh's equation, (6.20), written in the form

$$[(U - c)\tilde{\psi}_y - U_y\tilde{\psi}]_y + [\beta - k^2(U - c)]\tilde{\psi} = 0. \quad (6.24)$$

Integrating across the interface gives (6.23).

(ii) *Material interface condition*

At the interface, the normal velocity v is given by the kinematic condition

$$v = \frac{D\eta}{Dt} \quad (6.25)$$

where η is the interface displacement. The linear version of (6.25) is

$$\frac{\partial \eta'}{\partial t} + U \frac{\partial \eta'}{\partial x} = \frac{\partial \psi'}{\partial x}. \quad (6.26)$$

If the fluid itself is continuous then this equation must hold at either side of the interface, giving two equations and their normal-mode counterparts, namely,

$$\frac{\partial \eta'}{\partial t} + U_1 \frac{\partial \eta'}{\partial x} = \frac{\partial \psi'_1}{\partial x} \quad \rightarrow \quad (U_1 - c)\tilde{\eta} = \tilde{\psi}_1, \quad (6.27)$$

$$\frac{\partial \eta'}{\partial t} + U_2 \frac{\partial \eta'}{\partial x} = \frac{\partial \psi'_2}{\partial x} \quad \rightarrow \quad (U_2 - c)\tilde{\eta} = \tilde{\psi}_2. \quad (6.28)$$

Material continuity at the interface thus gives the second jump condition

$$\boxed{\Delta \left[\frac{\tilde{\psi}}{U - c} \right] = 0}. \quad (6.29)$$

That is, $\tilde{\psi}/(U - c)$ is continuous at the interface. Note that if U is continuous across the interface the condition becomes one of continuity of the normal velocity.

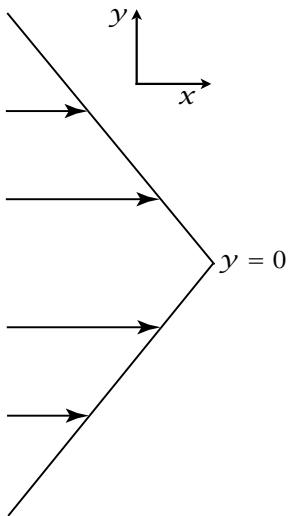


Fig. 6.3 Velocity profile of a point jet (so-called because the vorticity is concentrated at a point). Although the vorticity is discontinuous, a small perturbation gives rise only to *edge waves* centred at $y = 0$, and so the jet is stable.

6.2.2 Kelvin–Helmholtz instability, revisited

We now use Rayleigh's equation and the jump conditions to consider the situation illustrated in Fig. 6.1; that is, vorticity is everywhere zero except in a thin sheet at $y = 0$. On either side of the interface, Rayleigh's equation is simply

$$(U_i - c)(\partial_{yy}\tilde{\psi}_i - k^2\tilde{\psi}_i) = 0 \quad i = 1, 2 \quad (6.30)$$

or, assuming that $U_i \neq c$, $\partial_{yy}\tilde{\psi}_i - k^2\tilde{\psi}_i = 0$. (This is just Laplace's equation, coming from $\nabla^2\psi' = \zeta'$, with $\zeta' = 0$ everywhere except at the interface.) Solutions of this that decay away on either side of the interface are

$$y > 0 : \quad \tilde{\psi}_1 = \Psi_1 e^{-ky}, \quad (6.31a)$$

$$y < 0 : \quad \tilde{\psi}_2 = \Psi_2 e^{ky}, \quad (6.31b)$$

where Ψ_1 and Ψ_2 are constants. The boundary condition (6.23) gives

$$(U_1 - c)(-k)\Psi_1 = (U_2 - c)(k)\Psi_2, \quad (6.32)$$

and (6.29) gives

$$\frac{\Psi_1}{(U_1 - c)} = \frac{\Psi_2}{(U_2 - c)}. \quad (6.33)$$

The last two equations combine to give $(U_1 - c)^2 = -(U_2 - c)^2$, which, supposing that $U = U_1 = -U_2$ gives $c^2 = -U^2$. Thus, since U is purely real, $c = \pm iU$, and the disturbance grows exponentially as $\exp(kU_1 t)$, just as we obtained in section 6.1. All wavelengths are unstable, and indeed the shorter the wavelength the greater the instability. In reality, viscosity will damp the smallest waves, although the presence of viscosity would also mean that the initial profile is not an exact, steady solution of the equations of motion.

6.2.3 Edge waves

We now consider a case sketched in Fig. 6.3 in which the velocity is continuous, but the vorticity is discontinuous. Since on either side of the interface $U_{yy} = 0$, Rayleigh's equation is just

$$(U(y) - c)(\tilde{\psi}_{yy} - k^2 \tilde{\psi}) = 0. \quad (6.34)$$

Provided $c \neq U$ this has solutions,

$$\tilde{\psi} = \begin{cases} \Phi_1 e^{-ky} & y > 0 \\ \Phi_2 e^{ky} & y < 0. \end{cases} \quad (6.35)$$

The value of c is found by applying the jump conditions (6.23) and (6.29) at $y = 0$. Using (6.35) these give

$$-k(U_0 - c)\Phi_1 - \Phi_1 \partial_y U_1 = k(U_0 - c)\Phi_2 - \Phi_2 \partial_y U_2 \quad (6.36a)$$

$$\Phi_1 = \Phi_2, \quad (6.36b)$$

where U_1 and U_2 are the values of U at either side of the interface, and both are equal to U_0 at the interface. After a line of algebra the above equations give

$$c = U_0 + \frac{\partial_y U_1 - \partial_y U_2}{2k}. \quad (6.37)$$

This is the dispersion relationship for *edge waves* that propagate along the interface a speed equal to the sum of the fluid speed and a factor proportional to the difference in the vorticity between the two layers. No matter what the shear is on either side of the interface, the phase speed is purely real and there is no instability. Equation (6.37) is imperfectly analogous to the Rossby wave dispersion relation $c = U_0 - \beta/K^2$, and reflects a similarity in the physics — β is a planetary vorticity gradient, which in (6.37) is collapsed to a front and represented by the difference $\partial_y U_1 - \partial_y U_2 = -(Z_1 - Z_2)$, where Z_1 and Z_2 are the basic-state vorticities on either side of the interface.

6.2.4 Interacting edge waves producing instability

Now we consider a slightly more complicated case in which edge waves may interact giving rise, as we shall see, to an instability. The physical situation is illustrated in Fig. 6.4. We consider the simplest case, that of a shear layer (which we denote as region 2) sandwiched between two semi-infinite layers, regions 1 and 3, as in the left-hand panel of the figure. Thus, the basic state is

$$y > a: \quad U = U_1 = U_0 \text{ (a constant)}, \quad (6.38a)$$

$$-a < y < a: \quad U = U_2 = \frac{U_0}{a} y, \quad (6.38b)$$

$$y < -a: \quad U = U_3 = -U_0. \quad (6.38c)$$

We assume a solution of Rayleigh's equation of the form:

$$y > a: \quad \tilde{\psi}_1 = A e^{-k(y-a)}, \quad (6.39a)$$

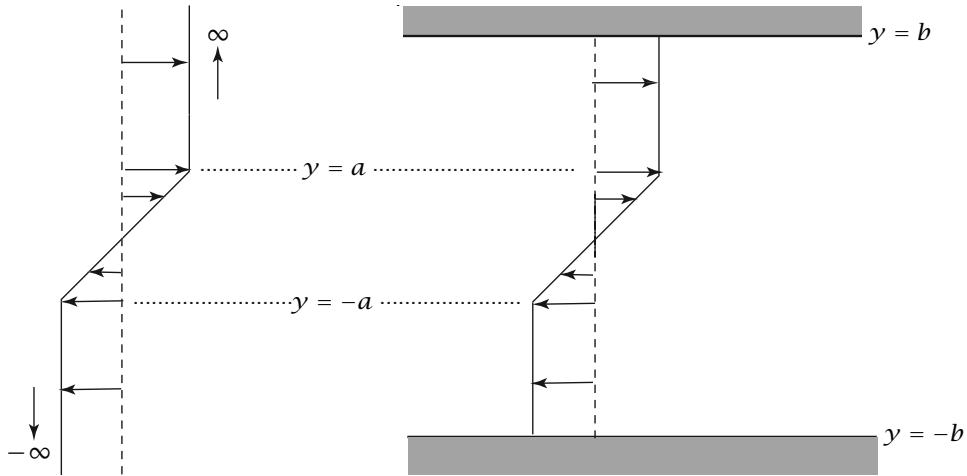


Fig. 6.4 Barotropically unstable velocity profiles. In the simplest case, on the left, a region of shear is sandwiched between two infinite regions of constant velocity. The edge waves at $y = \pm a$ interact to produce an instability. If $a = 0$, then the situation corresponds to that of Fig. 6.1, giving Kelvin-Helmholtz instability. In the case on the right, the flow is bounded at $y = \pm b$. It may be shown that the flow is still unstable, provided that b is sufficiently larger than a . If $b = a$ (plane Couette flow) the flow is stable to infinitesimal disturbances.

$$-a < y < a: \quad \tilde{\psi}_2 = B e^{-k(y-a)} + C e^{k(y+a)}, \quad (6.39b)$$

$$y < -a: \quad \tilde{\psi}_3 = D e^{k(y+a)}. \quad (6.39c)$$

Applying the jump conditions (6.23) and (6.29) at the interfaces at $y = a$ and $y = -a$ gives the following relations between the coefficients:

$$A[(U_0 - c)k] = B \left[(U_0 - c)k + \frac{U_0}{a} \right] + C e^{2ka} \left[\frac{U_0}{a} - (U_0 - c)k \right], \quad (6.40a)$$

$$A = B + C e^{2ka}, \quad (6.40b)$$

$$D[(U_0 + c)k] = B e^{2ka} \left[-(U_0 + c)k + \frac{U_0}{a} \right] + C \left[\frac{U_0}{a} + (U_0 + c)k \right], \quad (6.40c)$$

$$D = B e^{2ka} + C. \quad (6.40d)$$

These are a set of four homogeneous equations, with the unknown parameters A , B , C and D , which may be written in the form of a matrix equation,

$$\begin{pmatrix} k(U_0 - c) & -k(U_0 - c) - U_0/a & e^{2ka}[k(U_0 - c) - (U_0/a)] & 0 \\ 1 & -1 & -e^{2ka} & 0 \\ 0 & e^{2ka}[k(U_0 + c) - (U_0/a)] & -k(U_0 + c) - (U_0/a) & k(U_0 + c) \\ 0 & e^{2ka} & 1 & -1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = 0. \quad (6.41)$$

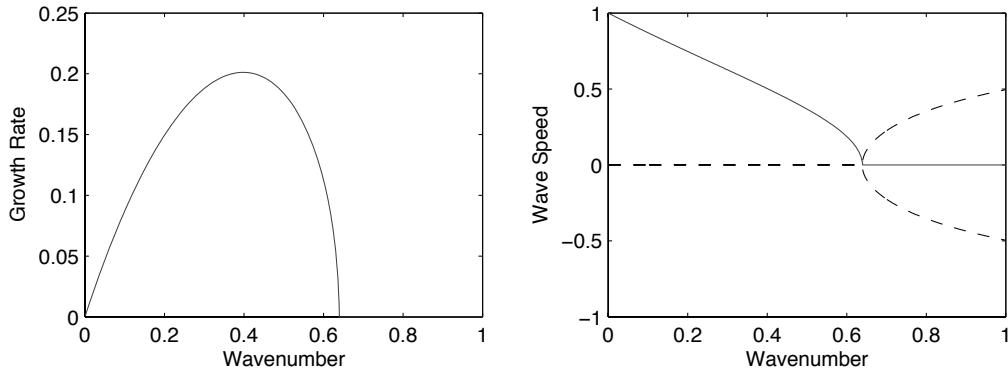


Fig. 6.5 Left: Growth rate ($\sigma = kc_i$) calculated from (6.42) with c non-dimensionalized by U_0 and k non-dimensionalized by $1/a$ (equivalent to setting $a = U_0 = 1$). Right: Real (c_r , dashed) and imaginary (c_i , solid) wave speeds. The flow is unstable for $k < 0.63$, with the maximum instability occurring at $k = 0.39$.

For non-trivial solutions the determinant of the matrix must be zero, and solving the ensuing equation gives the dispersion relationship⁵

$$c^2 = \left(\frac{U_0}{2ka}\right)^2 \left[(1 - 2ka)^2 - e^{-4ka} \right], \quad (6.42)$$

and this is plotted in Fig. 6.5. The flow is unstable for sufficiently long wavelengths, since then the right-hand side of (6.42) is negative. The critical wavenumber below which instability occurs is found by solving $(1 - 2ka)^2 = e^{-4ka}$, which gives instability for $ka < 0.63293$. A numerical solution of the initial value problem is illustrated in Figs. 6.6 and 6.7. Here, the initial perturbation is small and random, containing components at all wavenumbers. All the modes in the unstable range grow exponentially, and the pattern is soon dominated by the mode that grows fastest — a horizontal wavenumber of 3 in this problem. Eventually, the perturbation grows sufficiently that the linear equations are no longer valid and, as is seen in the second column of Fig. 6.6, vortices form and pinch off. The vortices interact and the flow develops into two-dimensional turbulence, as considered in chapter 8.

The mechanism of the instability — an informal view

(A similar mechanism is discussed in section 6.7, and the reader may wish to read the two descriptions in tandem.) We have seen that an edge wave in isolation is stable, with the instability arising when two edge waves have sufficient cross-stream extent that they can interact with each other. This occurs for sufficiently long wavelengths because the cross-stream decay scale is proportional to the along-stream wavelength — hence the high-wavenumber cut-off. To see the mechanism of the instability transparently, let us first suppose that the interfaces are, in fact, sufficiently far away that the edge waves at each interface do not interact. Using (6.37) the edge waves at $y = -a$ and $y = +a$ have dispersion relationships

$$c_{+a} = U_0 - \frac{U_0/a}{2k}, \quad c_{-a} = -U_0 + \frac{U_0/a}{2k}. \quad (6.43a,b)$$

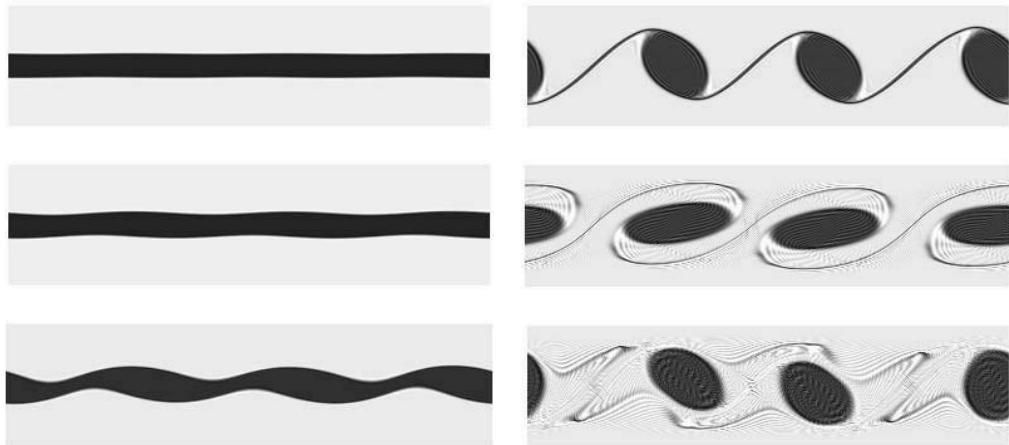


Fig. 6.6 A sequence of plots of the vorticity, at equal time intervals, from a numerical solution of the nonlinear vorticity equation (6.12), with initial conditions as in Fig. 6.4 with $a = 0.1$, plus a very small random perturbation. Time increases first down the left column and then down the right column. The solution is obtained in a rectangular (4×1) domain, with periodic conditions in the x -direction and slippery walls at $y = (0, 1)$. The maximum linear instability occurs for a wavelength of 1.57, which for a domain of length 4 corresponds to a wavenumber of 2.55. Since the periodic domain quantizes the allowable wavenumbers, the maximum instability is at wavenumber 3, and this is what emerges. Only in the first two or three frames is the linear approximation valid.

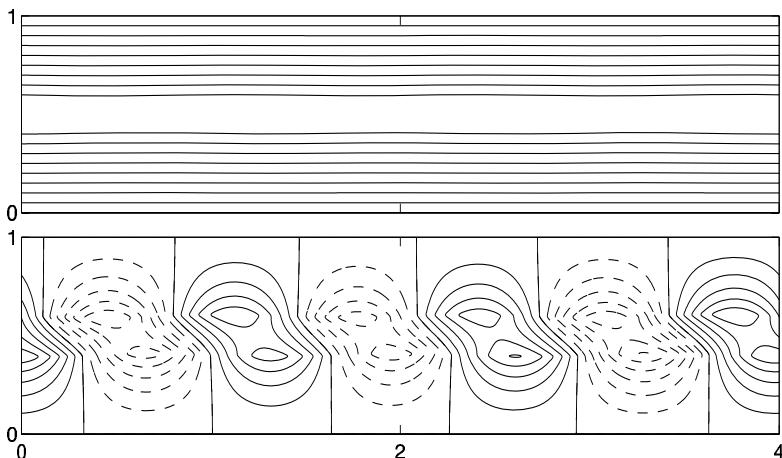


Fig. 6.7 The total streamfunction (top panel) and the perturbation streamfunction from the same numerical calculation as in Fig. 6.6, at a time corresponding to the second frame. Positive values are solid lines, and negative values are dashed. The perturbation pattern leans into the shear, and grows exponentially in place.

If the two waves are to interact these phase speeds must be equal, giving the condition

$$c = 0, \quad k = 1/(2a). \quad (6.44a,b)$$

That is, the waves are stationary, and their wavelength is proportional to the separation of the two edges. In fact, (6.44) approximately characterizes the conditions at the critical wavenumber $k = 0.63/a$ (see Fig. 6.5). In the region of the shear the two waves have the form

$$\psi_{+a} = \operatorname{Re} \tilde{\psi}_{+a}(t) e^{k(y-a)} e^{i\phi} e^{ikx}, \quad \psi_{-a} = \operatorname{Re} \tilde{\psi}_{-a}(t) e^{-k(y+a)} e^{ikx} \quad (6.45a,b)$$

where ϕ is the phase shift between the waves; in the case of pure edge waves we have $\tilde{\psi}_{\pm a} = A_{\pm a} e^{-ikct}$ where we may take $A_{\pm a}$ to be real.

Now consider how the wave generated at $y = -a$ might affect the wave at $y = +a$ and vice versa. The contribution of ψ_{-a} to the acceleration of ψ_{+a} is given by applying the x -momentum equation, (6.21), at either side of the interface at $y = +a$. Thus we take the kinematic solutions, (6.45), and use them in a dynamical equation, the momentum equation, to calculate the ensuing acceleration. At $y = +a$ we heuristically write

$$\frac{\partial u'_{+a}}{\partial t} = -[\nu'_{+a}(+a) + \nu'_{-a}(+a)] \frac{dU}{dy} \quad (6.46)$$

with a similar expression at $y = -a$. Here $\nu'_{-a}(+a)$ denotes the value of ν at $y = +a$ due to the edge wave generated at $-a$. It is the second term on the right-hand side that is necessary for any potential instability, as the first term gives only neutral edge waves. If the spatial dependence of the waves is given by (6.45), then at $y = +a$ we have, omitting the contribution from the edge waves,

$$\frac{\partial u'_{+a}}{\partial t} = -\nu'_{-a}(+a) \frac{dU}{dy} \quad \rightarrow \quad -k e^{i\phi} e^{ikx} \frac{d\tilde{\psi}_{+a}}{dt} = -ik e^{ikx} e^{-2ka} \frac{\partial U}{\partial y} \tilde{\psi}_{-a}, \quad (6.47)$$

with real part $\cos(kx + \phi) d\tilde{\psi}_{+a}/dt = -\sin kx e^{-2ka} dU/dy$. If the edge waves have the appropriate phase with respect to each other then the two edge waves can feed back on each other and couple to form a single growing mode. In particular if $\phi = \pi/2$ then the evolution equations for ψ_{+a} and ψ_{-a} become

$$\frac{d\tilde{\psi}_{+a}}{dt} = e^{-2ka} \frac{dU}{dy} \tilde{\psi}_{-a}, \quad \frac{d\tilde{\psi}_{-a}}{dt} = e^{-2ka} \frac{dU}{dy} \tilde{\psi}_{+a}, \quad (6.48a,b)$$

which give exponential growth. When $\phi = \pi/2$ the wave at $y = +a$ lags the wave at $y = -a$, and *the unstable perturbation tilts into the shear*; this property of the instability is seen in the full solution, Fig. 6.7.

6.3 NECESSARY CONDITIONS FOR INSTABILITY

6.3.1 Rayleigh's criterion

For simple profiles it may be possible to calculate, or even intuit, the instability properties, but for continuous profiles of $U(y)$ this is often impossible and it would be nice to have some general guidelines as to when a profile might be unstable. To this end, we will derive a

couple of *necessary* conditions for instability, or *sufficient* conditions for stability, that will at least tell us if a flow *might* be unstable. We begin by writing Rayleigh's equation, (6.20), as

$$\tilde{\psi}_{yy} - k^2 \tilde{\psi} + \frac{\beta - U_{yy}}{U - c} \tilde{\psi} = 0. \quad (6.49)$$

Multiply by $\tilde{\psi}^*$ (the complex conjugate of $\tilde{\psi}$) and integrate over the domain of interest. After integrating the first term by parts we obtain

$$\int_{y_1}^{y_2} \left(\left| \frac{\partial \tilde{\psi}}{\partial y} \right|^2 + k^2 |\tilde{\psi}|^2 \right) dy - \int_{y_1}^{y_2} \frac{\beta - U_{yy}}{U - c} |\tilde{\psi}|^2 dy = 0, \quad (6.50)$$

assuming that $\tilde{\psi}$ vanishes at the boundaries. (The limits to the integral may be infinite, in which case it is assumed that $\tilde{\psi}$ decays to zero as $|y|$ approaches ∞ .) The only variable in this expression that is complex is c , and thus the first integral is real. The imaginary component of the second integral is

$$c_i \int \frac{\beta - U_{yy}}{|U - c|^2} |\tilde{\psi}|^2 dy = 0. \quad (6.51)$$

Thus, *either* c_i vanishes or the integral does. For there to be an instability c_i must be non-zero. [Note also that the eigenvalues of Rayleigh's equation come in pairs, for each decaying mode (negative c_i) there is a corresponding growing mode (positive c_i).] If c_i is to be non-zero, the integrand of (6.51) must be zero and therefore

A necessary condition for instability is that the expression

$$\beta - U_{yy}$$

change sign somewhere in the domain.

Equivalently, a sufficient criterion for stability is that $\beta - U_{yy}$ not vanish in the domain interior. This condition is known as Rayleigh's inflection-point criterion, or when $\beta \neq 0$, the Rayleigh-Kuo inflection point criterion.⁶

An alternate, more general, derivation

Consider again the vorticity equation, linearized about a parallel shear flow [cf. (6.14) with a β -term],

$$\frac{\partial \zeta}{\partial t} + U \frac{\partial \zeta}{\partial x} + v \left(\frac{\partial z}{\partial y} + \beta \right) = 0, \quad (6.52)$$

(dropping the primes on the perturbation quantities). Multiply by ζ and divide by $\beta + z_y$ to obtain

$$\frac{\partial}{\partial t} \left(\frac{\zeta^2}{\beta + z_y} \right) + \frac{U}{\beta + z_y} \frac{\partial \zeta^2}{\partial x} + 2v\zeta = 0, \quad (6.53)$$

and then integrate with respect to x to give

$$\frac{\partial}{\partial t} \int \left(\frac{\zeta^2}{\beta + z_y} \right) dx = -2 \int v\zeta dx. \quad (6.54)$$

Now, using $\nabla \cdot \mathbf{u} = 0$, the vorticity flux may be written as

$$v\zeta = -\frac{\partial}{\partial y} (u v) + \frac{1}{2} \frac{\partial}{\partial x} (v^2 - u^2). \quad (6.55)$$

That is, *the flux of vorticity is the divergence of some quantity*. Its integral therefore vanishes provided there are no contributions from the boundary, and integrating (6.54) with respect to y gives

$$\frac{d}{dt} \int \left(\frac{\zeta^2}{\beta + \mathcal{Z}_y} \right) dx dy = 0. \quad (6.56)$$

If there is to be an instability ζ must grow, but the integral is identically zero. These two conditions can only be simultaneously satisfied if $\beta + \mathcal{Z}_y$, or equivalently $\beta - U_{yy}$, is zero somewhere in the domain.

This derivation shows that the inflection-point criterion applies even if disturbances are not of normal-mode form. The quantity $\zeta^2/(\beta + \mathcal{Z}_y)$ is an example of a *wave-activity density* — a wave activity being a conserved quantity, quadratic in the amplitude of the wave. Such quantities play an important role in instabilities, and we consider them further in chapter 7.

6.3.2 Fjørtoft's criterion

Another necessary condition for instability was obtained by Fjørtoft.⁷ In this section we will derive his condition for normal-mode disturbances, and provide a more general derivation in section 7.7. From the real part of (6.50) we find

$$\int_{y_1}^{y_2} (\beta - U_{yy}) \frac{(U - c_r)}{|U - c|^2} |\tilde{\psi}|^2 dy = \int_{y_1}^{y_2} \left| \frac{\partial \tilde{\psi}}{\partial y} \right|^2 + k^2 |\tilde{\psi}|^2 dy > 0. \quad (6.57)$$

Now, from (6.51), we know that for an instability we must have

$$\int_{y_1}^{y_2} \frac{\beta - U_{yy}}{|U - c|^2} |\tilde{\psi}|^2 dy = 0. \quad (6.58)$$

Using this equation and (6.57) we see that, for an instability,

$$\int_{y_1}^{y_2} (\beta - U_{yy}) \frac{(U - U_s)}{|U - c|^2} |\tilde{\psi}|^2 dy > 0 \quad (6.59)$$

where U_s is *any* real constant. It is most useful to choose this constant to be the value of $U(y)$ at which $\beta - U_{yy}$ vanishes. This leads directly to the criterion

A necessary condition for instability is that the expression

$$(\beta - U_{yy})(U - U_s)$$

where U_s is the value of $U(y)$ at which $\beta - U_{yy}$ vanishes, be positive somewhere in the domain.

This criterion is satisfied if the magnitude of the vorticity has an extremum inside the domain, and not at the boundary or at infinity (Fig. 6.8). Why choose U_s in the manner we did? Suppose we chose U_s to have a large positive value, so that $U - U_s$ is negative everywhere. Then (6.59) just implies that $\beta - U_{yy}$ must be negative somewhere, and this is already known from Rayleigh's criterion. If we choose U_s to be large and negative, we simply find that $\beta - U_{yy}$ must be positive somewhere. The most stringent criterion is obtained by choosing U_s to be the value of $U(y)$ at which $\beta - U_{yy}$ vanishes.

Interestingly, the β -effect can be either stabilizing or destabilizing: It can stabilize the middle two profiles of Fig. 6.8, because if it is large enough $\beta - U_{yy}$ will be one-signed.

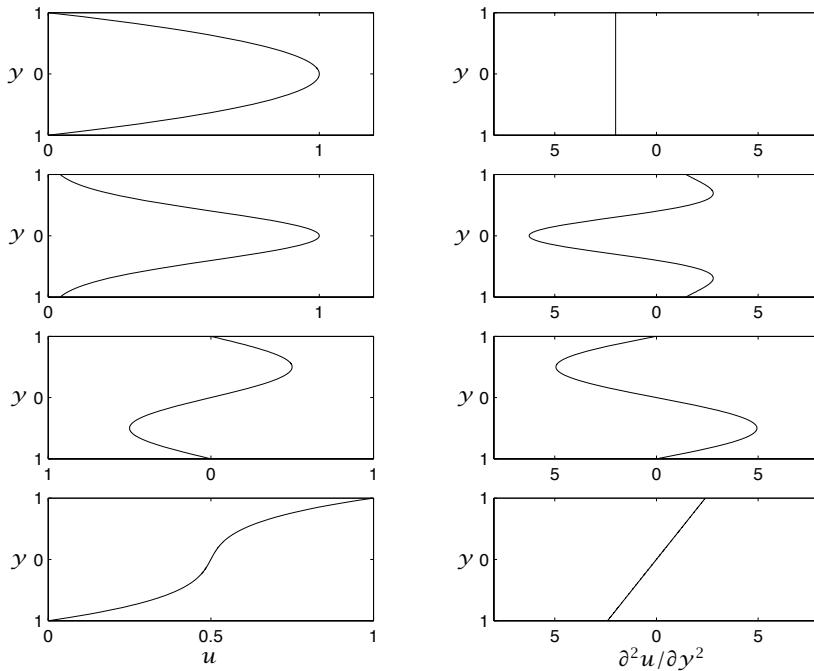


Fig. 6.8 Example parallel velocity profiles (left column) and their second derivatives (right column). From the top: Poiseuille flow ($u = 1 - y^2$); a Gaussian jet; a sinusoidal profile; a polynomial profile. By Rayleigh's criterion, the top profile is stable, whereas the lower three are potentially unstable. However, the bottom profile is stable by Fjørtoft's criterion (note that the vorticity maxima are at the boundaries). If the β -effect were present and large enough it would stabilize the middle two profiles.

However, the β -effect will destabilize a westward point jet, $U(y) = -(1 - |y|)$ (the negative of the jet in Fig. 6.3), because $\beta - U_{yy}$ is negative at $y = 0$ and positive elsewhere. An eastward point jet is stable, with or without β . Finally, we emphasize that both Fjørtoft's and Rayleigh's criteria are necessary conditions for instability, and examples exist that do satisfy their criterion, yet which are stable to infinitesimal perturbations.

6.4 BAROCLINIC INSTABILITY

Baroclinic instability is a hydrodynamic instability that occurs in stably stratified, rotating fluids, and it is ubiquitous in the Earth's atmosphere and oceans, and almost certainly occurs in other planetary atmospheres. It gives rise to weather, and so is perhaps the form of hydrodynamic instability that affects us most, and that certainly we talk about most.

6.4.1 A physical picture

We will first draw a picture of baroclinic instability as a form of 'sloping convection' in which the fluid, although statically stable, is able to release available potential energy when

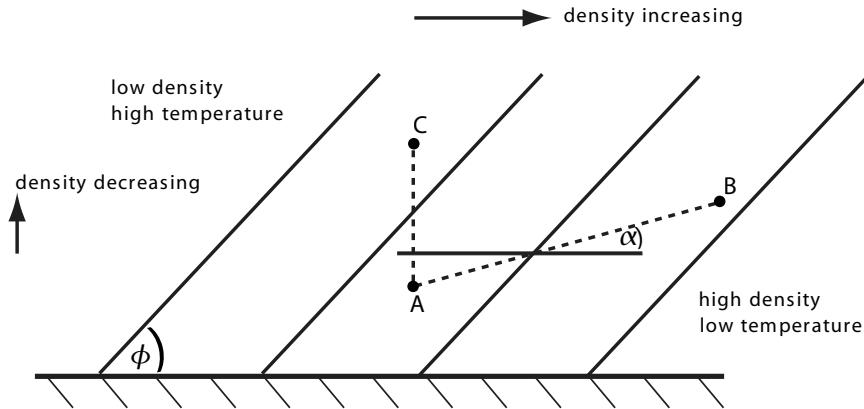


Fig. 6.9 A steady basic state giving rise to baroclinic instability. Potential density decreases upwards and equatorwards, and the associated horizontal pressure gradient is balanced by the Coriolis force. Parcel 'A' is heavier than 'C', and so statically stable, but it is lighter than 'B'. Hence, if 'A' and 'B' are interchanged there is a release of potential energy.

parcels move along a sloping path. To this end, let us first ask: what is the basic state that is baroclinically unstable? In a stably stratified fluid potential density decreases with height; we can also easily imagine a state in which the basic state temperature decreases, and the potential density increases, polewards. (We will couch most of our discussion in terms of the Boussinesq equations, and henceforth drop the qualifier 'potential' from density.) Can we construct a steady solution from these two conditions? The answer is yes, provided the fluid is also rotating; rotation is necessary because the meridional temperature gradient generally implies a meridional pressure gradient; there is nothing to balance this in the absence of rotation, and a fluid parcel would therefore accelerate. In a rotating fluid this pressure gradient can be balanced by the Coriolis force and a steady solution can be maintained even in the absence of viscosity. Consider a stably stratified Boussinesq fluid in geostrophic and hydrostatic balance on an *f*-plane, with buoyancy decreasing uniformly polewards. Then $fu = -\partial\phi/\partial y$ and $\partial\phi/\partial z = b$, where $b = -g\delta\rho/\rho_0$ is the buoyancy. These together give the thermal wind relation, $\partial u/\partial z = \partial b/\partial y$. If there is no variation of these fields in the zonal direction, then, for any variation of b with y , this is a steady solution to the primitive equations of motion, with $v = w = 0$.

The density structure corresponding to a uniform increase of density in the meridional direction is illustrated in Fig. 6.9. Is this structure stable to perturbations? The answer is no, although the perturbations must be a little special. Suppose the particle at 'A' is displaced upwards; then, since the fluid is (by assumption) stably stratified it will be denser than its surroundings and hence experience a restoring force, and similarly if displaced downwards. Suppose, however, we interchange the two parcels at positions 'A' and 'B'. Parcel 'A' finds itself surrounded by parcels of higher density than itself, and it is therefore buoyant; it is also higher than where it started. Parcel 'B' is negatively buoyant, and at a lower altitude than where it started. Thus, overall, the centre of gravity of the fluid has been lowered, and so its overall potential energy lowered. This loss in potential energy of the basic state

must be accompanied by a gain in kinetic energy of the perturbation. Thus, the perturbation amplifies and converts potential energy into kinetic energy.

The loss of potential energy is easily calculated. Since

$$PE = \int \rho g \, dz \quad (6.60)$$

the change in potential energy due to the interchange is

$$\Delta PE = g(\rho_A z_A + \rho_B z_B - \rho_A z_B - \rho_B z_A) = g(z_A - z_B)(\rho_A - \rho_B) = g \Delta \rho \Delta z. \quad (6.61)$$

If both $\rho_B > \rho_A$ and $z_B > z_A$ then the initial potential energy is larger than the final one, energy is released and the state is unstable. If the slope of the isopycnals is ϕ [so that $\phi = -(\partial_y \rho) / (\partial_z \rho)$] and the slope of the displacements is α , then for a displacement of horizontal distance L the change in potential energy is given by

$$\Delta PE = g \Delta \rho \Delta z = g \left(L \frac{\partial \rho}{\partial y} + L \alpha \frac{\partial \rho}{\partial z} \right) \alpha L = g L^2 \alpha \frac{\partial \rho}{\partial y} \left(1 - \frac{\alpha}{\phi} \right), \quad (6.62)$$

if α and ϕ are small. If $0 < \alpha < \phi$ then energy is released by the perturbation, and it is maximized when $\alpha = \phi/2$. For the atmosphere the actual slope of the isotherms is about 10^{-3} , so that the slope and potential parcel trajectories are indeed shallow.

Although intuitively appealing, the thermodynamic arguments presented in this section pay no attention to satisfying the dynamical constraints of the equations of motion, and we now turn our attention to that.

6.4.2 Linearized quasi-geostrophic equations

To explore the dynamics of baroclinic instability we use the quasi-geostrophic equations, specifically a potential vorticity equation for the fluid interior and a buoyancy or temperature equation at two vertical boundaries, one representing the ground and the other the tropopause. (The tropopause is the boundary between the troposphere and stratosphere at about 10 km; it is not a true rigid surface, but the higher static stability of the stratosphere inhibits vertical motion. We discuss the effects of that in section 6.9.) For a Boussinesq fluid the equations are

$$\begin{aligned} \frac{\partial q}{\partial t} + \mathbf{u} \cdot \nabla q &= 0, \quad 0 < z < H, \\ q &= \nabla^2 \psi + \beta y + \frac{\partial}{\partial z} \left(F \frac{\partial \psi}{\partial z} \right), \end{aligned} \quad (6.63)$$

where $F = f_0^2/N^2$, and the buoyancy equation with $w = 0$,

$$\begin{aligned} \frac{\partial b}{\partial t} + \mathbf{u} \cdot \nabla b &= 0, \quad z = 0, H, \\ b &= f_0 \frac{\partial \psi}{\partial z}. \end{aligned} \quad (6.64)$$

A solution of these equations is a purely zonal flow, $\mathbf{u} = U(y, z)\mathbf{i}$ with a corresponding temperature field given by thermal wind balance. The potential vorticity of this basic state is

$$Q = \beta y - \frac{\partial U}{\partial y} + \frac{\partial}{\partial z} F \frac{\partial \Psi}{\partial z} = \beta y + \frac{\partial^2 \Psi}{\partial y^2} + \frac{\partial}{\partial z} F \frac{\partial \Psi}{\partial z}, \quad (6.65)$$

where Ψ is the streamfunction of the basic state, related to U by $U = -\partial\Psi/\partial y$. Linearizing (6.63) about this zonal flow gives the potential vorticity equation for the interior,

$$\frac{\partial q'}{\partial t} + U \frac{\partial q'}{\partial x} + v' \frac{\partial Q}{\partial y} = 0, \quad 0 < z < H, \quad (6.66)$$

where $q' = \nabla^2 \psi' + \partial_z (F \partial_z \psi')$ and $v' = \partial_x \psi'$. Similarly, the linearized buoyancy equation at the boundary is

$$\frac{\partial b'}{\partial t} + U \frac{\partial b'}{\partial x} + v' \frac{\partial B}{\partial y} = 0, \quad z = 0, H, \quad (6.67)$$

where $b' = f_0 \partial_z \psi'$ and $\partial_y B = \partial_y (f_0 \partial_z \Psi) = -f_0 \partial U / \partial z$.

Just as for the barotropic problem, a standard way of proceeding is to seek normal-mode solutions. Since the coefficients of (6.66) and (6.67) are functions of y and z , but not of x , we seek solutions of the form

$$\psi'(x, y, z, t) = \text{Re } \tilde{\psi}(y, z) e^{ik(x-ct)}, \quad (6.68)$$

and similarly for the derived quantities u' , v' , b' and q' . In particular,

$$\tilde{q} = \frac{\partial^2 \tilde{\psi}}{\partial y^2} + \frac{\partial}{\partial z} F \frac{\partial \tilde{\psi}}{\partial z} - k^2 \tilde{\psi}. \quad (6.69)$$

Using (6.69) and (6.68) in (6.66) and (6.67) gives, with subscripts y and z denoting derivatives,

$$(U - c) \left(\tilde{\psi}_{yy} + (F \tilde{\psi}_z)_z - k^2 \tilde{\psi} \right) + Q_y \tilde{\psi} = 0 \quad 0 < z < H, \quad (6.70a)$$

$$(U - c) \tilde{\psi}_z - U_z \tilde{\psi} = 0 \quad z = 0, H. \quad (6.70b)$$

These equations are analogous to Rayleigh's equations for parallel shear flow, and illustrate the similarity between the baroclinic instability problem and that of a parallel shear flow.

6.4.3 Necessary conditions for baroclinic instability

Necessary conditions for instability may be obtained (as for parallel shear flows) by multiplying (6.70) by $\tilde{\psi}^*$ and integrating over the domain. Integrating by parts, we first note that

$$\int_{y_1}^{y_2} \tilde{\psi}^* \tilde{\psi}_{yy} dy = \left[\tilde{\psi}^* \tilde{\psi}_y \right]_{y_1}^{y_2} - \int_{y_1}^{y_2} |\tilde{\psi}_y|^2 dy. \quad (6.71)$$

If the integral is performed between two quiescent latitudes, or $\psi = 0$ at the meridional boundaries, then the first term on the right-hand side vanishes. Similarly,

$$\begin{aligned} \int_0^H \tilde{\psi}^* (F \tilde{\psi}_z)_z dz &= \left[F \tilde{\psi}^* \tilde{\psi}_z \right]_0^H - \int_0^H F |\tilde{\psi}_z|^2 dz \\ &= \left[\frac{FU_z |\tilde{\psi}|^2}{(U - c)} \right]_0^H - \int_0^H F |\tilde{\psi}_z|^2 dz, \end{aligned} \quad (6.72)$$

using (6.70b). Now, multiply (6.70a) by $\tilde{\psi}^*$ and integrate over y and z , and use (6.71) and (6.72) to obtain

$$\int_0^H \int_{y_1}^{y_2} [\psi_y|^2 + F|\tilde{\psi}_z|^2 + k^2|\tilde{\psi}|^2] dy dz - \int_{y_1}^{y_2} \left\{ \int_0^H \frac{Q_y}{U-c} |\tilde{\psi}|^2 dz + \left[\frac{FU_z|\tilde{\psi}|^2}{U-c} \right]_0^H \right\} dy = 0. \quad (6.73)$$

The term on the first line is purely real. The term on the second line is complex, and its imaginary component is given by

$$-c_i \int_{y_1}^{y_2} \left\{ \int_0^H \frac{Q_y}{|U-c|^2} |\tilde{\psi}|^2 dz + \left[\frac{FU_z|\tilde{\psi}|^2}{|U-c|^2} \right]_0^H \right\} dy = 0. \quad (6.74)$$

If there is to be instability c_i must be non-zero, and the integrand must therefore vanish. This gives the *Charney–Stern–Pedlosky* (CSP) necessary condition for instability,⁸ namely that one of the following criteria must be satisfied:

- (i) Q_y changes sign in the interior.
- (ii) Q_y is the opposite sign to U_z at the upper boundary, $z = H$.
- (iii) Q_y is the same sign as U_z at the lower boundary, $z = 0$.
- (iv) U_z is the same sign at the upper and lower boundaries, a condition that differs from (ii) or (iii) if $Q_y = 0$.

In the Earth's atmosphere, Q_y is often dominated by β , and is positive everywhere, as, frequently, is the shear. The instability criterion is then normally satisfied through (iii): that is, both Q_y and $U_z(0)$ are positive. A more general derivation that does not rely on normal-mode disturbances is given in section 7.7.2.

6.5 THE EADY PROBLEM

We now proceed to explicitly calculate the stability properties of a particular configuration that has become known as the *Eady problem*. This was one of the first two mathematical descriptions of baroclinic instability, the other being the *Charney problem*.⁹ The two problems were formulated independently, each being the (largely unsupervised) Ph.D. thesis of its author, and although the Charney problem is in some respects more complete (for example in allowing a β -effect) the Eady problem displays the instability in a more transparent form. The Charney problem in its entirety is also quite mathematically opaque,¹⁰ and so we will first consider the Eady problem. The β -effect can be incorporated relatively simply in the two-layer model of the next section, and in section 6.9.1 we look at some aspects of the Charney problem approximately. To begin, let us make the following simplifying assumptions.

- (i) The motion is on the f -plane ($\beta = 0$). This assumption, although not particularly realistic for the Earth's atmosphere, greatly simplifies the analysis.
- (ii) The fluid is uniformly stratified; that is, N^2 is a constant. This is a decent approximation for the atmosphere below the tropopause, but less so for the ocean where the stratification is quite non-uniform, being much larger in the upper ocean.

- (iii) The basic state has uniform shear; that is, $U_0(z) = \Lambda z = Uz/H$, where Λ is the (constant) shear and U is the zonal velocity at $z = H$, where H the domain depth. This profile is more appropriate for the atmosphere than the ocean — below the thermocline the ocean is relatively quiescent and the shear is small.
- (iv) The motion is contained between two rigid, flat horizontal surfaces. In the atmosphere this corresponds to the ground and a 'lid' at a constant-height tropopause.

Although, apart from (i), these assumptions are more appropriate for the atmosphere than the ocean, the qualitative nature of baroclinic instability is similar in both systems.

6.5.1 The linearized problem

With a basic state streamfunction of $\Psi = -\Lambda z y$, the basic state potential vorticity, Q , is

$$Q = \nabla^2 \Psi + \frac{H^2}{L_d^2} \frac{\partial}{\partial z} \left(\frac{\partial \Psi}{\partial z} \right) = 0. \quad (6.75)$$

The fact that $Q = 0$ makes the Eady problem a special case, albeit an illuminating one. The linearized potential vorticity equation is

$$\left(\frac{\partial}{\partial t} + \Lambda z \frac{\partial}{\partial x} \right) \left(\nabla^2 \psi' + \frac{H^2}{L_d^2} \frac{\partial^2 \psi'}{\partial z^2} \right) = 0. \quad (6.76)$$

This equation has no x -dependent coefficients and in a periodic channel we may seek solutions of the form $\psi'(x, y, z, t) = \text{Re } \tilde{\psi}(y, z) e^{ik(x-ct)}$, yielding

$$(\Lambda z - c) \left(\frac{\partial^2 \tilde{\psi}}{\partial y^2} + \frac{H^2}{L_d^2} \frac{\partial^2 \tilde{\psi}}{\partial z^2} - k^2 \tilde{\psi} \right) = 0. \quad (6.77)$$

This equation is (6.70a) applied to the Eady problem.

Boundary conditions

There are two sets of boundary conditions to satisfy, the vertical boundary conditions at $z = 0$ and $z = H$ and the lateral boundary conditions. In the horizontal plane we may either consider the flow to be confined to a channel, periodic in x and confined between two meridional walls, or, with a slightly greater degree of idealization but with little change to the essential dynamics, suppose that domain is doubly-periodic. Either case is dealt with easily enough by the choice of geometric basis function; we will choose a channel of width L and impose $\psi = 0$ at $y = +L/2$ and $y = -L/2$ and, to satisfy this, seek solutions of the form $\Psi = \Phi(z) \sin ly$ or, using (6.68)

$$\psi'(x, y, z, t) = \text{Re } \Phi(z) \sin ly e^{ik(x-ct)}, \quad (6.78)$$

where $l = n\pi/L$, with n being a positive integer.

The vertical boundary conditions are that $w = 0$ at $z = 0$ and $z = H$. We follow the procedure of section 6.4.2 and from (6.67) we obtain

$$\left(\frac{\partial}{\partial t} + \Lambda z \frac{\partial}{\partial x} \right) \frac{\partial \psi'}{\partial z} - \Lambda \frac{\partial \psi'}{\partial x} = 0, \quad \text{at } z = 0, H. \quad (6.79)$$

Solutions

Substituting (6.78) into (6.77) gives the interior potential vorticity equation

$$(\Lambda z - c) \left[\frac{H^2}{L_d^2} \frac{\partial^2 \Phi}{\partial z^2} - (k^2 + l^2) \Phi \right] = 0, \quad (6.80)$$

and substituting (6.78) into (6.79) gives, at $z = 0$ and $z = H$,

$$c \frac{d\Phi}{dz} + \Lambda \Phi = 0 \quad \text{and} \quad (c - \Lambda H) \frac{d\Phi}{dz} + \Lambda \Phi = 0. \quad (6.81\text{a,b})$$

These are equivalent to (6.70b) applied to the Eady problem. If $\Lambda z \neq c$ then (6.80) becomes¹¹

$$H^2 \frac{d^2 \Phi}{dz^2} - \mu^2 \Phi = 0, \quad (6.82)$$

where $\mu^2 = L_d^2(k^2 + l^2)$. The non-dimensional parameter μ is a horizontal wavenumber, scaled by the inverse of the Rossby radius of deformation. Solutions of (6.82) are

$$\Phi(z) = A \cosh \mu \hat{z} + B \sinh \mu \hat{z}, \quad (6.83)$$

where $\hat{z} = z/H$; thus, μ determines the vertical structure of the solution. The boundary conditions (6.81) are satisfied if

$$\begin{aligned} A [\Lambda H] + B [\mu c] &= 0, \\ A [(c - \Lambda H) \mu \sinh \mu + \Lambda H \cosh \mu] + B [(c - \Lambda H) \mu \cosh \mu + \Lambda H \sinh \mu] &= 0. \end{aligned} \quad (6.84)$$

Equations (6.84) are two coupled homogeneous equations in the two unknowns A and B . Non-trivial solutions will only exist if the determinant of their coefficients (the terms in square brackets) vanishes, and this leads to

$$c^2 - Uc + U^2(\mu^{-1} \coth \mu - \mu^{-2}) = 0, \quad (6.85)$$

where $U \equiv \Lambda H$ and $\coth \mu = \cosh \mu / \sinh \mu$. The solution of (6.85) is

$$c = \frac{U}{2} \pm \frac{U}{\mu} \left[\left(\frac{\mu}{2} - \coth \frac{\mu}{2} \right) \left(\frac{\mu}{2} - \tanh \frac{\mu}{2} \right) \right]^{1/2}. \quad (6.86)$$

The waves, being proportional to $\exp(-ikct)$, will grow exponentially if c has an imaginary part. Since $\mu/2 > \tanh(\mu/2)$ for all μ , for an instability we require that

$$\frac{\mu}{2} < \coth \frac{\mu}{2}, \quad (6.87)$$

which is satisfied when $\mu < \mu_c$, where $\mu_c = 2.399$. The growth rates of the instabilities themselves are given by the imaginary part of (6.86), multiplied by the x -wavenumber; that is

$$\sigma = kc_i = k \frac{U}{\mu} \left[\left(\coth \frac{\mu}{2} - \frac{\mu}{2} \right) \left(\frac{\mu}{2} - \tanh \frac{\mu}{2} \right) \right]^{1/2}. \quad (6.88)$$

These solutions suggest a natural non-dimensionalization: scale length by L_d , height by

H and time by L_d/U . The *Eady growth rate* is the inverse of the time scaling, and is defined by

$$\sigma_E \equiv \frac{\Lambda H}{L_d} = \frac{U}{L_d}. \quad (6.89)$$

Its inverse, the Eady time scale, may also be written as

$$T_E = \frac{L_d}{U} = \frac{NH}{f_0 U} = \frac{1}{Fr f_0} = \frac{\sqrt{Ri}}{f_0}, \quad (6.90)$$

where $Fr = U/(NH)$ and $Ri = N^2/\Lambda^2$ are the Froude and Richardson numbers for this problem.

From (6.88) we may determine that the maximum growth rate occurs when $\mu = \mu_m = 1.61$, with an associated non-dimensional growth rate of $kc_i/\sigma_E = 0.31$, and phase speed $c_r/U = 0.5$. Note that for any given x -wavenumber, the most unstable wavenumber has $l = 0$, so that $L_d k = \mu$. The unstable x -wavenumbers and corresponding wavelengths occur for

$$k < k_c = \frac{\mu_c}{L_d} = \frac{2.4}{L_d}, \quad \lambda > \lambda_c = \frac{2\pi L_d}{\mu_c} = 2.6 L_d. \quad (6.91a,b)$$

The wavenumber and wavelength at which the instability is greatest are:

$$k_m = \frac{1.6}{L_d}, \quad \lambda_m = \frac{2\pi L_d}{\mu_m} = 3.9 L_d. \quad (6.92a,b)$$

These properties are illustrated in the left-hand panels of Fig. 6.10 and in Fig. 6.11.

Given c , we may use (6.84) to determine the vertical structure of the Eady wave and this is, to within an arbitrary constant factor,

$$\Phi(\hat{z}) = \cosh \mu \hat{z} - \frac{U}{\mu c} \sinh \mu \hat{z} = \left(\cosh \mu \hat{z} - \frac{U c_r \sinh \mu \hat{z}}{\mu |c^2|} + \frac{i U c_i \sinh \mu \hat{z}}{\mu |c^2|} \right). \quad (6.93)$$

The wave therefore has a phase, $\theta(z)$, given by

$$\theta(\hat{z}) = \tan^{-1} \left(\frac{U c_i \sinh \mu \hat{z}}{\mu |c^2| \cosh \mu \hat{z} - U c_r \sinh \mu \hat{z}} \right). \quad (6.94)$$

The phase and amplitude of the Eady waves are plotted in the right panels of Fig. 6.10, and their overall structure in Fig. 6.12, where we see the unstable wave tilting into the shear.

6.5.2 Atmospheric and oceanic parameters

To get a qualitative sense of the nature of the instability we choose some typical parameters, as follows.

For the atmosphere

Let us choose

$$H \sim 10 \text{ km}, \quad U \sim 10 \text{ m s}^{-1}, \quad N \sim 10^{-2} \text{ s}^{-1}. \quad (6.95)$$

We then obtain:

$$\text{deformation radius: } L_d = \frac{NH}{f} \approx \frac{10^{-2} 10^4}{10^{-4}} \approx 1000 \text{ km}, \quad (6.96)$$

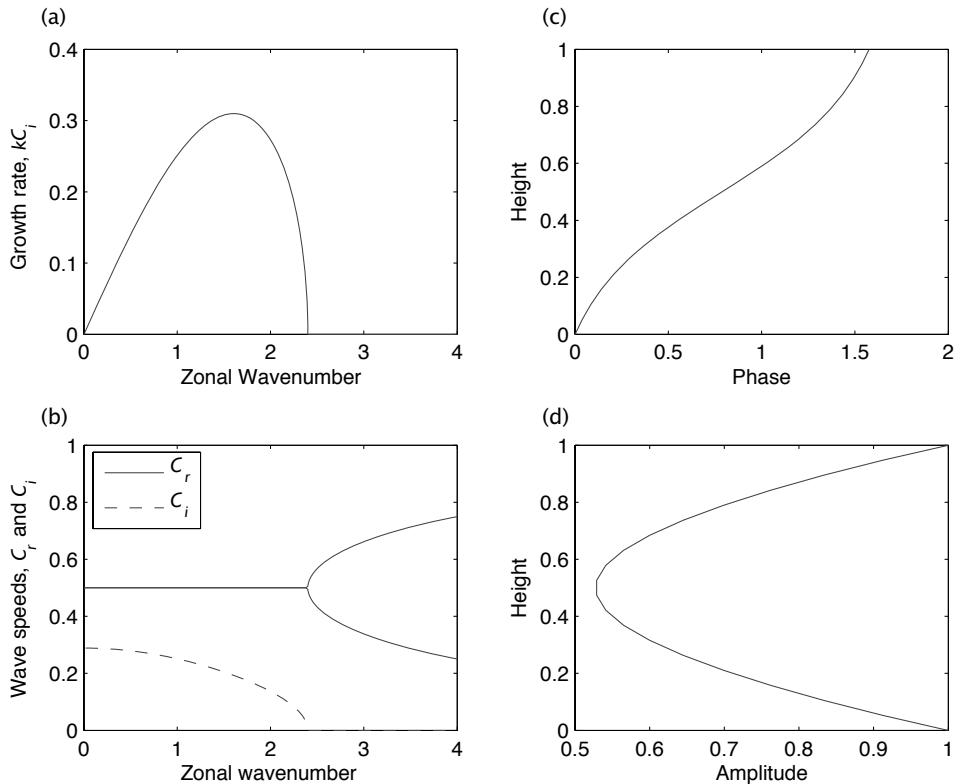


Fig. 6.10 Solution of the Eady problem, in non-dimensional units. (a) The growth rate, kc_i , of the most unstable Eady modes (i.e., those with the gravest meridional scale) as a function of scaled wavenumber μ , from (6.88) with $\Lambda = H = 1$. (b) The real (solid) and imaginary (dashed) wave speeds of those modes, as a function of horizontal wavenumber. (c) The phase of the single most unstable mode as a function of height. (d) The amplitude of that mode as a function of height. To obtain dimensional values, multiply the growth rate by $\Lambda H / L_d$ and the wavenumber by $1 / L_d$.

scale of maximum instability: $L_{\max} \approx 3.9L_d \approx 4000 \text{ km}$, (6.97)

$$\text{growth rate: } \sigma \approx 0.3 \frac{U}{L_d} \approx \frac{0.3 \times 10}{10^6} \text{ s}^{-1} \approx 0.26 \text{ day}^{-1}. \quad (6.98)$$

For the ocean

For the main thermocline in the ocean let us choose

$$H \sim 1 \text{ km}, \quad U \approx 0.1 \text{ m s}^{-1}, \quad N \sim 10^{-2} \text{ s}^{-1}. \quad (6.99)$$

We then obtain:

$$\text{deformation radius: } L_d = \frac{NH}{f} \approx \frac{10^{-2} \times 1000}{10^{-4}} = 100 \text{ km}, \quad (6.100)$$

$$\text{scale of maximum instability: } L_{\max} \approx 3.9 L_d \approx 400 \text{ km}, \quad (6.101)$$

Fig. 6.11 Contours of the growth rate, σ , in the Eady problem, in the $k-l$ plane using (6.88), non-dimensionalized as in Fig. 6.10. The growth rate peaks near the deformation scale, and for any given zonal wavenumber the most unstable wavenumber is that with the gravest meridional scale.

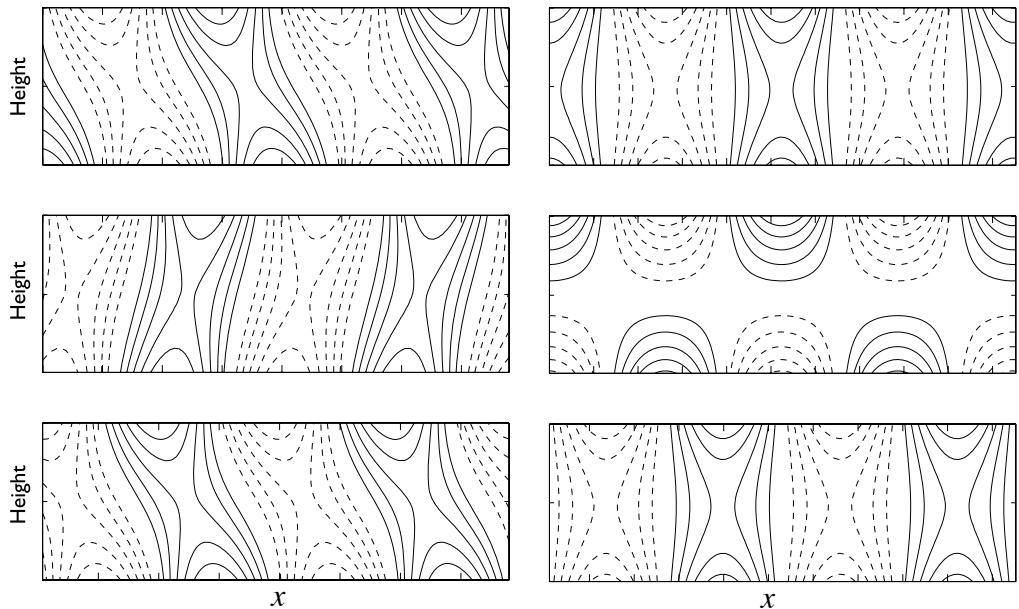
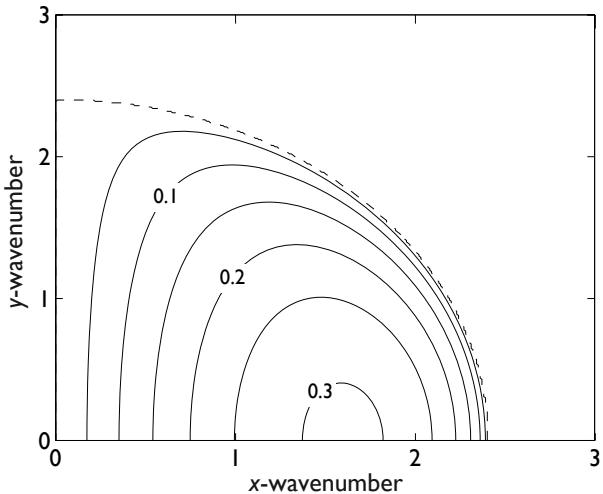


Fig. 6.12 Left column: vertical structure of the most unstable Eady mode. Top: contours of streamfunction. Middle: temperature, $\partial\psi/\partial z$. Bottom: meridional velocity, $\partial\psi/\partial x$. Negative contours are dashed, and two complete wavelengths are present in the horizontal direction. Polewards flowing (positive v) air is generally warmer than equatorwards flowing air. Right column: the same, but now for a wave just beyond the short-wave cut-off.

Summary of Results from the Eady Problem

- ★ The length scales of the instability are characterized by the deformation scale. The most unstable scale has a wavelength about four times the deformation radius L_d , where $L_d = NH/f_0$.
- ★ The growth rate of the instability is approximately

$$\sigma_E \approx \frac{U}{L_d} = \Lambda \frac{f_0}{N}. \quad (1)$$

That is, it is proportional to the shear, and scaled by the Prandtl ratio f_0/N . The value σ_E is known as the *Eady growth rate*.

- ★ The most unstable waves for a given zonal scale are those with the gravest meridional scale.
- ★ There is a *short-wave cutoff* beyond which (i.e., at higher wavenumber than) there is no instability. This occurs near the deformation radius.
- ★ The instability relies on an interaction between waves at the upper and lower boundaries. If either boundary is removed, the instability dies. This point is considered further in section 6.7.

$$\text{growth rate: } \sigma \approx 0.3 \frac{U}{L_d} \approx \frac{0.3 \times 0.1}{10^5} \text{ s}^{-1} \approx 0.026 \text{ day}^{-1}. \quad (6.102)$$

In the ocean, the Eady problem is not quantitatively applicable because of the non-uniformity of the stratification. Nevertheless, the above estimates give a qualitative sense of the scale and growth rate of the instability relative to the corresponding values in the atmosphere. A summary of the main points of the Eady problem is given in the shaded box above.

6.6 TWO-LAYER BAROCLINIC INSTABILITY

The eigenfunctions displaying the largest growth rates in the Eady problem have a relatively simple vertical structure. This suggests that an even simpler mathematical model of baroclinic instability might be constructed in which the vertical structure is a priori restricted to a very simple form, namely the two-layer or two-level QG model of sections 5.3.2 and 5.4.5. This instability problem is often called the ‘Phillips problem’.¹² One notable advantage over the Eady model is that it is possible to include the β -effect in a simple way.

6.6.1 Posing the problem

For two layers or two levels of equal thickness, we write the potential vorticity equations in the dimensional form,

$$\frac{D}{Dt} \left[\zeta_i + \beta y + \frac{k_d^2}{2} (\psi_j - \psi_i) \right] = 0, \quad i = 1, 2, \quad j = 3 - i, \quad (6.103)$$

where, using two-level notation for definiteness,

$$\frac{k_d^2}{2} = \left(\frac{2f_0}{NH} \right)^2 \rightarrow k_d = \frac{\sqrt{8}}{L_d}, \quad (6.104)$$

where H is the total depth of the domain, as in the Eady problem. The basic state we choose is

$$\Psi_1 = -U_1 y, \quad \Psi_2 = -U_2 y = +U_1 y. \quad (6.105)$$

It is possible to choose $U_2 = -U_1$ without loss of generality because there is no topography and the system is Galilean invariant. The basic state potential vorticity gradient is then given by

$$Q_1 = \beta y + k_d^2 U y, \quad Q_2 = \beta y - k_d^2 U y \quad (6.106)$$

where $U = U_1$. (Note that U differs by a constant multiplicative factor from the U in the Eady problem.) Even in the absence of β there is a non-zero potential vorticity gradient. Why should this be different from the Eady problem? — after all, the shear is uniform in both problems. The difference arises from the vertical boundary conditions. In the standard layered formulation the temperature gradient at the boundary is absorbed into the definition of the potential vorticity in the interior. This results in a non-zero interior potential vorticity gradient at the two levels adjacent to the boundary (the only layers in the two-layer problem), but with isothermal boundary conditions $D(\partial\psi/\partial z)/Dt = 0$. In the Eady problem we have a zero interior gradient of potential vorticity but a temperature gradient at the boundary. The two formulations are physically equivalent — a finite-difference example of the Bretherton boundary layer, encountered in section 5.4.3.

The linearized potential vorticity equation is, for each layer,

$$\frac{\partial q'_i}{\partial t} + U_i \frac{\partial q'_i}{\partial x} + v'_i \frac{\partial Q_i}{\partial y} = 0, \quad i = 1, 2 \quad (6.107)$$

or, more explicitly,

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \left[\nabla^2 \psi'_1 + \frac{k_d^2}{2} (\psi'_2 - \psi'_1) \right] + \frac{\partial \psi'_1}{\partial x} (\beta + k_d^2 U) = 0, \quad (6.108a)$$

$$\left(\frac{\partial}{\partial t} - U \frac{\partial}{\partial x} \right) \left[\nabla^2 \psi'_2 + \frac{k_d^2}{2} (\psi'_1 - \psi'_2) \right] + \frac{\partial \psi'_2}{\partial x} (\beta - k_d^2 U) = 0. \quad (6.108b)$$

For simplicity we will set the problem in a square, doubly periodic domain, and so seek solutions of the form,

$$\psi'_i = \operatorname{Re} \tilde{\psi}_i e^{i(kx+ly-\omega t)} = \operatorname{Re} \tilde{\psi}_i e^{ik(x-ct)} e^{ily}, \quad i = 1, 2. \quad (6.109)$$

Here, k and l are the x - and y -wavenumbers, and $(k, l) = (2\pi/L)(m, n)$, where L is the size of the domain, and m and n are integers. The constant $\tilde{\psi}_i$ is the complex amplitude.

6.6.2 The solution

Substituting (6.109) into (6.108) we obtain

$$[ik(U - c)] \left[-K^2 \tilde{\psi}_1 + k_d^2 (\tilde{\psi}_2 - \tilde{\psi}_1)/2 \right] + ik \tilde{\psi}_1 (\beta + k_d^2 U) = 0, \quad (6.110a)$$

$$[-ik(U + c)] \left[-K^2 \tilde{\psi}_2 + k_d^2 (\tilde{\psi}_1 - \tilde{\psi}_2)/2 \right] + ik \tilde{\psi}_2 (\beta - k_d^2 U) = 0, \quad (6.110b)$$

where $K^2 = k^2 + l^2$. Re-arranging these two equations gives

$$\left[(U - c)(k_d^2/2 + K^2) - (\beta + k_d^2 U) \right] \tilde{\psi}_1 - \left[k_d^2(U - c)/2 \right] \tilde{\psi}_2 = 0, \quad (6.111a)$$

$$- \left[k_d^2(U + c)/2 \right] \tilde{\psi}_1 + \left[(U + c)(k_d^2/2 + K^2) + (\beta - k_d^2 U) \right] \tilde{\psi}_2 = 0. \quad (6.111b)$$

These equations are of the form

$$[A]\tilde{\psi}_1 + [B]\tilde{\psi}_2 = 0, \quad [C]\tilde{\psi}_1 + [D]\tilde{\psi}_2 = 0, \quad (6.112)$$

where A, B, C, D correspond to the terms in square brackets in (6.111). For non-trivial solutions the determinant of coefficients must be zero, that is $AD - BC = 0$. This gives a quadratic equation in c and solving this we obtain

$$c = -\frac{\beta}{K^2 + k_d^2} \left\{ 1 + \frac{k_d^2}{2K^2} \pm \frac{k_d^2}{2K^2} \left[1 + \frac{4K^4(K^4 - k_d^4)}{k_\beta^4 k_d^4} \right]^{1/2} \right\}, \quad (6.113)$$

where $K^4 = (k^2 + l^2)^2$ and $k_\beta = \sqrt{\beta/U}$ (its inverse is known as the Kuo scale). We may non-dimensionalize this equation using the deformation radius L_d as the length scale and the shear velocity U as the velocity scale.¹³ Then, denoting non-dimensional parameters with hats, we have

$$k = \frac{\hat{k}}{L_d}, \quad c = \hat{c} U, \quad t = \frac{L_d}{U} \hat{t} \quad (6.114)$$

and the non-dimensional form of (6.113) is just

$$\hat{c} = -\frac{\hat{k}_\beta^2}{\hat{K}^2 + \hat{k}_d^2} \left\{ 1 + \frac{\hat{k}_d^2}{2\hat{K}^2} \pm \frac{\hat{k}_d^2}{2\hat{K}^2} \left[1 + \frac{4\hat{K}^4(\hat{K}^4 - \hat{k}_d^4)}{\hat{k}_\beta^4 \hat{k}_d^4} \right]^{1/2} \right\}, \quad (6.115)$$

where $\hat{k}_\beta = k_\beta L_d$ and $\hat{k}_d = \sqrt{8}$, as in (6.104). The non-dimensional parameter

$$\gamma = \frac{1}{4} \hat{k}_\beta^2 = \frac{\beta L_d^2}{4U}, \quad (6.116)$$

is often useful as a measure of the importance of β ; it is proportional to the square of the ratio of the deformation radius to the Kuo scale $\sqrt{U/\beta}$. (It is the two-layer version of the 'Charney-Green number' considered more in section 6.9.1.) Let us look at two special cases first, before considering the general solution to these equations.

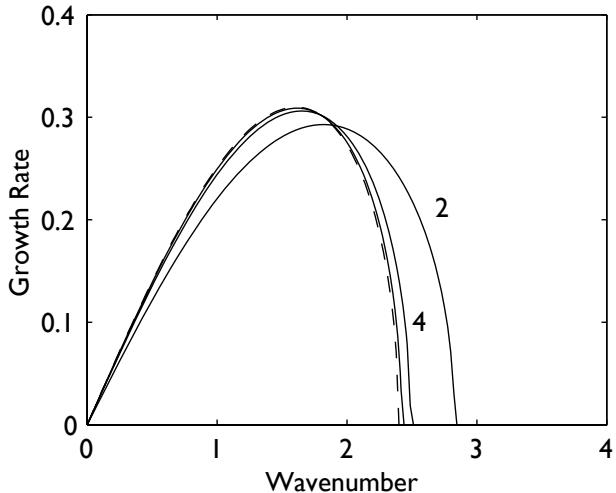
I. Zero shear, non-zero β

If there is no shear (i.e., $U = 0$) then (6.111a) and (6.111b) are identical and two roots of the equation give the purely real phase speeds c

$$c = -\frac{\beta}{K^2}, \quad c = -\frac{\beta}{K^2 + k_d^2}. \quad (6.117)$$

The first of these is the dispersion relationship for Rossby waves in a purely barotropic flow, and corresponds to the eigenfunction $\tilde{\psi}_1 = \tilde{\psi}_2$. The second solution corresponds to the baroclinic eigenfunction $\tilde{\psi}_1 + \tilde{\psi}_2 = 0$.

Fig. 6.13 Growth rates for models with varying numbers of vertical layers, all with $\beta = 0$ and the same uniform stratification and shear. The dashed line is the analytic solution to the continuous (Eady) problem, and the solid lines are results obtained using two, four and eight layers. The two- and four-layer results are labelled, and the eight-layer result is almost coincident with the dashed line.



II. Zero β , non-zero shear

If $\beta = 0$, then (6.111) yields, after a little algebra,

$$c = \pm U \left(\frac{K^2 - k_d^2}{K^2 + k_d^2} \right)^{1/2} \quad (6.118)$$

or, defining the growth rate σ by $\sigma = -i\omega$,

$$\sigma = U k \left(\frac{k_d^2 - K^2}{K^2 + k_d^2} \right)^{1/2}. \quad (6.119)$$

These expressions are very similar to those in the Eady problem. Indeed, as we increase the number of layers (using a numerical method to perform the calculation) the growth rate converges to that of the Eady problem (Fig. 6.13). We note the following.

- ★ There is an instability for *all* values of U .
- ★ There is a high-wavenumber cut-off, at a scale proportional to the radius of deformation. For the two-layer model, if $K > k_d = 2.82/L_d$ there is no growth. For the Eady problem, the high wavenumber cut-off occurs at $2.4/L_d$.
- ★ There is no low wavenumber cut-off.
- ★ For any given k , the highest growth rate occurs for $l = 0$. In the two-layer model, from (6.119), for $l = 0$ the maximum growth rate occurs when $k = 0.634k_d = 1.79/L_d$. For the Eady problem, the maximum growth rate occurs at $1.61/L_d$.

Solution in the general case: non-zero shear and non-zero β

Using (6.115), the growth rate and wave speeds as function of wavenumber are plotted in Fig. 6.14. We observe that there still appears to be a high-wavenumber cut-off and, for $\beta = 0$, there is a low-wavenumber cut-off. A little analysis elucidates the origin of these features.

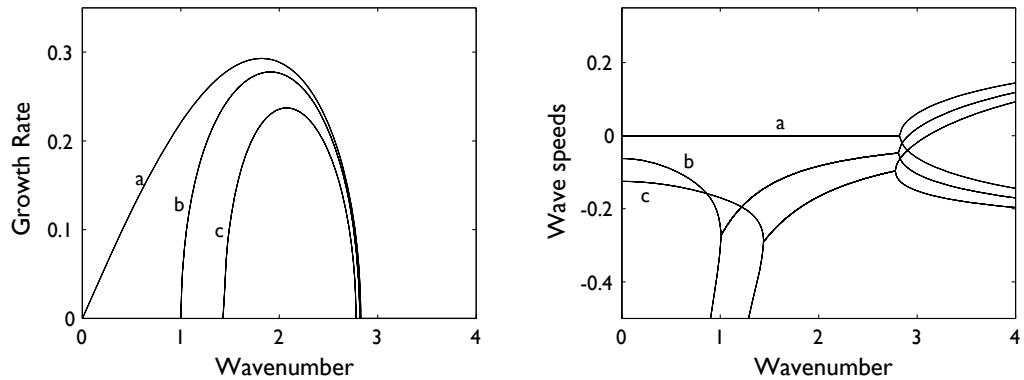


Fig. 6.14 Growth rates and wave speeds for the two-layer baroclinic instability problem, from (6.115), with three (non-dimensional) values of β : a, $\gamma = 0$ ($\hat{k}_\beta = 0$); b, $\gamma = 0.5$ ($\hat{k}_\beta = \sqrt{2}$); c, $\gamma = 1$ ($\hat{k}_\beta = 2$). As β increases, so does the low-wavenumber cut-off to instability, but the high-wavenumber cut-off is little changed. (The solutions are obtained from (6.115), with $\hat{k}_d = \sqrt{8}$ and $U_1 = -U_2 = 1/4$.)

The neutral curve

For instability, there must be an imaginary component to the phase speed in (6.113); that is, we require

$$k_\beta^4 k_d^4 + 4K^4(K^4 - k_d^4) < 0. \quad (6.120)$$

This is a quadratic equation in K^4 for the value of K , K_c say, at which the growth rate is zero. Solving, we find

$$K_c^4 = \frac{1}{2}k_d^4 \left(1 \pm \sqrt{1 - k_\beta^4/k_d^4} \right), \quad (6.121)$$

and this is plotted in Fig. 6.15. From (6.120) useful approximate expressions can be obtained for the critical shear as a function of wavenumber in the limits of small K and $K \approx k_d$, and these are left as exercises for the reader.

Minimum shear for instability

From (6.120), instability arises when $\beta^2 k_d^4 / U^2 < 4K^4(k_d^4 - K^4)$. The maximum value of the right-hand side of this expression arises when $K^4 = \hat{k}_d^4/2$; thus, instability arises only when

$$\frac{\beta^2 k_d^4}{U^2} < 4 \frac{k_d^4}{2} \frac{k_d^4}{2} \quad (6.122)$$

or

$$U_s > \frac{2\beta}{k_d^2} \quad (6.123)$$

where $U_s = U_1 - U_2 = 2U$. In terms of the deformation radius itself the minimum shear for instability is

$$U_s > \frac{1}{4}\beta L_d^2. \quad (6.124)$$

Figure 6.16 sketches how this might vary with latitude in the atmosphere and ocean. [In

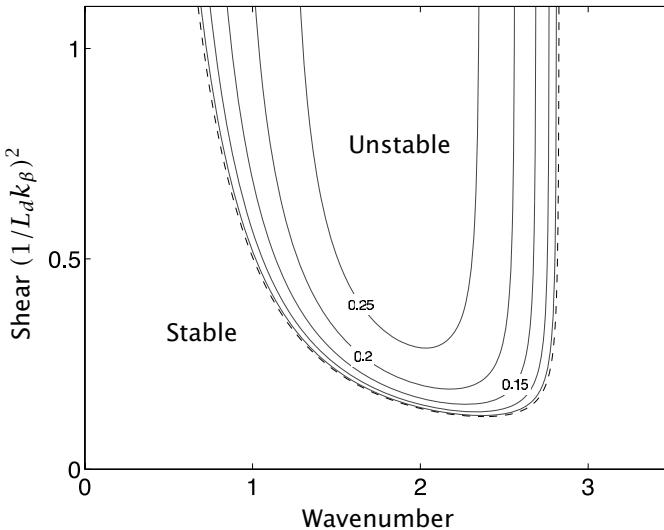


Fig. 6.15 Contours of growth rate in the two-layer baroclinic instability problem. The dashed line is the neutral stability curve obtained from (6.121), and the other curves are contours of growth rates obtained from (6.115). Outside of the dashed line, the flow is stable. The wavenumber is scaled by $1/L_d$ (i.e., by $k_d/\sqrt{8}$) and growth rates are scaled by the inverse of the Eady time scale (i.e., by U/L_d). Thus, for $L_d = 1000 \text{ km}$ and $U = 10 \text{ m s}^{-1}$, a non-dimensional growth rate of 0.25 corresponds to a dimensional growth rate of $0.25 \times 10^{-5} \text{ s}^{-1} = 0.216 \text{ day}^{-1}$.

(6.124), the shear is the difference in the velocity between level 1 and level 2, whereas the deformation radius, NH/f_0 , is based on the total height of the fluid. If we were to use half the depth of the fluid in the definition of the deformation radius, the factor of 4 would disappear.] If the shear is just this critical value, the instability occurs at $k = 2^{-1/4}k_d = 0.84k_d = 2.37/L_d$. As the shear increases, the wavenumber at which the growth rate is maximum decreases slightly (see Fig. 6.15), and for a sufficiently large shear the β -effect is negligible and the wavenumber of maximum instability is, as we saw earlier, $0.634k_d$ or $1.79/L_d$.

Note the relationship of the minimum shear to the basic state potential vorticity gradient in the respective layers. In the upper and lower layers the potential vorticity gradients are given by, respectively,

$$\frac{\partial Q_1}{\partial y} = \beta + k_d^2 U, \quad \frac{\partial Q_2}{\partial y} = \beta - k_d^2 U. \quad (6.125a,b)$$

Thus, the requirement for instability is exactly that which causes the potential vorticity gradient to change sign somewhere in the domain, in this case becoming negative in the lower layer. This is an example of the general rule that potential vorticity (suitably generalized to include the surface boundary conditions) must change sign somewhere in order for there to be an instability.

High-wavenumber cut-off

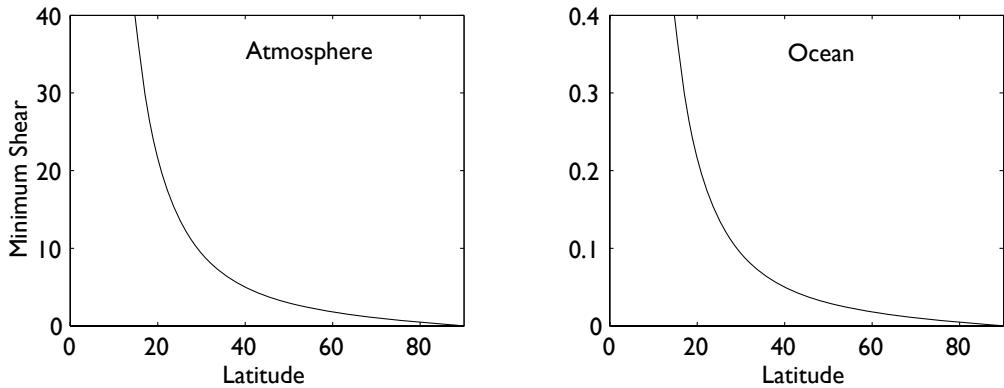


Fig. 6.16 The minimum shear ($U_s = U_1 - U_2$, in m s^{-1}) required for baroclinic instability in a two-layer model, calculated using (6.124), i.e., $U_s = \beta L_d^2/4$ where $\beta = 2\Omega a^{-1} \cos \vartheta$ and $L_d = NH/f$, with $f = 2\Omega \sin \vartheta$. The left panel uses $H = 10 \text{ km}$ and $N = 10^{-2} \text{ s}^{-1}$, and the right panel uses parameters representative of the main thermocline, $H = 1 \text{ km}$ and $N = 10^{-2} \text{ s}^{-1}$. The results are not quantitatively accurate, but the implications that the minimum shear is much less for the ocean, and that in both the atmosphere and the ocean the shear increases rapidly at low latitudes, are robust.

Instability can only arise when, from (6.120),

$$4K^4(k_d^4 - K^4) > k_\beta^4 k_d^4, \quad (6.126)$$

so that a necessary condition for instability is

$$k_d^2 > K^2. \quad (6.127)$$

Thus, waves shorter than the deformation radius are always stable, no matter what the value of β . We also see from Fig. 6.14 and Fig. 6.15 that the high wavenumber cut-off in fact varies little with β if $k_d \gg k_\beta$. Note that the critical shear required for instability approaches infinity as K approaches k_d .

Low-wavenumber cut-off

Suppose that $k \ll k_d$. Then (6.120) simplifies to $k_\beta^4 < 4K^4$. That is, for instability we require

$$K^2 > \frac{1}{2} k_\beta^2 = \frac{\beta}{2U}. \quad (6.128)$$

Thus, using (6.127) and (6.128) the unstable waves lie approximately in the interval $\beta/(\sqrt{2}U) < k < k_d$.

6.7 AN INFORMAL VIEW OF THE MECHANISM OF BAROCLINIC INSTABILITY

In this section we take a more intuitive look at baroclinic instability, trying to understand the mechanism without treating the problem in full generality or exactness. We will do this by way of a semi-kinematic argument that shows how the waves in each layer of a two-layer

model, or the waves on the top and bottom boundaries in the Eady model, can constructively interact to produce a growing instability. It is kinematic in the sense that we initially treat the waves independently, and only subsequently allow them to interact — but it is this dynamical interaction that gives the instability. We first revisit the two-layer model and simplify it to its bare essentials.

6.7.1 The two-layer model

A simple dynamical model

We first re-derive the instability ab initio from the equations of motion written in terms of the baroclinic streamfunction τ and the barotropic streamfunction ψ where

$$\tau \equiv \frac{1}{2}(\psi_1 - \psi_2), \quad \psi \equiv \frac{1}{2}(\psi_1 + \psi_2). \quad (6.129)$$

We linearize about a sheared basic state of with zero barotropic velocity and with $\beta = 0$. Thus, with $\psi = 0 + \psi'$ and $\tau = -Uy + \tau'$ the linearized equations of motion, equivalent to (6.108) with $\beta = 0$, are

$$\frac{\partial}{\partial t} \nabla^2 \psi' = -U \frac{\partial}{\partial x} \nabla^2 \tau', \quad (6.130a)$$

$$\frac{\partial}{\partial t} (\nabla^2 - k_d^2) \tau' = -U \frac{\partial}{\partial x} (\nabla^2 + k_d^2) \psi'. \quad (6.130b)$$

Neglecting the y -dependence for simplicity, we may seek solutions of the form $(\psi', \tau') = \text{Re}(\tilde{\psi}, \tilde{\tau}) \exp[ik(x - ct)]$, where $c = i c_i + c_r$, giving

$$c \tilde{\psi} - U \tilde{\tau} = 0, \quad (6.131a)$$

$$c(K^2 + k_d^2) \tilde{\tau} - U(K^2 - k_d^2) \tilde{\psi} = 0. \quad (6.131b)$$

These equations have non-trivial solutions if the determinant of the matrix of coefficients of $\tilde{\tau}$ and $\tilde{\psi}$ is zero, giving the quadratic equation $c^2(K^2 + k_d^2) - U^2(K^2 - k_d^2) = 0$. Solving this gives, repring (6.118),

$$c = \pm U \left(\frac{K^2 - k_d^2}{K^2 + k_d^2} \right)^{1/2}. \quad (6.132)$$

Instabilities occur for $K^2 < k_d^2$, for which $c_r = 0$; that is, the wave speed is purely imaginary. From (6.131) unstable modes have

$$\tilde{\tau} = i \frac{c_i}{U} \tilde{\psi} = e^{i\pi/2} \frac{c_i}{U} \tilde{\psi}. \quad (6.133)$$

That is, τ *lags* ψ by 90° for a growing wave ($c_i > 0$). Similarly, τ *leads* ψ by 90° for a decaying wave. Now, the temperature is proportional to τ , and in the two-level model is advected by the vertically averaged perturbation meridional velocity, v say (with Fourier amplitude \tilde{v}), where $v = \partial \psi / \partial x$. Thus, for growing or decaying waves,

$$\tilde{v} = \tilde{\tau} \frac{kU}{c_i}, \quad (6.134)$$

and the meridional velocity is exactly *in phase* with the temperature for growing modes,

and is *out of phase* with the temperature for decaying modes. That is, for unstable modes, poleward flow is correlated with high temperatures, and for decaying modes poleward flow is correlated with low temperatures. For neutral waves, $\tilde{\tau} = c_r \tilde{\psi} / U$ and so $\tilde{v} = ik\tilde{\tau}U/c_r$, and the meridional velocity and temperature are $\pi/2$ out of phase. Thus, to summarize:

- ★ growing waves transport heat (or buoyancy) polewards;
- ★ decaying waves transport heat equatorwards;
- ★ neutral waves do not transport heat.

Further simplifications to the two-layer model

First consider (6.130) for waves much larger than the deformation radius, $K^2 \ll k_d^2$; we obtain

$$\frac{\partial}{\partial t} \nabla^2 \psi = -U \frac{\partial}{\partial x} \nabla^2 \tau, \quad \frac{\partial}{\partial t} \tau = U \frac{\partial}{\partial x} \psi. \quad (6.135a,b)$$

for which we obtain [either directly or from (6.132)] $c = \pm iU$; that is, the flow is unstable. To see the mechanism, suppose that the initial perturbation is barotropic and sinusoidal in x , with no y variation. Polewards flowing fluid (i.e., $\partial \psi / \partial x > 0$) will, by (6.135b), generate a positive τ , and the baroclinic flow will be out of phase with the barotropic flow. Then, by (6.135a), the advection of τ by the mean shear produces growth of ψ that is in phase with the original disturbance. Contrast this case with that for very small disturbances, for which $K^2 \gg k_d^2$ and (6.130) becomes

$$\frac{\partial}{\partial t} \nabla^2 \psi = -U \frac{\partial}{\partial x} \nabla^2 \tau, \quad \frac{\partial}{\partial t} \nabla^2 \tau = -U \frac{\partial}{\partial x} \nabla^2 \psi, \quad (6.136a,b)$$

or, in terms of the equations for each layer,

$$\frac{\partial}{\partial t} \nabla^2 \psi_1 = -U \frac{\partial}{\partial x} \nabla^2 \psi_1, \quad \frac{\partial}{\partial t} \nabla^2 \psi_2 = +U \frac{\partial}{\partial x} \nabla^2 \psi_2. \quad (6.137a,b)$$

That is, the layers are completely decoupled and no instability can arise. Motivated by this, consider waves that propagate independently in each layer on the potential vorticity gradient caused by β (if non-zero) and shear. Thus, in (6.108) we keep the potential vorticity gradients but neglect k_d^2 where it appears alongside ∇^2 and find

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \nabla^2 \psi'_1 + \frac{\partial \psi'_1}{\partial x} \frac{\partial Q_1}{\partial y} = 0, \quad (6.138a)$$

$$\left(\frac{\partial}{\partial t} - U \frac{\partial}{\partial x} \right) \nabla^2 \psi'_2 + \frac{\partial \psi'_2}{\partial x} \frac{\partial Q_2}{\partial y} = 0, \quad (6.138b)$$

where $\partial Q_1 / \partial y = \beta + k_d^2 U$ and $\partial Q_2 / \partial y = \beta - k_d^2 U$. Seeking solutions of the form (6.109), the phase speeds of the associated waves are

$$c_1 = U - \frac{\partial_y Q_1}{K^2}, \quad c_2 = -U - \frac{\partial_y Q_2}{K^2}. \quad (6.139a,b)$$

In the upper layer the phase speed is a combination of an eastward advection and a fast westward wave propagation due to a strong potential vorticity gradient. In the lower layer the phase speed is a combination of a westward advection and a slow eastward wave propagation due to the weak potential vorticity gradient. The two phase speeds are, in general, not equal,

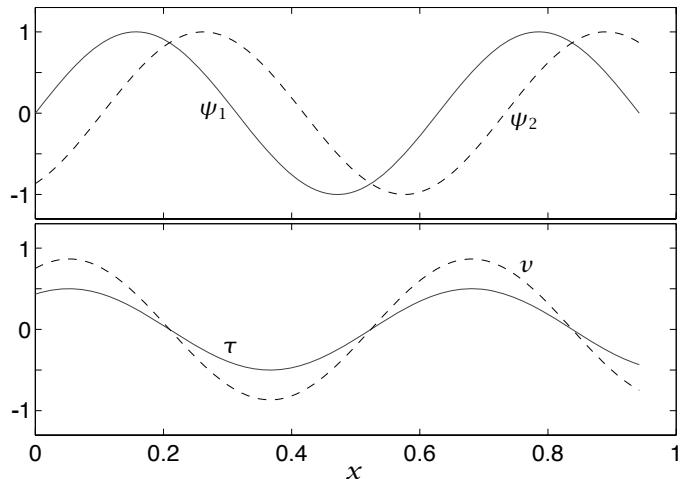


Fig. 6.17 Baroclinically unstable waves in a two layer model. The streamfunction is shown in the top panel, ψ_1 for the top layer and ψ_2 for the bottom layer. Given the westward tilt shown, the temperature, τ , and meridional velocity, v (bottom panel) are in phase, and the instability grows.

but they would need to be so if they were to combine to cause an instability. From (6.139) this occurs when $K^2 = k_d^2$ and $c_1 = c_2 = -\beta/k_d^2$. These conditions are just those occurring at the high-wavenumber cut-off to instability in the two-level model. At higher wavenumbers, the waves are unable to synchronize, whereas at lower wavenumbers they may become inextricably coupled.

Let us suppose that the phase of the wave in the upper layer lags that (i.e., is westward of) that in the lower layer, as illustrated in the top panel Fig. 6.17. The lower panel shows the temperature field, $\tau = (\psi_1 - \psi_2)/2$, and the average meridional velocity, $v = \partial_x(\psi_1 + \psi_2)/2$. In this configuration, the temperature field is *in phase* with the meridional velocity, meaning that warm fluid is advected polewards. Now, let us allow the waves in the two layers to interact by adding one dynamical equation, the thermodynamic equation, which in its simplest form is

$$\frac{\partial \tau}{\partial t} = -v \frac{\partial \bar{\tau}}{\partial y} = v U, \quad (6.140)$$

where $\bar{\tau}$ is proportional to the basic state temperature field. The temperature field, τ , grows in proportion to v , which is proportional to τ if the waves tilt westwards with height, and an instability results. This dynamical mechanism is just that which is compactly described by (6.135). It is a straightforward matter to show that if the streamfunction tilts eastwards with height, v is out of phase with τ and the waves decay. A similar description applies to the Eady problem, as we now see.

6.7.2 Interacting edge waves in the Eady problem

We now explore how the edge waves at the top and bottom surfaces in the Eady problem give rise to an instability, by way of semi-kinematic description that is very similar to that of the barotropic problem described on page 256. Let us first consider the case in which the bottom and top surfaces are essentially uncoupled. Instead of solutions of (6.82) that have the structure (6.83) (which satisfies both boundary conditions) consider solutions that *separately* satisfy the bottom and top boundary conditions and that decay into the interior.

These are

$$\psi_B = \operatorname{Re} A_B e^{ik(x-c_T t)} e^{-\mu z/H}, \quad \psi_T = \operatorname{Re} A_T e^{i\phi} e^{ik(x-c_B t)} e^{\mu(z-H)/H}, \quad (6.141a,b)$$

for the bottom and top surfaces, respectively, and ϕ is the phase shift, with A_B and A_T being real constants. The boundary conditions (6.81) then determine the phase speeds of the two systems and we find

$$c_B = \frac{\Lambda H}{\mu}, \quad c_T = \Lambda H \left(1 - \frac{1}{\mu}\right). \quad (6.142a,b)$$

These are the phase speeds of *edge waves* in the Eady problem; they are real and in general they are unequal. It must therefore be the *interaction* of the waves on the upper and lower boundaries that is necessary for instability, because the unstable wave has but a single phase speed. This interaction can occur when their phase speeds are equal and from (6.142) this occurs when $\mu = 2$, giving

$$k = \frac{2}{L_d} \quad \text{and} \quad c = \frac{\Lambda H}{2}. \quad (6.143a,b)$$

This phase speed is just that of the flow at mid-level, and at the critical wavenumber in the full Eady problem [$k_c = 2.4/L_d$, from (6.91)] the phase speed is purely real and equal to that of (6.143b) — see Fig. 6.10. Thus, (6.143) approximately characterizes the critical wavenumber in the full problem.

To turn this kinematic description into a dynamical instability, suppose that the two rigid surfaces are close enough so that the waves can interact, but still far enough so that their structure is approximately given by (6.141). (Note that if μ is too large, the waves decay rapidly away from the edges and will not intersect.) Specifically, let the buoyancy perturbation at a given boundary be advected by the total meridional velocity perturbation, including that arising from the perturbation at the other boundary, so that at the top and bottom boundaries

$$\frac{\partial b'_T}{\partial t} = -(\nu'_B + \nu'_T) \frac{\partial \bar{b}_T}{\partial y}, \quad \frac{\partial b'_B}{\partial t} = -(\nu'_B + \nu'_T) \frac{\partial \bar{b}_B}{\partial y}. \quad (6.144)$$

The waves will reinforce each other if ν'_T is in phase with b'_B at the lower boundary, and if ν'_B is in phase with b'_T at the upper boundary. Now, using (6.141), the velocity and buoyancy associated with the edge waves are given by

$$b_B = -\operatorname{Re} k N A_B e^{-\mu z/H} e^{ikx}, \quad b_T = \operatorname{Re} k N A_T e^{i\phi} e^{\mu(z-H)/H} e^{ikx}, \quad (6.145a)$$

$$\nu_B = \operatorname{Re} ik A_B e^{-\mu z/H} e^{ikx}, \quad \nu_T = \operatorname{Re} ik A_T e^{i\phi} e^{\mu(z-H)/H} e^{ikx}. \quad (6.145b)$$

The fields b_B and ν_T , and b_T and ν_B , will be positively correlated if $0 < \phi < \pi$, and will be exactly in phase if $\phi = \pi/2$, and this case is illustrated in Fig. 6.18. Just as in the two-layer case, this phase corresponds to a westward tilt with height, and it is this, in conjunction with geostrophic and hydrostatic balance, that allows warm fluid to move polewards and available potential energy to be released. From (6.144), the perturbation will grow and an instability will result. The analogy between baroclinic instability and barotropic instability should be evident from the similarity of this description and that of section 6.2.4, with z in the baroclinic problem playing the role of y in the barotropic problem, and b the role of

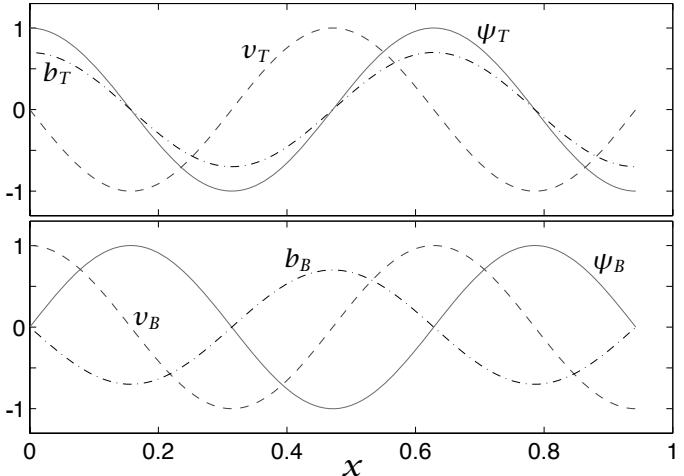


Fig. 6.18 Interacting edge waves in the Eady model. The upper panel shows waves on the top surface, and the lower panel show waves on the bottom. If the streamfunction tilts westwards with height, then the temperature on the top (bottom) is correlated with the meridional velocity on the bottom (top), the waves can reinforce each other. See also Fig. 6.12.

u ; note that (6.144) is almost identical to (6.46). However, the analogy of the two problems in full is not perfect because the boundary condition that $w = 0$ does not have an exact correspondence in the barotropic problem. More importantly, the nonlinear development of the baroclinic problem, discussed in chapter 9, is generally three-dimensional, which need not be the case in the barotropic problem.

6.8 * THE ENERGETICS OF LINEAR BAROCLINIC INSTABILITY

In baroclinic instability, warm parcels move polewards and cold parcels move equatorwards. This motion draws on the available potential energy of the mean state, because warm light parcels move upwards, and cold dense parcels move downwards and the height of the mean centre of gravity of the fluid falls, and the loss of potential energy is converted to kinetic energy of the perturbation. However, because the instability is growing, the energy of the perturbation is of course not conserved, and both the kinetic energy and the available potential energy of the perturbation will grow. However, we still expect a conversion of potential energy to kinetic energy, and the purpose of this section is to demonstrate that explicitly. For simplicity, we restrict attention to the flat-bottomed two-level model with $\beta = 0$.

As in section 5.6, the energy may be partitioned into kinetic energy and available potential energy. In a three-dimensional quasi-geostrophic flow the kinetic energy is given by, in general,

$$KE = \frac{1}{2} \int (\nabla \psi)^2 dV \quad (6.146)$$

which, in the case of the two-layer model becomes

$$KE = \frac{1}{2} \int (\nabla \psi_1)^2 + (\nabla \psi_2)^2 dA = \int (\nabla \psi)^2 + (\nabla \tau)^2 dA. \quad (6.147)$$

Restricting attention to a single Fourier mode this becomes

$$KE = k^2 \tilde{\psi}^2 + k^2 \tilde{\tau}^2. \quad (6.148)$$

The available potential energy in the continuous case is given by

$$APE = \frac{1}{2} \int \left(\frac{f_0}{N} \right)^2 \left(\frac{\partial \psi}{\partial z} \right)^2 dV. \quad (6.149)$$

For a single Fourier mode in a two-layer model this becomes

$$APE = k_d^2 \tilde{\tau}^2. \quad (6.150)$$

Now, the nonlinear vorticity equations for each level is

$$\frac{\partial}{\partial t} \nabla^2 \psi_1 + J(\psi_1, \nabla^2 \psi_1) = -2 \frac{f_0 w}{H} \quad (6.151a)$$

$$\frac{\partial}{\partial t} \nabla^2 \psi_2 + J(\psi_2, \nabla^2 \psi_2) = 2 \frac{f_0 w}{H}, \quad (6.151b)$$

where w is the vertical velocity between the levels. (These equations are the two-level analogues of the continuous vorticity equation, with the right-hand sides being finite difference versions of $f_0 \partial w / \partial z$.) Multiplying the two equations of (6.151) by $-\psi_1$ and $-\psi_2$, respectively, and adding we readily find

$$\frac{d}{dt} KE = \frac{4f_0}{H} \int w \tau dA. \quad (6.152a)$$

For a single Fourier mode this becomes

$$\frac{d}{dt} KE = \text{Re} \frac{4f_0}{H} \tilde{w} \tilde{\tau}^*, \quad (6.152b)$$

where $w = \tilde{w} \exp[i(kx - ct)] + \text{c.c.}$, and the asterisk denotes a complex conjugate.

The continuous thermodynamic equation is

$$\frac{Db}{Dt} + w N^2 = 0. \quad (6.153)$$

Using $b = f_0 \partial \psi / \partial z$ and finite-differencing [with $\partial \psi / \partial z \rightarrow (\psi_1 - \psi_2) / (H/2) = 4\tau / H$], we obtain the two-level thermodynamic equation:

$$\frac{\partial \tau}{\partial t} + J(\psi, \tau) + \frac{w N^2 H}{4f_0} = 0. \quad (6.154)$$

The change of available potential energy is obtained from this by multiplying by $k_d^2 \tau$ and integrating, giving

$$\int \left(\frac{1}{2} \frac{d}{dt} k_d^2 \tau^2 + \tau w \frac{2f_0}{H} \right) dA = 0 \quad (6.155)$$

or

$$\frac{d}{dt} APE = -\frac{4f_0}{H} \int w \tau dA \quad (6.156a)$$

or, for a single Fourier mode,

$$\frac{d}{dt} APE = -\text{Re} \frac{4f_0}{H} \tilde{w} \tilde{\tau}^*. \quad (6.156b)$$

From (6.152) and (6.156) it is clear that in the nonlinear equations the sum of the kinetic energy and the available potential energy is conserved.

We now specialize by obtaining w from the linear baroclinic instability problem. Using this in (6.152) and (6.156) will give us the conversion between kinetic energy and potential energy in the growing baroclinic wave. It is important to realize that the total energy of the disturbance will not be conserved — both the potential and kinetic energy are growing, exponentially in this problem, because they are extracting energy from the mean state. To calculate w we use the linearized thermodynamic equation. From (6.154) this is

$$\frac{\partial \tau}{\partial t} - U \frac{\partial \psi}{\partial x} + \frac{H w N^2}{4f_0} = 0, \quad (6.157)$$

omitting the primes on perturbation quantities. For a single Fourier mode, this gives

$$\frac{HN^2}{4f_0} \tilde{w} = ik(c\tilde{\tau} + U\tilde{\psi}). \quad (6.158)$$

But, from (6.131), $c\tilde{\psi} = U\tilde{\tau}$ in two-layer f -plane baroclinic instability and so

$$\frac{HN^2}{4f_0} \tilde{w} = ikc\tilde{\tau} \left(1 + \frac{U^2}{c^2} \right) = ikc\tilde{\tau} \left(\frac{2K^2}{K^2 - k_d^2} \right), \quad (6.159)$$

using (6.132). For stable waves, $K^2 > k_d^2$ and $c = c_r$ and in that case the vertical velocity is $\pi/2$ out of phase with the temperature, and there is no conversion of APE to KE. For unstable waves $c = ic_i$ and $K^2 < k_d^2$, and the vertical velocity is in phase with the temperature; that is, warm air is rising and so there is a conversion of APE to KE. To see this more formally, recall that the conversion from APE to KE is given by $4\tilde{w}\tilde{\tau}^* f_0/H$. Thus, using (6.159),

$$\frac{d}{dt} (APE \rightarrow KE) = \text{Re} 2ikck_d^2 \left(\frac{2K^2}{K^2 - k_d^2} \right) \tilde{\tau}^2, \quad (6.160)$$

using also the definition of k_d given in (6.104). If the wave is growing, then $K^2 < k_d^2$ and $c = ic_i$ and the right-hand side is real and positive. For neutral waves, if $c = c_r$ the right-hand side of (6.160) is pure imaginary, and so the conversion is zero. This completes our demonstration that baroclinic instability converts potential energy into kinetic energy.

6.9 * BETA, SHEAR AND STRATIFICATION IN A CONTINUOUS MODEL

The two-layer model of section 6.6 indicates that β has a number of important effects on baroclinic instability. Do these carry over to the continuously stratified case? The answer by and large is yes, but with some important qualifications that generally concern weak or shallow instabilities. In particular, we will find that there is no short-wave cut-off in the continuous model with non-zero beta, and that the instability determines its own depth scale. We will illustrate these properties first by way of scaling arguments and then by way of numerical calculations.¹⁴

6.9.1 Scaling arguments for growth rates, scales and depth

With finite density scale height and non-zero β , the quasi-geostrophic potential vorticity equation, linearized about a mean zonal velocity $U(z)$, is

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) q' + \frac{\partial \psi'}{\partial x} \frac{\partial Q}{\partial y} = 0, \quad (6.161)$$

where

$$q' = \nabla^2 \psi' + \frac{f_0^2}{\rho} \frac{\partial}{\partial z} \left(\frac{\rho}{N^2} \frac{\partial \psi'}{\partial z} \right), \quad \frac{\partial Q}{\partial y} = \beta - \frac{f_0^2}{\rho} \frac{\partial}{\partial z} \left(\frac{\rho}{N^2} \frac{\partial U}{\partial z} \right), \quad (6.162a,b)$$

and ρ is a specified density profile. If we assume that $U = \Lambda z$ where Λ is constant and that N is constant, and let $H_\rho^{-1} = -\rho^{-1} \partial \rho / \partial z$, then

$$\frac{\partial Q}{\partial y} = \beta + \frac{f_0^2 \Lambda}{N^2 H_\rho} = \beta(1 + \alpha), \quad \text{where } \alpha = \left(\frac{f_0^2 \Lambda}{\beta N^2 H_\rho} \right). \quad (6.163)$$

The boundary conditions on (6.161) are

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \frac{\partial \psi'}{\partial z} - \frac{\partial \psi'}{\partial x} \frac{\partial U}{\partial z} = 0, \quad \text{at } z = 0 \quad (6.164)$$

and that $\psi \rightarrow 0$ as $z \rightarrow \infty$. The problem we have defined essentially constitutes the Charney problem. We can reduce this to the Eady problem by setting $\beta = 0$ and $H_\rho = \infty$, and providing a lid that is some finite height above the ground.

As in the Eady problem, we seek solutions of the form

$$\psi = \text{Re } \tilde{\psi}(z) e^{i(kx+ly-kct)}, \quad (6.165)$$

and substituting into (6.161) gives

$$\left(\frac{f_0^2}{H_\rho^2 N^2} \right) \left(H_\rho^2 \frac{d^2 \tilde{\psi}}{dz^2} - H_\rho \frac{d \tilde{\psi}}{dz} \right) - \left(K^2 - \frac{\beta + \Lambda f_0^2 / (N^2 H_\rho)}{\Lambda z - c} \right) \tilde{\psi} = 0. \quad (6.166)$$

The Boussinesq version of this expression, for a fluid contained between two horizontal surfaces, is obtained by letting $H_\rho = \infty$, giving

$$\left(\frac{f_0^2}{N^2} \right) \frac{d^2 \tilde{\psi}}{dz^2} - \left(K^2 - \frac{\beta}{\Lambda z - c} \right) \tilde{\psi} = 0. \quad (6.167)$$

It seems natural to non-dimensionalize (6.166) using:

$$z = H_\rho \hat{z}, \quad c = \Lambda H_\rho \hat{c}, \quad K = \left(\frac{f_0}{N H_\rho} \right) \hat{K}, \quad (6.168)$$

whence it becomes

$$\frac{d^2 \tilde{\psi}}{d \hat{z}^2} - \frac{d \tilde{\psi}}{d \hat{z}} - \left(\hat{K}^2 - \frac{\gamma + 1}{\hat{z} - \hat{c}} \right) \tilde{\psi} = 0 \quad (6.169)$$

where

$$\gamma = \alpha^{-1} = \frac{\beta N^2 H_\rho}{f_0^2 \Lambda} = \frac{\beta L_d^2}{H_\rho \Lambda} = \frac{H_\rho}{h}. \quad (6.170)$$

where $h \equiv \Lambda f_0^2 / (\beta N^2)$. The non-dimensional parameter γ is known as the Charney–Green number.¹⁵ The Boussinesq version, (6.167), may be non-dimensionalized using H_D in place of H_ρ , where H_D is the depth of the fluid between two rigid surfaces. In that case

$$\frac{d^2 \tilde{\psi}}{d\hat{z}^2} - \left(\hat{K}^2 - \frac{\gamma}{\hat{z} - \hat{c}} \right) \tilde{\psi} = 0, \quad (6.171)$$

where here the non-dimensional variables are scaled with H_D .

Now, suppose that γ is large, for example if β or the static stability are large or the shear is weak. Equation (6.169) admits no non-trivial balance, suggesting that we rescale the variables using h instead of H_ρ as the vertical scale in (6.168). The rescaled version of (6.169) is then

$$\frac{d^2 \tilde{\psi}}{d\hat{z}^2} - \frac{1}{\gamma} \frac{d\tilde{\psi}}{d\hat{z}} - \left(\hat{K}^2 - \frac{1 + \gamma^{-1}}{\hat{z} - \hat{c}} \right) \tilde{\psi} = 0, \quad (6.172)$$

or, approximately,

$$\frac{d^2 \tilde{\psi}}{d\hat{z}^2} - \left(\hat{K}^2 - \frac{1}{\hat{z} - \hat{c}} \right) \tilde{\psi} = 0. \quad (6.173)$$

This is exactly the same equation as results from a similar rescaling of the Boussinesq system, (6.171), as we might have expected because now the dynamical vertical scale, h , is much smaller than the scale height H_ρ (or H_D) and the system is essentially Boussinesq. Thus, noting that (6.173) has the same non-dimensional form as (6.171) save that γ is replaced by unity, and that (6.173) with $\gamma = 1$ must produce the same scales and growth rates as in the Eady problem, we may deduce that:

- (i) the wavelength of the instability is $\mathcal{O}(Nh/f_0)$;
- (ii) the growth rate of the instability is $\mathcal{O}(Kc) = \mathcal{O}(f_0 \Lambda / N)$.
- (iii) the vertical scale of the instability is $\mathcal{O}(h) = \mathcal{O}(f_0^2 \Lambda / (\beta N^2))$.

These are the same as for the Eady problem, except with the dynamical height h replacing the geometric or scale height H_D . Effectively, the dynamics has determined its own vertical scale, h , which is much less than the scale height or geometric height, producing ‘shallow modes’.

In the limit $\gamma \ll 1$ (strong shear, weak β), the Boussinesq and compressible problems differ. The Boussinesq problem reduces to the Eady problem, considered previously, whereas (6.169) becomes, approximately,

$$\frac{d^2 \tilde{\psi}}{d\hat{z}^2} - \frac{d\tilde{\psi}}{d\hat{z}} - \left(\hat{K}^2 - \frac{1}{\hat{z} - \hat{c}} \right) \tilde{\psi} = 0, \quad (6.174)$$

and in this limit the appropriate vertical scale is the density scale height H_ρ . Because $H_\rho \gg h$ these are ‘deep modes’, occupying the entire vertical extent of the domain.

The scale h does not arise in the two-level model, but there is a connection between it and the critical shear for instability in the two-level model. The condition $\gamma \ll 1$, or $h \gg H$, may be written as

$$H\Lambda \gg \beta \left(\frac{NH}{f_0} \right)^2. \quad (6.175)$$

Compare this with the necessary condition for instability in a two-level model, (6.124), namely

$$(U_1 - U_2) > \beta \left(\frac{NH_\Delta}{f_0} \right)^2, \quad (6.176)$$

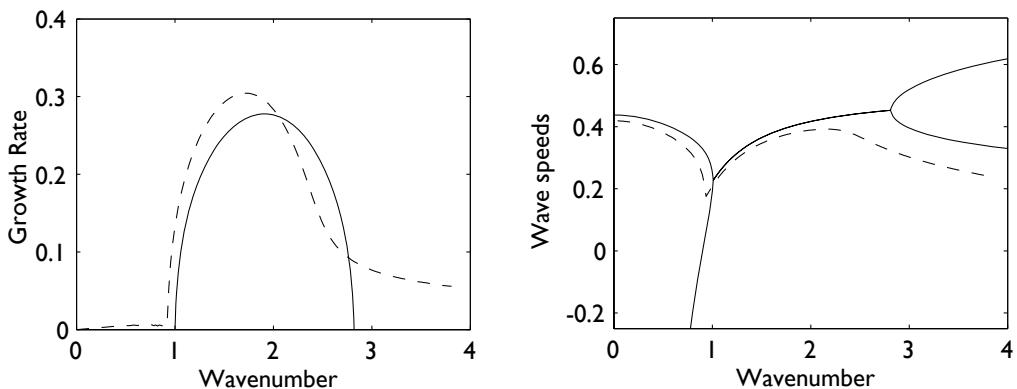


Fig. 6.19 Growth rates and wave speeds for the two-layer (solid) and continuous (dashed) models, with the same values of the Charney-Green number, γ , and uniform shear and stratification. (In the two-layer case $\gamma = \beta L_d^2 / [2(U_1 - U_2)] = 0.5$, and in the continuous case $\gamma = \beta L_d^2 / (H\Lambda) = 0.5$.) In the continuous case only the wave speed associated with the unstable mode is shown. In the two-layer case there are two real wave speeds which coalesce in the unstable region.

where H_Δ is the vertical distance between the two levels. Thus, essentially the same condition governs the onset of instability in the two-level model as governs the production of deep modes in the continuous model. This correspondence is a natural one, because in the two-level model *all* modes are ‘deep’, and the model fails (as it should) to capture the shallow modes of the continuous system. For similar reasons, there is a high-wavenumber cut-off in the two-level model: in the continuous model these modes are shallow and so cannot be captured by two-level dynamics. Somewhat counter-intuitively, for these modes the β -effect must be important, even though the modes have small horizontal scale: when $\beta = 0$ the instability arises via an interaction between edge waves at the top and bottom of the domain, whereas the shallow instability arises via an interaction of the edge waves at the surface with Rossby waves just above the surface.

6.9.2 Some numerical calculations

Adding β to the Eady model

Our first step is to add the β -effect to the Eady problem.¹⁶ That is, we suppose a Boussinesq fluid with uniform stratification, that the shear is zonal and constant and that the entire problem is sandwiched between two rigid surfaces. Growth rates and phase speeds of such an instability calculation are illustrated in Fig. 6.19 and the vertical structure is shown in Fig. 6.20. As in the two-layer problem, there is a low-wavenumber cut-off to the main instability, although there is now an additional weak instability at very large-scales. These so-called *Green modes* have no counterpart in the two-layer model — they are deep, slowly growing modes that will be dominated by faster growing modes in most real situations. (Also, the fact that the Green modes have a scale much larger than the deformation scale suggests that a degree of caution in the accuracy of the quasi-geostrophic calculation is warranted.) At high wavenumbers there is no cut-off to the instability in the continuous

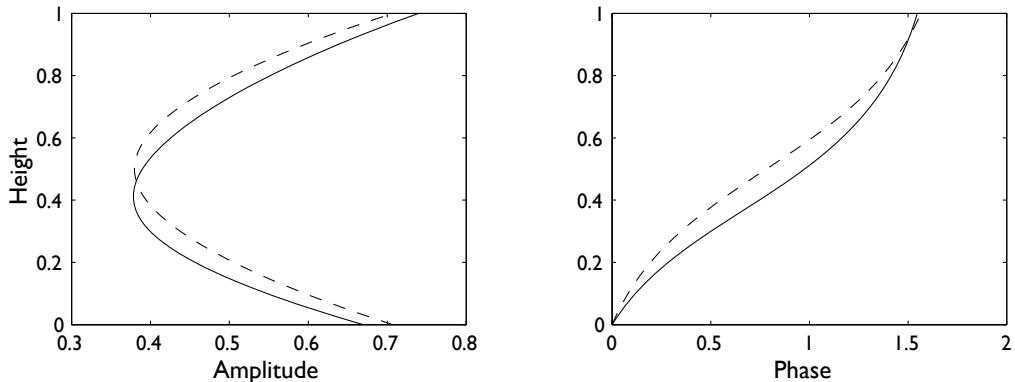


Fig. 6.20 Vertical structure of the most unstable modes in a continuously stratified instability calculation with $\beta = 0$ (dashed lines, the Eady problem) and $\beta \neq 0$ (solid lines), as in Fig. 6.19. The effect of beta is to depress the height of the maximum amplitude of the instability.

problem in the case of non-zero beta; the high-wavenumber modes are shallow and unstable via an interaction between edge waves at the lower boundary and Rossby waves in the lower atmosphere, and so have no counterpart in either the two-layer problem (where the modes are deep) or the Eady problem (which has no Rossby waves).

Effects of non-uniform shear and stratification

If the shear or stratification is non-uniform an analytic treatment is, even in problems without β , usually impossible and the resulting equations must be solved numerically. However, if we restrict attention to a *discontinuity* in the shear or the stratification, then the resulting problem is very similar to the problem with rigid boundaries, and this property provides much of the justification for using the Eady problem to model instabilities in the Earth's atmosphere: in the troposphere the stratification is (approximately) constant, and the rapid increase in stratification in the stratosphere can be approximated by a lid at tropopause. Heuristically, we can see this from the form of the thermodynamic equation, namely

$$\frac{Db}{Dt} + N^2 w = 0. \quad (6.177)$$

If N^2 is high this suggests w will be small, and a lid is the limiting case of this. The oceanic problem is rather more involved, because although both the stratification and the shear are concentrated in the upper ocean, they vary relatively smoothly; furthermore, the shear is high where the stratification is high, and the two have opposing effects.

To go one step further, consider the Boussinesq potential vorticity equation, linearized about a zonally uniform state $\Psi(y, z)$, with a rigid surface at $z = 0$. The normal-mode evolution equations are similar to (6.70), namely

$$(U - c) \left[\frac{\partial^2}{\partial y^2} - k^2 + \frac{\partial}{\partial z} \left(F \frac{\partial}{\partial z} \right) \right] \tilde{\Psi} + \frac{\partial Q}{\partial y} \tilde{\Psi} = 0, \quad z > 0, \quad (6.178a)$$

$$(U - c) \frac{\partial \tilde{\Psi}}{\partial z} - \frac{\partial U}{\partial z} \tilde{\Psi} = 0, \quad \text{at } z = 0, \quad (6.178b)$$

where $\partial_y Q = \beta - \partial_{yy} U - \partial_z (F \partial_z U)$. Now suppose that there is a discontinuity in the shear and/or the stratification in the interior of the fluid, at some level $z = z_c$. Integrating (6.178a) across the discontinuity, noting that $\tilde{\psi}$ is continuous in z , gives

$$(U - c) \left[F \frac{\partial \tilde{\psi}}{\partial z} \right]_{z_c^-}^{z_c^+} - \tilde{\psi}(y, z_c) \left[F \frac{\partial U}{\partial z} \right]_{z_c^-}^{z_c^+} = 0. \quad (6.179)$$

which has similar form to (6.178b). This construction is evocative of the equivalence of a delta-function sheet of potential vorticity at a rigid boundary, except that now a discontinuity in the potential vorticity in the *interior* has a similarity with a rigid boundary.

We can illustrate the effects of an interior discontinuity that crudely represents the tropopause by numerically solving the linear eigenvalue problem. We pose the problem on the f -plane, in a horizontally doubly-periodic domain, with no horizontal variation of shear, and between two horizontal rigid lids. The eigenvalue problem is defined by (6.70), and the numerical procedure then solves for the complex eigenvalue c and eigenfunction $\tilde{\psi}(z)$; various results are illustrated in Fig. 6.21. To parse this rather complex figure, first look at the solid curves in all the panels. These arise when the problem is solved with a uniform shear and a uniform stratification, with a lid at $z = 0$ and $z = 1$, hence simply giving the Eady problem. The familiar growth rates and vertical structure of the solution are given by the solid curves in panels (b), (c) and (d), and these are just the same as in Fig. 6.10. The various dotted and dashed curves show the results when the lid at $z = 1$ is replaced by a stratosphere stretching from $1 < z < 2$ either with high stratification, zero shear, or both, and in all of these cases the stratosphere acts qualitatively in the same way as a rigid lid. The vertical structure of the solution in the troposphere is, in all cases, quite similar, and the amplitude decays rapidly above the idealized tropopause, consistent with the almost uniform phase of the disturbance illustrated in panel (d) — recall that a tilting of the disturbance with height is necessary for instability. It is these properties that make the Eady problem of more general applicability to the Earth's atmosphere than might be first thought: the high stratification above the tropopause and consequent decay of the instability is mimicked by the imposition of a rigid lid.

In the ocean, the stratification is highest in the upper ocean where the shear is also strongest, and numerical calculations of the structure and growth rate of idealized profiles are illustrated in Fig. 6.22.¹⁷ The solid curve shows the Eady problem, and the various dashed curves show the phase speeds, growth rates and phase with combinations of the profiles illustrated in panel (a). Much of the ocean is characterized by having both a higher shear and a higher stratification in the upper 1 km or so, and this case is shown with the dotted line in Fig. 6.22. The amplitude of the instability is also largely confined to the upper ocean, and unlike the Eady problem it does not arise through the interaction of edge waves at the top and bottom: the potential vorticity changes sign because of the interior variations due to the non-uniform shear, mainly in the upper ocean. Consistently, the phase of the baroclinic waves is nearly constant in the lower ocean in those cases in which the shear is confined to the upper ocean. The real ocean is still more complicated, because the most unstable regions near intense western boundary currents are often also barotropically unstable, and the mean flow itself may be meridionally directed. Nevertheless, the result that linear baroclinic instability is primarily an *upper ocean* phenomenon is quite robust. However, we will find in chapter 9 that the nonlinear evolution of baroclinic instability leads to eddies throughout the water column.

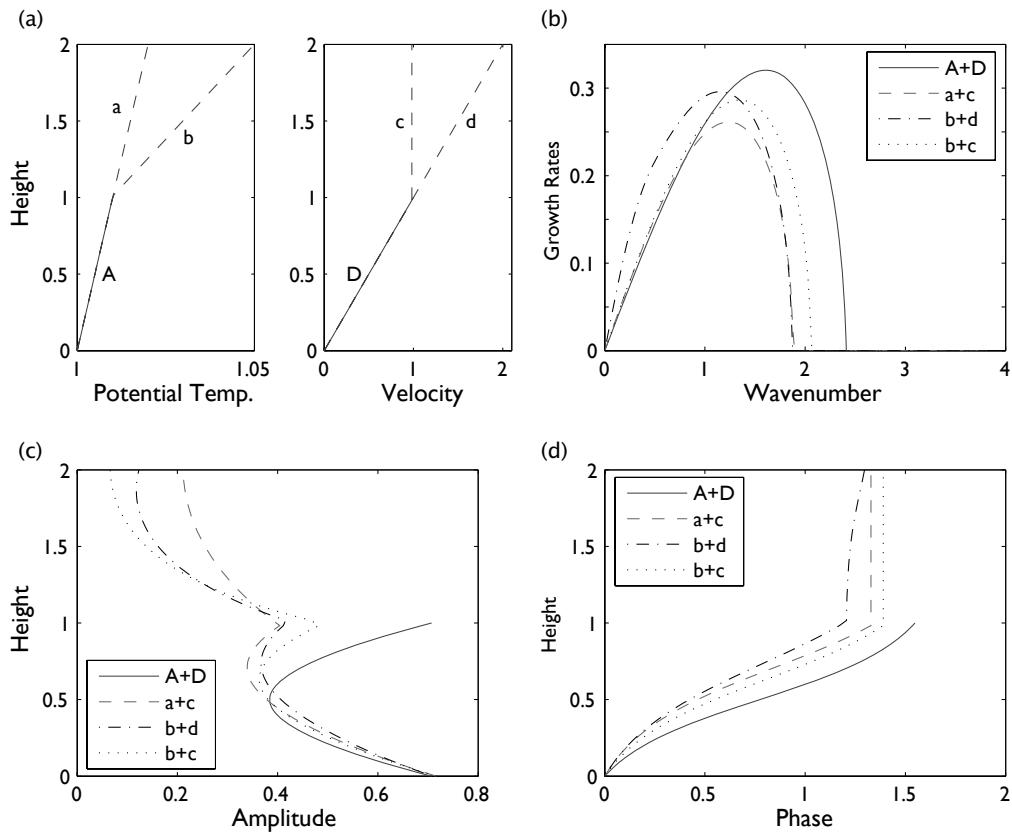


Fig. 6.21 The effect of a stratosphere on baroclinic instability: (a) the given profiles of shear and stratification; (b) the growth rate of the instabilities; (c) the amplitude of the most unstable mode as a function of height; (d) the phase of the most unstable mode. The instability problem is solved numerically with various profiles of stratification and shear. In each profile, in the idealized troposphere ($z < 1$) the shear and stratification are uniform and the same in each case. We consider four idealized stratospheres ($z \geq 1$): 1, A lid at $z = 1$, i.e., no stratosphere, giving the Eady problem itself (profiles A+D, solid lines); 2, stratospheric stratification as for the troposphere, but with zero shear (profiles a+c, dashed); 3, stratospheric shear as for troposphere, but with stratification (N^2) four times the tropospheric value (b+d, dot-dashed); 4, zero shear and high stratification in the stratosphere (b+c, dotted). In the troposphere the amplitude and structure of the instability are similar in all cases, illustrating the similarity of a rigid lid and abrupt changes in shear or stratification. Either a high stratification or a low shear (or both) will result in weak stratospheric instability.

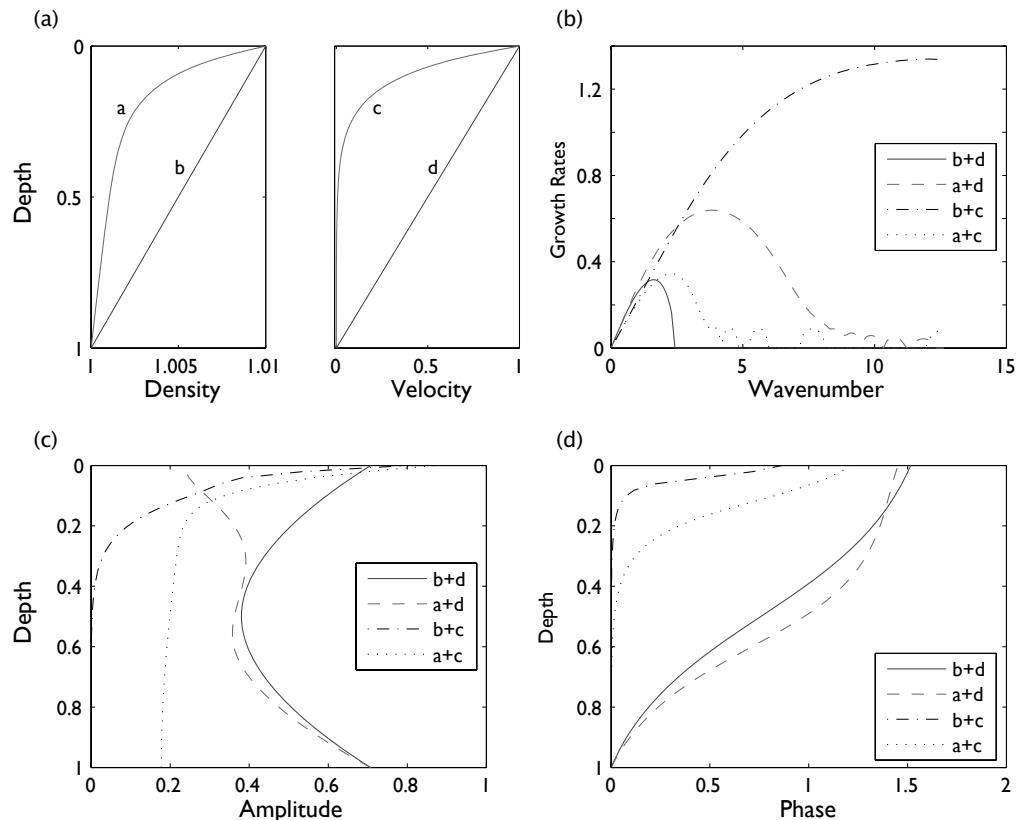


Fig. 6.22 The baroclinic instability in an idealized ocean, with four different profiles of shear or stratification. The panels are: (a) the profiles of velocity and density (and so N^2) used; (b) the growth rates of the various cases; (c) the vertical structure of the amplitude of the most unstable models; (d) the phase in the vertical of the most unstable modes. The instability is calculated numerically with four combinations of shear and stratification: 1, uniform stratification and shear i.e., the Eady problem, (profiles b+d, solid lines); 2, uniform shear, upper-ocean enhanced stratification (a+d, dashed); 3, uniform stratification, upper ocean enhanced shear (b+c, dot-dashed); 4, both stratification and shear enhanced in the upper ocean (a+c, dotted). Case 2 (a+d, dashed) is really more like an atmosphere with a stratosphere (see Fig. 6.21), and the amplitude of the disturbance falls off, rather unrealistically, in the upper ocean. Case 4 (a+c, dotted) is the most oceanographically relevant.

Notes

1 Thomson (1871), Helmholtz (1868). Thomson later became Lord Kelvin.

2 See Drazin & Reid (1981) or Chandrasekhar (1961) for more detail.

3 This is Squire's theorem, which states that for every three-dimensional disturbance to a plane-parallel flow there exists a more unstable two-dimensional one (Squire 1933). This means there is no need to consider three dimensional effects to determine whether such a flow is unstable.

- 4 Rayleigh (1880).
- 5 First obtained by Rayleigh (1894). .
- 6 Rayleigh, Lord (1880) and, for the case with β , Kuo (1949).
- 7 Fjørtoft (1950).
- 8 Charney & Stern (1962), Pedlosky (1964).
- 9 Eady (1949), Charney (1947). Eric Eady (1915–1966) is best remembered today as the author of the iconic ‘Eady model’ of baroclinic instability, which describes the fundamental hydrodynamic instability mechanism that gives rise to weather systems. After an undergraduate education in mathematics he joined the U. K. Meteorological Office in 1937, becoming a forecaster and upper air analyst, in which capacity he served throughout the war. In 1946 he joined the Department of Mathematics at Imperial College, presenting his Ph.D. thesis in 1948 on ‘The theory of development in dynamical meteorology’, subsequently summarized in *Tellus* (Eady 1949). This work, masterly in its combination of austerity and relevance, provides a mathematical description of the essential aspects of cyclone development that stands to this day as a canonical model in the field. It also includes, rather obliquely, a derivation of the stratified quasi-geostrophic equations, albeit in a special form. The impact of the work was immediate and it led to visits to Bergen (in 1947 with J. Bjerknes), Stockholm (in 1952 with C.-G. Rossby) and Princeton (in 1953 with J. von Neumann and Charney). Eady followed his baroclinic instability work with prescient discussions of the general circulation of the atmosphere (Eady 1950, Eady & Sawyer 1951, Eady 1954). A perfectionist who sought to understand it all, Eady’s subsequent published output was small and he later turned his attention to fundamental problems in other areas of fluid mechanics, the dynamics of the sun and the Earth’s interior, and biochemistry. He became increasingly divorced from normal scientific and human discourse, and finally took his own life. There is little published about him, save for the obituary by Charnock *et al.* (1966).

Jule Charney (1917–1981) played a defining role in dynamical meteorology in the second half of the twentieth century. He made seminal contributions in many areas including: the theory of baroclinic instability (Charney 1947); a systematic scaling theory for large-scale atmospheric motions and the derivation of the quasi-geostrophic equations (Charney 1948); a theory of stationary waves in the atmosphere (Charney & Eliassen 1949); the demonstration of the feasibility of numerical weather forecasts (Charney *et al.* 1950); planetary wave propagation into the stratosphere (Charney & Drazin 1961); a criterion for baroclinic instability (Charney & Stern 1962); a theory for hurricane growth (Charney & Eliassen 1964); and the concept of geostrophic turbulence (Charney 1971). His Ph.D. from UCLA in 1946 was entitled ‘Dynamics of long waves in a baroclinic westerly current’ and this became his well-known 1947 paper. After this he spent a year at Chicago and another at Oslo, and in 1948 joined the Institute of Advanced Study in Princeton where he stayed until 1956 (and where Eady visited for a while). He spent most of his subsequent career at MIT, interspersed with many visits to Europe, especially Norway. For a more complete picture of Charney, see Lindzen *et al.* (1990) and a brief biography by N. Phillips, available at <http://www.nap.edu/readingroom/books/biomems/jcharney.html>.

- 10 At least I find it so. My treatment of the Eady problem draws from unpublished notes by J.S.A. Green, as well as from Eady (1949) itself.
- 11 If c is real (and the waves are neutral), then there exists the possibility that $\Lambda z - c = 0$, and the equation for Φ is

$$\frac{d^2\Phi}{dz^2} - \mu^2\Phi = C\delta(z - z_c), \quad z_c = c/\Lambda. \quad (6.180)$$

where C is a constant. Because z_c is continuous in the interval $[0, 1]$ so is c , and these solutions have a *continuous spectrum* of eigenvalues. The associated eigenfunctions provide formal completeness to the normal modes, enabling any function to be represented as their superposition.

- 12 After Phillips (1954).
- 13 Our non-dimensionalization of the two-layer system is such as to be in correspondence with that for the continuous system. Thus we choose H to be the total depth of the domain. This choice produces growth rates and wavenumbers that are equivalent to those in the Eady problem.
- 14 Green (1960) and Branscome (1983). Lindzen & Farrell (1980) also provide an approximate calculation of growth rates in the Charney problem.
- 15 After Charney (1947), in whose problem it appears, and Green (1960), who appreciated its importance.
- 16 Our numerical procedure is to assume a wave-like solution in the horizontal direction of the form $\tilde{\psi} \exp[i(kx + ly - \omega t)]$, and to finite difference the equations in the vertical direction. The resulting eigenvalue equations are solved by standard matrix methods, for each horizontal wavenumber. See Smith & Vallis (1998).
- 17 Gill *et al.* (1974) and Robinson & McWilliams (1974) were among the first to look at baroclinic instability in the ocean.

Further reading

Chandrasekhar, S., 1961. *Hydrodynamic and Hydromagnetic Stability*.

A classic text discussing many forms of instability but not, alas, baroclinic instability.

Drazin P. & Reid, W. H., 1981. *Hydrodynamic Stability*.

A standard text on hydrodynamic instability theory. It discusses nearly all the classic cases in a straightforward and clear fashion. It includes a more extensive discussion of the linear instability of parallel shear flow than is contained here, although the treatment of baroclinic instability is rather brief.

Pierrehumbert & Swanson (1995) review many aspects of baroclinic instability.

Problems

- 6.1 Derive the jump condition (6.29) without directly considering the motion of the interface. In particular, from the momentum equation along the interface show that

$$\frac{\partial}{\partial y} \left(\frac{\tilde{\psi}}{U - c} \right) = - \frac{\tilde{p}}{(U - c)^2} \quad (\text{P6.1})$$

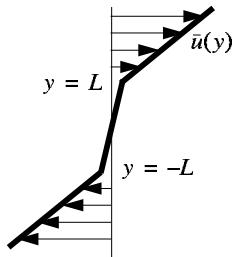
and show that (6.29) follows. Be explicit about the conditions under which the right-hand side vanishes when integrated across the interface. (For help see Drazin & Reid 1981).

- 6.2 By applying the matching conditions (6.23) and (6.29) at $y = \pm a$ to Rayleigh's equation, explicitly derive the dispersion relationship (6.42).
- 6.3 Show that for very long waves, or as the shear layer becomes thinner, the growth rate given by (6.42) reduces to that of Kelvin-Helmholtz instability of a vortex sheet.
- 6.4 Obtain the stability properties of the triangular jet, with a basic state velocity given by

$$U(y) = \begin{cases} 0 & \text{for } y \geq 1 \\ 1 - |y| & \text{for } -1 \leq y \leq 1 \\ 0 & \text{for } y \leq -1. \end{cases} \quad (\text{P6.2})$$

In particular, obtain the eigenfunctions and eigenvalues of the problem, and show that each eigenfunction is either even or odd. Perturbations with even ψ' are known as ‘sinuous modes’ and those with odd ψ' are ‘varicose modes’. Show that sinuous waves are unstable for sufficiently long wavelengths in the y -direction, but that all varicose modes are stable.

- 6.5 Obtain the stability properties of the jet illustrated in Fig. 6.3, but in the presence of a β -effect [using, for example, Rayleigh’s equation (6.20)]. Repeat the problem for a jet going in the opposite direction. Are your answers are consistent with the Rayleigh–Kuo criterion?
- 6.6 Consider the incompressible piecewise linear shear flow below:



$$u = \begin{cases} AL + B(y - L), & y \geq L \\ Ay, & -L \leq y \leq L \\ -AL + B(y + L), & y \leq -L. \end{cases} \quad (\text{P6.3})$$

The flow is two-dimensional, and A and B are constants with $B > 0$.

- (a) Find the two normal-mode frequencies as a function of zonal wavenumber k .
 (b) Find the stability boundaries in terms of k and A and provide a physical interpretation. If $A = B$ is the flow stable or unstable? Why?
 (If the algebra defeats you, explain carefully the method for solving the problem.)

- 6.7 Show numerically or analytically that, in the Eady problem: (a) instability occurs for $\mu < 2.399$; (b) the wavenumber at which the instability is greatest is $\mu = 1.61$; (c) the non-dimensional growth rate at that wavenumber is 0.31.

6.8 ♦ *Vertical modes of continuously stratified problems*

- (a) When solving the continuous form of the eigenvalue stability problem (as in the Eady problem, for example) the differential equation typically seems to have just one pair of eigenvalues. However, if the equation is solved on a vertical grid with N levels, the resulting difference equation has N roots. Does this mean that $N - 2$ roots are spurious, and if so how might the ‘correct’ eigenvalues be identified? Alternatively, are there corresponding additional roots in the continuously stratified problem?
 (b) The N -level problem is equivalent to a physically realizable N -layer system, in which there are presumably N physically meaningful eigenvalues. As N becomes large, with the density differences and thicknesses of each layer chosen to become smaller in a consistent way, the equations describing the layered system presumably converge to those describing the continuous system, yet there are N eigenvalues in the former. What is the physical nature of the eigenvalues in the layered system, and how do they relate to those of the continuous system?

- 6.9 Show, using the two-layer model (or otherwise) that the presence of β reduces the efficiency of baroclinic instability. For example, show that it makes the meridional velocity slightly out of phase with the temperature.

- 6.10 ♦ Consider the baroclinic instability problem with a discontinuity in the stratification, but a uniform shear. For example, suppose the shear is uniform for $z \in (0, 1)$ with an abrupt change in stratification at $z = 0.5$. How does the amplitude of the instability vary on either side of the discontinuity? Your answer may be an analytical or a numerical calculation, or both.

Another advantage of a mathematical statement is that it is so definite that it might be definitely wrong... Some verbal statements have not this merit.

L. F. Richardson (1881–1953).

CHAPTER SEVEN

Wave–Mean Flow Interaction

WAVE–MEAN FLOW INTERACTION is concerned with how some mean flow, perhaps a time or zonal average, interacts with a departure from that mean, and this chapter provides an elementary introduction to a number of topics in this area. It is ‘elementary’ because our derivations and discussion are obtained by direct and straightforward manipulations of the equations of motion, often in the simplest case that will illustrate the relevant principle. It is implicit in what we do that it is a sensible thing to decompose the fields into a mean plus some departure, and one case when this is so is when the departure is of small amplitude. Departures from the mean — generically called *eddies* — are of course not always small; for example, in the mid-latitude troposphere the eddies are often of similar amplitude to the mean flow, and chapters 9 and 10 will explore this from the standpoint of turbulence. However, in this chapter we will generally assume that eddies are indeed of small amplitude, and, in particular, that eddy–mean flow interaction is larger than eddy–eddy interaction.

A *wave* is an eddy that satisfies, at least approximately, a dispersion relation. It is the presence of such a dispersion relation that enables a number of results to be obtained that would otherwise be out of our reach, and that gives rise to the appellation ‘wave–mean flow’. In mid-latitudes the relevant waves are usually Rossby waves, as introduced in chapter 5, although gravity waves also interact with the mean flow. It is implicit in defining waves this way that they are generally of small amplitude, for it is this that allows the equations of motion to be sensibly linearized and a dispersion relation to be obtained (although a monochromatic wave may have finite amplitude and still satisfy a dispersion relation). However, this does not mean that the waves do not interact with each other and with the mean flow; we may expect, or at least hope, that the qualitative nature of such interactions, as calculated by wave–mean flow interaction theory, will carry over and provide insights

into the finite-amplitude problem. Thus, one goal of this chapter is to provide a way of qualitatively understanding more realistic situations, and to suggest diagnostics that might be used to analyse both observations and numerical solutions of the fully nonlinear problem. We will almost exclusively concern ourselves with a *zonal* mean, since this is the simplest and often most useful case because of the presence of simple — i.e., periodic — boundary conditions. (With care some of our results can be extended to the case of a temporal mean.) We will also be mainly concerned with quasi-geostrophic dynamics on a β -plane.

7.1 QUASI-GEOSTROPHIC PRELIMINARIES

To fix our dynamical system and our notation, we write down the quasi-geostrophic potential vorticity equation

$$\frac{\partial q}{\partial t} + J(\psi, q) = D, \quad (7.1)$$

where D represents any non-conservative terms and the potential vorticity in a Boussinesq system is

$$q = \beta y + \zeta + \frac{\partial}{\partial z} \left(\frac{f_0}{N^2} b \right), \quad (7.2)$$

where ζ is the relative vorticity and b is the buoyancy perturbation from the background state characterized by N^2 . [In an ideal gas $q = \beta y + \zeta + (f_0/\rho_s) \partial_z (\rho_s b/N^2)$, and most of our derivations can be extended to that case.] We will refer to lines of constant b as isentropes. In terms of the streamfunction, the variables are

$$\zeta = \nabla^2 \psi, \quad b = f_0 \frac{\partial \psi}{\partial z}, \quad q = \beta y + \left[\nabla^2 + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial}{\partial z} \right) \right] \psi. \quad (7.3)$$

where $\nabla^2 \equiv (\partial_x^2 + \partial_y^2)$. The potential vorticity equation holds in the fluid interior; the boundary conditions on (7.3) are provided by the thermodynamic equation

$$\frac{\partial b}{\partial t} + J(\psi, b) + w N^2 = H, \quad (7.4)$$

where H represents heating terms. The vertical velocity at the boundary, w , is zero in the absence of topography and Ekman friction, and if H is also zero the boundary condition is just:

$$\frac{\partial b}{\partial t} + J(\psi, b) = 0. \quad (7.5)$$

Equations (7.1) and (7.5) are the evolution equations for the system and if both D and H are zero they conserve both the total energy, \hat{E} and the total enstrophy, \hat{Z} :

$$\begin{aligned} \frac{d\hat{E}}{dt} &= 0, & \hat{E} &= \frac{1}{2} \int_V (\nabla \psi)^2 + \frac{f_0^2}{N^2} \left(\frac{\partial \psi}{\partial z} \right)^2 dV, \\ \frac{d\hat{Z}}{dt} &= 0, & \hat{Z} &= \frac{1}{2} \int_V q^2 dV. \end{aligned} \quad (7.6)$$

where V is a volume bounded by surfaces at which the normal velocity is zero, or that has periodic boundary conditions. The enstrophy is also conserved layerwise; that is, the horizontal integral of q^2 is conserved at every level.

7.1.1 Potential vorticity flux in the linear equations

Let us decompose the fields into a mean (to be denoted with an overbar) plus a perturbation (denoted with a prime), and let us suppose the perturbation fields are of small amplitude. (In linear problems, such as those considered in chapter 6, we decomposed the flow into a ‘basic state’ plus a perturbation, with the basic state fixed in time. Our approach here is similar, but later we allow the mean state to evolve.) If the mean is temporally steady then the linearized quasi-geostrophic potential vorticity equation is

$$\frac{\partial q'}{\partial t} + \bar{u} \frac{\partial q'}{\partial x} + u' \frac{\partial \bar{q}}{\partial x} + \bar{v} \frac{\partial q'}{\partial y} + v' \frac{\partial \bar{q}}{\partial y} = D'. \quad (7.7)$$

where D' represents eddy forcing and dissipation. If the mean is a zonal mean then $\bar{v} = 0$ (because v is purely geostrophic) and (7.7) simplifies to

$$\frac{\partial q'}{\partial t} + \bar{u} \frac{\partial q'}{\partial x} + v' \frac{\partial \bar{q}}{\partial y} = D', \quad (7.8)$$

where

$$\frac{\partial \bar{q}}{\partial y} = \beta - \frac{\partial^2 \bar{u}}{\partial y^2} - \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial \bar{u}}{\partial z} \right), \quad (7.9)$$

and we take this to be non-zero. Multiplying by q' and zonally averaging gives the enstrophy equation:

$$\frac{1}{2} \frac{\partial}{\partial t} \overline{q'^2} = -\bar{v}' q' \frac{\partial \bar{q}}{\partial y} + \overline{D' q'}. \quad (7.10)$$

The quantity $\overline{v' q'}$ is the meridional flux of potential vorticity; this is downgradient (by definition) when the first term on the right-hand side is positive (i.e., $\overline{v' q' \partial \bar{q} / \partial y} < 0$), and it then acts to increase the variance of the perturbation. (This occurs, for example, when the flux is diffusive so that $\overline{v' q'} = -\kappa \partial \bar{q} / \partial y$, where κ may vary but is everywhere positive.) This argument may be inverted: for inviscid flow ($D = 0$), if the waves are growing, as for example in the canonical models of baroclinic instability discussed in chapter 6, then *the potential vorticity flux is downgradient*.

If the second term on the right-hand side of (7.10) is negative, as it will be if D' is a dissipative process (e.g., if $D' = A \nabla^2 q'$ or if $D' = -r q'$, where A and r are positive) then a statistical balance can be achieved between enstrophy production via downgradient transport, and dissipation. If the waves are steady (by which we mean statistically steady, neither growing nor decaying in amplitude) and conservative (i.e., $D' = 0$) then we must have:

$$\overline{v' q'} = 0. \quad (7.11)$$

Similar results follow for the buoyancy at the boundary; we start by linearizing the thermodynamic equation (7.5) to give

$$\frac{\partial b'}{\partial t} + \bar{u} \frac{\partial b'}{\partial x} + v' \frac{\partial \bar{b}}{\partial y} = H', \quad (7.12)$$

where H' is a diabatic source term. Multiplying (7.12) by b' and averaging gives

$$\frac{1}{2} \frac{\partial}{\partial t} \overline{b'^2} = -\bar{v}' b' \frac{\partial \bar{b}}{\partial y} + \overline{H' b'}. \quad (7.13)$$

Thus growing adiabatic waves have a downgradient flux of buoyancy at the boundary. In the Eady problem there is no interior gradient of basic-state potential vorticity and all the terms in (7.10) are zero, but the perturbation grows at the boundary. If the waves are steady and adiabatic then, analogously to (7.11),

$$\overline{v' b'} = 0. \quad (7.14)$$

The boundary conditions and fluxes may be absorbed into the interior definition of potential vorticity and its fluxes by way of the delta-function boundary layer construction, described in section 5.4.3.

7.2 THE ELIASSEN-PALM FLUX

The eddy flux of potential vorticity may be expressed in terms of vorticity and buoyancy fluxes as

$$v' q' = v' \zeta' + f_0 v' \frac{\partial}{\partial z} \left(\frac{b'}{N^2} \right). \quad (7.15)$$

The second term on the right-hand side can be written as

$$\begin{aligned} f_0 v' \frac{\partial}{\partial z} \left(\frac{b'}{N^2} \right) &= f_0 \frac{\partial}{\partial z} \left(\frac{v' b'}{N^2} \right) - f_0 \frac{\partial v'}{\partial z} \frac{b'}{N^2} \\ &= f_0 \frac{\partial}{\partial z} \left(\frac{v' b'}{N^2} \right) - f_0 \frac{\partial}{\partial x} \left(\frac{\partial \psi'}{\partial z} \right) \frac{b'}{N^2} \\ &= f_0 \frac{\partial}{\partial z} \left(\frac{v' b'}{N^2} \right) - \frac{f_0^2}{2N^2} \frac{\partial}{\partial x} \left(\frac{\partial \psi'}{\partial z} \right)^2, \end{aligned} \quad (7.16)$$

using $b' = f_0 \partial \psi' / \partial z$.

Similarly, the flux of relative vorticity can be written

$$v' \zeta' = -\frac{\partial}{\partial y} (u' v') + \frac{1}{2} \frac{\partial}{\partial x} (v'^2 - u'^2) \quad (7.17)$$

Using (7.16) and (7.17), (7.15) becomes

$$v' q' = -\frac{\partial}{\partial y} (u' v') + \frac{\partial}{\partial z} \left(\frac{f_0}{N^2} v' b' \right) + \frac{1}{2} \frac{\partial}{\partial x} \left((v'^2 - u'^2) - \frac{b'^2}{N^2} \right). \quad (7.18)$$

Thus the meridional potential vorticity flux, in the quasi-geostrophic approximation, can be written as the divergence of a vector: $v' q' = \nabla \cdot \mathbf{F}$ where

$$\mathbf{F} \equiv \frac{1}{2} \left((v'^2 - u'^2) - \frac{b'^2}{N^2} \right) \mathbf{i} - (u' v') \mathbf{j} + \left(\frac{f_0}{N^2} v' b' \right) \mathbf{k}. \quad (7.19)$$

A particularly useful form of this arises after zonally averaging, for then (7.18) becomes

$$\overline{v' q'} = -\frac{\partial}{\partial y} \overline{u' v'} + \frac{\partial}{\partial z} \left(\frac{f_0}{N^2} \overline{v' b'} \right). \quad (7.20)$$

The vector defined by

$$\mathbf{F} \equiv -\overline{u'v'}\mathbf{j} + \frac{f_0}{N^2}\overline{v'b'}\mathbf{k} \quad (7.21)$$

is called the (quasi-geostrophic) *Eliassen-Palm (EP) flux*,¹ and its divergence, given by (7.20), gives the poleward flux of potential vorticity:

$$\overline{v'q'} = \nabla_x \cdot \mathbf{F}, \quad (7.22)$$

where $\nabla_x \cdot \equiv (\partial/\partial y, \partial/\partial z) \cdot$ is the divergence in the meridional plane. Unless the meaning is unclear, the subscript x on the meridional divergence will be dropped.

7.2.1 The Eliassen-Palm relation

On dividing by $\partial \bar{q}/\partial y$ and using (7.22), the enstrophy equation (7.10) becomes

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla \cdot \mathbf{F} = \mathcal{D},$$

(7.23a)

where

$$\mathcal{A} = \frac{\overline{q'^2}}{2\partial \bar{q}/\partial y}, \quad \mathcal{D} = \frac{\overline{D'q'}}{\partial \bar{q}/\partial y}. \quad (7.23b)$$

Equation (7.23a) is known as the *Eliassen-Palm relation*, and it is a conservation law for the *wave activity density* \mathcal{A} . If we integrate (7.23b) over a meridional area A bounded by walls where the eddy activity vanishes, and if $\mathcal{D} = 0$, we obtain

$$\frac{d}{dt} \int_A \mathcal{A} dA = 0. \quad (7.24)$$

The integral is a wave activity — a quantity that is quadratic in the amplitude of the perturbation and that is conserved in the absence of forcing and dissipation. More specifically, \mathcal{A} is the negative of the *pseudomomentum*, for reasons we will encounter later. Note that neither the perturbation energy nor the perturbation enstrophy are wave activities of the linearized equations, because there can be an exchange of energy or enstrophy between mean and perturbation — indeed, this is how a perturbation grows in baroclinic or barotropic instability! This is already evident from (7.10), or in general take (7.7) with $D' = 0$ and multiply by q' to give the enstrophy equation

$$\frac{1}{2} \frac{\partial q'^2}{\partial t} + \frac{1}{2} \overline{\mathbf{u} \cdot \nabla q'^2} + \overline{\mathbf{u}' q' \cdot \nabla \bar{q}} = 0, \quad (7.25)$$

where here the overbar is an average (although it need not be a zonal average). Integrating this over a volume V gives

$$\frac{d\hat{Z}'}{dt} \equiv \frac{d}{dt} \int_V \frac{1}{2} q'^2 dV = - \int_V \overline{\mathbf{u}' q' \cdot \nabla \bar{q}} dV. \quad (7.26)$$

The right-hand side does not, in general, vanish and so \hat{Z}' is not in general conserved.

7.2.2 The group velocity property

The vector \mathcal{F} describes how the wave activity propagates. In general, we cannot express it simply in terms of \mathcal{A} , but in the case in which the disturbance is composed of plane or almost plane waves that satisfy a dispersion relation, then $\mathcal{F} = \mathbf{c}_g \mathcal{A}$, where \mathbf{c}_g is the group velocity and (7.23a) becomes

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla \cdot (\mathcal{A} \mathbf{c}_g) = 0. \quad (7.27)$$

This is a useful property, because if we can diagnose \mathbf{c}_g from observations we can use (7.23a) to determine how wave activity density propagates. We shall first demonstrate this when the waves in question are plane Rossby waves, and then give a more general derivation.

Group velocity property for Rossby waves

The Boussinesq quasi-geostrophic equation on the β -plane, linearized around a uniform zonal flow and with constant static stability, is

$$\frac{\partial q'}{\partial t} + \bar{u} \frac{\partial q'}{\partial x} + v' \frac{\partial \bar{q}}{\partial y} = 0, \quad (7.28)$$

where $q' = [\nabla^2 + (f_0^2/N^2)\partial^2/\partial z^2]\psi'$ and, if \bar{u} is constant, $\partial \bar{q}/\partial y = \beta$. Thus we have

$$\left(\frac{\partial}{\partial t} + \bar{u} \frac{\partial}{\partial x} \right) \left[\nabla^2 \psi' + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial \psi'}{\partial z} \right) \right] + \beta \frac{\partial \psi'}{\partial x} = 0. \quad (7.29)$$

Seeking solutions of the form

$$\psi' = \text{Re } \tilde{\psi} e^{i(kx + ly + mz - \omega t)}, \quad (7.30)$$

we find the dispersion relation,

$$\omega = \bar{u}k - \frac{\beta k}{\kappa^2}. \quad (7.31)$$

where $\kappa^2 = (k^2 + l^2 + m^2 f_0^2/N^2)$, and the group velocity components:

$$c_g^y = \frac{2\beta kl}{\kappa^4}, \quad c_g^z = \frac{2\beta km f_0^2/N^2}{\kappa^4}. \quad (7.32)$$

Also, if $u' = \text{Re } \tilde{u} \exp[i(kx + ly + mz - \omega t)]$, and similarly for the other fields, then

$$\begin{aligned} \tilde{u} &= -\text{Re } il\tilde{\psi}, & \tilde{v} &= \text{Re } ik\tilde{\psi}, \\ \tilde{b} &= \text{Re } imf_0\tilde{\psi}, & \tilde{q} &= -\text{Re } \kappa^2 \tilde{\psi}, \end{aligned} \quad (7.33)$$

The wave activity density is then

$$\mathcal{A} = \frac{1}{2} \frac{\overline{q'^2}}{\beta} = \frac{\kappa^4}{4\beta} |\tilde{\psi}^2|, \quad (7.34)$$

where the additional factor of 2 in the denominator arises from the averaging. Using (7.33) the EP flux, (7.21), is

$$\mathcal{F}^y = -\overline{u'v'} = \frac{1}{2} kl |\tilde{\psi}^2|, \quad \mathcal{F}^z = \frac{f_0}{N^2} \overline{v'b'} = \frac{f_0^2}{2N^2} km |\tilde{\psi}^2|. \quad (7.35)$$

Using (7.32), (7.34) and (7.35) we obtain

$$\mathbf{F} = (\mathcal{F}^y, \mathcal{F}^z) = \mathbf{c}_g \mathcal{A}. \quad (7.36)$$

If the properties of the medium are slowly varying, so that a (spatially varying) group velocity can still be defined, then this is a useful expression to estimate how the wave activity propagates in the atmosphere and in numerical simulations.

** Group velocity property: a general derivation*

We now demonstrate the group velocity property in a rather more general way.² Our derivation holds generally for waves and wave activities that satisfy the following three assumptions.

- (i) The wave activity, A , and flux, \mathbf{F} , (which are not necessarily pseudomomentum and EP flux) obey the general conservation relation

$$\frac{\partial A}{\partial t} + \nabla \cdot \mathbf{F} = 0. \quad (7.37)$$

(ii) Both the wave activity and the flux are quadratic functions of the wave amplitude.

(iii) The waves themselves are of the general form

$$\psi = \tilde{\psi} e^{i\theta(\mathbf{x}, t)} + \text{c.c.}, \quad \theta = \mathbf{k} \cdot \mathbf{x} - \omega t, \quad \omega = \omega(\mathbf{k}), \quad (7.38a,b,c)$$

where (7.38c) is the dispersion relation, and ψ is any wave field. We will restrict attention to the case in which $\tilde{\psi}$ is a constant, but the derivation may be extended to the case in which it varies slowly over a wavelength.

Given assumption (ii), the wave activity must have the general form

$$A = b + a e^{2i(\mathbf{k} \cdot \mathbf{x} - \omega t)} + a^* e^{-2i(\mathbf{k} \cdot \mathbf{x} - \omega t)}, \quad (7.39a)$$

where the asterisk, $*$, denotes complex conjugacy, and b is a real constant and a is a complex constant. For example, suppose that $A = \psi^2$ and $\psi = c e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} + c^* e^{-i(\mathbf{k} \cdot \mathbf{x} - \omega t)}$, then we find that (7.39a) is satisfied with $a = c^2$ and $b = 2cc^*$. Similarly, the flux has the general form

$$\mathbf{F} = \mathbf{g} + \mathbf{f} e^{2i(\mathbf{k} \cdot \mathbf{x} - \omega t)} + \mathbf{f}^* e^{-2i(\mathbf{k} \cdot \mathbf{x} - \omega t)}. \quad (7.39b)$$

where \mathbf{g} is a real constant vector and \mathbf{f} is a complex constant vector. The mean activity and mean flux are obtained by averaging over a cycle; the oscillating terms vanish on integration and therefore the wave activity and flux are given by

$$\bar{A} = b, \quad \bar{\mathbf{F}} = \mathbf{g}, \quad (7.40)$$

where the overbar denotes the mean.

Now formally consider a wave with a slightly different phase, $\theta + i\delta\theta$, where $\delta\theta$ is small compared with θ . Thus, we formally replace \mathbf{k} by $\mathbf{k} + i\delta\mathbf{k}$ and ω by $\omega + i\delta\omega$ where, to satisfy the dispersion relation, we have

$$\omega + i\delta\omega = \omega(\mathbf{k} + i\delta\mathbf{k}) \approx \omega(\mathbf{k}) + i\delta\mathbf{k} \cdot \frac{\partial\omega}{\partial\mathbf{k}}, \quad (7.41)$$

and therefore

$$\delta\omega = \boldsymbol{\delta k} \cdot \frac{\partial \omega}{\partial \mathbf{k}} = \boldsymbol{\delta k} \cdot \mathbf{c}_g, \quad (7.42)$$

where $\mathbf{c}_g \equiv \partial\omega/\partial\mathbf{k}$ is the group velocity.

The new wave has the general form

$$\psi' = (\tilde{\psi} + \delta\tilde{\psi}) e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} e^{-\delta\mathbf{k} \cdot \mathbf{x} + \delta\omega t} + \text{c.c.}, \quad (7.43)$$

and, analogously to (7.39), the associated wave activity and flux have the forms:

$$A' = [b + \delta b + (a + \delta a) e^{2i(\mathbf{k} \cdot \mathbf{x} - \omega t)} + (a^* + \delta a^*) e^{-2i(\mathbf{k} \cdot \mathbf{x} - \omega t)}] e^{-2\delta\mathbf{k} \cdot \mathbf{x} + 2\delta\omega t} \quad (7.44a)$$

$$F' = [\mathbf{g} + \delta\mathbf{g} + (\mathbf{f} + \delta\mathbf{f}) e^{2i(\mathbf{k} \cdot \mathbf{x} - \omega t)} + (\mathbf{f}^* + \delta\mathbf{f}^*) e^{-2i(\mathbf{k} \cdot \mathbf{x} - \omega t)}] e^{-2\delta\mathbf{k} \cdot \mathbf{x} + 2\delta\omega t}, \quad (7.44b)$$

where the δ quantities are small. If we now demand that A' and F' satisfy assumption (i), then substituting (7.44) into (7.37) gives, after a little algebra,

$$(\mathbf{g} + \delta\mathbf{g}) \cdot \boldsymbol{\delta k} = (b + \delta b) \delta\omega \quad (7.45)$$

and therefore at first order in δ quantities, $\mathbf{g} \cdot \boldsymbol{\delta k} = b \delta\omega$. Using (7.42) and (7.40) we obtain

$$\mathbf{c}_g = \frac{\mathbf{g}}{b} = \frac{\overline{\mathbf{F}}}{\overline{A}}, \quad (7.46)$$

and using this the conservation law, (7.37), becomes

$$\frac{\partial \overline{A}}{\partial t} + \nabla \cdot (\mathbf{c}_g \overline{A}) = 0. \quad (7.47)$$

Thus, for waves satisfying our three assumptions, the flux velocity — that is, the propagation velocity of the wave activity — is equal to the group velocity.

7.2.3 * The orthogonality of modes

It is a direct consequence of the conservation of wave activity that disturbance modes are orthogonal in the ‘wave activity norm’, defined later on, and thus are a useful measure of the amplitude of a particular mode.³ To explore this, we start with the linearized potential vorticity equation,

$$\frac{\partial q'}{\partial t} + \overline{u} \frac{\partial q'}{\partial x} + \nu' \frac{\partial \overline{q}}{\partial y} = 0. \quad (7.48)$$

Let us formally seek solutions of the form $\psi' = \text{Re } \Psi \exp(i k x)$ where Ψ is the sum of *modes*,

$$\Psi = \sum_n \tilde{\psi}_n(\gamma, z) e^{-ikc_n t}, \quad (7.49)$$

where n is an identifier of the modes. The modes satisfy

$$(\overline{u} \Delta_k^2 + \overline{q}_y) \tilde{\psi}_n = c_n \Delta_k^2 \tilde{\psi}_n, \quad (7.50)$$

where

$$\Delta_k^2 = \frac{\partial^2}{\partial \gamma^2} + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial}{\partial z} \right) - k^2. \quad (7.51)$$

The upper and lower boundary conditions (at $z = 0, -H$) are given by the thermodynamic equation

$$\frac{\partial b'}{\partial t} + \bar{u} \frac{\partial b'}{\partial x} + v' \frac{\partial \bar{b}}{\partial y} = 0, \quad (7.52)$$

and if we simplify further by supposing $\partial \bar{u} / \partial z = 0$ then the boundary condition becomes

$$\frac{\partial \psi'_z}{\partial t} + \bar{u} \frac{\partial \psi'_z}{\partial x} = 0. \quad (7.53)$$

There are no meridional buoyancy fluxes at the boundary. If N^2 is a constant (a simplifying but not essential assumption) then we can let $\tilde{\psi}_n(y, z) = \psi_n(y) \cos pz$, with $p = j\pi/H$ where j is an integer and the mode n now labels only the meridional modes. The corresponding potential vorticity modes are given by

$$q_n = \Delta_{k,m}^2 \psi_n, \quad \Delta_{k,m}^2 = \frac{\partial^2}{\partial y^2} - (f_0^2/N^2)m^2 - k^2, \quad (7.54)$$

and the boundary conditions are then built in to any solution we construct from (7.50) and (7.54).⁴ We may then consider a single zonal and a single vertical wavenumber. (If there is no horizontal variation of the shear, the meridional modes are harmonic functions, for example $\psi_n \propto \sin(n\pi y/L)$ for a channel of width L .)

For a given basic state we may imagine solving (7.50), numerically or analytically, and determining the modes. However, these modes are not orthogonal in the sense of either energy or enstrophy. That is, denoting the inner product by

$$\langle a, b \rangle \equiv (2L)^{-1} \int_L ab \, dy, \quad (7.55)$$

then, in general,

$$I_E = \langle \psi_n, q_m \rangle \neq 0, \quad I_Z = \langle q_n, q_m \rangle \neq 0, \quad (7.56a,b)$$

for $n \neq m$, where $q_n = \Delta_{k,p}^2 \psi_n$. Perturbation energy and enstrophy are thus not wave activities of the linearized equations, and it is not meaningful to talk about the energy or enstrophy of a particular mode. However, by the same token we may expect orthogonality in the wave activity norm. To prove this and understand what it means, suppose that at $t = 0$ the disturbance consists of two modes, n and m , so that at a later time $q = (q_n e^{-ikc_n t} + q_m e^{-ikc_m t} + \text{c.c.})$, where $c_m \neq c_n$ and we assume that both are real. The wave activity is

$$P \equiv \int \mathcal{A} \, dy \, dz = \left\langle q_n, q_m^* / \bar{q}_y \right\rangle e^{-ik(c_n - c_m)t} + \left\langle q_m, q_m^* / \bar{q}_y \right\rangle + \left\langle q_n, q_n^* / \bar{q}_y \right\rangle + \text{c.c.} \quad (7.57)$$

The second and third terms on the right-hand side are the wave activities of each mode, and these are constants (to see this, consider the case when the disturbance is just a single mode). Now, because $dP/dt = 0$ the first term must vanish if $c_n \neq c_m$, implying the modes are orthogonal and, in particular,

$$\text{Re} \int \frac{1}{\bar{q}_y} q_n q_m^* \, dy = 0, \quad (7.58)$$

for $n \neq m$. The inner product weighted by $1/\bar{q}_y$ defines the wave activit norm. Readers

who would prefer a more direct derivation of the orthogonality condition directly from the eigenvalue equation (7.50) should see problem 7.3. Orthogonality is a useful result, for it means that the wave activity is a proper measure of the amplitude of a given mode unlike, for example, energy. The conservation of wave activity will lead to a derivation of the necessary conditions for stability, in section 7.6.

7.3 THE TRANSFORMED EULERIAN MEAN

The so-called *transformed Eulerian mean*, or TEM, is a transformation of the equation of motion that provides a useful framework for discussing eddy effects under a wide range of conditions.⁵ It is useful because, as we shall see, it is equivalent to a very natural form of averaging the equations that serves to eliminate eddy fluxes in the thermodynamic equation and collect them together, in a simple form, in the momentum equation, highlighting the role of potential vorticity fluxes. It also provides a natural separation between diabatic and adiabatic effects or between advective and diffusive fluxes and, in the case in which the flow is adiabatic, a pleasing simplification of the equations. In later chapters we will use the TEM to better understand the mid-latitude troposphere and the dynamics of the Antarctic Circumpolar Current, and as a framework for the parameterization of eddy fluxes. Of course, there being no free lunch, the TEM brings with it its own difficulties, and in particular the implementation of boundary conditions can cause difficulties, especially in the actual numerical integration of the TEM equations.

7.3.1 Quasi-geostrophic form

For simplicity we will use the Boussinesq equations on the beta-plane, and the zonally averaged Eulerian mean equations for the zonally averaged zonal velocity and the buoyancy may then be written as (see section 2.2.6)

$$\frac{\partial \bar{u}}{\partial t} - (f + \bar{\zeta})\bar{v} + \bar{w}\frac{\partial \bar{u}}{\partial z} = -\frac{\partial}{\partial y}\bar{u}'\bar{v}' - \frac{\partial}{\partial z}\bar{u}'\bar{w}' + \bar{F}, \quad (7.59a)$$

$$\frac{\partial \bar{b}}{\partial t} + \bar{w}\frac{\partial \bar{b}}{\partial z} = -\frac{\partial}{\partial y}\bar{v}'\bar{b}' - \frac{\partial}{\partial z}\bar{w}'\bar{b}' + \bar{S}, \quad (7.59b)$$

where \bar{F} and \bar{S} represent frictional and heating terms, respectively. Note that the meridional velocity, \bar{v} , is purely ageostrophic. Using quasi-geostrophic scaling we neglect the vertical eddy flux divergences and all ageostrophic velocities except when multiplied by f_0 or N^2 . The above equations then become

$$\frac{\partial \bar{u}}{\partial t} = f_0\bar{v} - \frac{\partial}{\partial y}\bar{u}'\bar{v}' + \bar{F}, \quad (7.60a)$$

$$\frac{\partial \bar{b}}{\partial t} = -N^2\bar{w} - \frac{\partial}{\partial y}\bar{v}'\bar{b}' + \bar{S}. \quad (7.60b)$$

These two equations are connected by the thermal wind relation,

$$f_0\frac{\partial \bar{u}}{\partial z} = -\frac{\partial \bar{b}}{\partial y}, \quad (7.61)$$

which is a combination of the geostrophic v -momentum equation ($f_0 \bar{u} = -\partial \bar{\phi} / \partial y$) and hydrostasy ($\partial \bar{\phi} / \partial z = \bar{b}$). One less than ideal aspect of (7.60) is that in the extratropics the dominant balance is usually between the first two terms on the right-hand sides of each equation, even in time-dependent cases. Thus, the Coriolis force closely balances the divergence of the eddy momentum fluxes, and the advection of the mean stratification ($N^2 w$, or ‘adiabatic cooling’) often balances the divergence of eddy heat flux, with heating being a small residual. This may lead to an underestimation of the importance of diabatic heating, as this is ultimately responsible for the mean meridional circulation. Furthermore, the link between \bar{u} and b via thermal wind dynamically couples buoyancy and momentum, and obscures the understanding of how the eddy fluxes influence these fields — is it through the eddy heat fluxes or momentum fluxes, or some combination?

To address this issue (at least for some problems), we combine the terms $N^2 w$ and the eddy flux in (7.60b) into a single total or *residual* (so recognizing the cancellation between the mean and eddy terms) heat transport term that in a steady state is balanced by the diabatic term \bar{S} . To do this, we first note that because \bar{v} and \bar{w} are related by mass conservation we can define a mean meridional streamfunction ψ_m such that

$$(\bar{v}, \bar{w}) = \left(-\frac{\partial \psi_m}{\partial z}, \frac{\partial \psi_m}{\partial y} \right). \quad (7.62)$$

The velocities then satisfy $\partial \bar{v} / \partial y + \partial \bar{w} / \partial z = 0$ automatically. If we define a *residual streamfunction* by

$$\psi^* \equiv \psi_m + \frac{1}{N^2} \bar{v}' \bar{b}', \quad (7.63a)$$

the components of the *residual mean meridional circulation* are then given by

$$(\bar{v}^*, \bar{w}^*) = \left(-\frac{\partial \psi^*}{\partial z}, \frac{\partial \psi^*}{\partial y} \right), \quad (7.63b)$$

and

$$\bar{v}^* = \bar{v} - \frac{\partial}{\partial z} \left(\frac{1}{N^2} \bar{v}' \bar{b}' \right), \quad \bar{w}^* = \bar{w} + \frac{\partial}{\partial y} \left(\frac{1}{N^2} \bar{v}' \bar{b}' \right). \quad (7.64)$$

Note that by construction, the residual overturning circulation satisfies

$$\frac{\partial \bar{v}^*}{\partial y} + \frac{\partial \bar{w}^*}{\partial z} = 0. \quad (7.65)$$

Substituting (7.64) into (7.60a) and (7.60b) the zonal momentum and buoyancy equations then take the simple forms

$$\begin{aligned} \frac{\partial \bar{u}}{\partial t} &= f_0 \bar{v}^* + \bar{v}' \bar{q}' + \bar{F} \\ \frac{\partial \bar{b}}{\partial t} &= -N^2 \bar{w}^* + \bar{S} \end{aligned} \quad (7.66a,b)$$

which are known as the (quasi-geostrophic) *transformed Eulerian mean equations*, or TEM equations. The potential vorticity flux, $\bar{v}' \bar{q}'$, is given in terms of the heat and vorticity fluxes by (7.15) or (7.20), and is equal to the divergence of the Eliassen–Palm flux as in (7.22).

The TEM equations make it apparent that we may consider the potential vorticity fluxes,

rather than the separate contributions of the vorticity and heat fluxes, to force the circulation. If we know the potential vorticity flux as well as \bar{F} and \bar{S} , then (7.65) and (7.66), along with thermal wind balance

$$f_0 \frac{\partial \bar{u}}{\partial z} = - \frac{\partial \bar{b}}{\partial y} \quad (7.67)$$

form a complete set. The meridional overturning circulation is obtained by eliminating time derivatives from (7.66) using (7.67), giving

$$f_0^2 \frac{\partial^2 \psi^*}{\partial z^2} + N^2 \frac{\partial^2 \psi^*}{\partial y^2} = f_0 \frac{\partial}{\partial z} \bar{v}' \bar{q}' + f_0 \frac{\partial \bar{F}}{\partial z} + \frac{\partial \bar{S}}{\partial y}. \quad (7.68)$$

Thus, the residual or net overturning circulation is driven by the (vertical derivative of the) potential vorticity fluxes and the diabatic terms — driven in the sense that if we know those terms we can calculate the overturning circulation. (Of course the fluxes themselves depend on the circulation). Note that this equation applies at every instant, even if the equations are not in a steady state.

7.3.2 The TEM in isentropic coordinates

The residual circulation has an illuminating interpretation if we think of the fluid as comprising multiple layers of shallow water, or equivalently if we cast the problem in isentropic coordinates (section 3.9). Using the notation of a shallow water system, the momentum and mass conservation equation can then be written as

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} - f \mathbf{v} = \mathbf{F}, \quad \frac{\partial h}{\partial t} + \nabla \cdot (h \mathbf{u}) = S. \quad (7.69a,b)$$

The quantity h is the thickness between two isentropic surfaces and S is a thickness source term. (The field h plays the same role as σ in section 3.9.) With quasi-geostrophic scaling, so that variations in Coriolis parameter and layer thickness are small, zonally averaging in a conventional way gives

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v} = \bar{v}' \bar{\zeta}' + \bar{F}, \quad \frac{\partial \bar{h}}{\partial t} + H \frac{\partial \bar{v}}{\partial y} = - \frac{\partial}{\partial y} \bar{v}' \bar{h}' + \bar{S}. \quad (7.70a,b)$$

The overbars in these equations denote averages taken along isentropes — i.e., they are averages for a given layer — but are otherwise conventional, and the meridional velocity is purely ageostrophic. By analogy to (7.64), we define the residual circulation by

$$\bar{v}^* \equiv \bar{v} + \frac{1}{H} \bar{v}' \bar{h}', \quad (7.71)$$

where H is the mean thickness of the layer. Using (7.71) in (7.70) gives

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v}^* = \bar{v}' \bar{q}' + \bar{F}, \quad \frac{\partial \bar{h}}{\partial t} + H \frac{\partial \bar{v}^*}{\partial y} = \bar{S}, \quad (7.72a,b)$$

where

$$\bar{v}' \bar{q}' = \bar{v}' \bar{\zeta}' - \frac{f_0}{H} \bar{v}' \bar{h}', \quad (7.73)$$

is the meridional potential vorticity flux in a shallow water system. From (7.71) we see that the residual velocity is a measure of the *total meridional thickness flux*, eddy plus mean, in an isentropic layer. This is often a more useful quantity than the Eulerian velocity \bar{v} because it is generally the former, not the latter, that is constrained by the external forcing. What we have done, of course, is to effectively use a thickness-weighted mean in (7.69b); to see this, define the thickness-weighted mean by

$$\bar{v}_* \equiv \frac{\bar{h}v}{\bar{h}}. \quad (7.74)$$

(We use \bar{v}_* to denote a thickness- or mass-weighted mean, and \bar{v}^* to denote a residual velocity; the quantities are closely related, as we will see.) From (7.74) we have

$$\bar{v}_* = \bar{v} + \frac{1}{\bar{h}}\bar{v}'\bar{h}', \quad (7.75)$$

then the zonal average of (7.69b) is just

$$\frac{\partial \bar{h}}{\partial t} + \frac{\partial}{\partial y}(\bar{h}\bar{v}_*) = \bar{S}, \quad (7.76)$$

which is the same as (7.72b) if we take $H = \bar{h}$. Similarly, if we use the thickness weighted velocity (7.75) in the momentum equation (7.70a) we obtain (7.72a).

Evidently, if the mass-weighted meridional velocity is used in the momentum and thickness equations then the eddy mass flux does not enter the equations explicitly: the only eddy flux in (7.72) is that of potential vorticity. That is, in isentropic coordinates the equations in TEM form are equivalent to the equations that arise from a particular form of averaging — thickness weighted averaging — rather than the conventional Eulerian averaging. A similar correspondence occurs in height coordinates, as we now see.

7.3.3 Connection between the residual and thickness-weighted circulation

It is evident from the above arguments that, in a shallow water system or in isentropic coordinates, the residual velocity is a measure of the total (i.e., mean plus eddy) thickness transport. In height coordinates, the definition of residual velocity, (7.63) does not lend itself so easily to such an interpretation. However, the residual velocity in height coordinates is, in fact, also a measure of the total thickness transport, or equivalently of the mass transport between two isentropic surfaces, as we now discover. Specifically, we show that averaging the total transport in isentropic layers is equivalent to the mass transport evaluated by the TEM formalism in height coordinates, and specifically that the thickness-weighted mean, \bar{v}_* is equivalent to the residual velocity, \bar{v}^* in height coordinates. Our demonstration is for a Boussinesq system, but the extension to a compressible gas is reasonably straightforward.⁶

Consider two isentropic surfaces, η_1 and η_2 with mean positions $\bar{\eta}_1$ and $\bar{\eta}_2$, as in Fig. 7.1. (We use z to denote the vertical coordinate, and η to denote the location of isentropic surfaces.) The meridional transport between these surfaces is given by

$$T = \int_{\eta_2}^{\eta_1} v \, dz. \quad (7.77)$$

If the velocity does not vary with height within the layer (and in the limit of layer thickness

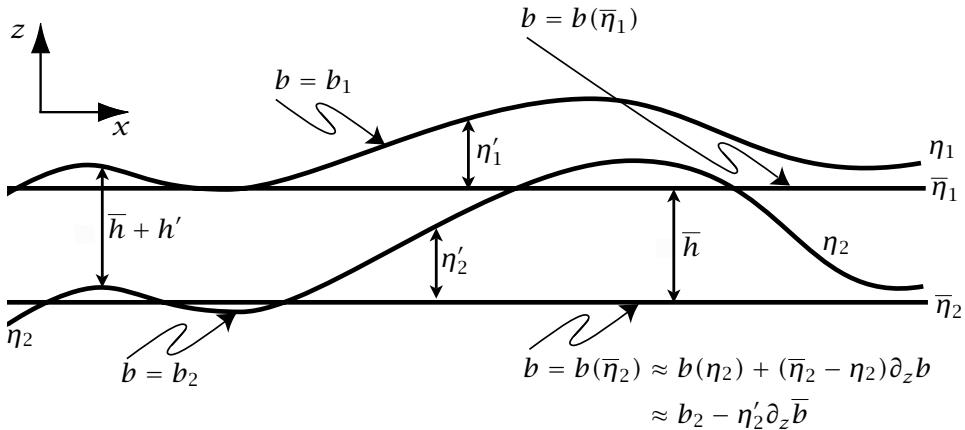


Fig. 7.1 Two isentropic surfaces, η_1 and η_2 , and their mean positions, $\bar{\eta}_1$ and $\bar{\eta}_2$. The departure of an isentrope from its mean position is proportional to the temperature perturbation at the mean position of the isentrope, and the variations in thickness (h') of the isentropic layer are proportional to the vertical derivative of this.

going to zero this is the case) then $T = vh$ where $h = \eta_1 - \eta_2$ is the thickness of the isentropic layer. The zonally averaged transport is then given by

$$\bar{T} = \frac{1}{L} \int_L T \, dx = \frac{1}{L} \int_L \left(\int_{\eta_2}^{\eta_1} v \, dz \right) \, dx = \overline{\int_{\eta_2}^{\eta_1} v \, dz} = \overline{vh} = \bar{v}\bar{h} + \bar{v}'h', \quad (7.78)$$

with obvious notation, and with an overbar denoting a zonal average. Letting the distance between isentropes shrink to zero this result allows us to write

$$\bar{v}_* \equiv \frac{\bar{v}\bar{\sigma}^b}{\bar{\sigma}} = \bar{v}^b + \frac{\bar{v}'\bar{\sigma}'^b}{\bar{\sigma}}, \quad (7.79)$$

where $(\cdot)^b$ denotes an average along an isentrope and $\bar{\sigma} = \overline{\partial z / \partial b}$ is the thickness density, a measure of the thickness between two isentropes. Equation (7.79) is analogous to (7.75), for a continuously stratified system. The averaged quantity \bar{v}_* is not proportional to the average of the velocity at constant height, or even to the average along an isentrope; rather, it is the *thickness-weighted* zonal average of the velocity *between* two isentropic surfaces, Δb apart, of mean separation proportional to $\bar{\sigma}\Delta b$. Our goal is to express this transport in terms of Eulerian-averaged quantities, at a constant height z .

Let us first connect an average along an isentrope of some variable χ to its average at constant height by writing, for small isentropic displacements,

$$\bar{\chi}^b = \overline{\chi(z + \eta')}^z \approx \overline{\chi(z) + \eta' \partial \chi / \partial z}^z, \quad (7.80)$$

where the superscript explicitly denotes how the zonal average is taken, and η' is the displacement of the isotherm from its mean position. This can be expressed in terms of the temperature perturbation at the location of the mean isentrope by Taylor expanding b around its value on that mean isentrope. That is,

$$b(\eta) = b(\bar{\eta}) + \left(\frac{\partial b}{\partial z} \right)_{z=\bar{\eta}} (\eta - \bar{\eta}) + \dots, \quad (7.81)$$

where $\bar{\eta} = \bar{\eta}(z)$, giving

$$\eta' \approx \frac{-b'}{\partial_z b(\bar{\eta})} \approx -\frac{b'}{\partial_z \bar{b}}, \quad (7.82)$$

where $\eta' = \eta - \bar{\eta}$ and $b' = b(\bar{\eta}) - b(\eta)$. Using (7.82) in (7.80) (and omitting the superscript z on $\partial_z \bar{b}$) we obtain, with $\chi = v$,

$$\bar{v}^b = \bar{v}^z - \frac{\bar{b}' \partial_z \bar{v}'^z}{\partial_z \bar{b}}. \quad (7.83)$$

Note that if v is in thermal wind balance with b then the second term vanishes identically, but we will not invoke this.

We now transform the second term on the right-hand side (7.79) to an average at constant z . The variations in thickness of an isothermal layer are given by

$$\sigma' \approx \bar{\sigma} \frac{\partial \eta'}{\partial z} = -\bar{\sigma} \frac{\partial}{\partial z} \left(\frac{b'}{\partial_z \bar{b}} \right), \quad (7.84)$$

using (7.82). Thus, neglecting terms that are third-order in amplitude,

$$\bar{v}' \bar{\sigma}'^b = -\bar{\sigma} v' \frac{\partial}{\partial z} \left(\frac{b'}{\partial_z \bar{b}} \right)^z. \quad (7.85)$$

Using both (7.83) and (7.85), (7.79) becomes

$$\bar{v}_* = \bar{v}^z - \frac{\bar{b}' \partial_z \bar{v}'^z}{\partial_z \bar{b}} - v' \frac{\partial}{\partial z} \left(\frac{b'}{\partial_z \bar{b}} \right)^z = \bar{v}^z - \frac{\partial}{\partial z} \left(\frac{v' b'}{\partial_z \bar{b}} \right)^z. \quad (7.86)$$

The right-hand side of the last equation is the TEM form of the residual velocity; thus, we have shown that

$$\bar{v}_* \equiv \frac{\bar{v} \bar{\sigma}}{\bar{\sigma}} = \bar{v}^b + \frac{\bar{v}' \bar{\sigma}'^b}{\bar{\sigma}} \approx \bar{v}^z - \frac{\partial}{\partial z} \left(\frac{\bar{v}' b'^z}{\partial_z \bar{b}} \right) \equiv \bar{v}^*. \quad (7.87)$$

We see the equivalence of the thickness-weighted mean velocity on the left-hand side and the residual velocity on the right-hand side. In the quasi-geostrophic limit $N^2 = \partial_z \bar{b}$ and $\bar{\sigma}$ is a reference thickness.

7.3.4 * The TEM in the primitive equations

We now look at the TEM equations in a slightly more general way, although we will largely restrict our attention to the thermodynamic budget and, for simplicity of exposition, the Boussinesq equations.⁸ The basic idea, as before, is that we seek to replace the advecting velocity in the equations of motion by some residual velocity \bar{v}^* where

$$\bar{v}^* = v + \tilde{v} = v + \nabla \times \psi. \quad (7.88)$$

We will choose the ‘eddy induced velocity’, \tilde{v} , or equivalently the associated vector streamfunction ψ , to simplify the resulting equations (as far as possible) by incorporating the eddy

Aspects of the TEM Formulation

Properties and features

- ★ The residual mean circulation is equivalent to the total mass-weighted (eddy plus Eulerian mean) circulation, and it is this circulation that is driven by the diabatic forcing.
- ★ There are no explicit eddy fluxes in the buoyancy budget; the only eddy term is the flux of potential vorticity, and this is divergence of the Eliassen-Palm flux; that is $\bar{v}'\bar{q}' = \nabla_x \cdot \mathcal{F}$.
- ★ The residual circulation, \bar{v}^* , becomes part of the solution, just as \bar{v} is part of the solution in an Eulerian mean formulation.

But note

- ★ The TEM formulation does not solve the parameterization problem, and eddy fluxes are still present in the equations.
- ★ The theory and practice are well developed for a zonal average, but less so for three-dimensional, non-zonal flow. This is because the geometry enforces simple boundary conditions in the zonal mean case.⁷
- ★ The boundary conditions on the residual circulation are neither necessarily simple nor easily determined; for example, at a horizontal boundary \bar{w}^* is not zero if there are horizontal buoyancy fluxes.

Examples of the use of the TEM in the general circulation of the atmosphere and ocean arise in sections 12.2, 12.4 and 16.6.

fluxes that have an advective nature into the prognostic equations. Now, (7.88) is just a kinematic transformation, and we cannot hope to completely eliminate the effects of the eddy fluxes no matter how we choose ψ ; however, it may be easier to parameterize the cross-gradient fluxes than the along-gradient fluxes and we can choose our transformation to facilitate this. (The parameterization issue is discussed more in section 10.6 and those following.) These statements may well seem obscure at the moment, so for now let us just regard the transformation (7.88) as a formal one, and proceed.

Decomposing the velocity and buoyancy into Eulerian mean and eddy terms, denoted with an overbar and a prime respectively, the momentum and thermodynamic equations may be written as

$$\frac{\partial \bar{\mathbf{u}}}{\partial t} + \bar{\mathbf{v}} \cdot \nabla \bar{\mathbf{u}} + \mathbf{f} \times \bar{\mathbf{u}} = -\nabla \bar{\phi} + \mathbf{R}, \quad (7.89a)$$

$$\frac{\partial \bar{b}}{\partial t} + \bar{\mathbf{v}} \cdot \nabla \bar{b} = S[b] - \nabla \cdot \mathbf{F}[b], \quad (7.89b)$$

and a tracer, φ , obeys

$$\frac{\partial \bar{\varphi}}{\partial t} + \bar{\mathbf{v}} \cdot \nabla \bar{\varphi} = S[\varphi] - \nabla \cdot \mathbf{F}[\varphi]. \quad (7.89c)$$

The equations of motion are completed by way of the hydrostatic equation ($\partial \bar{\phi} / \partial z = \bar{b}$) and mass continuity equation ($\nabla \cdot \bar{\mathbf{v}} = 0$). The term \mathbf{R} represents the various Reynolds stress terms, S represents diabatic sources such as diffusion and heating, and the $\nabla \cdot \mathbf{F}$ terms are eddy fluxes of the tracer indicated in the square brackets. Thus,

$$\nabla \cdot \mathbf{F}[\varphi] = \frac{\partial \bar{u}'\varphi'}{\partial x} + \frac{\partial \bar{v}'\varphi'}{\partial y} + \frac{\partial \bar{w}'\varphi'}{\partial z} = \nabla \cdot \bar{\mathbf{v}}'\varphi', \quad (7.90)$$

and similarly for $\nabla \cdot \mathbf{F}[b]$. (We will omit the identifying label in square brackets if there is no ambiguity. Also, in this section, the divergence and related operators are three-dimensional unless noted otherwise.) The fluxes may be divided into components along and across the tracer gradient,

$$\mathbf{F} = \mathbf{F}_\perp + \mathbf{F}_\parallel = (\mathbf{n} \cdot \mathbf{F})\mathbf{n} + (\mathbf{n} \times \mathbf{F}) \times \mathbf{n}, \quad (7.91)$$

where $\mathbf{n} = \nabla \bar{\varphi} / |\nabla \bar{\varphi}|$. Thus,

$$\mathbf{F}_\perp = \frac{\mathbf{F} \cdot \nabla \bar{\varphi}}{|\nabla \bar{\varphi}|^2} \nabla \bar{\varphi}, \quad \mathbf{F}_\parallel = \frac{\nabla \bar{\varphi} \times \mathbf{F}}{|\nabla \bar{\varphi}|^2} \times \nabla \bar{\varphi} \quad (7.92)$$

so that \mathbf{F}_\perp (\mathbf{F}_\parallel) is perpendicular to (along) iso-surfaces of the tracer. Such a decomposition is possible for any vector, of course. The component \mathbf{F}_\parallel is known as a *skew flux*, and this may be written as if it were an advection by some divergence-free velocity $\tilde{\mathbf{v}}$; that is,

$$\nabla \cdot \mathbf{F}_\parallel = \nabla \cdot \left(\frac{\nabla \bar{\varphi} \times \mathbf{F}}{|\nabla \bar{\varphi}|^2} \times \nabla \bar{\varphi} \right) = \left(\nabla \times \frac{\nabla \bar{\varphi} \times \mathbf{F}}{|\nabla \bar{\varphi}|^2} \right) \cdot \nabla \bar{\varphi} \equiv \tilde{\mathbf{v}} \cdot \nabla \bar{\varphi}, \quad (7.93)$$

using the vector identity $\nabla \cdot (\mathbf{A} \times \nabla \phi) = \nabla \phi \cdot \nabla \times \mathbf{A}$. Furthermore, $\tilde{\mathbf{v}} \cdot \nabla \bar{\varphi} = \nabla \cdot (\tilde{\mathbf{v}} \bar{\varphi})$ so that the skew flux is conservative: that is, it serves to redistribute φ , but, just like advection by a true incompressible velocity, it does not of itself change the distribution, or the census, of φ . (If φ were buoyancy or entropy, we would say that the skew flux is adiabatic. A skew flux may nonetheless create a situation in which true dissipation is more effective, by the production of small scales and enhanced tracer gradients.)

If we use (7.88) in (7.89c) then we obtain

$$\frac{\partial \bar{\varphi}}{\partial t} + \bar{\mathbf{v}}^* \cdot \nabla \bar{\varphi} = S[\varphi] - \nabla \cdot \mathbf{F}^*[\varphi], \quad (7.94)$$

where

$$\nabla \cdot \mathbf{F}^*[\varphi] = \nabla \cdot (\mathbf{F}[\varphi] - \bar{\varphi} \nabla \times \boldsymbol{\psi}) = \nabla \cdot (\mathbf{F}[\varphi] + \nabla \bar{\varphi} \times \boldsymbol{\psi}), \quad (7.95)$$

using the vector identity $\nabla \cdot (\bar{\varphi} \nabla \times \boldsymbol{\psi}) = \nabla \cdot (\boldsymbol{\psi} \times \nabla \bar{\varphi})$. Thus, we may define the *residual flux* with respect to φ by

$$\mathbf{F}^*[\varphi] \equiv \mathbf{F}[\varphi] + \nabla \bar{\varphi} \times \boldsymbol{\psi}, \quad (7.96)$$

remembering that the flux is only relevant up to a non-divergent additive factor. From (7.96)

we see that, no matter how we choose ψ , the components of the residual and raw flux along the mean tracer gradient, or across isosurfaces of the tracer, are the same:

$$\mathbf{F}^*[\varphi] \cdot \nabla \bar{\varphi} = \mathbf{F}[\varphi] \cdot \nabla \bar{\varphi}. \quad (7.97)$$

However, the two fluxes along the tracer contours — that is, the skew fluxes — differ and we may choose ψ in such a way that this is zero for the residual flux. Why is this a useful choice? It is because a skew flux is already present in the advective term, $\bar{\mathbf{v}}^* \cdot \nabla \varphi$ on the left-hand side and so it is useful for the right-hand side to contain only digradient fluxes.

To remove the residual skew flux, we make the choice

$$\psi = -\frac{\mathbf{F}[\varphi] \times \nabla \bar{\varphi}}{|\nabla \bar{\varphi}|^2}. \quad (7.98)$$

The flux associated with ψ is just the negative of the raw skew flux, because from (7.96) this is

$$\nabla \bar{\varphi} \times \psi = \frac{\nabla \bar{\varphi} \times \mathbf{F}[\varphi]}{|\nabla \bar{\varphi}|^2} \times \nabla \bar{\varphi} = -\mathbf{F}_{\parallel}. \quad (7.99)$$

Making this choice, the tracer evolution equation may be written, with $S[\varphi] = 0$, as

$$\frac{\partial \bar{\varphi}}{\partial t} + \bar{\mathbf{v}}^* \cdot \nabla \bar{\varphi} = -\nabla \cdot \mathbf{F}^*[\varphi] \quad (7.100a)$$

where

$$\bar{\mathbf{v}}^* = \bar{\mathbf{v}} + \nabla \times \psi = \bar{\mathbf{v}} - \nabla \times \frac{\mathbf{F}[\varphi] \times \nabla \bar{\varphi}}{|\nabla \bar{\varphi}|^2} \quad (7.100b)$$

and

$$\mathbf{F}^*[\varphi] = \frac{\mathbf{F}[\varphi] \cdot \nabla \bar{\varphi}}{|\nabla \bar{\varphi}|^2} \nabla \bar{\varphi}. \quad (7.100c)$$

Evidently there is a different residual flux and a TEM formulation for any advected tracer, be it a passive tracer or an active one like potential temperature or potential vorticity. We must choose just one, and in doing so the skew flux of the other tracers will not be removed. The one most commonly chosen, for both oceanographic and atmospheric applications, is potential temperature, or buoyancy. For diagnostic purposes the choice of scalar may not in fact be critical: the various eddy fluxes are in reality caused by advection by small-scale turbulent motion, and such motion often acts similarly on all conserved tracers. Thus, removing the skew flux in potential temperature is likely to nearly remove it for all nearly conserved tracers. Potential vorticity might be another suitable choice in some circumstances.

Suppose that the eddying flow is adiabatic, which is a particularly good approximation in the ocean interior and, to a lesser degree, in the stratosphere. This means that there is no eddy buoyancy flux across isopycnals (or heat flux across isotherms) and the digradient residual flux with respect to buoyancy is zero. Using buoyancy to define the transformation in (7.100), the TEM form of the thermodynamic equation is

$$\frac{\partial \bar{b}}{\partial t} + \bar{\mathbf{v}}^* \cdot \nabla \bar{b} = S[b]. \quad (7.101)$$

Thus, in much the same way as arose in the quasi-geostrophic equations, the transformed thermodynamic equation appears devoid of explicit eddy terms. However, the (usually numerical) solution of the transformed equations is not necessarily straightforward, because the residual velocity does not vanish at the boundary. Thus, for example, $\bar{w}^* \neq 0$ at a surface where there are non-zero heat fluxes.

A zonal average

Let the averages (as denoted by the overbars) in the above expressions be zonal average; the residual streamfunction is then a scalar, ψ , where $\psi = -\mathbf{i}\psi$ and

$$\bar{v}^* = -\frac{\partial \psi}{\partial z}, \quad \bar{w}^* = \frac{\partial \psi}{\partial y}. \quad (7.102)$$

The residual streamfunction of (7.98) becomes

$$\psi = \frac{\bar{v}' \bar{b}' \bar{b}_z - \bar{w}' \bar{b}' \bar{b}_y}{|\nabla_x \bar{b}|^2} \quad (7.103)$$

and the residual flux is

$$\mathbf{F}^*[\bar{b}] = \frac{\bar{v}' \bar{b}' \bar{b}_y + \bar{w}' \bar{b}' \bar{b}_z}{|\nabla_x \bar{b}|^2} \nabla \bar{b}, \quad (7.104)$$

where $\nabla_x \bar{b} = \mathbf{j} \partial \bar{b} / \partial y + \mathbf{k} \partial \bar{b} / \partial z$.

If the isotherms are nearly flat then the above expressions reduce to the corresponding ones in quasi-geostrophic TEM theory. Specifically: (i) $|\nabla \bar{b}|$ is approximated by $\bar{b}_z = N^2$, and (ii) vertical eddy thermal fluxes are neglected. The streamfunction (7.103) then reduces to

$$\psi = \frac{\bar{v}' \bar{b}'}{|\bar{b}_z|}, \quad (7.105)$$

which is the same as the eddy part of (7.63a).

† More general forms

Transformations more general than that of (7.98) are possible that still lead to a transformed thermodynamic budget with the same general form, and, in particular, still partition the flux into a skew component and a diagradient component, and that may be particularly useful as boundaries are approached. In particular, we may choose

$$\psi = -\frac{\mathbf{F}[\varphi] \times \nabla \bar{\varphi}}{|\nabla \bar{\varphi}|^2} - \frac{\mathbf{F}[\varphi] \cdot \nabla \bar{\varphi}}{|\nabla \bar{\varphi}|^2} \boldsymbol{\alpha}, \quad (7.106)$$

where $\boldsymbol{\alpha}$ is an arbitrary vector. The transformed evolution equation for $\bar{\varphi}$ is then

$$\frac{\partial \bar{\varphi}}{\partial t} + \bar{\mathbf{v}}^* \cdot \nabla \bar{\varphi} = -\nabla \cdot \mathbf{F}^*[\varphi], \quad (7.107)$$

where

$$\mathbf{F}^*[\varphi] = \frac{\mathbf{F}[\varphi] \cdot \nabla \bar{\varphi}}{|\nabla \bar{\varphi}|^2} (\nabla \bar{\varphi} + \boldsymbol{\alpha} \times \nabla \bar{\varphi}) \quad (7.108)$$

and $\bar{\mathbf{v}}^* = \mathbf{v} + \nabla \times \psi$. The residual flux is still proportional to the diagradient flux, and will be zero if the eddies are adiabatic. However, unlike (7.100c) the flux is not directed parallel to the gradient.

Although the case with $\boldsymbol{\alpha} = 0$ (considered above) provides a clean separation between along-gradient (e.g., isopycnal) and diagradient fluxes, another useful choice is to set $\boldsymbol{\alpha}$ to be proportional to the isothermal slope, and pointing across the slope; that is, with $\varphi = b$, we choose

$$\boldsymbol{\alpha} = -\frac{\mathbf{k} \times \nabla \bar{b}}{\partial_z \bar{b}}, \quad (7.109)$$

which in the zonally averaged case is $\boldsymbol{\alpha} = \mathbf{i}\alpha = -\mathbf{i}\bar{b}_y/\bar{b}_z$. After some algebra we find

$$\boldsymbol{\psi} = \frac{\mathbf{k} \times \mathbf{F}[b]}{\partial_z \bar{b}} = \frac{-1}{\partial_z \bar{b}} \left(\mathbf{i} \bar{v}' \bar{b}' - \mathbf{j} \bar{u}' \bar{b}' \right) \quad (7.110)$$

and

$$\mathbf{F}^*[b] = \frac{\mathbf{F}[b] \cdot \nabla \bar{b}}{\partial_z \bar{b}} \mathbf{k}. \quad (7.111)$$

Using (7.110) and (7.111) in (7.100), with $\phi = b$, gives the TEM thermodynamic equation (with $S[b] = 0$) in the form:

$$\frac{\partial \bar{b}}{\partial t} + (\bar{\mathbf{u}} + \tilde{\mathbf{u}}) \cdot \nabla \bar{b} + (\bar{w} + \tilde{w}) \frac{\partial \bar{b}}{\partial z} = \frac{\partial G}{\partial z}, \quad (7.112a)$$

where

$$G = \frac{\bar{v}' \bar{b}' \cdot \nabla \bar{b}}{\partial_z \bar{b}} = \frac{1}{\partial_z \bar{b}} \left(\bar{\mathbf{u}}' \bar{b}' \cdot \nabla_z \bar{b} + \bar{w}' \bar{b}' \frac{\partial \bar{b}}{\partial z} \right). \quad (7.112b)$$

and, from $\tilde{\mathbf{v}} = +\nabla \times \boldsymbol{\psi}$,

$$\tilde{u} = -\frac{\partial}{\partial z} \left(\frac{\bar{u}' \bar{b}'}{\partial_z \bar{b}} \right), \quad \tilde{v} = -\frac{\partial}{\partial z} \left(\frac{\bar{v}' \bar{b}'}{\partial_z \bar{b}} \right), \quad \tilde{w} = \nabla_z \cdot \left(\frac{\bar{\mathbf{u}}' \bar{b}'}{\partial_z \bar{b}} \right). \quad (7.112c)$$

It may be verified by direct manipulation that (7.112) is equivalent to the usual, Eulerian averaged, thermodynamic equation. The corresponding Eliassen-Palm fluxes (that is, the fluxes in the transformed momentum equation) are given in the appendix to chapter 12.

The TEM equations in the form (7.112) are particularly useful in the diagnosis of numerical experiments and observations in the interior of the ocean or atmosphere where isopycnal slopes are small, and here they differ little from (7.100) but are somewhat easier to calculate and interpret — note for example that the quasi-geostrophic TEM equations follow particularly easily from (7.112) by setting $G = 0$ and $\bar{b}_z = N^2$. On the other hand, one difficulty with this form is that, from (7.112c), the vertical component of the residual velocity does not vanish at a horizontal boundary if there is a non-zero horizontal buoyancy flux. Another form of the TEM equations, and one that may be useful in regions of low stratification and near boundaries, arises by choosing $\boldsymbol{\alpha} = \mathbf{i}\bar{b}_z/\bar{b}_y$, but we refer the reader to the references given to explore this topic further.

7.4 THE NON-ACCELERATION RESULT

For the rest of this chapter we return to quasi-geostrophic dynamics, and consider further the interpretation and application of the potential vorticity flux and its relatives. We first consider an important result in wave-mean flow dynamics, the non-acceleration result.⁹ This result shows that under certain conditions, to be made explicit below, waves have no net effect on the zonally averaged flow, an important and somewhat counter-intuitive result.

7.4.1 A derivation from the potential vorticity equation

Consider how the potential vorticity fluxes affect the mean fields. If the waves are steady and conservative then, from (7.11) or (7.23a), the EP flux is non-divergent and the meridional

potential vorticity flux, $\overline{v'q'}$, is zero. Now, the zonally averaged potential vorticity equation is

$$\frac{\partial \overline{q}}{\partial t} + \frac{\partial \overline{v'q'}}{\partial y} = \overline{D}. \quad (7.113)$$

Thus, if the above conditions are satisfied we have $\partial \overline{q}/\partial t = 0$. Now, in quasi-geostrophic theory the buoyancy and geostrophically balanced velocity can be determined from the potential vorticity via an elliptic equation, and in particular

$$\overline{q} - \beta y = \frac{\partial^2 \overline{\psi}}{\partial y^2} + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial \overline{\psi}}{\partial z} \right), \quad (7.114)$$

and so differentiating (7.113) with respect to y and using (7.22) we obtain

$$\left[\frac{\partial^2}{\partial y^2} + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial}{\partial z} \right) \right] \frac{\partial \overline{u}}{\partial t} = (\nabla \cdot \mathcal{F})_{yy} - \overline{D}_y. \quad (7.115)$$

Thus, if the terms on the right-hand side are zero then a solution is

$$\frac{\partial \overline{u}}{\partial t} = 0, \quad (7.116)$$

and this is the unique solution if there is no acceleration at the boundaries. This is a *non-acceleration result*. That is to say, under the non-acceleration conditions the tendency of the mean fields, and in particular of the zonally averaged zonal flow, are independent of the waves. Explicitly, the non-acceleration conditions are the following.

- (i) The waves are steady [so that, using the Eliassen–Palm relation (7.23a), \mathcal{A} does not vary].
- (ii) The waves are conservative [i.e., $\mathcal{D} = 0$ in (7.23a)]. Given this and item (i), the Eliassen–Palm relation implies that $\nabla \cdot \mathcal{F} = 0$; that is, the potential vorticity flux is zero.
- (iii) The waves are of small amplitude (all of our analysis has neglected terms that are cubic in perturbation amplitude).
- (iv) The waves do not affect the boundary conditions (so there are no boundary contributions to the acceleration).

The result applies to the buoyancy and velocity fields that are directly invertible from potential vorticity, and not to the ageostrophic velocities. Given the way we have derived it, it does not seem a surprising result; however, it can be powerful and counter-intuitive, for it means that steady waves (i.e., those whose amplitude does not vary) do not affect the zonal flow. However, they do affect the meridional overturning circulation, and the relative vorticity flux may also be non-zero. In fact, the non-acceleration theorem is telling us that the changes in the vorticity flux are exactly compensated for by changes in the meridional circulation, and there is no net effect on the zonally averaged zonal flow. It is *irreversibility*, often manifested by the breaking of waves, that leads to permanent changes in the mean flow.

The derivation of this result by way of the momentum equation, which one might expect to be more natural, is rather awkward because one must consider momentum and buoyancy fluxes separately. Furthermore, the zonally averaged meridional circulation comes into play: for example, meridional velocity, \overline{v} , is, although small because it is purely ageostrophic, not zero and we cannot neglect it because it is multiplied by the Coriolis parameter, which

is large. Thus, the eddy vorticity fluxes can affect both the meridional circulation and the acceleration of the zonal mean flow, and it might seem impossible to disentangle the two effects without completely solving the equations of motion. Fortunately, the situation is simplified via the use of the TEM, as we now see.

7.4.2 Using TEM to give the non-acceleration result

A two-dimensional case

Consider two-dimensional incompressible flow on the β -plane, for which there is no buoyancy flux. The linearized vorticity equation is

$$\frac{\partial \zeta'}{\partial t} + \bar{u} \frac{\partial \zeta'}{\partial x} + v' \frac{\partial \bar{\zeta}}{\partial y} = D', \quad (7.117)$$

from which we derive, analogously to (7.23a), the Eliassen-Palm relation

$$\frac{\partial \mathcal{A}}{\partial t} + \frac{\partial \mathcal{F}}{\partial y} = \mathcal{D}, \quad (7.118)$$

where $\mathcal{F} = -\bar{u}'v'$, \mathcal{D} represents non-conservative forces, and

$$\mathcal{A} = \frac{\bar{\zeta}'^2}{2\partial_y \bar{\zeta}} = \frac{1}{2} \bar{\eta}'^2 \frac{\partial \bar{\zeta}}{\partial y}. \quad (7.119)$$

The quantity $\eta' \equiv -\zeta'/\partial_y \bar{\zeta}$ is proportional to the meridional particle displacement in a disturbance. Now consider the x -momentum equation

$$\frac{\partial u}{\partial t} = -\frac{\partial u^2}{\partial x} - \frac{\partial uv}{\partial y} - \frac{\partial \phi}{\partial x} + fv. \quad (7.120)$$

Zonally averaging, noting that $\bar{v} = 0$, gives

$$\frac{\partial \bar{u}}{\partial t} = -\frac{\partial \bar{u}v}{\partial y} = \bar{v}'\bar{\zeta}' = \frac{\partial \mathcal{F}}{\partial y}. \quad (7.121)$$

Finally, combining (7.118) and (7.121) gives

$$\frac{\partial}{\partial t} (\bar{u} + \mathcal{A}) = \mathcal{D}. \quad (7.122)$$

In the absence of non-conservative terms (i.e., if $\mathcal{D} = 0$) the quantity $\bar{u} + \mathcal{A}$ is constant.¹⁰ Further, if the waves are steady and conservative then \mathcal{A} is constant and, therefore, so is \bar{u} .

The stratified case

In the stratified case we can use the TEM form of the momentum equation to derive a similar result. The unforced zonally averaged zonal momentum equation can be written as

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v}^* = \nabla \cdot \mathbf{F}, \quad (7.123)$$

and using the Eliassen–Palm relation this may be written as

$$\frac{\partial}{\partial t}(\bar{u} + \mathcal{A}) - f_0 \bar{v}^* = \mathcal{D}, \quad (7.124)$$

and so again \mathcal{A} is related to the momentum of the flow. If, furthermore, the waves are steady ($\partial \mathcal{A} / \partial t = 0$) and conservative ($\mathcal{D} = 0$), then $\partial \bar{u} / \partial t - f_0 \bar{v}^* = 0$. However, under these same conditions the residual circulation will also be zero. This is because the residual meridional circulation (\bar{v}^*, \bar{w}^*) arises via the necessity to keep the temperature and velocity fields in thermal wind balance, and is thus determined by an elliptic equation, namely (7.68). If the waves are steady and adiabatic then the right-hand side of this equation is zero and

$$f_0^2 \frac{\partial^2 \psi^*}{\partial z^2} + N^2 \frac{\partial^2 \psi^*}{\partial y^2} = 0. \quad (7.125)$$

If $\psi^* = 0$ at the boundaries, then the unique solution of this is $\psi^* = 0$ everywhere. At the meridional boundaries we may certainly suppose that ψ^* vanishes if these are quiescent latitudes, and at the horizontal boundaries the buoyancy flux will vanish if the waves there are steady, because from (7.13) we have

$$\bar{v}' \bar{b}' \frac{\partial \bar{b}}{\partial y} = -\frac{1}{2} \frac{\partial}{\partial t} \bar{b}'^2 = 0. \quad (7.126)$$

Under these circumstances, then, the residual meridional circulation vanishes in the interior, and from (7.123), the mean flow is steady.

Compare (7.123) with the momentum equation in conventional Eulerian form, namely

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v} = \bar{v}' \bar{\zeta}'. \quad (7.127)$$

There is no reason that the vorticity flux should vanish when waves are present, even if they are steady. However, such a flux is (under non-acceleration conditions) precisely compensated by the meridional circulation $f_0 \bar{v}$, something that is hard to infer or intuit directly from (7.127); even when non-acceleration conditions do not apply there will be a significant cancellation between the Coriolis and eddy terms. Unlike the proof of the non-acceleration result given in section 7.4.1, the above argument does not use the invertibility property of potential vorticity so directly, suggesting an extension to the primitive equations, but we do not pursue that here.¹¹ Various results regarding the TEM and non-acceleration are summarized in the shaded box on the following page.

7.4.3 The EP flux and form drag

It may seem a little magical that the zonal flow is driven by the Eliassen–Palm flux via (7.123). The poleward vorticity flux is clearly related to the momentum flux convergence, but why should a poleward buoyancy flux affect the momentum? The TEM form of the momentum equation is

$$\frac{\partial \bar{u}}{\partial t} = \frac{\partial}{\partial z} \left(\frac{f_0}{N^2} \bar{v}' \bar{b}' \right) + F_m, \quad (7.128)$$

where $F_m = \bar{v}' \bar{\zeta}' + f_0 \bar{v}^*$ represents forces from the momentum flux and Coriolis force. The first term on the right-hand side certainly does not look like a force; however, it turns out to

TEM, Residual Velocities, Non-acceleration, and All That

For a Boussinesq quasi-geostrophic system, the TEM form of the unforced momentum equation and the thermodynamic equation are:

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v}^* = \nabla \cdot \mathcal{F}, \quad \frac{\partial \bar{b}}{\partial t} + \bar{w}^* \frac{\partial}{\partial z} \bar{b}_0 = \bar{S}, \quad (\text{T.1})$$

where $\partial \bar{b}_0 / \partial z = N^2$, \bar{S} represents diabatic effects, \mathcal{F} is the Eliassen-Palm (EP) flux (so that $\nabla_x \cdot \mathcal{F} = \bar{v}' \bar{q}'$), and the residual velocities are

$$\bar{v}^* = \bar{v} - \frac{\partial}{\partial z} \left(\frac{1}{N^2} \bar{v}' \bar{b}' \right), \quad \bar{w}^* = \bar{w} + \frac{\partial}{\partial y} \left(\frac{1}{N^2} \bar{v}' \bar{b}' \right). \quad (\text{T.2})$$

Spherical coordinate and ideal gas versions of these take a similar form. We may define a meridional overturning streamfunction such that $(\bar{v}^*, \bar{w}^*) = (-\partial \psi^* / \partial z, \partial \psi^* / \partial y)$, and using thermal wind to eliminate time-derivatives in (T.1) we obtain

$$f_0^2 \frac{\partial^2 \psi^*}{\partial z^2} + N^2 \frac{\partial^2 \psi^*}{\partial y^2} = f_0 \frac{\partial}{\partial z} \bar{v}' \bar{q}' + \frac{\partial \bar{S}}{\partial y}. \quad (\text{T.3})$$

The above manipulations may seem formal, in that they simply transform the momentum and thermodynamic equation from one form to another. However, the resulting equations have two potential advantages over the untransfomed ones.

- (i) The residual meridional velocity is approximately equal to the average thickness-weighted velocity between two neighbouring isentropic surfaces, and so is a measure of the total (Eulerian mean plus eddy) meridional transport of thickness or buoyancy.
- (ii) The EP flux has certain attractive features, directly related to the physical properties of waves. The divergence of the EP flux is the meridional flux of potential vorticity:

$$\mathcal{F} = -(\bar{u}' \bar{v}') \mathbf{j} + \left(\frac{f_0}{N^2} \bar{v}' \bar{b}' \right) \mathbf{k}, \quad \nabla \cdot \mathcal{F} = \bar{v}' \bar{q}'. \quad (\text{T.4})$$

Furthermore, the EP flux satisfies, to second order in wave amplitude,

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla \cdot \mathcal{F} = \mathcal{D}, \quad \text{where} \quad \mathcal{A} = \frac{\bar{q}'^2}{2\partial \bar{q}/\partial y}, \quad \mathcal{D} = \frac{\bar{D}' \bar{q}'}{\partial \bar{q}/\partial y}. \quad (\text{T.5})$$

The quantity \mathcal{A} is a *wave activity density*, and \mathcal{D} is its dissipation. For nearly plane waves, \mathcal{A} and \mathcal{F} are connected by the *group velocity property*,

$$\mathcal{F} = (\mathcal{F}^y, \mathcal{F}^z) = \mathbf{c}_g \mathcal{A}, \quad (\text{T.6})$$

where \mathbf{c}_g is the group velocity of the waves. If the waves are steady ($\partial \mathcal{A} / \partial t = 0$) and dissipationless ($\mathcal{D} = 0$) then $\nabla \cdot \mathcal{F} = 0$ and, using (T.1) and (T.3) there is no wave-induced acceleration of the mean flow. More commonly there is enstrophy dissipation, or wave-breaking, and $\nabla \cdot \mathcal{F} < 0$; such *wave drag* leads to flow deceleration and/or a poleward residual meridional velocity.

be directly proportional to the *form drag* between isentropic layers. Recall from section 3.5 that the form drag, τ_d , at an interface between two layers of shallow water is

$$\tau_d = -\overline{\eta' \frac{\partial p'}{\partial x}}, \quad (7.129)$$

where η is the interfacial displacement. But from (7.82) $\eta' = -b'/N^2$ and with this and geostrophic balance we have

$$\tau_d = \frac{\rho_0 f_0}{N^2} \overline{v' b'}. \quad (7.130)$$

Thus, the vertical component of the EP flux (i.e., the meridional buoyancy flux) is in fact a real stress acting on a fluid layer and equal to the momentum flux caused by the wavy interface. The net momentum convergence into an infinitesimal layer of mean thickness \bar{h} is then [cf. (3.63)],

$$F_d = \bar{h} \frac{\partial \tau_d}{\partial z} = \bar{h} \rho_0 f_0 \frac{\partial}{\partial z} \left(\frac{\overline{v' b'}}{N^2} \right), \quad (7.131)$$

and a layer of mean thickness \bar{h} is accelerated according to

$$\frac{\partial \bar{u}}{\partial t} = f_0 \frac{\partial}{\partial z} \left(\frac{\overline{v' b'}}{\partial_z \bar{b}} \right) + F_m. \quad (7.132)$$

The appearance of the buoyancy flux is really a consequence of the way we have chosen to average the equations: obtaining (7.132) involved averaging the forces over an isentropic layer, and given this it can only be the residual circulation that contributes to the Coriolis force. One might say that the vertical component of the EP flux is a force in drag, masquerading as a buoyancy flux.

7.5 INFLUENCE OF EDDIES ON THE MEAN FLOW IN THE EADY PROBLEM

We now consider the eddy fluxes in the Eady problem, and, in particular, how these might feed back on to the mean flow. Because of the simplicity of the setting the problem can be fully solved in both the Eulerian or residual frameworks, and it is therefore a very instructive example.¹²

7.5.1 Formulation

Let us first distinguish between the basic flow, the zonal mean fields, and the perturbation. The basic flow is the flow around which the equations of motion are linearized; this flow is unstable, and the perturbations, assumed to be small, grow exponentially with time. Because the perturbations are formally always small they do not affect the basic flow, but they do produce changes in the zonal mean velocity and buoyancy fields. In Eulerian form this is represented by,

$$\frac{\partial \bar{u}}{\partial t} = f_0 \bar{v} - \frac{\partial \bar{u} \bar{v}'}{\partial y}, \quad \frac{\partial \bar{b}}{\partial t} = -N^2 \bar{w} - \frac{\partial \bar{b} \bar{v}'}{\partial y}, \quad (7.133)$$

and the TEM version of these equation is

$$\frac{\partial \bar{u}}{\partial t} = f_0 \bar{v}^* + \bar{v}' \bar{q}', \quad \frac{\partial \bar{b}}{\partial t} = -N^2 \bar{w}^*, \quad (7.134)$$

where in the Eady problem $\partial_y(\bar{u}'v')$ and $\bar{v}'q'$ are both zero. We can calculate the perturbation quantities from the solution to the Eady problem (e.g., calculate $\bar{v}'b'$) and thus infer the structure of the mean flow tendencies $\partial\bar{u}/\partial t$ and $\partial\bar{b}/\partial t$ and the meridional circulation, (\bar{v}, \bar{w}) or (\bar{v}^*, \bar{w}^*) . All of these fields are perturbation quantities and all are exponentially growing, and so in reality they will eventually have a finite effect on the pre-existing zonal flow, but in the Eady problem, or any similar linear problem, such rectification is assumed to be small and is neglected.

Using the thermal wind relation, $f_0\partial_z\bar{u} = -\partial_y\bar{b}$ to eliminate time derivatives in (7.133) gives an equation for the meridional streamfunction ψ_E , namely,

$$\frac{L^2}{L_d^2} \frac{\partial^2 \psi_E}{\partial z^2} + \frac{\partial^2 \psi_E}{\partial y^2} = -\frac{1}{N^2} \frac{\partial^2 \bar{b}'v'}{\partial y^2}, \quad (7.135)$$

where $(\bar{v}, \bar{w}) = (-\partial\psi_E/\partial z, \partial\psi_E/\partial y)$ and we have non-dimensionalized z with D and y with L . The boundary conditions are that $\psi_E = 0$ at $y = 0, L$ and $z = 0, D$. Similarly, and analogously to (7.68), we obtain an equation for the residual streamfunction, ψ^* , namely

$$\frac{L^2}{L_d^2} \frac{\partial^2 \psi^*}{\partial z^2} + \frac{\partial^2 \psi^*}{\partial y^2} = 0, \quad (7.136)$$

where now the boundary conditions are that $N^2\bar{w}^* = \partial\bar{v}'b'/\partial y$ at the upper and lower boundaries, and $\bar{v} = 0$ at the lateral boundaries. In terms of the residual streamfunction this is

$$\psi^* = \frac{1}{N^2} \bar{v}'b', \text{ at } z = 0, 1, \quad \psi^* = 0, \text{ at } y = 0, 1. \quad (7.137)$$

The residual and overturning circulations are related by (7.63a), and (7.136) and (7.137) are, at one level, simply different representations of the same problem, connected by a simple mathematical transformation. However, the residual streamfunction better represents the total transport of the fluid. Equation (7.136) is particularly simple, because of the absence of potential vorticity fluxes in the interior, and it is apparent that the residual circulation is driven by boundary sources. We care only about the spatial structure of the right-hand sides of (7.135) and the boundary conditions of (7.137). The former is given by

$$-\frac{\partial^2 \bar{b}'v'}{\partial y^2} \propto -\frac{\partial^2}{\partial y^2} \sin^2 ly = -2l^2 \cos 2ly. \quad (7.138)$$

The eddy heat fluxes in the Eady problem are independent of height, as may be calculated explicitly from the solutions of chapter 6. In fact, the result follows without detailed calculation, by first noting that the eddy potential vorticity flux is zero because the basic state has zero QG potential vorticity and therefore none may be generated. Further, because the basic state does not vary in y there can be no momentum flux convergence in the y -direction, and so the momentum flux itself is zero if it is zero on the boundary. Thus [using for example (7.21) and (7.22)] the eddy heat flux is independent of height and the EP vectors are directed purely vertically (Fig. 7.2).

The boundary conditions for the residual circulation are

$$\psi^*(y, 0) = \psi^*(y, 1) \propto \sin^2 ly. \quad (7.139)$$

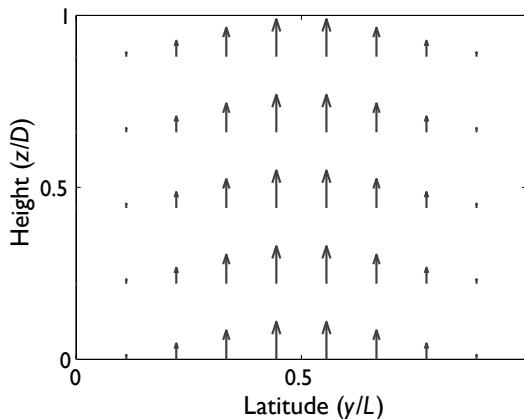


Fig. 7.2 The Eliassen-Palm vector in the Eady problem.

7.5.2 Solution

The solutions to (7.135) and (7.136) may be obtained either analytically or numerically. In a domain $0 < y < 1$ and $0 < z < 1$ the residual streamfunction for $l = \pi$ is given by:

$$\begin{aligned} \psi^* &= \sum_{n=1}^{\infty} A_n \sin[(2n-1)\pi y] \frac{\cosh[L_d \pi(2n-1)(z-0.5)/L]}{\cosh[L_d \pi(2n-1)/2L]}, \\ A_n &= \frac{2}{\pi(2n-1)} - \frac{1}{\pi(2n-1)-2l} - \frac{1}{\pi(2n-1)+2l}. \end{aligned} \quad (7.140)$$

The solution is obtained by first projecting the boundary conditions [proportional to $\sin^2 ly$, or $(1 - \cos 2ly)/2$] on to the eigenfunctions of the horizontal part of the Laplacian (i.e., sine functions), and this gives the coefficients of A_n . The vertical structure is then obtained by solving $(L/L_d)^2 \partial_z^2 \psi^* = -\partial_y^2 \psi^*$, which gives the cosh functions. The series converges very quickly, and the first term in the series captures the dominant structure of the solution, essentially because, for $l = \pi$, $\sin ly$ is not unlike $\sin^2 ly$ on the interval $[0, 1]$.

The Eulerian circulation is obtained from the residual circulation using (7.63a) and so by the addition of a field independent of z and proportional to $\sin^2 ly$. The resulting structure is dominated by this and the first term of (7.140) (proportional to $\sin ly$) and, noting that the circulation is symmetric about $z = 0.5$, we obtain a circulation dominated by a single cell, with equatorward motion aloft and poleward motion near the surface (Fig. 7.3). The heat flux convergence in high latitudes is leading to mean rising motion, with the precise shape of the streamfunction determined by the boundary conditions. Although this is true, the heat flux arises *because* of the motion of fluid parcels, so it may be a little misleading to infer, as one might from the Eulerian streamfunction, that the heat flux *causes* the individual parcels to rise or sink in this fashion. The residual streamfunction is a better indicator of the total mass transport and, perhaps as one might intuitively expect, these show parcels rising in the low latitudes and sinking in high latitudes, providing a tendency to flatten the isopycnals and to reduce the meridional temperature gradient.

The residual circulation also shows fluid entering or leaving the domain at the boundary — what does this represent? Suppose that instead of solving the continuous problem we had posed the problem in a finite number of layers (and we explicitly consider the two-layer

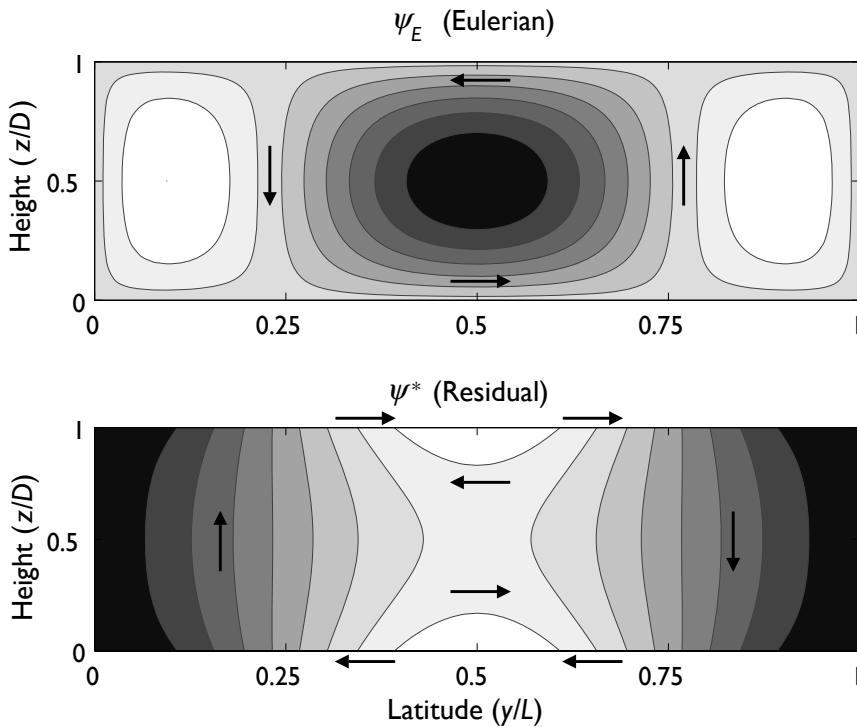


Fig. 7.3 The Eulerian streamfunction (top) and the residual streamfunction for the Eady problem, calculated using (7.135) and (7.136), with $L^2/L_d^2 = 9$.

problem below). As the number of layers increases the solutions to the linear baroclinic instability problem approaches that of the Eady problem (e.g., Fig. 6.13); however, as we saw in section 7.3 the residual circulation is closed in the layered model, and the sum over all the layers of the meridional transport vanishes. Now, in the layered model the vertical boundary conditions are built in to the representation by way of a redefinition of the potential vorticity of the top and bottom layers, so that, in the layered version of the Eady problem there appears to be a potential vorticity gradient in these two layers, instead of a buoyancy gradient at the boundary. The residual circulation is then closed by a return flow that occurs only in the top and bottom layers, and as the number of layers increases this flow is confined to a thinner and thinner layer, and to a delta-function in the continuous limit. To indicate this we have placed arrows just above and below the domain in Fig. 7.3. (This equivalence between boundary conditions and delta-function sources is the same as that giving rise to the delta-function boundary layer of section 5.4.3.)

The effect on the mean flow is inferred directly from the residual circulation: the mean flow acceleration is proportional to \bar{v}^* and the buoyancy tendency is proportional to $-\bar{w}^*$, and these are plotted in Figs. 7.4 and 7.5. Because there is no momentum flux convergence in the problem the zonal flow tendency is entirely baroclinic — its vertical integral is zero — and over most of the domain is such as to reduce the mean shear. Consistently (using thermal wind) the buoyancy tendency is such as to reduce the meridional temperature gradient; that

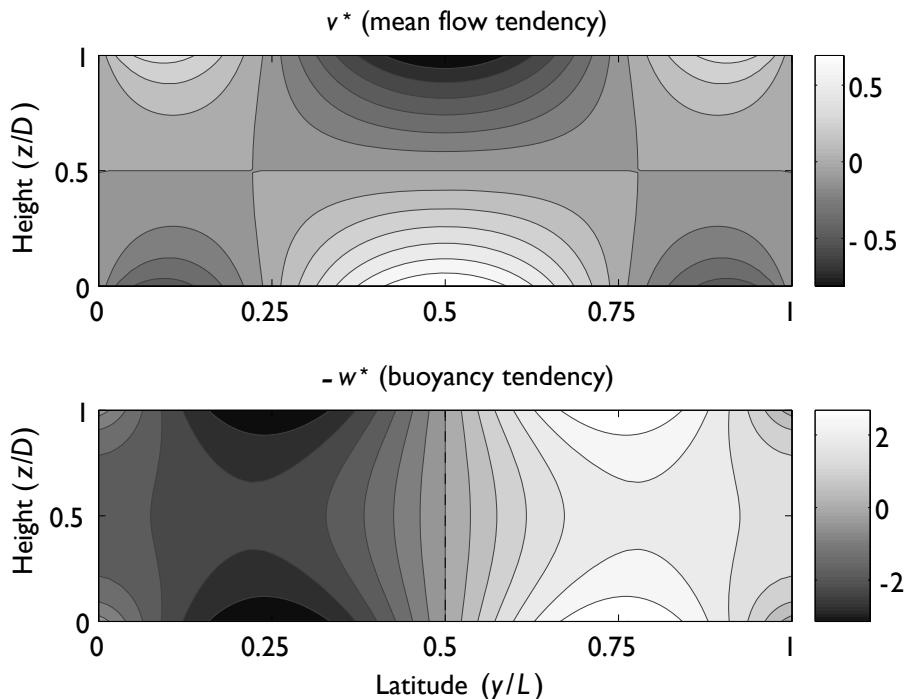


Fig. 7.4 The tendency of the zonal mean flow ($\partial \bar{u} / \partial t$) and the buoyancy ($\partial \bar{b} / \partial t$) for the Eady problem. Lighter (darker) shading means a positive (negative) tendency, but the units themselves are arbitrary.

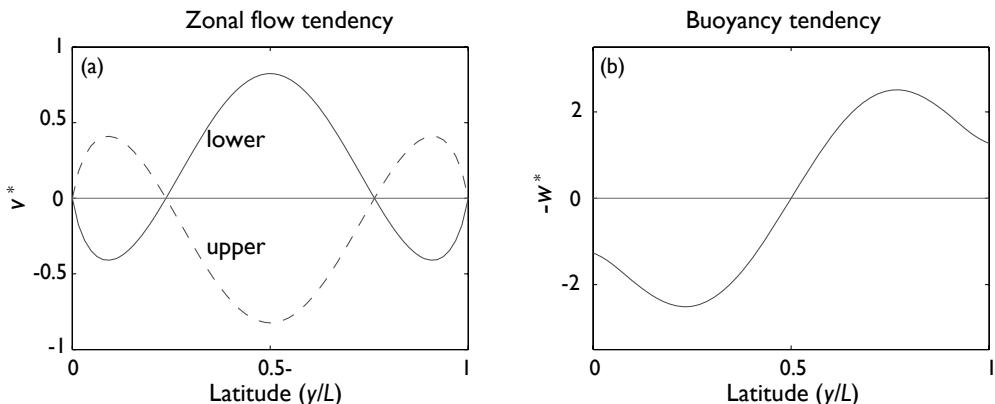


Fig. 7.5 (a) The tendency of the zonal mean flow ($\partial \bar{u} / \partial t$) just below the upper lid (dashed) and just above the surface (solid) in the Eady problem. The vertically integrated tendency is zero. (b) The vertically averaged buoyancy tendency.

is, the instabilities act to transport heat polewards and so reduce the instability of the mean flow.

7.5.3 The two-level problem

The residual circulation and mean-flow tendencies can also be calculated for the two-level (Phillips) problem, with the β -effect. The potential vorticity fluxes in each layer are non-zero and the mean flow equations are, for $i = 1, 2$,

$$\frac{\partial \bar{u}_i}{\partial t} = f_0 \bar{v}_i^* + \bar{v}'_i \bar{q}'_i, \quad \frac{\partial \bar{b}}{\partial t} = -N^2 \bar{w}^*. \quad (7.141)$$

The vertical velocity and buoyancy are evaluated at mid-depth, and the thermal wind equation is $\bar{u}_1 - \bar{u}_2 = -(D/2) \partial_y \bar{b}$ and, by mass conservation, $\bar{v}_1^* = -\bar{v}_2^*$. If we define a residual streamfunction ψ^* such that

$$\bar{v}_1^* = -\bar{v}_2^* = \psi^*, \quad \bar{w}^* = \frac{\partial \psi^*}{\partial y}, \quad (7.142)$$

then eliminating time derivatives in (7.141) gives an equation for the residual streamfunction,

$$\frac{\partial^2 \psi^*}{\partial y^2} - \frac{k_d^2}{2} \psi^* = \frac{2f_0 L^2}{N^2 D} (\bar{v}'_1 \bar{q}'_1 - \bar{v}'_2 \bar{q}'_2), \quad (7.143)$$

where $k_d^2/2 = [2f_0/(NH)]^2$ and D is the total depth of the fluid, and we have non-dimensionalized vertical scales by D and horizontal scales by L . As in the Eady problem it is only the spatial structure of the terms on the right-hand side that are relevant, and these may be calculated from the solutions to the two-level instability problem. The main difference from the Eady problem is that the potential vorticity fluxes are non-zero, even in the case with $\beta = 0$: effectively, the boundary fluxes of the Eady problem are absorbed into the potential vorticity fluxes of the two layers. Solving for the residual circulation and interpreting the mean-flow tendencies is left as an exercise for the reader (problem 7.8).

7.6 * NECESSARY CONDITIONS FOR INSTABILITY

Let's take a taxi to the finish line.

Chris Garrett, Ocean Science Meeting, Hawaii 2002.

As we noted in chapter 6, necessary conditions for instability, or sufficient conditions for stability, can be very useful because when satisfied they obviate the need to perform a detailed calculation. In the remainder of this chapter we use the conservation of wave activities — pseudomomentum and pseudoenergy — to derive such conditions. In sections 6.3 and 6.4.3 we derived such conditions assuming the instability to be of normal-mode form. Here we give derivations that are both more general and, in some ways, simpler; they rely on the fact that the potential vorticity flux may be written as a divergence of a vector and therefore vanishes when integrated over a domain, aside from possible boundary contributions.

7.6.1 Stability conditions from pseudomomentum conservation

Consider the perturbation enstrophy equation,

$$\frac{1}{2} \frac{\partial}{\partial t} \overline{q'^2} = - \frac{\partial \overline{q}}{\partial y} \nabla_x \cdot \mathcal{F}, \quad (7.144)$$

where \mathcal{F} is the Eliassen-Palm flux given by (7.21), the overbar is a zonal mean and the divergence is in y - z plane. Dividing by $\partial \overline{q}/\partial y$ and integrating over a domain A which is such that the Eliassen-Palm flux vanishes at the boundaries gives the pseudomomentum conservation law,

$$\int_A \frac{\partial}{\partial t} \left(\frac{\overline{q'^2}}{\partial_y \overline{q}} \right) dy dz = 0. \quad (7.145)$$

Equation (7.145) implies that, in the *norm* $[\overline{q'^2}/\partial_y \overline{q}]$, the perturbation cannot grow unless $\partial \overline{q}/\partial y$ changes sign somewhere in the domain, or at the boundaries. This result does not depend upon the instability being of normal-mode form. The simplest result of all occurs in a barotropic problem with no vertical variation. Then $\partial \overline{q}/\partial y = \partial \overline{\zeta}_a/\partial y = \beta - \partial^2 \overline{u}/\partial y^2$, and demanding that this must change sign for an instability reprises the inflection point (Rayleigh-Kuo) condition. In the more general case, if $\partial \overline{q}/\partial y$ changes sign along a vertical line then the instability is called a baroclinic instability, and if it changes sign along a horizontal line the instability is barotropic — these may be taken as the definitions of those terms. A mixed instability has a change of sign along both horizontal and vertical lines.

7.6.2 Inclusion of boundary terms

Suppose now the flow is contained between two flat boundaries, at $z = 0$ and $z = H$. The relevant equations of motion are the potential vorticity evolution in the interior, supplemented by the thermodynamic equation at the boundary. For unforced and inviscid flow these give [cf. (7.10) and (7.13)]

$$\frac{\partial}{\partial t} \left(\frac{1}{2} \frac{\overline{q'^2}}{\partial_y \overline{q}} \right) = - \overline{v' q'}, \quad 0 < z < H, \quad (7.146)$$

and

$$\frac{\partial}{\partial t} \left(\frac{1}{2} \frac{\overline{b'^2}}{\partial_y \overline{b}} \right) = - \overline{v' b'}, \quad z = 0, H. \quad (7.147)$$

The poleward flux of potential vorticity is

$$\overline{v' q'} = - \frac{\partial}{\partial y} \overline{u' v'} + \frac{\partial}{\partial z} \left(\frac{f_0}{N^2} \overline{v' b'} \right), \quad (7.148)$$

and integrating this expression with respect to both y and z gives

$$\int_A \overline{v' q'} dy dz = \left[\frac{f_0}{N^2} \overline{v' b'} \right]_0^H, \quad (7.149)$$

assuming that the meridional boundaries are at quiescent latitudes. Integrating (7.146) over y and z , and using (7.149) gives

$$\frac{\partial}{\partial t} \iint \frac{1}{2} \frac{\overline{q'^2}}{\partial_y \overline{q}} dy dz = - \left[\frac{f_0}{N^2} \overline{v' b'} \right]_0^H. \quad (7.150)$$

Using (7.147) to eliminate $\overline{v' b'}$ finally gives

$$\frac{\partial}{\partial t} \left\{ \iint \frac{1}{2} \frac{\overline{q'^2}}{\partial_y \overline{q}} dy dz - \int \left[\frac{1}{2} \frac{f_0}{N^2} \frac{\overline{b'^2}}{\partial_y \overline{b}} \right]_0^H dy \right\} = 0. \quad (7.151)$$

If this expression is positive or negative definite the perturbation cannot grow and therefore the basic state is stable. Stability thus depends on the meridional gradient of potential vorticity in the interior, and the meridional gradient of buoyancy at the boundary. If $\partial \overline{q} / \partial y$ changes sign in the interior, or $\partial \overline{b} / \partial y$ changes sign at the boundary, we have the potential for instability. If these are both one signed, then various possibilities exist, and using the thermal wind relation ($f_0 \partial \overline{u} / \partial z = -\partial \overline{b} / \partial y$) we obtain.

I. A stable case:

$$\frac{\partial \overline{q}}{\partial y} > 0 \text{ and } \frac{\partial u}{\partial z} \Big|_{z=0} < 0 \text{ and } \frac{\partial u}{\partial z} \Big|_{z=H} > 0 \implies \text{stability.} \quad (7.152)$$

Stability also ensues if all inequalities are switched.

II. Instability via interior-surface interactions:

$$\frac{\partial \overline{q}}{\partial y} > 0 \text{ and } \frac{\partial u}{\partial z} \Big|_{z=0} > 0 \text{ or } \frac{\partial u}{\partial z} \Big|_{z=H} < 0 \implies \text{potential instability.} \quad (7.153)$$

The condition $\partial \overline{q} / \partial y > 0$ and $(\partial u / \partial z)_{z=0} > 0$ is the most common criterion for instability that is met in the atmosphere. In the troposphere we can sometimes ignore contributions of the buoyancy fluxes at the tropopause ($z = H$), and stability is then determined by the interior potential vorticity gradient and the surface buoyancy gradient. Similarly, in the ocean contributions from the ocean floor are normally very small.

III. Instability via edge wave interaction:

$$\frac{\partial u}{\partial z} \Big|_{z=0} > 0 \text{ and } \frac{\partial u}{\partial z} \Big|_{z=H} > 0 \implies \text{potential instability.} \quad (7.154)$$

(And similarly, with both inequalities switched.) Such an instability may occur where the troposphere acts like a lid, as for example in the Eady problem. If $\partial \overline{q} / \partial y = 0$ and there is no lid at $z = H$ (e.g., the Eady problem with no lid) then the instability disappears.

One consequence of the upper boundary condition is that it provides a condition on the depth of the disturbance. In the Eady problem the evolution of the system is determined by temperature evolution at the surface,

$$\frac{Db}{Dt} = 0 \quad \text{at } z = 0, H, \quad (7.155)$$

(where $b = f_0 \partial \psi / \partial z$) and zero potential vorticity in the interior, which implies that

$$\nabla^2 \psi + k_d^2 H^2 \frac{\partial^2 \psi}{\partial z^2} = 0 \quad 0 < z < H, \quad (7.156)$$

where $k_d = f_0/(HN)$. Assuming a solution of the form $b \sim \sin kx$ then the Poisson equation (7.156) becomes

$$H^2 k_d^2 \frac{\partial^2 \psi}{\partial z^2} = k^2 \psi, \quad (7.157)$$

with solutions $\psi = A \exp(-\alpha z) + B \exp(\alpha z)$, where $\alpha^2 = k^2 N^2 / f_0^2$. The scale height of the disturbance is thus

$$h \sim \frac{f_0 L}{2\pi N}. \quad (7.158)$$

where $L \sim 2\pi/k$ is the horizontal scale of the disturbance. If the upper boundary is higher than this, it cannot interact strongly with the surface, because the disturbances at either boundary decay before reaching the other. Put another way, if the structure of the disturbance is such that it is shallower than H , the presence of the upper boundary is not felt. In the Eady problem, we know that the upper boundary must be important, because it is only by its presence that the flow can be unstable. Thus, all unstable modes in the Eady problem must be ‘deep’ in this sense, which can be verified by direct calculation. This condition gives rise to a physical interpretation of the high-wavenumber cut-off: if L is too small, the modes are too shallow to span the full depth of the fluid, and from (7.158) the condition for stability is thus

$$L < L_c = 2\pi \frac{NH}{f_0} \quad \text{or} \quad K > K_c = \frac{f_0}{NH} = L_d. \quad (7.159)$$

where L_c and K_c are the critical length scales and wavenumbers. Wavenumbers larger than the deformation radius are stable in the Eady problem. If β is non-zero, this condition does not apply, because the necessary condition for instability can be satisfied by a combination of a surface temperature gradient and an interior gradient of potential vorticity provided by β , as in condition (II) in section 7.6.2. Thus, we may expect that, if $\beta \neq 0$, higher wavenumbers ($k > k_d$) may be unstable but if so they will be shallow, and this may be confirmed by explicit calculation (see Figs. 6.12 and 6.19). In the two-level model shallow modes are, by construction, not allowed so that high wavenumbers will be stable, with or without beta.

7.7 * NECESSARY CONDITIONS FOR INSTABILITY: USE OF PSEUDOENERGY

In this section we derive another necessary condition for instability, sometimes called an ‘Arnold condition’, that is based on the conservation properties of energy and enstrophy. Such conditions can be derived more generally by variational methods, and these lead to somewhat stronger results (in particular, nonlinear results that do not require the perturbation to be small) but our derivations will be elementary and direct.¹³

7.7.1 Two-dimensional flow

First consider inviscid, incompressible two-dimensional flow governed by the equation of motion

$$\frac{\partial q}{\partial t} + J(\psi, q) = 0, \quad (7.160)$$

where $q = \zeta + f = \nabla^2 \psi + f$ is the absolute vorticity and ψ is the streamfunction. In a steady state, the streamfunction and the potential vorticity are functions of each other so

that

$$q = Q(\Psi) \quad \text{and} \quad \psi = \Psi(Q), \quad (7.161)$$

where Q is a differentiable but otherwise arbitrary function of its argument, and Ψ its functional inverse. Equation (7.160) is then

$$\frac{\partial q}{\partial t} = -\frac{dQ}{d\Psi} J(\Psi, \Psi) = 0 \quad (7.162)$$

and all steady solutions are of the form (7.161). We shall prove that if $d\Psi/dQ > 0$ then the flow is stable, in a sense to be made explicit below. Consider the evolution of perturbations about such a steady state, so that

$$q = Q + q', \quad \psi = \Psi + \psi', \quad (7.163)$$

and we suppose that the perturbation vanishes at the domain boundary or that the boundary conditions are periodic. The potential vorticity perturbation satisfies, in the linear approximation,

$$\frac{\partial q'}{\partial t} + J(\psi', Q) + J(\Psi, q') = 0. \quad (7.164)$$

Now, because potential vorticity is conserved on parcels, any function of potential vorticity is also materially conserved, and in particular

$$\frac{D\Psi(q)}{Dt} = \frac{\partial \Psi}{\partial t} + J(\psi, \Psi) = 0. \quad (7.165)$$

Linearizing this using (7.163) gives

$$\frac{d\Psi}{dQ} \frac{\partial q'}{\partial t} + J(\psi', \Psi) + J\left(\Psi, \frac{d\Psi}{dQ} q'\right) = 0. \quad (7.166)$$

We now form an energy equation from (7.164) by multiplying by $-\psi'$ and integrating over the domain. Integrating the first term by parts we find

$$\frac{d}{dt} \int \frac{1}{2} (\nabla \psi')^2 dA = \int \psi' J(\Psi, q') dA. \quad (7.167)$$

Similarly, from (7.166) we obtain

$$\frac{d}{dt} \int \frac{1}{2} \frac{d\Psi}{dQ} q'^2 dA = - \int \left[q' J(\psi', \Psi) + q' J\left(\Psi, \frac{d\Psi}{dQ} q'\right) \right] dA. \quad (7.168)$$

The second term in square brackets vanishes. This follows using the property of Jacobians, obtained by integrating by parts, that

$$\langle aJ(b, c) \rangle = \langle bJ(c, a) \rangle = \langle cJ(a, b) \rangle = -\langle cJ(b, a) \rangle, \quad (7.169)$$

where the angle brackets denote horizontal integration. Using this we have

$$\begin{aligned} \left\langle q' J\left(\Psi, \frac{d\Psi}{dQ} q'\right) \right\rangle &= - \left\langle \frac{d\Psi}{dQ} q' J(\Psi, q') \right\rangle = -\frac{1}{2} \left\langle \frac{d\Psi}{dQ} J(\Psi, q'^2) \right\rangle \\ &= -\frac{1}{2} \left\langle q'^2 J\left(\frac{d\Psi}{dQ}, \Psi\right) \right\rangle = 0. \end{aligned} \quad (7.170)$$

Adding (7.167) and (7.168) the remaining nonlinear terms cancel and we obtain the conservation law,

$$\boxed{\begin{aligned}\hat{H} &= \frac{1}{2} \int \left[(\nabla \psi')^2 + \frac{d\Psi}{dQ} q'^2 \right] dA \\ \frac{d\hat{H}}{dt} &= 0\end{aligned}}. \quad (7.171)$$

The quantity \hat{H} is known as the *pseudoenergy* of the disturbance and because it is a conserved quantity, quadratic in the wave amplitude, it is (like pseudomomentum) a wave activity. Its conservation holds whether the disturbance is growing, decaying or neutral.

If $d\Psi/dQ$ is positive everywhere the pseudoenergy is a positive-definite quantity, and the growth of the disturbance is then strictly limited, and the basic state is said to be *stable in the sense of Liapunov*. This means that the magnitude of the perturbation, as measured by some norm, is bounded by its initial magnitude. In this case we define the norm

$$\|\psi\|^2 \equiv \int \left[(\nabla \psi)^2 + \frac{d\Psi}{dQ} (\nabla^2 \psi)^2 \right] dA, \quad (7.172)$$

so that

$$\|\psi'(t)\|^2 = \|\psi'(0)\|^2. \quad (7.173)$$

If $d\Psi/dQ > 0$ then, although the energy of the disturbance can grow, its final amplitude is bounded by the initial value of the pseudoenergy, because if perturbation energy is to grow perturbation enstrophy must shrink but it cannot shrink past zero. Normal-mode instability, in which modes grow exponentially, is completely precluded.

If the pseudoenergy is *negative definite* then stability is also assured, but this is a less common situation for it demands that $d\Psi/dQ$ be sufficiently negative so that the (negative of the) enstrophy contribution is always larger than the energy contribution, and this can usually only be satisfied in a sufficiently small domain. To see this, suppose that $q' = \nabla^2 \psi'$, and that in the domain under consideration the Laplacian operator has eigenvalues $-k^2$ where

$$\nabla^2 \psi' = -k^2 \psi' \quad (7.174)$$

and the smallest eigenvalue, by magnitude, is k_0^2 . Then, using Poincaré's inequality,

$$\int (\nabla^2 \psi')^2 dA \geq k_0^2 \int (\nabla \psi')^2 dA, \quad (7.175)$$

a sufficient condition to make \hat{H} negative definite is that

$$\frac{d\Psi}{dQ} < -\frac{1}{k_0^2}. \quad (7.176)$$

As the domain gets bigger, k_0 diminishes and this condition becomes harder to satisfy.¹⁴

Parallel shear flow and Fjørtoft's condition

Consider the stability of a zonal flow (i.e., a flow in the x -direction), that varies only with y . The flow stability condition is then

$$\frac{d\Psi}{dQ} = \frac{d\Psi/dy}{dQ/dy} = -\frac{U - U_s}{\beta - U_{yy}} > 0, \quad (7.177)$$

where U_s is a constant, representing an arbitrary, constant, zonal flow. The last equality follows because the problem is Galilean invariant, and we are therefore at liberty to choose U_s arbitrarily. To connect this with Fjørtoft's condition (chapter 6) multiply the top and bottom by $(\beta - U_{yy})$, whence we see that a sufficient condition for stability is that $(U - U_s)(\beta - U_{yy})$ is everywhere negative. The derivation here, unlike our earlier one in section 6.3.2, makes it clear that the condition does not only apply to normal-mode instabilities.

7.7.2 * Stratified quasi-geostrophic flow

The extension of the pseudoenergy arguments to quasi-geostrophic flow is mostly straightforward, but with a complication from the vertical boundary conditions at the surface and at an upper boundary, and the trusting reader may wish to skip straight to the results, (7.182)–(7.184).¹⁵ For definiteness, we consider Boussinesq, β -plane quasi-geostrophic flow confined between flat rigid surfaces at $z = 0$ and $z = H$. The interior flow is governed by the familiar potential vorticity equation $Dq/Dt = 0$ and the buoyancy equation $Db/Dt = 0$ at the two boundaries, where

$$q = \nabla^2 \psi + \beta y + \frac{\partial}{\partial z} \left(S(z) \frac{\partial \psi}{\partial z} \right), \quad b = f_0 \frac{\partial \psi}{\partial z}, \quad (7.178)$$

and $S(z) = f_0^2/N^2$ is positive. The basic state ($\psi = \Psi, q = Q, b = B_1, B_2$) satisfies

$$\begin{aligned} \psi &= \Psi(Q), \quad 0 < z < H, \\ \psi &= \Psi_1(B_1), \quad z = 0 \quad \text{and} \quad \psi = \Psi_2(B_2), \quad z = H. \end{aligned} \quad (7.179)$$

Analogous to the barotropic case, we obtain the equations of motion for the interior perturbation

$$\frac{\partial q'}{\partial t} + J(\psi', Q) + J(\Psi, q') = 0, \quad (7.180a)$$

$$\frac{d\Psi}{dQ} \frac{\partial q'}{\partial t} + J(\psi', \Psi) + J\left(\Psi, \frac{d\Psi}{dQ} q'\right) = 0, \quad (7.180b)$$

and at the two boundaries

$$\frac{\partial b'}{\partial t} + J(\psi', B_i) + J(\Psi_i, b') = 0, \quad (7.181a)$$

$$\frac{d\Psi_i}{dB_i} \frac{\partial b'}{\partial t} + J(\psi', \Psi_i) + J\left(\Psi_i, \frac{d\Psi_i}{dB_i} b'\right) = 0, \quad (7.181b)$$

for $i = 1, 2$. (By $d\Psi_i/dB_i$ we mean the derivative of Ψ_i with respect to its argument, evaluated at B_i .) From these equations, we form the pseudoenergy by multiplying (7.180a) by $-\psi'$, (7.180b) by q' , and (7.181a) by ψ' , (7.181b) by b' . After some manipulation we obtain the pseudoenergy conservation law:

$$\hat{H} = \mathcal{E} + \mathcal{Z} + \mathcal{B}_1 + \mathcal{B}_2$$

$$\frac{d\hat{H}}{dt} = 0$$

(7.182)

where

$$\begin{aligned}\mathcal{E} &= \frac{1}{2} \left\{ (\nabla \psi')^2 + S \left(\frac{\partial \psi'}{\partial z} \right)^2 \right\}, & \mathcal{Z} &= \frac{1}{2} \left\{ \frac{d\Psi}{dQ} q'^2 \right\}, \\ \mathcal{B}_1 &= \frac{1}{2} \left\langle \frac{S(0)}{f_0} \frac{d\Psi_1}{dB_1} b'(0)^2 \right\rangle, & \mathcal{B}_2 &= -\frac{1}{2} \left\langle \frac{S(H)}{f_0} \frac{d\Psi_2}{dB_2} b'(H)^2 \right\rangle.\end{aligned}\quad (7.183)$$

where the curly brackets denote a three-dimensional integration over the fluid interior, and the angle brackets denote a horizontal integration over the boundary surfaces at 0 and H . The pseudoenergy \hat{H} is positive-definite, and therefore stability is assured in that norm, if all of the following conditions are satisfied:

$$\frac{d\Psi}{dQ} > 0, \quad \frac{1}{f_0} \frac{d\Psi_1}{dB_1} > 0, \quad \frac{1}{f_0} \frac{d\Psi_2}{dB_2} < 0. \quad (7.184)$$

If the flow is compressible, the potential vorticity is $q = \nabla^2 \psi + \beta y + \rho^{-1} \partial_z (\rho S \partial_z \psi)$, where $\rho = \rho(z)$, but the final stability conditions are unaltered. If the upper boundary is then removed to infinity where $\rho(z) = 0$ then only the lower boundary condition contributes to (7.184). In the layered form of the quasi-geostrophic equations the vertical boundary conditions are built in to the definitions of potential vorticity in the top and bottom layers. In this case, a sufficient condition for stability is that $d\Psi/dQ > 0$ in each layer. Indeed, an alternate derivation of (7.182)–(7.184) would be to incorporate the boundary conditions on buoyancy into the definition of potential vorticity by the delta-function construction of section 5.4.3.

Zonal shear flow

Consider now zonally uniform zonal flows, such as might give rise to baroclinic instability in a channel. The fields are then functions of y and z only, and the sufficient conditions for stability are:

$$\begin{aligned}\frac{d\Psi}{dQ} &= \frac{\partial \Psi / \partial y}{\partial Q / \partial y} = -\frac{U}{dQ/dy} > 0, \\ \frac{d\Psi_1}{dB_1} &= \frac{d\Psi_1 / dy}{dB_1 / dy} = \frac{U(0)}{dU(0)/dz} > 0, \\ \frac{d\Psi_2}{dB_2} &= \frac{d\Psi_2 / dy}{dB_2 / dy} = \frac{U(H)}{dU(H)/dz} < 0.\end{aligned}\quad (7.185)$$

using the thermal wind relation, and setting $f_0 = 1$ (its value is irrelevant). These results generalize Fjørtoft's condition to the stratified case,¹⁶ and as in that case we are at liberty to add a uniform zonal flow to all the velocities.

7.7.3 * Applications to baroclinic instability

We may use the stability conditions derived above to provide a few more results about baroclinic instability, including an alternate derivation of the minimum shear criterion in two-layer flow, and a derivation of the high-wavenumber cut-off to instability. In what follows we do not derive any new criteria; rather, derivations make it apparent that the criteria are not restricted to perturbations of normal-mode form.

Minimum shear in two-layer flow

We consider two layers of equal depth, on a flat-bottomed β -plane with basic state

$$\Psi_1 = -U_1 y, \quad \Psi_2 = -U_2 y \quad (7.186a)$$

$$Q_1 = \beta y - \frac{k_d^2}{2} (U_2 - U_1) y, \quad Q_2 = \beta y - \frac{k_d^2}{2} (U_1 - U_2) y. \quad (7.186b)$$

This state is characterized by $Q_i = \gamma_i \Psi_i$ where

$$\gamma_1 = -\frac{(\beta + k_d^2 \hat{U})}{(\bar{U} + \hat{U})}, \quad \gamma_2 = -\frac{(\beta - k_d^2 \hat{U})}{(\bar{U} - \hat{U})}, \quad (7.187)$$

with $\bar{U} = (U_1 + U_2)/2$ and $\hat{U} = (U_1 - U_2)/2$. The barotropic flow does not affect the stability properties, so without loss of generality we may choose $\bar{U} < -\hat{U}$, and this makes $\gamma_1 > 0$. Then γ_2 is also positive if $\beta > k_d^2 \hat{U}/2$. Thus, a sufficient condition for stability is that

$$\hat{U} < \frac{\beta}{k_d^2}, \quad (7.188)$$

as obtained in chapter 6. The derivation here shows that the condition is not restricted to normal-mode instabilities.¹⁷

Use of pseudomomentum conservation provides an alternative derivation of the same result. The flow will also be stable if in both layers $\partial Q/\partial y > 0$, for then the conserved pseudomomentum will be positive definite. If $U_1 > U_2$ then, from (7.186) $dQ_1/dy > 0$. The flow will be stable if $dQ_2/dy > 0$, and this gives

$$\hat{U} = \frac{1}{2} (U_1 - U_2) < \frac{\beta}{k_d^2}, \quad (7.189)$$

as in (7.188).

The high-wavenumber cut-off in two-layer baroclinic instability

We can use a pseudoenergy argument to show that there is a high-wavenumber cut-off to two-layer baroclinic instability, with the basic state (7.186). The conserved pseudoenergy analogous to (7.182) and (7.183) is readily found to be

$$\hat{H} = \left\langle (\nabla \psi'_1)^2 + (\nabla \psi'_2)^2 + \frac{1}{2} k_d^2 (\psi'_1 - \psi'_2)^2 + \frac{q'_1^2}{\gamma_1} + \frac{q'_2^2}{\gamma_2} \right\rangle = 0. \quad (7.190)$$

Let us choose (without loss of generality) the barotropic flow to be $\bar{U} = \beta/k_d^2$. We then have $\gamma_1 = \gamma_2 = -1/k_d^2$, and the pseudoenergy is then just the actual energy minus k_d^{-2} times the total enstrophy. If we define $\psi = (\psi'_1 + \psi'_2)/2$ and $\tau = (\psi'_1 - \psi'_2)/2$ then, using (9.31a) and (9.34), (7.190) may be expressed as

$$\hat{H} = \left\langle (\nabla \psi)^2 + (\nabla \tau)^2 + k_d^2 \tau^2 - k_d^{-2} \left\{ (\nabla^2 \psi)^2 + [(\nabla^2 - k_d^2) \tau]^2 \right\} \right\rangle. \quad (7.191)$$

Now, let us express the fields as Fourier sums,

$$(\tau, \psi) = \sum_{k,l} (\tilde{\tau}_{k,l}, \tilde{\psi}_{k,l}) e^{i(kx+ly)}. \quad (7.192)$$

(This expression assumes a doubly-periodic domain; essentially the same end-result is obtained in a channel.) The pseudoenergy may then be written as

$$\hat{H} = \sum_{k,l} \left[K^2 \tilde{\psi}_{k,l}^2 (k_d^2 - K^2) + K'^2 \tilde{\tau}_{k,l}^2 (k_d^2 - K'^2) \right] \quad (7.193)$$

where $K^2 = k^2 + l^2$ and $K'^2 = K^2 + k_d^2$. If the deformation radius is sufficiently large (or the domain sufficiently small) that $K^2 > k_d^2$, then the pseudoenergy is *negative-definite*, so the flow is stable, no matter what the shear may be. Such a situation might arise on a planet whose circumference was less than the deformation radius, or in a small ocean basin. In the linear problem, in which perturbation modes do not interact, horizontal wavenumbers with $k^2 > k_d^2$ are stable and there is thus a high-wavenumber cut-off to instability, as was found in chapter 6 by direct calculation.

Notes

- 1 After Eliassen & Palm (1961).
- 2 The form of this derivation was originally given by Hayes (1977) in the context of wave energy. See Vanneste & Shepherd (1998) for extensions.
- 3 Andrews & McIntyre (1976), Ripa (1981) and Held (1985). See also problem 7.3.
- 4 These restrictions on the basic state are not necessary to prove orthogonality, but they make the algebra simpler. Also, we pay no attention here to the nature of the eigenvalues of (7.50), which, in general, consist of both a discrete and a continuous spectrum. See Farrell (1984) and McIntyre & Shepherd (1987).
- 5 The TEM was introduced by Andrews & McIntyre (1976, 1978) and Boyd (1976). A precursor is the paper of Riehl & Fultz (1957), who noted the shortcomings of zonal averaging in uncovering the meaning of indirect cells in laboratory experiments, and by extension the atmosphere.
- 6 The main result of this subsection was originally obtained by McIntosh & McDougall (1996). I thank A. Plumb for a discussion about the derivation given here. See also de Szoeke & Bennett (1993) for related earlier work, and Juckes (2001) and Nurser & Lee (2004) for generalizations.
- 7 This problem can be worked around in some cases (Plumb 1990, Greatbatch 1998).
- 8 Plumb & Ferrari (2005), Ferrari & Plumb (2003), who also discuss the momentum budget. See also Held & Schneider (1999).
- 9 Non-acceleration arguments have a long history, with contributions from Charney & Drazin (1961), Eliassen & Palm (1961), Holton (1974) and Boyd (1976). Andrews & McIntyre (1978) put these results in the context of the EP flux and the TEM formalism, and Dunkerton (1980) reviews and provides examples.
- 10 Conservation laws of this ilk, their connection to the underlying symmetries of the basic state and (relatedly) their finite-amplitude extension, are discussed by McIntyre & Shepherd (1987) and Shepherd (1990). Conservation of momentum is related to the translational invariance of the medium; conservation of \mathcal{A} may be shown to be related to the translational invariance of the basic state, and hence the appellation ‘pseudomomentum’.
- 11 See Andrews & McIntyre (1978).

- 12 Steve Garner and Raffaele Ferrari provided very helpful input to this section.
- 13 The original papers are Arnold (1965, 1966), and a number of results were developed by Holm *et al.* (1985). See Shepherd (1990) for a review.
- 14 The stability criterion is sometimes referred to as ‘Arnold’s second condition’. More discussion, especially with regard to boundary conditions, is given in McIntyre & Shepherd (1987).
- 15 Blumen (1968), but the method we use is more direct.
- 16 Pedlosky (1964) derived these conditions by a normal-mode approach.
- 17 Pierini & Vulpiani (1981) and Vallis (1985) consider the finite-amplitude case.

Further reading

Andrews, D. G., Holton, J. R. & Leovy, C.B., 1987. *Middle Atmosphere Dynamics*.

Provides a discussion of a number of topics in wave dynamics and wave-mean flow interaction, including the TEM, mainly in the context of stratospheric dynamics.

Problems

- 7.1 Prove that

$$\langle aJ(b,c) \rangle = \langle bJ(c,a) \rangle = \langle cJ(a,b) \rangle = -\langle cJ(b,a) \rangle \quad (\text{P7.1})$$

where the angle brackets denote a horizontal integration, and the boundary conditions correspond to no-normal-flow or periodic.

- 7.2 Consider an axisymmetric barotropic shear flow given by

$$\zeta = \begin{cases} 2\Omega & r \leq R \\ 0 & r > R \end{cases} \quad (\text{P7.2})$$

where Ω and R are constants. Thus, the inner region is in solid body rotation and the outer region is irrotational, and we suppose that the velocity is continuous. [Is this implied by (P7.2) or is it an extra condition?] Suppose that the boundary between the two regions is perturbed. Find the phase speed of this disturbance in terms of its azimuthal wavenumber. Show that for large wavenumbers this reduces to the dispersion relation for a point jet [i.e., $c = U_0 + (\zeta_1 - \zeta_2)/2k$ — see (6.37)], and define what ‘large’ means in this context.

- 7.3 ♦ Show that perturbations to a horizontally sheared flow are orthogonal with respect to the wave activity norm. You may restrict attention to two-dimensional (barotropic) flow on the β -plane.

Partial solution. The eigenvalue value equation is

$$(\bar{u}\nabla_k^2 + \partial_y \bar{q})\psi = c\nabla_k^2\psi \quad (\text{P7.3})$$

where $\nabla_k^2 = \partial_{yy} - k^2$. If $q = \nabla_k^2\psi$ then the eigenvalue equation may be written

$$Mq \equiv (\bar{u} + \partial_y \bar{q} \nabla_k^{-2})q = cq \quad (\text{P7.4})$$

or

$$Nq \equiv \left(\frac{\bar{u}}{\partial_y \bar{q}} + \nabla_k^{-2} \right) q = c \frac{q}{\partial_y \bar{q}} \quad (\text{P7.5})$$

The operator N on the left-hand side of (P7.5) is self-adjoint (show this) so that the eigenfunctions associated with two different eigenvalues are orthogonal with respect to $\partial_y \bar{q}$; that is, for $n \neq m$,

$$\int \int \frac{q_n q_m}{\partial_y \bar{q}} \, dy = 0. \quad (\text{P7.6})$$

If this derivation holds, why is it not simply the case that, from (P7.4), that

$$\int q_n q_m \, dy = 0, \quad (P7.7)$$

for $n \neq m$? (Is M self-adjoint?)

- 7.4 Obtain an expression for the EP flux due to equatorial Kelvin waves. What is the sign of the wave drag in the regions of Kelvin wave generation and dissipation?
- 7.5 ♦ Can the high-wavenumber cut-off to instability in the Eady problem be obtained by wave-activity arguments (e.g., by proving the pseudoenergy is negative definite, as in the two-layer problem). If so, do so.
- 7.6 ♦ *Stability conditions in the continuously stratified QC model*

Consider the modified Eady problem (Boussinesq, uniform stratification, flow contained between two flat horizontal surfaces at 0 and H), but instead of a uniform shear suppose that the basic state is given by

$$U = -U_0 \cos(\pi p z / H) \quad (P7.8)$$

where $U_0 > 0$, and allow β to be non-zero.

- (a) Show there is a critical shear, and that this diminishes as p increases.
 (b) Show there is a high-wavenumber cut-off to instability.

Compare the results to those of the two-layer model.

Sketch of solution.

- (a) The basic state has no temperature gradient at the boundary, and so stability is assured if $\partial Q / \partial y$ is positive everywhere. The basic state potential vorticity is $Q = \nabla^2 \Psi + (f_0^2 / N^2) \partial^2 \Psi / \partial z^2 + \beta y$, so that

$$\frac{dQ}{dy} = \beta - k_d^2 \pi^2 p^2 U_0 \cos(\pi p z / H), \quad (P7.9)$$

- where $k_d^2 = f_0^2 H^2 / N^2$ and $U = -\partial \Psi / \partial y$. Thus, stability is assured if $U_0 < (\beta / k_d^2 \pi^2 p^2)$.
 (b) Choose a barotropic flow equal to $-\beta / (k_d^2 \pi^2 p^2)$ so that the mean flow is given by $U = U_0 \cos(\pi p z / H) - \beta / (k_d^2 \pi^2 p^2)$. Then $y(z) \equiv Q / \Psi = -k_d^2 \pi^2 p^2$. Expand the perturbation streamfunction as $\psi = \sum_{k,\alpha} \psi_{k,\alpha} e^{ik \cdot x} \cos \alpha z$ and obtain an expression for the pseudoenergy analogous to (7.193), and find the conditions under which it is sign-definite.

- 7.7 Obtain, or at least verify, the solution (7.140). Plot it for various values of deformation radii (using appropriate computer software).
- 7.8 ♦ Obtain and plot the residual circulation in the linear two-level (Phillips) baroclinic instability problem, both for $\beta = 0$ and $\beta \neq 0$. Also obtain and plot the corresponding tendencies of the zonally averaged zonal wind and buoyancy fields, and interpret your results. A good answer will include a comparison of the solutions with and without beta, and a comparison of the solutions with those of the Eady problem. [Reading Shepherd (1983) may be helpful.]
- 7.9 ♦ *Balance of terms in the mean flow equations*

- (a) In the Eady problem the mean flow evolves according to (7.133). To what degree is there an instantaneous balance between the terms on the right-hand side? That is, is the mean-flow evolution a residual between two larger terms? (The answer is trivial for the zonal flow evolution, but less so for buoyancy.)
 (b) Repeat this problem for the two-layer problem, in both Eulerian and TEM forms. For the latter, calculate the balance between the potential vorticity flux, the residual meridional flow and the zonal flow tendency.

I shall not today attempt further to define the kinds of material...embraced within that shorthand definition.... But I know it when I see it.

Potter Stewart, Supreme Court Associate Justice, *Jacobellis vs. Ohio*, 1964.

CHAPTER EIGHT

Basic Theory of Incompressible Turbulence

TURBULENCE IS HIGH REYNOLDS NUMBER FLUID FLOW, dominated by nonlinearity, containing both spatial and temporal disorder. No definition is perfect, and it is hard to disentangle a definition from a property, but this statement captures the essential aspects. A turbulent flow has eddies with a spectrum of sizes between some upper and lower bounds, the former often determined by the forcing scale or the domain scale, and the latter usually by viscosity. The individual eddies come and go, and are inherently unpredictable.

The circulation of the atmosphere and ocean is, *inter alia*, the motion of a forced-dissipative fluid subject to various constraints such as rotation and stratification. The larger scales are orders of magnitude larger than the dissipation scale (the scale at which molecular viscosity becomes important) and at many if not all scales the motion is highly nonlinear and quite unpredictable. Thus, we can justifiably say that the atmosphere and ocean are turbulent fluids. Note that here we are not simply talking about the small-scale flows traditionally regarded as turbulent; rather, our main focus will be the large-scale flows associated with baroclinic instability and greatly influenced by rotation and stratification, a kind of turbulence known as geostrophic turbulence. However, before discussing turbulence in the atmosphere and ocean, in this chapter we consider from a fairly elementary standpoint the basic theory of two- and three-dimensional turbulence, and in particular the theory of inertial ranges. We do not provide a comprehensive discussion of turbulence; rather, we provide an introduction to those aspects of most interest or relevance to the dynamical oceanographer or meteorologist. In the next chapter we consider the effects of rotation and stratification and the problem of geostrophic turbulence.

8.1 THE FUNDAMENTAL PROBLEM OF TURBULENCE

Turbulence is a difficult subject because it is nonlinear, and because (and relatedly) there are interactions between scales of motion. Let us first see what difficulties these bring.¹

8.1.1 The closure problem

Although in a turbulent flow it may be virtually impossible to predict the detailed motion of each eddy, the statistical properties — time averages for example — might not be changing and we might like to predict such averages. Effectively, we accept that we cannot predict the weather, but we can try to predict the climate. Even though we know which equations determine the system, this task proves to be very difficult because the equations are nonlinear, and we come up against the *closure problem*. To see what this is, let us decompose the velocity field into mean and fluctuating components,

$$\mathbf{v} = \bar{\mathbf{v}} + \mathbf{v}', \quad (8.1)$$

Here $\bar{\mathbf{v}}$ is the mean velocity field, and \mathbf{v}' is the deviation from that mean. The mean might be a time average, in which case $\bar{\mathbf{v}}$ is a function only of space and not of time, or it might be a time mean over a finite period (e.g., a season if we are dealing with the weather), or it might be some form of ensemble mean. The average of the deviation is, by definition, zero; that is $\bar{\mathbf{v}'} = 0$. The idea is to substitute (8.1) into the momentum equation and try to obtain a closed equation for the mean quantity $\bar{\mathbf{v}}$. Rather than dealing with the full Navier-Stokes equations, let us carry out this program for a model nonlinear system that obeys

$$\frac{du}{dt} + uu + ru = 0, \quad (8.2)$$

where r is a constant. The average of this equation is:

$$\frac{d\bar{u}}{dt} + \bar{u}\bar{u} + r\bar{u} = 0. \quad (8.3)$$

The value of the term $\bar{u}\bar{u}$ is not deducible simply by knowing \bar{u} , since it involves correlations between eddy quantities $\bar{u}'\bar{u}'$. That is, $\bar{u}\bar{u} = \bar{u}\bar{u} + \bar{u}'\bar{u}' \neq \bar{u}\bar{u}$. We can go to the next order to try (vainly!) to obtain an equation for $\bar{u}\bar{u}$. First multiply (8.2) by u to obtain an equation for u^2 , and then average it to yield

$$\frac{1}{2} \frac{d\bar{u}^2}{dt} + \bar{u}\bar{u}\bar{u} + r\bar{u}^2 = 0. \quad (8.4)$$

This equation contains the undetermined cubic term $\bar{u}\bar{u}\bar{u}$. An equation determining this would contain a quartic term, and so on in an unclosed hierarchy. Many methods of closing the hierarchy make assumptions about the relationship of $(n + 1)$ th order terms to n th order terms, for example by supposing that

$$\bar{u}\bar{u}\bar{u}\bar{u} = \alpha\bar{u}\bar{u}\bar{u}\bar{u} + \beta\bar{u}\bar{u}\bar{u}\bar{u}, \quad (8.5)$$

where α and β are some parameters, and closures set in physical space or in spectral space (i.e., acting on the Fourier transformed variables) both exist. If we know that the variables

are distributed normally then such closures can be made exact, but this is not generally the case in fluid turbulence.

This same closure problem arises in the Navier-Stokes equations. If density is constant (say $\rho = 1$) the x -momentum equation for an averaged flow is

$$\frac{\partial \bar{u}}{\partial t} + (\bar{\mathbf{v}} \cdot \nabla) \bar{u} = -\frac{\partial \bar{p}}{\partial x} - \nabla \cdot \bar{\mathbf{v}}' \bar{u}'. \quad (8.6)$$

Written out in full in Cartesian coordinates, the last term is

$$\nabla \cdot \bar{\mathbf{v}}' \bar{u}' = \frac{\partial}{\partial x} \bar{u}' \bar{u}' + \frac{\partial}{\partial y} \bar{u}' \bar{v}' + \frac{\partial}{\partial z} \bar{u}' \bar{w}'. \quad (8.7)$$

These terms, and the similar ones in the y - and z -momentum equations, represent the effects of eddies on the mean flow and are known as *Reynolds stress* terms. The ‘closure problem of turbulence’ may be thought of as finding a representation of such Reynolds stress terms in terms of mean flow quantities. Nobody has been able to close the system, in any useful way, without introducing physical assumptions not directly deducible from the equations of motion themselves. Indeed, not only has the problem not been solved, it is not clear that in general a useful closed-form solution actually exists.

8.1.2 Triad interactions in turbulence

The nonlinear term in the equations of motion not only leads to difficulties in closing the equations, but it leads to interactions among different length scales, and in this section we write the equations of motion in a form that makes this explicit. For algebraic simplicity we will restrict our attention to two-dimensional flows, but very similar considerations also apply in three dimensions, and the details of the algebra following are not of themselves important to subsequent sections.²

The equation of motion for an incompressible fluid in two dimensions [see for example (4.67) or (5.119)] may be written as

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) = F + \nu \nabla^2 \zeta, \quad \zeta = \nabla^2 \psi. \quad (8.8)$$

We include a forcing and viscous term but no Coriolis term. Let us suppose that the fluid is contained in a square, doubly-periodic domain of side L , and let us expand the streamfunction and vorticity in Fourier series so that, with a tilde denoting a Fourier coefficient,

$$\psi(x, y, t) = \sum_{\mathbf{k}} \tilde{\psi}(\mathbf{k}, t) e^{i\mathbf{k} \cdot \mathbf{x}}, \quad \zeta(x, y, t) = \sum_{\mathbf{k}} \tilde{\zeta}(\mathbf{k}, t) e^{i\mathbf{k} \cdot \mathbf{x}}, \quad (8.9)$$

where $\mathbf{k} = i k^x + j k^y$, $\tilde{\zeta} = -k^2 \tilde{\psi}$ where $k^2 = k^{x^2} + k^{y^2}$ and, to ensure that ψ is real, $\tilde{\psi}(k^x, k^y, t) = \tilde{\psi}^*(-k^x, -k^y, t)$, where $*$ denotes the complex conjugate, and this property is known as conjugate symmetry. The summations are over all positive and negative x - and y -wavenumbers, and $\tilde{\psi}(\mathbf{k}, t)$ is shorthand for $\tilde{\psi}(k^x, k^y, t)$. Substituting (8.9) in (8.8) gives, with (for the moment) F and ν both zero,

$$\begin{aligned} \frac{\partial}{\partial t} \sum_{\mathbf{k}} \tilde{\zeta}(\mathbf{k}, t) e^{i\mathbf{k} \cdot \mathbf{x}} &= \sum_{\mathbf{p}} p^x \tilde{\psi}(\mathbf{p}, t) e^{i\mathbf{p} \cdot \mathbf{x}} \times \sum_{\mathbf{q}} q^y \tilde{\zeta}(\mathbf{q}, t) e^{i\mathbf{q} \cdot \mathbf{x}} \\ &\quad - \sum_{\mathbf{p}} p^y \tilde{\psi}(\mathbf{p}, t) e^{i\mathbf{p} \cdot \mathbf{x}} \times \sum_{\mathbf{q}} q^x \tilde{\zeta}(\mathbf{q}, t) e^{i\mathbf{q} \cdot \mathbf{x}}, \end{aligned} \quad (8.10)$$

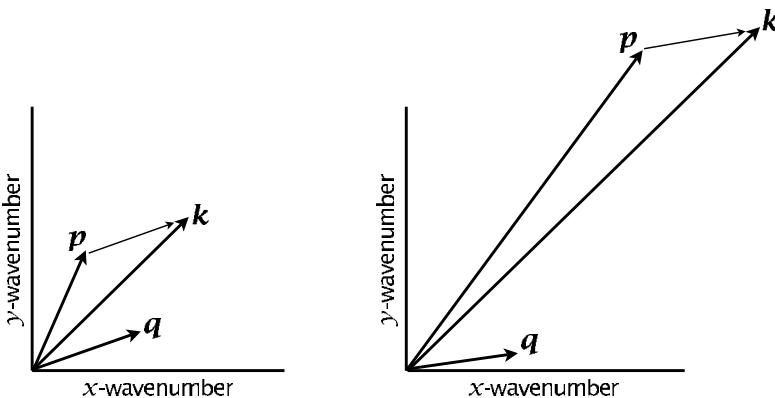


Fig. 8.1 Two interacting triads, each with $\mathbf{k} = \mathbf{p} + \mathbf{q}$. On the left, a local triad with $k \sim p \sim q$. On the right, a non-local triad with $k \sim p \gg q$.

where \mathbf{p} and \mathbf{q} are, like \mathbf{k} , horizontal wave vectors. We may obtain an evolution equation for the wavevector \mathbf{k} by multiplying (8.10) by $\exp(-i\mathbf{k} \cdot \mathbf{x})$ and integrating over the domain, and using the fact that the Fourier modes are orthogonal; that is

$$\int e^{i\mathbf{p} \cdot \mathbf{x}} e^{i\mathbf{q} \cdot \mathbf{x}} dA = L^2 \delta(\mathbf{p} + \mathbf{q}), \quad (8.11)$$

where $\delta(\mathbf{p} + \mathbf{q})$ equals unity if $\mathbf{p} = -\mathbf{q}$ and is zero otherwise. Using this, (8.10) becomes, restoring the forcing and dissipation terms,

$$\frac{\partial}{\partial t} \tilde{\psi}(\mathbf{k}, t) = \sum_{\mathbf{p}, \mathbf{q}} A(\mathbf{k}, \mathbf{p}, \mathbf{q}) \tilde{\psi}(\mathbf{p}, t) \tilde{\psi}(\mathbf{q}, t) + \tilde{F}(\mathbf{k}) - \nu k^2 \tilde{\psi}(\mathbf{k}, t), \quad (8.12)$$

where $A(\mathbf{k}, \mathbf{p}, \mathbf{q}) = (q^2/k^2)(p^x q^y - p^y q^x) \delta(\mathbf{p} + \mathbf{q} - \mathbf{k})$ is an ‘interaction coefficient’, and the summation is over all \mathbf{p} and \mathbf{q} ; however, note that only those wavevector triads with $\mathbf{p} + \mathbf{q} = \mathbf{k}$ make a non-zero contribution, because of presence of the delta function.

Consider, then, a fluid in which just two Fourier modes are initially excited, with wavevectors \mathbf{p} and \mathbf{q} say, along with their conjugate-symmetric partners at $-\mathbf{p}$ and $-\mathbf{q}$. These modes interact [obeying (8.12)] to generate third and fourth wavenumbers, $\mathbf{k} = \mathbf{p} + \mathbf{q}$ and $\mathbf{m} = \mathbf{p} - \mathbf{q}$ (along with their conjugate-symmetric partners). These four wavenumbers can interact among themselves to generate several additional wavenumbers, $\mathbf{k} + \mathbf{p}$, $\mathbf{k} + \mathbf{m}$ and so on, and these in turn lead to still more interactions so potentially filling out the entire spectrum of wavenumbers. The individual interactions are called *triad interactions*, and it is by way of such interactions that energy is transferred between scales in turbulent flows, in both two and three dimensions. The dissipation term does not lead to interactions between modes with different wavevectors; rather, it acts like a drag on each Fourier mode, with a coefficient that increases with wavenumber and therefore that preferentially affects small scales.

The selection rule for triad interactions — that $\mathbf{k} = \mathbf{p} + \mathbf{q}$ — does not restrict the scales of these interacting wavevectors, and the types of triad interactions fall between two extremes:

- (i) local interactions, in which $k \sim p \sim q$;

(ii) non-local interactions, in which $k \sim p \gg q$.

These two kinds of triads are schematically illustrated in Fig. 8.1. Without very detailed analysis of the solutions of the equations of motion — an analysis that is in general impossible for fully-developed turbulence — we cannot say with certainty whether one particular kind of triad interaction dominates. The theory of Kolmogorov considered below, and its two-dimensional analogue, assume that it is the *local* triads that are most important in transferring energy; this is a reasonable assumption because, from the perspective of a small eddy, large eddies appear as a nearly-uniform flow and so simply translate the small eddies around without distorting them and thus without transferring energy between scales.

8.2 THE KOLMOGOROV THEORY

The foundation of many theories of turbulence is the spectral theory of Kolmogorov.³ This theory does not close the equations in quite as explicit a manner as (8.5), but it does provide a prediction for the energy spectrum of a turbulent flow (loosely speaking, how much energy is present at a particular spatial scale) and it does this by suggesting a relationship between the energy spectrum (a second-order quantity in velocity) and the spectral energy flux (a third-order quantity).

8.2.1 The physical picture

Consider high Reynolds number (Re) incompressible flow that is being maintained by some external force. Then the evolution of a system that has $\rho = 1$ is governed by

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\nabla p + \mathbf{F} + \nu \nabla^2 \mathbf{v} \quad (8.13)$$

and

$$\nabla \cdot \mathbf{v} = 0. \quad (8.14)$$

Here, \mathbf{F} is some force we apply to maintain fluid motion — for example, we stir the fluid with a spoon. (One might argue that such stirring is not a force, like gravity, but a continuous changing of the boundary conditions. Having noted this, we treat it as a force.) A simple scale analysis of these equations seems to indicate that the ratio of the size of the inertial terms on the left-hand side to the viscous term is the Reynolds number VL/ν , where V and L are velocity and length scales. To be explicit let us consider the ocean, and take $V = 0.1 \text{ m s}^{-1}$, $L = 1000 \text{ km}$ and $\nu = 10^{-6} \text{ m}^2 \text{ s}^{-1}$. Then $Re = VL/\nu \approx 10^{11}$, and it seems that we can neglect the viscous term on the right-hand side of (8.13). But this can lead to a paradox, as if the fluid is being forced this forcing is likely to put energy into the fluid. To see this, we obtain the energy budget for (8.13) by multiplying by \mathbf{v} and integrating over a domain. If there is no flow into or out of our domain, the inertial terms in the momentum equation conserve energy and, recalling section 1.10, the energy equation is

$$\frac{d\hat{E}}{dt} = \frac{d}{dt} \int \frac{1}{2} \mathbf{v}^2 dV = \int (\mathbf{F} \cdot \mathbf{v} + \nu \mathbf{v} \cdot \nabla^2 \mathbf{v}) dV = \int (\mathbf{F} \cdot \mathbf{v} - \nu \mathbf{w}^2) dV, \quad (8.15)$$

where \hat{E} is the total energy. If we neglect the viscous term we are led to an inconsistency, since the forcing term is a source of energy ($\overline{\mathbf{F} \cdot \mathbf{v}} > 0$), because a force will normally, on

average, produce a velocity that is correlated with the force itself. Without viscosity, energy keeps on increasing.

What is amiss? It is true that for motion with a 1000 km length scale and a velocity of a few centimetres per second we can neglect viscosity when considering the balance of forces in the momentum equation. But this does not mean that there is no motion at much smaller length scales — indeed we seem to be led to the inescapable conclusion that there must be some motion at smaller scales in order to remove energy. Scale analysis of the momentum equation suggests that viscous terms will be comparable with the inertial terms at a scale L_v , where the Reynolds number based on that scale is of order unity, giving

$$L_v \sim \frac{V}{V}. \quad (8.16)$$

This is a very small scale for geophysical flows, of order millimetres or less. Where and how are such small scales generated? Boundaries are one important region; if there is high Reynolds number flow above a solid boundary, for example the wind above the ground, then viscosity *must* become important in bringing the velocity to zero in order that it can satisfy the no-slip condition at the surface, as illustrated in Fig. 2.10.

Motion on very small scales may also be generated in the fluid interior. How might this happen? Suppose the forcing acts only at large scales, and its direct action is to set up some correspondingly large-scale flow, composed of eddies and shear flows and such-like. Then typically there will be an instability in the flow, and a smaller eddy will grow: initially, the large-scale flow may be treated as an unchanging shear flow, and the disturbance while small will obey linear equations of motion similar to those applicable in an idealized Kelvin-Helmholtz instability. This instability clearly must draw from the large scale quasi-stationary flow, and it will eventually saturate at some finite amplitude. Although it has grown in intensity, it is still typically smaller than the large scale flow that fostered it (remember how the growth rate of the shear instability gets larger as the wavelength of the perturbation decreased). As it reaches finite amplitude, the perturbation itself may become unstable, and smaller eddies will feed off its energy and grow, and so on.⁴ Vortex stretching plays an important role in all this, stretching line elements and creating eddies, and energy, at small scales. The general picture that emerges is of a large-scale flow that is unstable to eddies somewhat smaller in scale. These eddies grow, and develop still smaller eddies, and energy is transferred to smaller and smaller scales in a cascade-like process, sketched in Fig. 8.2. Finally, eddies are generated that are sufficiently small that they feel the effects of viscosity, and energy is drained away. Thus, there is a flux of kinetic energy from the large to the small scales, where it is dissipated into heat.

8.2.2 Inertial-range theory

Given the above picture it becomes possible to predict what the energy spectrum is. Let us suppose that the flow is statistically isotropic (i.e., the same in all directions) and homogeneous (i.e., the same everywhere; all isotropic flows are homogeneous, but not vice versa). Homogeneity precludes the presence of solid boundaries but can be achieved in a periodic domain, and of course finite domain puts an upper limit, sometimes called the outer scale, on the size of eddies.

If we decompose the velocity field into Fourier components, then in a finite domain we

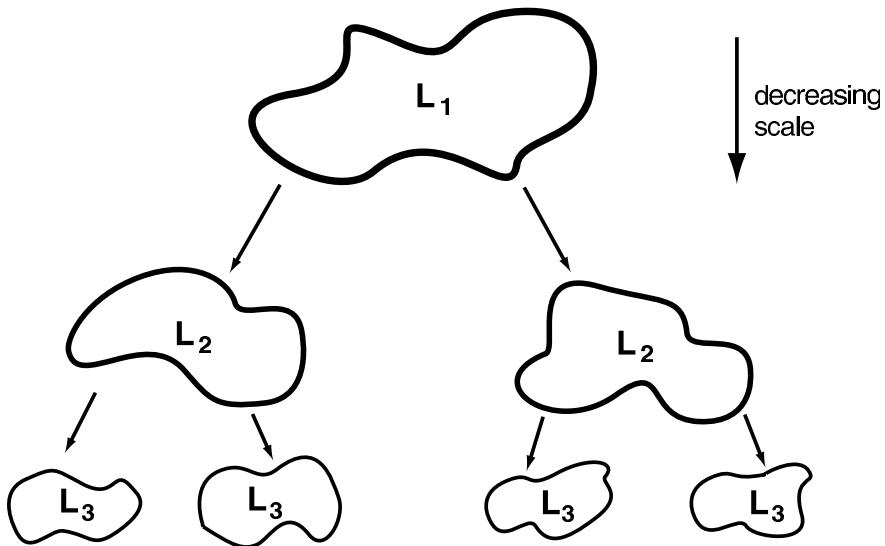


Fig. 8.2 Schema of the passage of energy to smaller scales: eddies at a large scale break up into smaller scale eddies, thereby transferring energy to smaller scales. If the passage occurs between eddies of similar sizes (i.e., if it is spectrally local) the transfer is said to be a cascade. The eddies in reality are embedded within each other.

may write

$$u(x, y, z, t) = \sum_{\mathbf{k}} \tilde{u}(\mathbf{k}) e^{i(k^x x + k^y y + k^z z)}, \quad (8.17)$$

where \tilde{u} is the Fourier transformed field of u , with similar identities for v and w , and $\mathbf{k} = (k^x, k^y, k^z)$. The sum is a triple sum over all wavenumbers (k^x, k^y, k^z) , and in a finite domain these wavenumbers are quantized. Finally, to ensure that u is real we require that $\tilde{u}(-\mathbf{k}) = \tilde{u}^*(\mathbf{k})$, where the asterisk denotes the complex conjugate. Using Parseval's theorem (and assuming density is unity, as we shall throughout this chapter) the energy in the fluid is given by

$$\frac{1}{V} \int_V E \, dV = \frac{1}{2V} \int_V (u^2 + v^2 + w^2) \, dV = \frac{1}{2} \sum_{\mathbf{k}} (|\tilde{u}|^2 + |\tilde{v}|^2 + |\tilde{w}|^2) \equiv \sum_{\mathbf{k}} \mathcal{E}_{\mathbf{k}}, \quad (8.18)$$

where E is the energy density per unit mass, V is the volume of the domain, and the last equality serves to define the discrete energy spectrum $\mathcal{E}_{\mathbf{k}}$. We will now assume that the turbulence is isotropic, and that the domain is sufficiently large that the sums in the above equations may be replaced by integrals. We may then write

$$\bar{E} = \frac{1}{V} \hat{E} = \frac{1}{2V} \int_V \mathbf{v}^2 \, dV = \int \mathcal{E}(k) \, dk, \quad (8.19)$$

where \bar{E} is the average energy, \hat{E} is the total energy and $\mathcal{E}(k)$ is the energy spectral density, or the energy spectrum, so that $\mathcal{E}(k)\delta k$ is the energy in the small wavenumber interval δk . Because of the assumed isotropy, the energy is a function only of the scalar wavenumber k , where $k^2 = k^x^2 + k^y^2 + k^z^2$. The units of $\mathcal{E}(k)$ are L^3/T^2 and the units of \bar{E} are L^2/T^2 .

Dimensions and the Kolmogorov Spectrum

Quantity	Dimension
Wavenumber, k	$1/L$
Energy per unit mass, E	$U^2 = L^2/T^2$
Energy spectrum, $\mathcal{E}(k)$	$EL = L^3/T^2$
Energy Flux, ε	$E/T = L^2/T^3$

If $\mathcal{E} = f(\varepsilon, k)$ then the only dimensionally consistent relation for the energy spectrum is

$$\mathcal{E} = \mathcal{K}\varepsilon^{2/3}k^{-5/3},$$

where \mathcal{K} is a dimensionless constant.

We now suppose that the fluid is stirred at large scales and, via the nonlinear terms in the momentum equation, that this energy is transferred to small scales where it is dissipated by viscosity. The key assumption is to suppose that, if forcing scale is sufficiently larger than the dissipation scale, there exists a range of scales that is intermediate between the large scale and the dissipation scale and where neither forcing nor dissipation are explicitly important to the dynamics. This assumption, known as the *locality hypothesis*, depends on the nonlinear transfer of energy being sufficiently local (in spectral space). This intermediate range is known as the *inertial range*, because the inertial terms and not forcing or dissipation must dominate in the momentum balance. If the rate of energy input per unit volume by stirring is equal to ε , then if we are in a steady state there must be a flux of energy from large to small scales that is also equal to ε , and an energy dissipation rate, also ε .

Now, we have no general theory for the energy spectrum of a turbulent fluid, but we might suppose it takes the general form

$$\mathcal{E}(k) = g(\varepsilon, k, k_0, k_v), \quad (8.20)$$

where the right-hand side denotes a function of the spectral energy flux or cascade rate ε , the wavenumber k , the forcing wavenumber k_0 and the wavenumber at which dissipation acts, k_v (and $k_v \sim L_v^{-1}$). The function g will of course depend on the particular nature of the forcing. Now, the locality hypothesis essentially says that at some scale within the inertial range the flux of energy to smaller scales depends only on processes occurring at or near that scale. That is to say, the energy flux is only a function of \mathcal{E} and k , or equivalently that the energy spectrum can be a function *only* of the energy flux ε and the wavenumber itself. From a physical point of view, as energy cascades to smaller scales the details of the forcing are forgotten but the effects of viscosity are not yet apparent, and the energy spectrum takes the form,

$$\mathcal{E}(k) = g(\varepsilon, k). \quad (8.21)$$

The function g is assumed to be *universal*, the same for every turbulent flow.

Let us now use dimensional analysis to give us the form of the function $g(\varepsilon, k)$ (see the shaded box above). In (8.21), the left-hand side has dimensions L^3/T^2 ; the factor T^{-2}

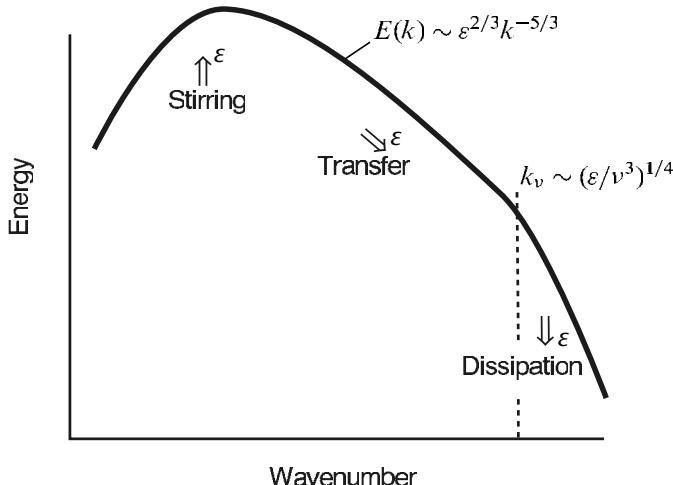


Fig. 8.3 Schema of energy spectrum in three-dimensional turbulence, in the theory of Kolmogorov. Energy is supplied at some rate ε ; it is cascaded to small scales, where it is ultimately dissipated by viscosity. There is no systematic energy transfer to scales larger than the forcing scale, so here the energy falls off.

can only be balanced by $\varepsilon^{2/3}$ because k has no time dependence; that is, (8.21), and its dimensions, must take the form

$$\mathcal{E}(k) = \varepsilon^{2/3} g(k), \quad (8.22a)$$

$$\frac{L^3}{T^2} \sim \frac{L^{4/3}}{T^2} g(k), \quad (8.22b)$$

where $g(k)$ is some function. Evidently $g(k)$ must have dimensions $L^{5/3}$, and the functional relationship we must have, if the physical assumptions are right, is

$$\boxed{\mathcal{E}(k) = \mathcal{K} \varepsilon^{2/3} k^{-5/3}}. \quad (8.23)$$

This is the famous ‘Kolmogorov -5/3 spectrum’, enshrined as one of the cornerstones of turbulence theory. It is sketched in Fig. 8.3, and some experimental results are shown in Fig. 8.4. The parameter \mathcal{K} is a dimensionless constant, undetermined by the theory; it is known as Kolmogorov’s constant and experimentally its value is found to be about 1.5.⁵

An equivalent, perhaps slightly more intuitive, way to derive this is to first define an eddy turnover time $τ_k$, which is the time taken for a parcel with velocity v_k to move a distance $1/k$, v_k being the velocity associated with the (inverse) scale k . On dimensional considerations $v_k = [\mathcal{E}(k)k]^{1/2}$ so that

$$τ_k = [k^3 \mathcal{E}(k)]^{-1/2}. \quad (8.24)$$

Kolmogorov’s assumptions are then equivalent to setting

$$\varepsilon \sim \frac{v_k^2}{τ_k} = \frac{k \mathcal{E}(k)}{τ_k}. \quad (8.25)$$

If we demand that ε be a constant then (8.24) and (8.25) yield (8.23).

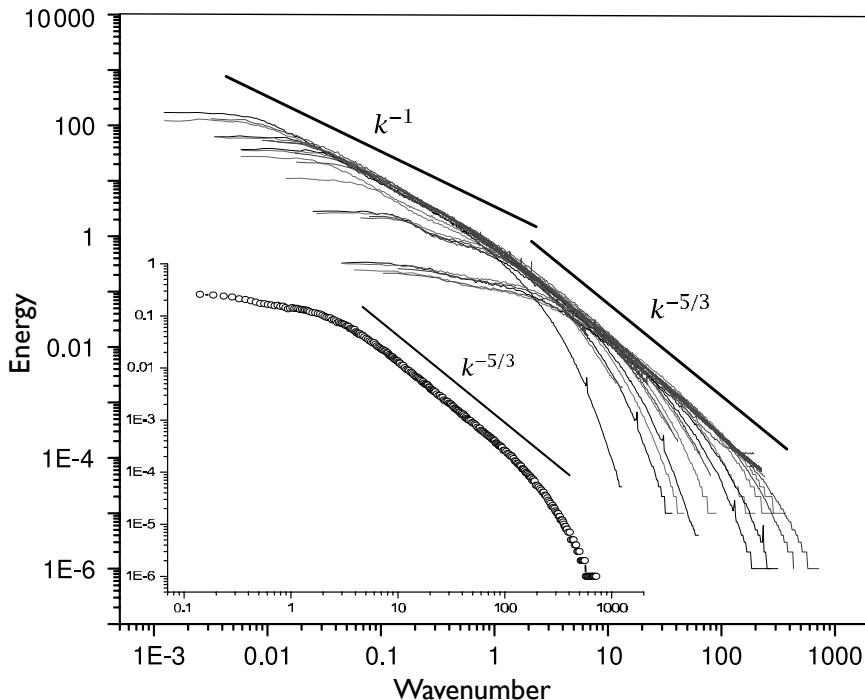


Fig. 8.4 The energy spectrum of 3D turbulence measured in some experiments at the Princeton Superpipe facility.⁶ The outer plot shows the spectra from a large number of experiments at different Reynolds numbers up to 10^6 , with the magnitude of their spectra appropriately rescaled. Smaller scales show a good $-5/3$ spectrum, whereas at larger scales the eddies feel the effects of the pipe wall and the spectra are a little shallower. The inner plot shows the spectrum in the centre of the pipe in a single experiment at $Re \approx 10^6$.

The viscous scale and energy dissipation

At some small length scale we should expect viscosity to become important and the scaling theory we have just set up will fail. What is that scale? In the inertial range friction is unimportant because the time scales on which it acts are too long for it to be important and dynamical effects dominate. In the momentum equation the viscous term is $\nu \nabla^2 u$ so that a viscous or dissipation time scale at a scale k^{-1} , τ_k^ν , is

$$\tau_k^\nu \sim \frac{1}{k^2 \nu}, \quad (8.26)$$

so that the viscous time scale decreases with scale. The eddy turnover time, τ_k — that is, the inertial time scale — in the Kolmogorov spectrum is

$$\tau_k = \varepsilon^{-1/3} k^{-2/3}. \quad (8.27)$$

The wavenumber at which dissipation becomes important is then given by equating the above two time scales, yielding the dissipation wavenumber, k_ν and the associated length

scale, L_v ,

$$k_v \sim \left(\frac{\varepsilon}{\nu^3} \right)^{1/4}, \quad L_v \sim \left(\frac{\nu^3}{\varepsilon} \right)^{1/4}. \quad (8.28a,b)$$

L_v is called the *Kolmogorov scale*. It is the *only* quantity which can be created from the quantities ν and ε that has the dimensions of length. (It is the same as the scale given by (8.16) provided that in that expression V is the velocity magnitude at the Kolmogorov scale.) Thus, for $L \gg L_v$, $\tau_k \ll \tau_k^\nu$ and inertial effects dominate. For $L \ll L_v$, $\tau_k^\nu \ll \tau_k$ and frictional effects dominate. In fact for length scales smaller than the dissipation scale, (8.27) is inaccurate; the energy spectrum falls off more rapidly than $k^{-5/3}$ and the inertial time scale falls off less rapidly than (8.27) implies, and dissipation dominates even more.

Given the dissipation scale, let us estimate the average energy dissipation rate, $\dot{\bar{E}}$. This is given by

$$\dot{\bar{E}} = \frac{1}{V} \int \nu \mathbf{v} \cdot \nabla^2 \mathbf{v} \, dV. \quad (8.29)$$

The length at which dissipation acts is the Kolmogorov scale and, noting that $\nu_k^2 \sim \varepsilon^{2/3} k^{-2/3}$ and using (8.28a), the average energy dissipation rate scales as

$$\dot{\bar{E}} \sim \nu k_v^2 \nu_{k_v}^2 \sim \nu k_v^2 \frac{\varepsilon^{2/3}}{k_v^{2/3}} \sim \varepsilon. \quad (8.30)$$

That is, the energy dissipation rate is equal to the energy cascade rate. On the one hand this seems sensible, but on the other hand it is *independent of the viscosity*. In particular, in the limit of viscosity tending to zero the energy dissipation remains finite! Surely the energy dissipation rate must go to zero if viscosity goes to zero? To see that this is not the case, consider that energy is input at some large scales, and the magnitude of the stirring largely determines the energy input and cascade rate. The scale at which viscous effects then become important is determined by the viscous scale, L_v , given by (8.28b). As viscosity tends to zero L_v becomes smaller in just such a way as to preserve the constancy of the energy dissipation. This is one of the most important results in three-dimensional turbulence. Now, we established in section 1.10 that the Euler equations (i.e., the fluid equations with the viscous term omitted from the outset) do conserve energy. This means that the Euler equations are a *singular limit* of the Navier-Stokes equations: the behaviour of the Navier-Stokes equations as viscosity tends to zero is different from the behaviour resulting from simply omitting the viscous term from the equations ab initio.

How big is L_v in the atmosphere? A crude estimate, perhaps wrong by an order of magnitude, comes from noting that ε has units of U^3/L , and that at length scales of the order of 100 m in the atmospheric boundary layer (where there might be a three-dimensional energy cascade to small scales) velocity fluctuations are of the order of 1 cm s^{-1} , giving $\varepsilon \approx 10^{-8} \text{ m}^2 \text{ s}^{-3}$. Using (8.28b) we then find the dissipation scale to be of the order of a millimetre or so. In the ocean the dissipation scale is also of the order of millimetres. Various inertial range properties, in both three and two dimensions, are summarized in the shaded box on page 350.

Degrees of freedom

How many degrees of freedom does a turbulent fluid like the atmosphere potentially have? We might estimate this number, N say, by the expression

$$N \sim \left(\frac{L}{L_v} \right)^3, \quad (8.31)$$

where L is the length scale of the energy-containing eddies at the large scale. If we take $L = 1000$ km and $L_v = 1$ mm this gives about 10^{27} ! On a rather more general basis, we can obtain an expression for N using (8.28b), to give

$$N \sim L^3 \left(\frac{\varepsilon}{\nu^3} \right)^{3/4}, \quad (8.32)$$

or, using $\varepsilon \sim U^3/L$,

$$N \sim \left(\frac{UL}{\nu} \right)^{9/4} = Re^{9/4}, \quad (8.33)$$

where Re is the Reynolds number based on the large-scale flow. For typical large-scale atmospheric flows with $U \sim 10 \text{ m s}^{-1}$, $L \sim 10^6 \text{ m}$ and $\nu = 10^{-5} \text{ m}^2 \text{ s}^{-1}$, $Re \sim 10^{12}$ and again $N \sim 10^{27}$. Obviously, this number is very approximate, but nevertheless the number of potential degrees of freedom in the atmosphere is truly enormous, greater than Avogadro's number. Thus trying to explicitly model the turbulent atmosphere explicitly is akin to trying to model the gas in a room by following the motion of each individual molecule, and it seems unnecessary. How *should* we model it? That, in a nutshell, is the (unsolved) problem of turbulence.

8.2.3 * Another expression of the inertial-range scaling argument

Kolmogorov's spectrum, as well as some other useful scaling relationships, can be obtained in a slightly different way as follows. If we for the moment ignore viscosity, the Euler equations are invariant under the following scaling transformation:

$$x \rightarrow x\lambda \quad v \rightarrow v\lambda^r \quad t \rightarrow t\lambda^{1-r}, \quad (8.34)$$

where r is an arbitrary scaling exponent. So far there is minimal physics. Now make the following *physical* assumptions about the behaviour of a turbulent fluid, *with viscosity*:

- (i) That the flux of energy from large to small scales (i.e., ε) is finite and constant.
- (ii) That the scale invariance (8.34) holds, on a time-average, in the intermediate scales between the forcing and dissipation scales.

The second assumption plays the role of the locality hypothesis. Dimensional analysis then tells us that the energy flux at some wavenumber k scales as

$$\varepsilon_k \sim \frac{v_k^3}{l_k} \sim \lambda^{3r-1}, \quad (8.35)$$

where v_k and l_k are the velocity and length scales at wavenumber k . Invoking assumption (i), that ε is independent of scale, gives $r = 1/3$. The velocity then scales as

$$v_k \sim \varepsilon^{1/3} k^{-1/3}, \quad (8.36)$$

and the velocity gradient (and so vorticity) scales as $k\nu_k \sim \varepsilon^{1/3}k^{2/3}$. (This becomes infinite at very small scales, but this behaviour is avoided in any real physical situation by the presence of viscosity.) We can now recover (8.23) because, on dimensional grounds,

$$\mathcal{E}(k) \sim \nu_k^2 k^{-1} \sim \varepsilon^{2/3} k^{-2/3} k^{-1} \sim \varepsilon^{2/3} k^{-5/3}. \quad (8.37)$$

In general, the slope of the energy spectrum, k^n , is related to the scaling exponent by $n = -(2r + 1)$. The 'structure functions' S_m , which are the average of the m th power of the velocity difference over distances $l \sim 1/k$, scale as $S_m \sim (\nu_k)^m \sim \varepsilon^{m/3} k^{-m/3}$. In particular the second-order structure function, which is the Fourier transform of the energy spectra, scales as $S_2 \sim \varepsilon^{2/3} k^{-2/3}$. Other results of the Kolmogorov theory follow similarly.

8.2.4 A final note on our assumptions

The assumptions of homogeneity and isotropy that are made in the Kolmogorov theory are ansatzes, in that we make them because we want to have a tractable model of turbulence (and certainly we can conceive of an experiment in which turbulence is for most practical purposes homogeneous and isotropic). The essential *physical* assumptions are: (i) that there exists an inertial range in which the energy flux is constant, and (ii) that the energy is cascaded from large to small scales in a series of small steps, as the energy spectra will then be determined by spectrally local quantities. The second assumption is the locality assumption and without it we could have, instead of (8.23),

$$\mathcal{E}(k) = C\varepsilon^{2/3}k^{-5/3}g(k/k_0)h(k/k_v), \quad (8.38)$$

where g and h are unknown functions. We essentially postulate that there exists a range of intermediate wavenumbers over which $g(k/k_0) = h(k/k_v) = 1$.

The first, and less obvious, assumption might be called the *non-intermittency* assumption, and it demands that rare events (in time or space) with large amplitudes do not dominate the energy flux or the dissipation rate. If they were to do so, then the flux would fluctuate strongly, the turbulent statistics would not be completely characterized by ε and Kolmogorov's theory would not be exactly right. (Note that in the theory ε is the mean energy cascade rate, and $\varepsilon^{2/3}$ is the two-thirds power of the mean, which is not equal to the mean of the two-thirds power.) In fact, in high Reynolds turbulence the $-5/3$ spectra is often observed to a fairly high degree of accuracy (e.g., as in Fig. 8.4), although the higher-order statistics (e.g., higher-order structure functions) predicted by the theory are often found to be in error, and it is generally believed that Kolmogorov's theory is not exact.⁷

8.3 TWO-DIMENSIONAL TURBULENCE

Two-dimensional turbulence behaves in a profoundly different way from three-dimensional turbulence, largely because of the presence of another quadratic invariant, the enstrophy (defined below; see also section 5.6.3). In two dimensions, the vorticity equation for incompressible flow is:

$$\frac{\partial \zeta}{\partial t} + \mathbf{u} \cdot \nabla \zeta = F + \nu \nabla^2 \zeta \quad (8.39)$$

Inertial Range Properties in 3D and 2D Turbulence

For reference, a few inertial range properties are listed below, omitting non-dimensional constants.

	3D Energy Range	2D Enstrophy Range	
Energy spectrum	$\varepsilon^{2/3} k^{-5/3}$	$\eta^{2/3} k^{-3}$	(T.1)
Turnover time	$\varepsilon^{-1/3} k^{-2/3}$	$\eta^{-1/3}$	(T.2)
Viscous scale, L_v	$(\nu^3/\varepsilon)^{1/4}$	$(\nu^3/\eta)^{1/6}$	(T.3)
Passive tracer spectrum	$\chi \varepsilon^{-1/3} k^{-5/3}$	$\chi \eta^{-1/3} k^{-1}$	(T.4)

In these expressions:

ν = viscosity, k = wavenumber, ε = energy cascade rate,

η = enstrophy cascade rate, χ = tracer variance cascade rate.

where $\mathbf{u} = u\mathbf{i} + v\mathbf{j}$ and $\zeta = \mathbf{k} \cdot \nabla \times \mathbf{u}$ and F is a stirring term. In terms of a streamfunction, $u = -\partial\psi/\partial y$, $v = \partial\psi/\partial x$, and $\zeta = \nabla^2\psi$, and (8.39) may be written as

$$\frac{\partial \nabla^2\psi}{\partial t} + J(\psi, \nabla^2\psi) = F + \nu \nabla^4\psi. \quad (8.40)$$

We obtain an energy equation by multiplying by $-\psi$ and integrating over the domain, and an enstrophy equation by multiplying by ζ and integrating. When $F = \nu = 0$ we find

$$\hat{E} = \frac{1}{2} \int_A (u^2 + v^2) dA = \frac{1}{2} \int_A (\nabla\psi)^2 dA, \quad \frac{d\hat{E}}{dt} = 0, \quad (8.41a)$$

$$\hat{Z} = \frac{1}{2} \int_A \zeta^2 dA = \frac{1}{2} \int_A (\nabla^2\psi)^2 dA, \quad \frac{d\hat{Z}}{dt} = 0, \quad (8.41b)$$

where the integral is over a finite area with either no-normal flow or periodic boundary conditions. The quantity \hat{E} is the energy, and \hat{Z} is known as the *enstrophy*. The enstrophy invariant arises because the vortex stretching term, so important in three-dimensional turbulence, vanishes identically in two dimensions. In fact, because vorticity is conserved on parcels it is clear that the integral of *any* function of vorticity is zero when integrated over A ; that is, from (8.39)

$$\frac{Dg(\zeta)}{Dt} = 0 \quad \text{and} \quad \frac{d}{dt} \int_A g(\zeta) dA = 0. \quad (8.42)$$

where $g(\zeta)$ is an arbitrary function. Of this infinity of conservation properties, enstrophy conservation (with $g(\zeta) = \zeta^2$) in particular has been found to have enormous consequences to the flow of energy between scales, as we will soon discover.⁸

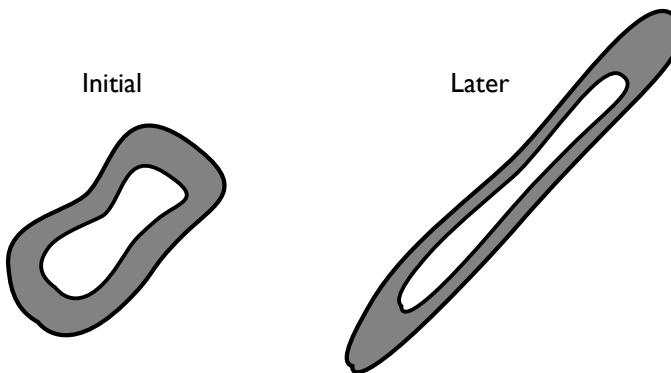


Fig. 8.5 In incompressible two-dimensional flow, a band of fluid is elongated, but its area is preserved. Elongation is followed by ‘folding’ and more elongation, producing filaments as in Fig. 8.8. As vorticity is tied to fluid parcels, the values of the vorticity in the shaded area (and in the hole) are maintained; thus, vorticity gradients increase and the enstrophy is thereby, on average, moved to smaller scales.

8.3.1 Energy and enstrophy transfer

In three-dimensional turbulence we posited that energy is cascaded to small scales via vortex stretching. In two dimensions that mechanism is absent, and it turns out that it is more reasonable to expect energy to be transferred to *larger scales*. This counter-intuitive behaviour arises from the twin integral constraints of energy and enstrophy conservation, and the following three arguments illustrate why this should be so.

I. Vorticity elongation

Consider a band or a patch of vorticity, as in Fig. 8.5, in a nearly inviscid fluid. The vorticity of each element of fluid is conserved as the fluid moves. Now, we should expect that the quasi-random motion of the fluid will act to elongate the band but, as its area must be preserved, the band narrows and so vorticity gradients will increase. This is equivalent to the enstrophy moving to smaller scales. Now, the energy in the fluid is

$$\hat{E} = -\frac{1}{2} \int \psi \zeta \, dA, \quad (8.43)$$

where the streamfunction is obtained by solving the Poisson equation $\nabla^2 \psi = \zeta$. If the vorticity is locally elongated primarily only in one direction (as it must be to preserve area), the integration involved in solving the Poisson equation will lead to the scale of the streamfunction becoming larger in the direction of stretching, but virtually no smaller in the perpendicular direction. Because stretching occurs, on average, in all directions, the overall scale of the streamfunction will increase in all directions, and the cascade of enstrophy to small scales will be accompanied by a transfer of energy to large scales.

II. An energy-enstrophy conservation argument

A moment’s thought will reveal that the distribution of energy and enstrophy in wavenumber space are respectively analogous to the distribution of mass and moment of inertia of a

lever, with wavenumber playing the role of distance from the fulcrum. Any re-arrangement of mass such that its distribution also becomes wider must be such that the centre of mass moves toward the fulcrum. Thus, analogously, any rearrangement of a flow that preserves both energy and enstrophy, and that causes the distribution to spread out in wavenumber space, will tend to move energy to small wavenumbers and enstrophy to large. To prove this we begin with expressions for the average energy and enstrophy:

$$\bar{E} = \int \mathcal{E}(k) dk, \quad \bar{Z} = \int \mathcal{Z}(k) dk = \int k^2 \mathcal{E}(k) dk, \quad (8.44)$$

where $\mathcal{E}(k)$ and $\mathcal{Z}(k)$ are the energy and enstrophy spectra. A wavenumber characterizing the spectral location of the energy is the centroid,

$$k_e = \frac{\int k \mathcal{E}(k) dk}{\int \mathcal{E}(k) dk} \quad (8.45)$$

and, for simplicity, we normalize units so that the denominator is unity. The spreading out of the energy distribution is formalized by setting

$$I \equiv \int (k - k_e)^2 \mathcal{E}(k) dk, \quad \frac{dI}{dt} > 0. \quad (8.46)$$

Here, I measures the width of the energy distribution, and this is assumed to increase. Expanding out the integral gives

$$\begin{aligned} I &= \int k^2 \mathcal{E}(k) dk - 2k_e \int k \mathcal{E}(k) dk + k_e^2 \int \mathcal{E}(k) dk \\ &= \int k^2 \mathcal{E}(k) dk - k_e^2 \int \mathcal{E}(k) dk, \end{aligned} \quad (8.47)$$

where the last equation follows because $k_e = \int k \mathcal{E}(k) dk$ is, from (8.45), the energy-weighted centroid. Because both energy and enstrophy are conserved, (8.47) gives

$$\frac{dk_e^2}{dt} = -\frac{1}{\bar{E}} \frac{dI}{dt} < 0. \quad (8.48)$$

Thus, the centroid of the distribution moves to smaller wavenumbers and to larger scales (see Fig. 8.6).

An appropriately defined measure of the centre of the enstrophy distribution, on the other hand, moves to higher wavenumbers. The demonstration follows easily if we work with the inverse wavenumber, which is a direct measure of length. Let $q = 1/k$ and assume that the enstrophy distribution spreads out by nonlinear interactions, so that, analogously to (8.46),

$$J = \int (q - q_e)^2 \mathcal{Y}(q) dq, \quad \frac{dJ}{dt} > 0, \quad (8.49)$$

where $\mathcal{Y}(q)$ is such that the enstrophy is $\int \mathcal{Y}(q) dq$ and

$$q_e = \frac{\int q \mathcal{Y}(q) dq}{\int \mathcal{Y}(q) dq}. \quad (8.50)$$

Expanding the integrand in (8.49) and using (8.50) gives

$$J = \int q^2 \mathcal{Y}(q) dq - q_e^2 \int \mathcal{Y}(q) dq, \quad (8.51)$$

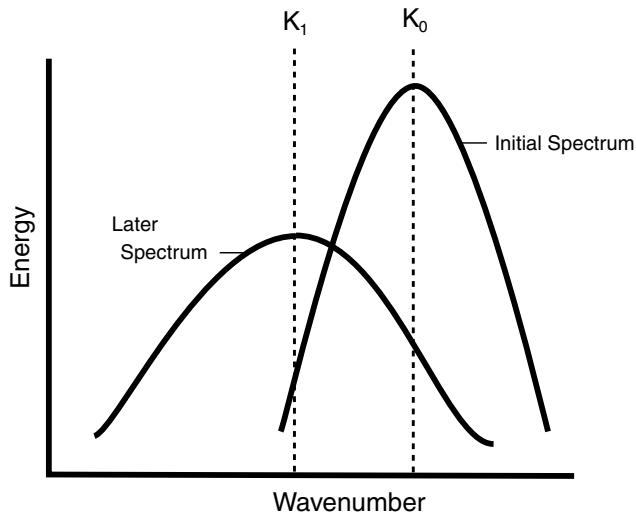


Fig. 8.6 In two-dimensional flow, the centroid of the energy spectrum will move to large scales (smaller wavenumber) provided that the width of the distribution increases — as can be expected in a nonlinear, eddying flow.

But $\int q^2 \mathcal{Y}(q) dq$ is conserved, because this is the energy. Thus,

$$\frac{dJ}{dt} = - \frac{d}{dt} q_e^2 \int \mathcal{Y}(q) dq \quad (8.52)$$

whence

$$\frac{dq_e^2}{dt} = - \frac{1}{Z} \frac{dJ}{dt} < 0. \quad (8.53)$$

Thus, the length scale characterizing the enstrophy distribution gets smaller, and the corresponding wavenumber gets larger.

III. A similarity argument

Consider an initial value problem, in which a fluid with some initial distribution of energy is allowed to freely evolve, unencumbered by boundaries. We note two aspects of the problem:

- (i) there is no externally imposed length scale (because of the way the problem is posed);
- (ii) the energy is conserved (this being an assumption).

It is the second condition that limits the argument to two dimensions, for in three dimensions energy is quickly cascaded to small scales and dissipated, but let us here posit that this does not occur. These two assumptions are then sufficient to infer the general direction of transfer of energy, using a rather general similarity argument. To begin, write the energy per unit mass of the fluid as

$$\bar{E} = U^2 = \int \mathcal{E}(k, t) dk, \quad (8.54)$$

where $\mathcal{E}(k, t)$ is the energy spectrum and U is measure of the total energy, with units of velocity. Now, solely on dimensional considerations we can write

$$\mathcal{E}(k, t) = U^2 L \hat{\mathcal{E}}(\hat{k}, \hat{t}), \quad (8.55)$$

where $\hat{\mathcal{E}}$, and its arguments, are non-dimensional quantities, and L is some length scale. However, on physical considerations, the only parameters available to determine the energy

spectrum are U , t and k , the wavenumber. A little thought reveals that the most general form for the energy spectrum with no L dependence is

$$\mathcal{E}(k, t) = U^3 t \hat{\mathcal{E}} = U^3 t g(Ukt), \quad (8.56)$$

where g is an arbitrary function of its arguments. The argument of g is the only non-dimensional grouping of U , t and k , and $U^3 t$ provides the proper dimensions for \mathcal{E} . Conservation of energy now implies that the integral

$$I = \int_0^\infty t g(Ukt) dk \quad (8.57)$$

is *not* a function of time. Defining $\vartheta = Ukt$, this requirement is met if

$$\int_0^\infty g(\vartheta) d\vartheta = \text{constant}. \quad (8.58)$$

Now, the spectrum is a function of k only through the combination $\vartheta = Ukt$. Thus, as time proceeds features in the spectrum move to smaller k . Suppose, for example, that the energy is initially peaked at some wavenumber k_p ; the product tk_p is preserved, so k_p must diminish with time and the energy must move to larger scales. Similarly, the energy weighted mean wavenumber, k_e , moves to smaller wavenumbers, or larger scales. To see this explicitly, we have

$$k_e = \frac{\int k \mathcal{E} dk}{\int \mathcal{E} dk} = \frac{\int k \mathcal{E} dk}{U^2} = \int k Ut g(Ukt) dk = \int \frac{\vartheta g(\vartheta)}{Ut} d\vartheta = \frac{C}{Ut} \quad (8.59)$$

where all the integrals are over the interval $(0, \infty)$ and $C = \int \vartheta g(\vartheta) d\vartheta$ is a constant. Thus, the wavenumber centroid of the energy distribution decreases with time, and the characteristic scale of the flow, $1/k_e$, increases with time. Interestingly, the enstrophy does not explicitly enter this argument, and in general it is not conserved; rather, it is the requirement that energy be conserved that limits the argument to two dimensions; if we accept ab initio that energy *is* conserved, it must be transferred to larger scales.⁹

8.3.2 Inertial ranges in two-dimensional turbulence

If, unlike the case in three dimensions, energy is transferred to larger scales in inviscid, nonlinear, two-dimensional flow then we might expect the inertial ranges of two-dimensional turbulence to be quite different from their three-dimensional counterparts. But before looking in detail at the inertial ranges themselves, let us establish a couple of general properties of forced-dissipative flow in two dimensions.

Some properties of forced-dissipative flow

We will first show that, unlike the case in three dimensions, energy dissipation goes to zero as the Reynolds number rises. In the absence of forcing terms, the total dissipation of energy is, from (8.39)),

$$\frac{d\hat{\mathcal{E}}}{dt} = -\nu \int \zeta^2 dA. \quad (8.60)$$

Energy dissipation can only remain finite as $\nu \rightarrow 0$ if vorticity becomes infinite. However, this cannot happen because vorticity is conserved on parcels except for the action of viscosity,

meaning that $D\zeta/Dt = \nu \nabla^2 \zeta$. However, the viscous term can only *reduce* the value of vorticity on a parcel, and so vorticity can never become infinite if it is not so initially, and therefore using (8.60) energy dissipation goes to zero with ν . (In three dimensions vorticity becomes infinite as viscosity goes to zero because of the effect of vortex stretching.) This conservation of energy is related to the fact that energy is trapped at large scales, even in forced-dissipative flow. On the other hand, enstrophy is transferred to small scales and therefore we expect it to be dissipated at large wavenumbers, even as the Reynolds number becomes very large.

We can show that energy is trapped at large scales in forced-dissipative two-dimensional flow (in a sense that will be made explicit) by the following argument.¹⁰ Suppose that the forcing of the fluid is confined to a particular scale, characterized by the wavenumber k_f , and that dissipation is effected by a linear drag and a small viscosity. The equation of motion is

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) = F - r\zeta + \nu \nabla^2 \zeta, \quad (8.61)$$

where F is the stirring and r and ν are positive constants. This leads to the following energy and enstrophy equations:

$$\frac{d\hat{E}}{dt} = -2r\hat{E} - \int \psi F dA - \int \nu \zeta^2 dA \approx -2r\hat{E} - \int \psi F dA, \quad (8.62a)$$

$$\frac{d\hat{Z}}{dt} = -2r\hat{Z} + \int \zeta F dA - D_Z \approx -2r\hat{Z} - k_f^2 \int \psi F dA - D_Z, \quad (8.62b)$$

where $D_Z = \int \nu(\nabla \zeta)^2 dA$ is the enstrophy dissipation. To obtain the right-most expressions, in (8.62a) we assume there is no dissipation of energy by the viscous term, and in (8.62b) we assume that the forcing is confined to wavenumbers near k_f . In a statistically steady state we may eliminate the integral involving ψF between (8.62a) and (8.62b), and writing $\hat{E} = \int E(k) dk$ and $\hat{Z} = \int k^2 \mathcal{E}(k) dk$ (for a domain of unit area) we obtain

$$\int k^2 \mathcal{E}(k) dk + \frac{D_Z}{2r} = \int k_f^2 \mathcal{E}(k) dk, \quad (8.63)$$

where the integrations are over all wavenumbers. Now, from the obvious inequality $\int (k - k_e)^2 \mathcal{E}(k) dk \geq 0$, where k_e is the energy centroid defined in (8.45), we obtain

$$\int (k^2 - k_e^2) \mathcal{E}(k) dk \geq 0. \quad (8.64)$$

Combining (8.63) and (8.64) gives

$$\int (k_f^2 - k_e^2) \mathcal{E}(k) dk \geq \frac{D_Z}{2r} > 0. \quad (8.65)$$

Thus, in a statistically steady state, the energy containing scale, as characterized by k_e^{-1} , must be larger than the forcing scale k_f^{-1} . This demonstration (like argument II in section 8.3.1) relies both on the conservation of energy and enstrophy by the nonlinear terms and on the particular relationship between the energy and enstrophy spectra.

This result, and (especially) the arguments of section 8.3.1, suggest that in a forced-dissipative two-dimensional fluid, energy is transferred to larger scales and enstrophy is transferred to small scales. To obtain a statistically steady state friction [such as the Rayleigh

drag of (8.61)] is necessary to remove energy at large scales, and enstrophy must be removed at small scales, but if the forcing scale is sufficiently well separated in spectral space from such frictional effects then two inertial ranges may form — an *energy inertial range* carrying energy to larger scales, and an *enstrophy inertial range* carrying enstrophy to small scales (Fig. 8.7). These ranges are analogous to the three-dimensional inertial range of section 8.2, and similar conditions must apply if the ranges are to be truly inertial — in particular we must assume spectral locality of the energy or enstrophy transfer. Given that, we can then calculate their properties, as follows.

The enstrophy inertial range

In the enstrophy inertial range the enstrophy cascade rate η , equal to the rate at which enstrophy is supplied by stirring, is assumed constant. By analogy with (8.25) we may assume that this rate is given by

$$\eta \sim \frac{k^3 \mathcal{E}(k)}{\tau_k}. \quad (8.66)$$

With τ_k (still) given by (8.24) we obtain

$$\boxed{\mathcal{E}(k) = \mathcal{K}_\eta \eta^{2/3} k^{-3}}, \quad (8.67)$$

where \mathcal{K}_η is, we presume, a universal constant, analogous to the Kolmogorov constant of (8.23). This, and various other properties in two- and three-dimensional turbulence, are summarized in the shaded box on page 350.

It is also possible to obtain (8.67) from scaling arguments similar to those in section 8.2.3. The scaling transformation (8.34) still holds, but now instead of (8.35) we assume that the enstrophy flux is constant with wavenumber. Dimensionally, and analogously to (8.35), we have

$$\eta \sim \frac{v_k^3}{l_k^3} \sim \lambda^{3r-3}, \quad (8.68)$$

and the constancy of η gives $r = 1$ for the scaling exponent. The exponent n determining the slope of the inertial range is given, as before, by $n = -(2r + 1)$ yielding the -3 spectra of (8.67). The velocity at a particular wavenumber then scales as

$$v_k \sim \eta^{1/3} k^{-1}, \quad (8.69)$$

and the time scales as

$$t_k \sim l_k / v_k \sim \eta^{-1/3}. \quad (8.70)$$

We may also obtain (8.70) by substituting (8.67) into (8.24). Thus, *the eddy turnover time in the enstrophy range of two-dimensional turbulence is length-scale invariant*. The appropriate viscous scale is given by equating the inertial and viscous terms in (8.39). Using (8.69) we obtain, analogously to (8.28a),

$$k_v \sim \left(\frac{\eta^{1/3}}{\nu} \right)^{1/2}. \quad (8.71)$$

The enstrophy dissipation, analogously to (8.30) goes to a finite limit given by

$$\dot{\bar{Z}} = \nu \int_A \zeta \nabla^2 \zeta \, dA \sim \nu k_v^4 v_{k_v}^2 \sim \eta, \quad (8.72)$$

using (8.69) and (8.71). Thus, the enstrophy dissipation in two-dimensional turbulence is (at least according to this theory) independent of the viscosity.

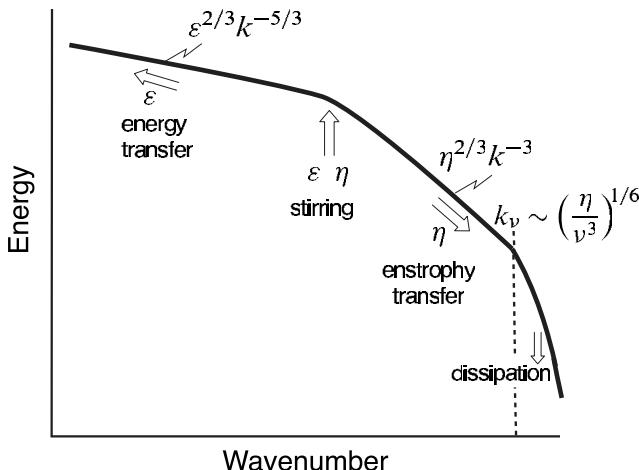


Fig. 8.7 The energy spectrum of two-dimensional turbulence. (Compare with Fig. 8.3.) Energy supplied at some rate ε is transferred to large scales, whereas enstrophy supplied at some rate η is transferred to small scales, where it may be dissipated by viscosity. If the forcing is localized at a scale k_f^{-1} then $\eta \approx k_f^2 \varepsilon$.

Energy inertial range

The energy inertial range of two-dimensional turbulence is quite similar to that of three-dimensional turbulence, except in one major respect: the energy flows from smaller to larger scales! Because the atmosphere and ocean behave in some ways as two-dimensional fluids, this has profound consequences on their behaviour, and is something we return to in the next chapter. The upscale energy flow is known as the *inverse cascade*, and the associated energy spectrum is, as in the three-dimensional case,

$$E(k) = \mathcal{K}_\varepsilon \varepsilon^{2/3} k^{-5/3}, \quad (8.73)$$

where \mathcal{K}_ε is a non-dimensional constant [sometimes called the Kolmogorov-Kraichnan constant, and not necessarily equal to \mathcal{K} in (8.23)], and ε is the rate of energy transfer to larger scales. Of course we now need a mechanism to remove energy at large scales, otherwise it will pile up at the scale of the domain and a statistical steady state will not be achieved. Introducing a linear drag, $-r\zeta$, into the vorticity equation, as in (8.61), is one means of removing energy, and such a term may be physically justified by appeal to Ekman layer theory (section 2.12). Although such a term appears to be scale invariant, its effects will be felt primarily at large scales because at smaller scales the time scale of the turbulence is much shorter than that of the friction, and we may estimate the scale at which the drag becomes important by equating the two time scales. The turbulent timescale is given by (8.27), and equating this to the frictional time scale r^{-1} gives

$$r^{-1} = \varepsilon^{-1/3} k_r^{-2/3} \quad \rightarrow \quad k_r = \left(\frac{r^3}{\varepsilon} \right)^{1/2}, \quad (8.74)$$

where k_r is the frictional wavenumber. Frictional effects are important at scales *larger* than k_r^{-1} .

8.3.3 † More about the phenomenology

The phenomenology of two-dimensional turbulence is not quite as settled as the above arguments imply. Note, for example, that time scale (8.70) is independent of length scale,

whereas in three-dimensional turbulence the time scale decreases with length scale, which seems more physical and more conducive to spectrally local interactions. A useful heuristic measure of this locality is given by estimating the contributions to the straining rate, $S(k)$, from motions at all scales larger than k^{-1} .¹¹ The strain rate scales like the shear, so that an estimate of the total strain rate is given by

$$S(k) = \left[\int_{k_0}^k \mathcal{E}(p) p^2 dp \right]^{1/2}, \quad (8.75)$$

where k_0 is the wavenumber of the largest scale present. The contributions to the integrand from a given wavenumber octave are given by

$$\int_p^{2p} \mathcal{E}(p') p'^3 d\log p' \sim \mathcal{E}(p) p^3. \quad (8.76)$$

In three dimensions, use of the $-5/3$ spectrum indicates that the contributions from each octave below a given wavenumber k increase with wavenumber, being a maximum close to k , and this is a posteriori consistent with the locality hypothesis. However, in two-dimensional turbulence with a -3 spectrum each octave makes the same contribution. That is to say, the contributions to the strain rate at a given wavenumber, as defined by (8.75), are not spectrally local. This does not prove that the enstrophy transfer is spectrally non-local, but nor does it build confidence in the theory.

Dimensionally the strain rate is the inverse of a time, and if this is a spectrally non-local quantity then, instead of (8.24), we might use the inverse of the strain rate as an eddy turnover time giving

$$\tau_k = \left[\int_{k_0}^k p^2 \mathcal{E}(p) dp \right]^{-1/2}. \quad (8.77)$$

This has the advantage over (8.24) in that it is a non-increasing function of wavenumber, whereas if the spectrum is steeper than k^{-3} (8.24) implies a time scale increasing with wavenumber. Using this in (8.66) gives a prediction for the enstrophy inertial range, namely

$$\mathcal{E}(k) = \mathcal{K}_\eta \eta^{2/3} [\log(k/k_0)]^{-1/3} k^{-3}, \quad (8.78)$$

which is similar to (8.67) except for a logarithmic correction. This expression is, of course, spectrally non-local, in contradiction to our original assumption: it has arisen by noting the spectral locality inherent in (8.75), and proposing a reasonable, although ad hoc, solution.

The discussion above suggests that the phenomenology of the forward enstrophy cascade is on the verge of being internally inconsistent, and that the k^{-3} spectral slope might be the shallowest limit that is likely to be actually achieved in nature or in any particular computer simulation or laboratory experiment rather than a robust, universal slope. To see this, suppose the detailed fluid dynamics attempts in some way to produce a slope shallower than k^{-3} ; then, using (8.76), the strain is local and the shallow slope is forbidden by the Kolmogorovian scaling results. However, if the dynamics organizes itself into structures with a slope steeper than k^{-3} the strain is quite non-local. The fundamental assumption of Kolmogorov scaling is not satisfied, and there is no internal inconsistency; the theory simply does not apply. The k^{-3} slope itself is at the margin.

There are two other potential problems with the theory of two-dimensional turbulence

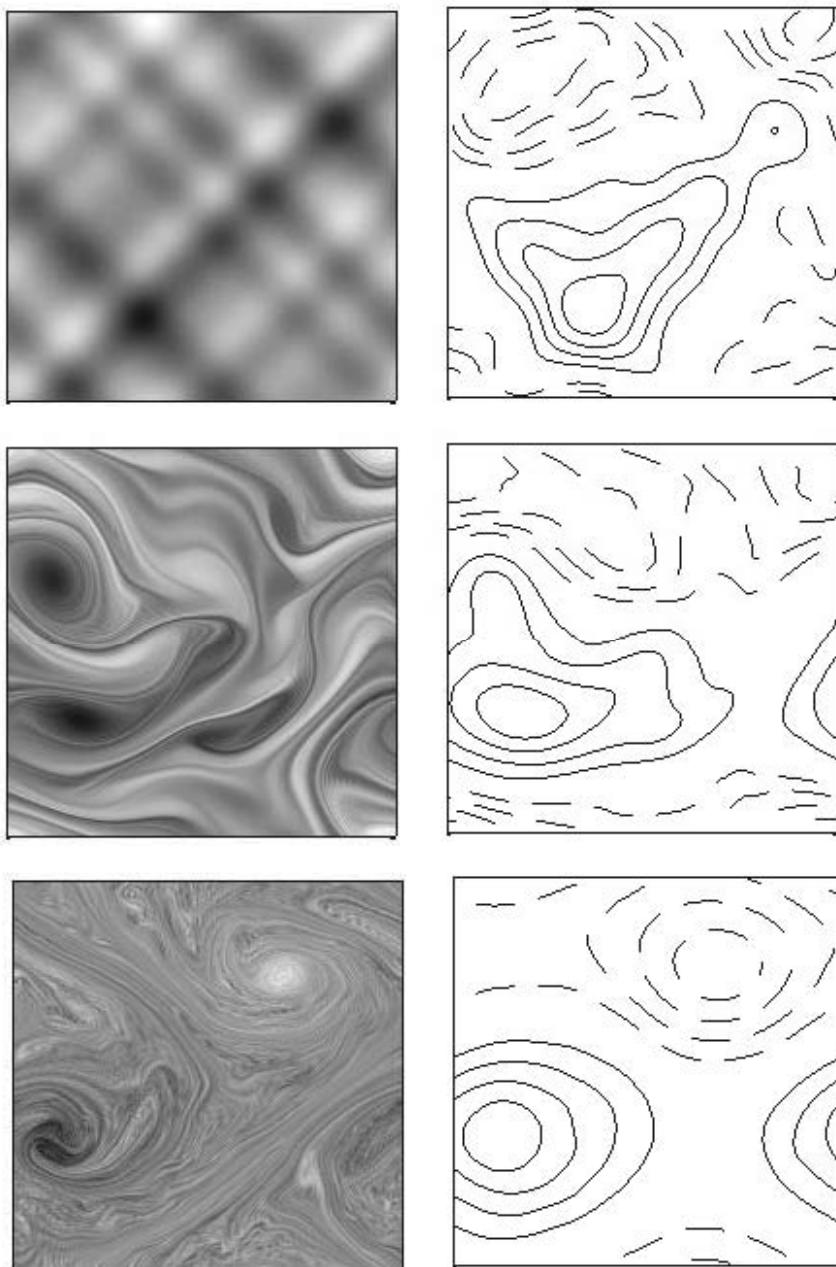


Fig. 8.8 Nearly-free evolution of vorticity (left column) and streamfunction (right column) in a doubly-periodic square domain (of length 2π) at times (from the top, and in units of inverse vorticity) $t = 0, 50$ and 260 , obeying the two-dimensional vorticity equation with no forcing but with a weak viscous term. The initial conditions have just a few non-zero Fourier modes with randomly generated phases, producing a maximum value of vorticity of about three. Kelvin-Helmholtz instability leads to vortex formation and roll-up (as in Fig. 6.6), and like-signed vortices merge, ultimately leading to a state of just two oppositely-signed vortices. Between the vortices, enstrophy cascades to smaller scales. The scale of the streamfunction grows larger, reflecting the transfer of energy to larger scales.

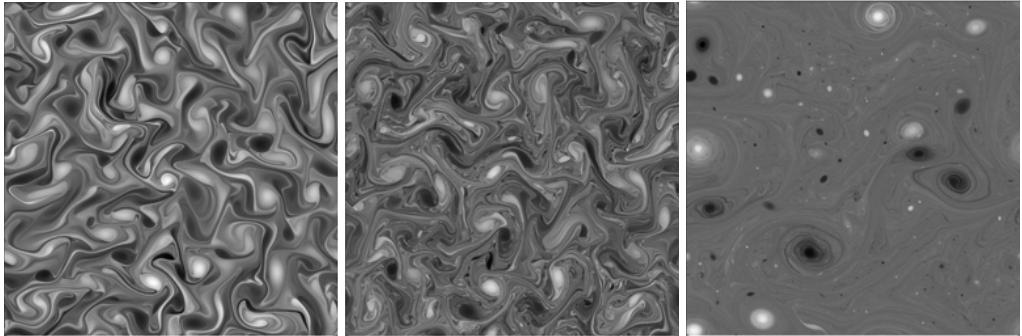


Fig. 8.9 Snapshots of the vorticity field in decaying two-dimensional turbulence, similar to Fig. 8.8, with time increasing left to right. The flow ultimately consists of a small number of vortices whose trajectories are similar to that of interacting point vortices, with occasional close encounters leading to vortex merger.

that we have described. One problem is that enstrophy is only one of an infinity of invariants of inviscid two-dimensional flow, and the theory takes no account of the presence of others. The second is that, as in three-dimensional turbulence, if there is strong intermittency the flow cannot be fully characterized by single enstrophy and energy cascade rates. In two-dimensional turbulence the main form of intermittency may be the formation of coherent vortices, discussed more below. In spite of these problems, the notions of a forward transfer of enstrophy and an inverse transfer of energy are quite robust, and have considerable numerical support.¹²

8.3.4 Numerical illustrations

Numerical simulations nicely illustrate both the classical phenomenology and its shortcomings. In the simulations shown in Figs. 8.8 and 8.9 the vorticity field is initialized ‘randomly’, with little structure in the initial field, and with only a few non-zero Fourier components; the flow then freely evolves, save for the effects of a weak viscosity.

Vortices soon form, and between them enstrophy is cascaded to small scales where it is dissipated, producing a flat and nearly featureless landscape. The energy cascade to larger scales is reflected in the streamfunction field (right column of Fig. 8.8), the length scale of which slowly grows larger with time. The vortices themselves form through a roll-up mechanism, similar to that illustrated in Fig. 6.6, and their presence provides problems to the phenomenology. Because circular vortices are nearly exact, stable solutions of the inviscid equations they can ‘store’ enstrophy, disrupting the relationship between enstrophy flux and enstrophy itself that is assumed in the Kolmogorov-Kraichnan phenomenology. When vortices merge, the enstrophy dissipation rate increases rapidly for a short period of time, so providing a source of intermittency.

Nevertheless, some forced-dissipative numerical simulations suggest that the presence of vortices may be confined to scales close to that of the forcing, and if the resolution is sufficiently high then the $-5/3$ inverse cascade and -3 forward enstrophy cascade appear. Certainly, if the forcing is spectrally localized, then a well-defined $-5/3$ spectrum robustly

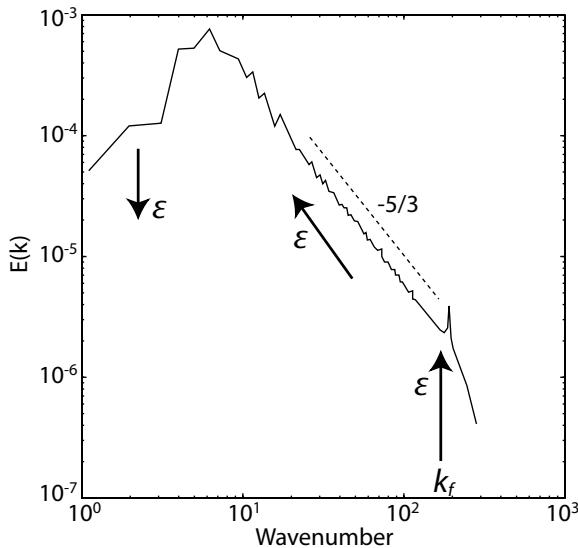


Fig. 8.10 The energy spectrum in a numerical simulation of forced-dissipative two-dimensional turbulence. The fluid is stirred at wavenumber k_f and dissipated at large scales with a linear drag, and there is an $k^{-5/3}$ spectrum at intermediate scales. The arrows indicate the direction of the energy flux, ε .¹³

forms, as illustrated in Fig. 8.10. The forward k^{-3} spectrum is typically more delicate, being influenced by the presence of coherent vortices, but it does arise in some numerical simulations when the resolution is sufficiently high.¹⁴ The atmosphere itself is not observed to have an inverse $-5/3$ spectrum at large scales; indeed there is no well-defined inverse energy cascade in the sense described above. The atmosphere does have a -3 cascade (see Fig. 9.7 in the next chapter), but whether we should attribute this to a classical forward enstrophy cascade is not settled.

8.4 PREDICTABILITY OF TURBULENCE

Small differences in the initial conditions may produce very great ones in the final phenomenon... Prediction becomes impossible... A tenth of a degree more or less at any given point, and the cyclone will burst here and not there, and extend its ravages over districts it might otherwise have spared.

Henri Poincaré, *Science and Method*, 1908.

It is a truth universally acknowledged that weather is hard to predict. That this is so stems from the fact that the atmosphere is chaotic, and chaotic systems are unpredictable virtually by definition. However, the atmosphere (and turbulence in general) is certainly not a *low-dimensional* chaotic system (meaning a system with only a few degrees of freedom), and the connection between atmospheric unpredictability and the so-called ‘sensitive dependence on initial conditions’ of low-dimensional systems is not as straightforward as it might seem. In this section we expand on and clarify these issues, beginning with an informal discussion of a few aspects of low-dimensional dynamical systems.

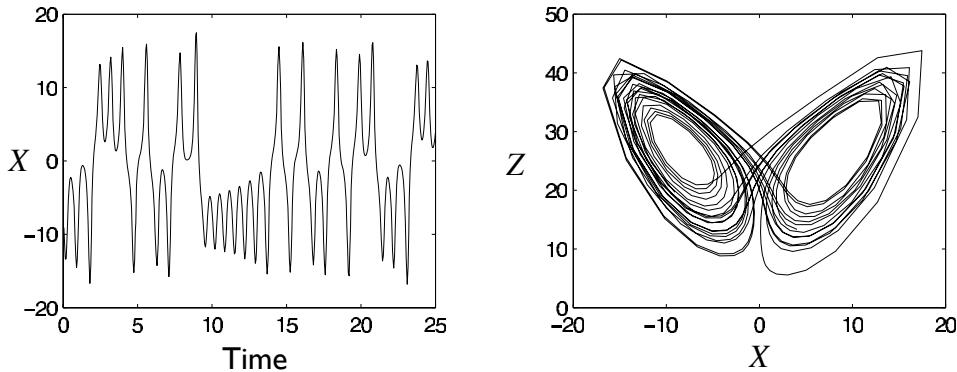


Fig. 8.11 A solution of the Lorenz equations, with $\sigma = 10$, $r = 28$ and $b = 8/3$. The left panel plots X as a function of time. The right panel plots X against Z .

8.4.1 Low-dimensional chaos and unpredictability

Chaos, or temporal disorder leading to effective indeterminism, is a ubiquitous property of nonlinear dynamical systems. Much of this was known to Poincaré, but in its modern reincarnation it stems in part from the ‘Lorenz equations’.¹⁵ These are a set of three coupled nonlinear ordinary differential equations, originally derived by way of a rather *ad hoc* truncation of the fluid equations governing a two-dimensional convective system: the streamfunction of a convective role is written as $\psi(x, z, t) = X(t) \sin kx \sin \pi z$, and the temperature perturbation as $\theta(x, z, t) = Y(t) \cos kx \sin \pi z + Z(t) \sin 2\pi z$, where x and z are the horizontal and vertical coordinates in physical space, k is a wavenumber, and X, Y, Z are amplitudes (not coordinates). Thus, X represents the rotational speed of the convection roll, Y the temperature difference horizontally across the roll, and Z the deviation temperature from a background vertical stratification. The resulting equations are, in notation standard for them,

$$\frac{dX}{dt} = \sigma(Y - X), \quad \frac{dY}{dt} = rX - Y - XZ, \quad \frac{dZ}{dt} = XY - bZ, \quad (8.79a,b,c)$$

where the parameters are σ , the Prandtl number; r , proportional to the Rayleigh number; and b , a wavenumber dependent dissipation coefficient. The behaviour of the system varies with the parameters, and a well studied set uses $\sigma = 10$, $r = 28$ and $b = 8/3$. A typical solution of the system, a ‘flow’, is given in Fig. 8.11, and evidently the behaviour is quite complex. It is aperiodic, and the frequency spectrum (not shown) is quite broad.

Now, suppose that at any given instant the flow is perturbed slightly. Or to put it another way, suppose that we are trying to predict the future behaviour of the system by integrating the equations of motion but that we have inaccurate knowledge about the system at some particular time. We find that the evolution of the original flow and that of the perturbed flow diverge from each other, and after a little while the two systems are completely different (Fig. 8.12). Because we can never expect to have completely accurate information about the state of the system, the system is thus unpredictable; that is, the details of the evolution depend sensitively on the initial conditions, and small errors in the initial conditions grow to finite amplitude. Let us make two other points.

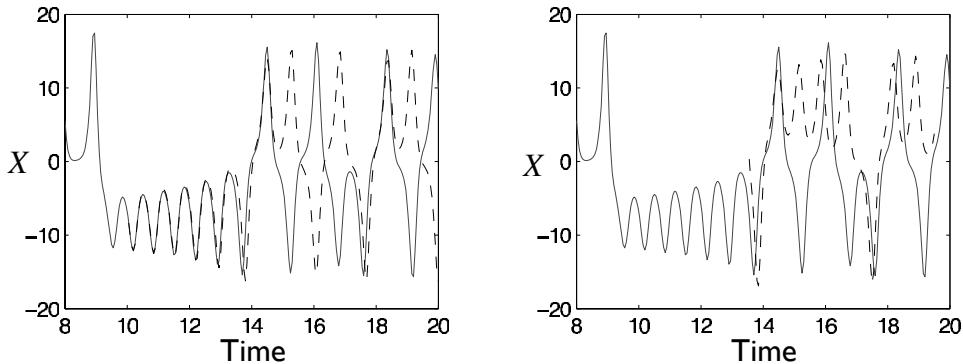


Fig. 8.12 Examples of the evolution of the Lorenz model subject to a small perturbation at time 10 (left panel) and time 13.5 (right panel). The original and perturbed systems are the solid and dashed lines, respectively.

- (i) The time taken for the trajectories to diverge depends on the magnitude of the initial perturbation. Small perturbations grow exponentially at first, and at any given point in the trajectory, the smaller the perturbation the longer the predictability.
- (ii) Once the perturbation has reached finite amplitude, the predictability time — the time for the error to become as large as the solution itself — will typically be of the order of the characteristic advective time of the system, the time for a convective roll to overturn.

Deterministic unpredictability is in fact a common feature of nonlinear dynamical systems, and may be taken as an informal definition of a chaotic system.

8.4.2 * Predictability of a turbulent flow

A turbulent flow will be unpredictable if turbulence is chaotic. Because it is our common experience is that turbulence is both spatially and temporally disordered, it seems, perhaps with the benefit of hindsight, that turbulence must be chaotic and unpredictable. However, this has not always been evident, and it relies on the non-trivial recognition that chaotic systems exist.¹⁶

Presuming that a turbulent flow *is* unpredictable, can we estimate its predictability time, namely the time taken for an initially small error to completely contaminate the system? Let us first note that turbulent flows in general contain multiple scales of motion, and let us suppose that the initial error is confined to small scales. The predictability time of the system may be taken as the time taken for the error to contaminate all scales of motion. There are two possible routes that the error may take in affecting the larger scales. In the first we suppose, following classical turbulence phenomenology, that errors on a small scale will mostly contaminate the motion on the next larger scale (in a logarithmic sense), and that this contamination occurs on the local eddy turnover time. Eddies on this larger scale then grow and affect the next larger scale, and the error field is so cascaded upscale via local triad interactions finally reaching the largest scales of the fluid. Note that this mechanism does not rely on there being an inverse cascade of energy — it is only the error,

or the contamination, that is cascaded upscale. In the second route we suppose that errors occurring on the small scale immediately contaminate the largest scales, with an initial error equal to the amplitude of the small scale, and that the large-scale error then grows exponentially.

I. Error growth via a local cascade

Let us suppose that the error is initially confined to some small scale characterized by the (inverse of) the wavenumber k_1 , as determined by the resolution of our observing network. For modes at that scale the error may be considered finite rather than infinitesimal, and it will saturate and contaminate the next largest scale in a time scale comparable to the eddy turnover time at that scale. Thus, in general, errors initially confined to a scale k will contaminate the scale $2k$ after a time τ_k , with τ_k given by (8.24). The total time taken for errors to propagate from the small scale k_1 to the largest scale k_0 is then given by

$$T = \int_{k_0}^{k_1} \tau_k d(\ln k) = \int_{k_0}^{k_1} [k^3 E(k)]^{-1/2} d(\ln k), \quad (8.80)$$

treating the wavenumber spectrum as continuous. The logarithmic integral arises because the cascade proceeds logarithmically — error cascades from k to $2k$ in a time τ_k . For an energy spectrum of the form $E = Ak^{-n}$ this becomes

$$T = \frac{2}{A^{1/2}(n-3)} \left[k^{(n-3)/2} \right]_{k_0}^{k_1}, \quad (8.81)$$

for $n \neq 3$, and $T = A^{-1/2} \ln(k_1/k_0)$ for $n = 3$. If in two-dimensional turbulence we have $n = 3$ and $A = \eta^{2/3}$, and if in three-dimensional turbulence we have $n = 5/3$ and $A = \varepsilon^{2/3}$, then the respective predictability times are given by:

$$T_{2d} \sim \eta^{-1/3} \ln(k_1/k_0),$$

$$T_{3d} \sim \varepsilon^{-1/3} k_0^{-2/3}$$

(8.82a,b)

As $k_1 \rightarrow \infty$, that is as the initial error is confined to smaller and smaller scales, predictability time grows larger for two-dimensional turbulence (and for $n \geq 3$ in general), but remains finite for three-dimensional turbulence.

II. Error growth via a direct interaction

Let us now assume that the small scale error directly affects the large scales, where the error then grows exponentially until it saturates. That is, if ϕ is a measure of the amplitude of the large-scale error then

$$\phi \sim \phi_0 \exp(\sigma t), \quad (8.83)$$

where σ is the inverse of the eddy turnover time at the large scale, and ϕ_0 is the amplitude of its initial error and this, we assume, is equal to the amplitude of the motion at the poorly-observed small scales at wavenumber k_1 . In the enstrophy cascade of two-dimensional turbulence the eddy turnover time is given by $\tau_k \sim \eta^{-1/3}$ and $A \sim \eta^{2/3}$, and so we take

$$\sigma = \eta^{1/3} = A^{1/2}, \quad \phi_0 = Ak_1^{-n}. \quad (8.84)$$

where $n = 3$. The time, T'_{2d} , needed for the error to saturate the large scales, k_0 , is then approximately given by the solution of

$$Ak_0^{-n} = Ak_1^{-n} \exp(A^{1/2}T'_{2d}), \quad (8.85)$$

giving

$$T'_{2d} \sim \eta^{-1/3} \ln(k_1/k_0). \quad (8.86)$$

In three-dimensional turbulence the eddy turnover time is given by $\tau_k \sim \varepsilon^{-1/3} k^{-2/3}$ and $A \sim \varepsilon^{2/3}$, and so we take

$$\sigma = k_0^{2/3} A^{1/2}, \quad \phi_0 = Ak_1^{-n} \quad (8.87)$$

where $n = 5/3$. The time, T'_{3d} , needed for an error to saturate the large scale is then approximately given by the solution of

$$Ak_0^{-n} = Ak_1^{-n} \exp(k_0^{2/3} A^{1/2} T'_{3d}), \quad (8.88)$$

giving

$$T'_{3d} \sim k_0^{-2/3} \varepsilon^{-1/3} \ln(k_1/k_0). \quad (8.89)$$

The estimates (8.86) and (8.89) are to be compared with (8.82). For two-dimensional turbulence, the estimates are equal (reflecting the scale independence of the eddy turnover time), whereas for three-dimensional turbulence the estimate from (8.82) is much shorter than that from (8.89) if $k_1 \gg k_0$, meaning that the local cascade mechanism of error growth will dominate.

8.4.3 Implications and weather predictability

In two-dimensional flow the predictability time, (8.82a) increases without bound as the scale of the initial error decreases. This is consistent with what has been rigorously proven about the two-dimensional Navier-Stokes equations, namely that provided the initial conditions are sufficiently smooth, the solutions have a continuous dependence on the initial conditions, and a change in solution at some later time may be bounded by reducing the magnitude of the change in the initial conditions. This does not mean that two-dimensional flow is in practice necessarily predictable: a small error or small amount of noise in the system will still render a flow truly unpredictable sometime in the future, but we can put off that time indefinitely if we know the initial conditions well enough.

In three dimensions, with a spectrum of $k^{-5/3}$, the predictability-time estimate from (8.82b) is not dependent on the scale of the initial error. Thus, even if the initial error is confined to smaller and smaller scales, the predictability time is bounded. The time it takes for such errors to spread to the largest scales is simply a few large eddy turnover times, essentially because the eddy turnover times of the small scales are so small. For such a fluid, there is no unique error doubling time, because the error growth rate is a function of scale.

In the troposphere the large-scale flow behaves more like a two-dimensional fluid than a three-dimensional fluid and from scales from a few hundred to a few thousand kilometres it has, roughly, a k^{-3} spectrum (look ahead to Fig. 9.7). If this spectrum extended indefinitely to small scales the predictability time would be correspondingly large, but at scales smaller than about 100 km or so, the atmosphere starts to behave more three dimensionally and the predictability time cannot be significantly extended by making observations at still finer

scales. That is, the effective limit to predictability is governed by the horizontal scale at which the atmosphere turns three dimensional. (Hypothetically, we might be able to increase the predictability time if we could observe scales well into the viscous regime where the spectrum steepens again, but this is not a practicable proposition.) Putting in the numbers gives a predictability limit of about 12 days (but there is at least a factor of two uncertainty in such a calculation), and small perturbations that are impossible to observe will change the course of the large-scale weather systems on this time scale. The ‘butterfly effect’ has its origins in this argument: a butterfly flapping its wings is, so it goes, able to change the course of the weather a week or so later.¹⁷

Thus, as regards our attempts to predict weather, as the atmosphere becomes observed more and more accurately, the initial error will become concentrated at smaller and smaller scales, eventually reaching the scale at which the atmosphere ceases to behave as a quasi-two-dimensional fluid and at which its spectrum flattens. The initial error growth rate will then increase, indicating the unavoidable onset of diminishing returns in adding resolution to our observing systems. Note that, unlike the situation in a low-order chaotic system, the growth rate of errors in a turbulent flow is not, in general, exponential (except for a pure -3 spectrum) even for a small initial error. This is because the initial error will never be properly infinitesimal, in that a given error will nearly always project on to some scale at finite amplitude.¹⁸

8.5 * SPECTRA OF PASSIVE TRACERS

In fluid dynamics we are often concerned with the transport of tracers by turbulent flows. An *active tracer* is one that affects the flow itself (potential vorticity and salt are examples), whereas a *passive tracer* does not affect the flow field (a neutrally buoyant dye, for example), and its dynamics are often simpler. In some circumstances an active tracer can be considered to be approximately passive; the atmospheric temperature field at very large scales turns out to be one example, and its transport is obviously a key determinant of our climate. Let us therefore spend some time discussing the dynamics of a passive tracer; in this section we will consider their spectra, and in chapter 10 we will consider their diffusive transport.

We consider a tracer that obeys

$$\frac{D\phi}{Dt} = F[\phi] + \kappa \nabla^2 \phi, \quad (8.90)$$

where $F[\phi]$ is the stirring of the dye and κ is its diffusivity; κ in general differs from the kinematic molecular viscosity ν . If ϕ is temperature, the ratio of viscosity to diffusivity is called the *Prandtl number* and denoted σ , so that $\sigma \equiv \nu/\kappa$. If ϕ is a passive tracer, the ratio is sometimes called the *Schmidt number*, but we shall call it the Prandtl number in all cases. We assume that the tracer variance is created at some well-defined scale k_0 , and that κ is sufficiently small that dissipation only occurs at very small scales. (Note that dissipation only reduces the tracer *variance*, not the amount of tracer itself.) The turbulent flow will generically tend to stretch patches of dye into elongated filaments, in much the same way as vorticity in two-dimensional turbulence is filamented — note that Fig. 8.5 applies just as well to a passive tracer in either two or three dimensions as it does to vorticity in two dimensions. Thus we expect a transfer of tracer variance from large to small scales. If the dye is stirred

at a rate χ then, by analogy with our treatment of the cascade of energy, we posit that

$$\mathcal{K}_\chi \chi \propto \frac{\mathcal{P}(k)k}{\tau_k}, \quad (8.91)$$

where $\mathcal{P}(k)$ is the spectrum of the tracer, k is the wavenumber, τ_k is an eddy time scale and \mathcal{K}_χ is a constant, not necessarily the same constant in all cases. (In the rest of the section, Kolmogorov-like constants will be denoted by \mathcal{K} , differentiated by miscellaneous superscripts or subscripts.) We will also assume that τ_k is given by

$$\tau_k = [k^3 \mathcal{E}(k)]^{-1/2}. \quad (8.92)$$

Suppose that the turbulent spectrum is given by $\mathcal{E}(k) = Ak^{-n}$. Using (8.92), (8.91) becomes

$$\mathcal{K}_\chi \chi = \frac{\mathcal{P}(k)k}{[Ak^{3-n}]^{-1/2}}, \quad (8.93)$$

and

$$\boxed{\mathcal{P}(k) = \mathcal{K}_\chi A^{-1/2} \chi k^{(n-5)/2}}. \quad (8.94)$$

Note that the steeper the energy spectrum the shallower the tracer spectrum. If the energy spectrum is steeper than -3 then (8.92) may not be a good estimate of the eddy turnover time, and we use instead

$$\tau_k = \left[\int_{k_0}^k p^2 \mathcal{E}(p) dp \right]^{-1/2}, \quad (8.95)$$

where k_0 is the low-wavenumber limit of the spectrum. If the energy spectrum is shallower than -3 , then the integrand is dominated by the contributions from high wavenumbers and (8.95) effectively reduces to (8.92). If the energy spectrum is steeper than -3 , then the integrand is dominated by contributions from low wavenumbers. For $k \gg k_0$ we can approximate the integral by $[k_0^3 \mathcal{E}(k_0)]^{-1/2}$, that is the eddy-turnover time at large scales, τ_{k_0} , given by (8.92). The tracer spectrum then becomes

$$\boxed{\mathcal{P}(k) = \mathcal{K}'_\chi \chi \tau_{k_0} k^{-1}}, \quad (8.96)$$

where \mathcal{K}'_χ is a constant. In all these cases the tracer cascade is to smaller scales even if, as may happen in two-dimensional turbulence, energy is cascading to larger scales.

The scale at which diffusion becomes important is given by equating the turbulent time scale τ_k to the diffusive time scale $(\kappa k^2)^{-1}$. This is independent of the flux of tracer, χ , essentially because the equation for the tracer is linear. Determination of expressions for these scales in two and three dimensions are left as problems for the reader.

8.5.1 Examples of tracer spectra

Energy inertial range flow in three dimensions

Consider a range of wavenumbers over which neither viscosity nor diffusivity directly influence the turbulent motion and the tracer. Then, in (8.94), $A = \mathcal{K} \varepsilon^{2/3}$ where ε is the rate

of energy transfer to small scales, \mathcal{K} is the Kolmogorov constant, and $n = 5/3$. The tracer spectrum becomes¹⁹

$$\mathcal{P}(k) = \mathcal{K}_\chi^{3d} \varepsilon^{-1/3} \chi k^{-5/3}. \quad (8.97)$$

where \mathcal{K}_χ^{3d} is a (putatively universal) constant. It is interesting that the $-5/3$ exponent appears in both the energy spectrum and the passive tracer spectrum. Using (8.92), this is the only spectral slope for which this occurs. Experiments show that this range does, at least approximately, exist with a value of \mathcal{K}_χ^{3d} of about 0.5–0.6 in three dimensions.

Inverse energy-cascade range in two-dimensional turbulence

Suppose that the energy injection occurs at a smaller scale than the tracer injection, so that there exists a range of wavenumbers over which energy is cascading to larger scales while tracer variance is simultaneously cascading to smaller scales. The tracer spectrum is then

$$\mathcal{P}(k) = \mathcal{K}_\chi^{2d} \varepsilon^{-1/3} \chi k^{-5/3}, \quad (8.98)$$

the same as (8.97), although ε is now the energy cascade rate to larger scales and the constant \mathcal{K}_χ^{2d} does not necessarily equal \mathcal{K}_χ^{3d} .

Enstrophy inertial range in two-dimensional turbulence

In the forward enstrophy inertial range the eddy time scale is $\tau_k = \eta^{-1/3}$ (assuming of course that the classical phenomenology holds). Directly from (8.91) the corresponding tracer spectrum is then

$$\mathcal{P}(k) = \mathcal{K}_\chi^{2d*} \eta^{-1/3} \chi k^{-1}. \quad (8.99)$$

The passive tracer spectrum now has the same slope as the spectrum of vorticity variance (i.e., the enstrophy spectrum), which is perhaps reassuring since the tracer and vorticity obey similar equations in two dimensions.

The viscous-advection range of large Prandtl number flow

If the Prandtl number $\sigma = \nu/\kappa \gg 1$ (and in seawater $\sigma \approx 7$) then there may exist a range of wavenumbers in which viscosity is important, but not tracer diffusion. The energy spectrum is then very steep, and (8.96) will apply. The straining then comes from wavenumbers near the viscous scale, so that for three dimensional flow the appropriate k_0 to use in (8.96) is the viscous wavenumber, and $k_0 = k_\nu = (\varepsilon/\nu^3)^{1/4}$. The dynamical time scale at this wavenumber is given by

$$\tau_{k_\nu} = \left(\frac{\nu}{\varepsilon} \right)^{1/2}, \quad (8.100)$$

and using this and (8.96) the tracer spectrum in this viscous-advection range becomes

$$\mathcal{P}(k) = \mathcal{K}'_B \left(\frac{\nu}{\varepsilon} \right)^{1/2} \chi k^{-1}. \quad (8.101)$$

This spectral form applies for $k_\nu < k < k_\kappa$, where k_κ is the wavenumber at which diffusion becomes important, found by equating the eddy turnover time given by (8.100) with the diffusive time scale $(\kappa k^2)^{-1}$. This gives

$$k_\kappa = \left(\frac{\varepsilon}{\nu \kappa^2} \right)^{1/4}, \quad (8.102)$$

and k_κ is known as the Batchelor wavenumber (and its inverse is the Batchelor scale). Beyond k_κ , the diffusive flux is not constant and the tracer spectrum can be expected to decay as wavenumber increases. A heuristic way to calculate the spectrum in the diffusive range is to first note that the flux of the tracer is not constant but diminishes according to

$$\frac{d\chi'(k)}{dk} = -2\kappa k^2 \mathcal{P}(k), \quad (8.103)$$

where χ' is the wavenumber-dependent rate of tracer transfer. Let us still assume that χ' and $\mathcal{P}(k)$ are related by an equation of the form (8.91), where now τ_k is a constant, given by (8.100). Thus,

$$\mathcal{K}_B \chi' = \frac{\mathcal{P}(k)k}{\tau_{k_\kappa}} = \frac{\mathcal{P}(k)k}{(\nu/\varepsilon)^{1/2}}, \quad (8.104)$$

where \mathcal{K}_B is a constant. Using (8.103) and (8.104) we obtain

$$\frac{d\chi'}{dk} = -2\mathcal{K}_B \kappa k \left(\frac{\nu}{\varepsilon}\right)^{1/2} \chi'. \quad (8.105)$$

Solving this, using $\chi' = \chi$ (where χ is a constant) for small k , gives

$$\mathcal{P}(k) = \mathcal{K}_B \left(\frac{\nu}{\varepsilon}\right)^{1/2} \chi k^{-1} \exp[-\mathcal{K}_B (k/k_\kappa)^2]. \quad (8.106)$$

This reduces to (8.101) if $k \ll k_\kappa$, and is known as the Batchelor spectrum.²⁰ Its high-wavenumber part $k > k_\kappa$ is known as the viscous-diffusive subrange. The spectrum, and its two-dimensional analogue, is illustrated in Fig. 8.13

In two dimensions the viscous-advection range occurs for wavenumbers greater than $k_v = (\eta/\nu^3)^{1/6}$. The appropriate time scale within this subrange is $\eta^{-1/3}$, and therefore gives a spectrum with the precisely the same form as (8.99). At sufficiently high wavenumbers tracer diffusion becomes important, with the diffusive scale now given by equating the eddy turnover time $\eta^{-1/3}$ with the viscous time scale $(\kappa k^2)^{-1}$. This gives the diffusive wavenumber, analogous to (8.102), of $k_\kappa = (\eta/\kappa^3)^{1/6}$. Using (8.105) and the procedure above we then obtain an expression for the spectrum in the region $k > k_\kappa$, that is a two-dimensional analogue of (8.106), namely

$$\mathcal{P}(k) = \mathcal{K}'_B \eta^{-1/3} \chi k^{-1} \exp[-\mathcal{K}'_B (k/k_\kappa)^2]. \quad (8.107)$$

For $k \ll k_\kappa$ this reduces to (8.99), possibly with a different value of the Kolmogorov-like constant.

† The inertial-diffusive range of small Prandtl number flow

For small Prandtl number ($\nu/\kappa \ll 1$) the energy inertial range may coexist with a range over which tracer variance is being dissipated, giving us the so-called inertial-diffusive range. The tracer will begin to be dissipated at a wavenumber obtained by equating a dynamical eddy turnover time with a diffusive time, and this gives a diffusive wavenumber

$$k'_\kappa = \begin{cases} (\varepsilon/\kappa^3)^{1/4} & \text{in three dimensions,} \\ (\eta/\kappa^3)^{1/6} & \text{in two dimensions.} \end{cases} \quad (8.108)$$

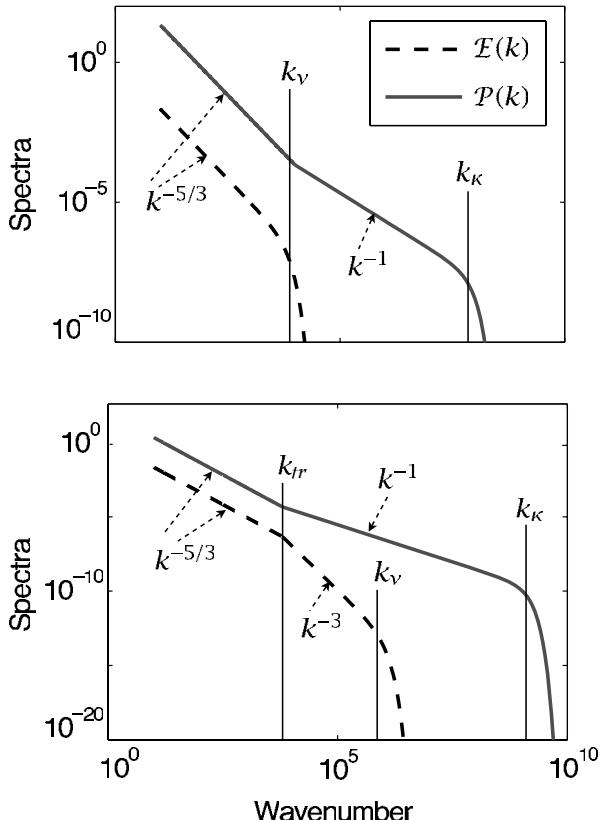


Fig. 8.13 The energy spectra, $\mathcal{E}(k)$ and passive tracer spectra $\mathcal{P}(k)$ in large Prandtl number three-dimensional (top) and two-dimensional (bottom) turbulence. In three dimensions $\mathcal{P}(k)$ is given by (8.97) for $k < k_v$ and by (8.106) for $k > k_v$. In two dimensions, if k_{tr} marks the transition between a $k^{-5/3}$ inverse energy cascade and a k^{-1} forward enstrophy cascade, then $\mathcal{P}(k)$ is given by (8.98) for $k < k_{tr}$ and by (8.107) for $k > k_{tr}$. In both two and three dimensions the tracer spectra falls off rapidly for $k > k_k$.

Beyond the diffusive wavenumber the flux of the tracer is no longer constant but diminishes according to (8.103).

Given a non-constant flux and an eddy-turnover time that varies with wavenumber there is no self-evidently correct way to proceed. One way is to assume that χ and $\mathcal{P}(k)$ are related by $\mathcal{K}_\chi'' \chi = \mathcal{P}(k)k/\tau_k$ [as in (8.104), but with a potentially different proportionality constant] and with τ_k given by (8.92); that is, $\tau_k = \varepsilon^{-1/3}k^{-2/3}$ in three dimensional turbulence. Using this in (8.103) leads to

$$\mathcal{P}(k) = \mathcal{K}_\chi'' \chi \varepsilon^{-1/3} k^{-5/3} \exp[-(\mathcal{K}_\chi'' 3/2)(k/k'_k)^{4/3}], \quad (8.109)$$

where χ is the tracer flux at the beginning of the tracer dissipation range. (A similar expression emerges in two-dimensional turbulence.) However, given such a steep spectrum an argument based on spectral locality is sometimes thought to be suspect. Another argument posits a particular relationship between the tracer spectrum and energy spectrum in the inertial-diffusive range, and this leads to

$$\mathcal{P}(k) = \frac{\mathcal{K}_B''}{3} \chi_0 \varepsilon^{2/3} \kappa^{-3} k^{-17/3} = \mathcal{K}_B'' \chi_0 \varepsilon^{-1/3} k^{-5/3} g(k/k_k), \quad (8.110)$$

where $g(\alpha) = \alpha^{-4}/3$ and \mathcal{K}_B'' is a constant.²¹

Notes

- 1 Horace Lamb has been quoted as saying that when he died and went to heaven he hoped for enlightenment on two things, quantum electrodynamics and turbulence, although he was only optimistic about the former. Quite consistently, it has also been said that turbulence is the invention of the Devil, put on Earth to torment us.
- 2 The algebra of the three-dimensional case is more complicated because of the pressure term and because the momentum equation is a vector equation. Nevertheless, in incompressible flow we can take the divergence of the momentum equation to obtain an elliptic equation for pressure of the form $\nabla^2 p = Q(\mathbf{v})$ where the right-hand side is a quadratic function of velocity and its derivatives, Fourier transform this and then proceed much as in the two-dimensional case.
- 3 A. N. Kolmogorov (1903–1987) was a Russian mathematician and theoretical physicist, who made seminal contributions to turbulence (with famous papers in 1941 and in 1962), to probability and statistics, and to classical mechanics (e.g., the Kolmogorov–Arnold–Moser theorem). Yaglom (1994) provides more details on both the man and his scientific contributions.
- 4 The process has been encapsulated in the following ditty by Lewis Fry (L.F.) Richardson (1881–1953), his own summary of Richardson (1920):

*Big whorls have little whorls, that feed on their velocity;
And little whorls have lesser whorls, and so on to viscosity.*

The verse follows a well-known one by the mathematician Augustus De Morgan (in *A Budget of Paradoxes* 1872), '*Great fleas have little fleas upon their backs to bite 'em...*', which in turn is a parody on a poem by Jonathan Swift. Richardson himself was a British scientist best known as the person who (following earlier work by V. Bjerknes) envisioned weather forecasting in its current form — that is, numerical weather prediction. However, as described in his 1922 book, instead of an electronic computer performing the calculations, he imagined, perhaps fancifully, a hall full of people performing calculations in unison all directed by a conductor at the front. His first numerical forecast, calculated by hand, was wildly inaccurate both because he failed to initialize his atmosphere properly and because his timestep was too long and did not satisfy the CFL condition, and unrealistic gravity waves dominated the solution. However, it was a prescient and important effort. He also worked on turbulence, and seems to have envisioned the turbulent cascade prior to Kolmogorov (to wit the verse above), and the 'Richardson number', a measure of fluid stratification, is named after him. He also made contributions to the theory of war and was known as a pacifist — he was a conscientious objector and drove ambulances in the first World War, and resigned from the U. K. Meteorological Office because it became part of the Air Ministry.

- 5 Kolmogorov (1941) obtained the result in a slightly different way, using distances in real space rather than wavenumber and deriving the equivalent result for the longitudinal structure function, $D(r) \equiv \langle [\mathbf{u}(\mathbf{x} + \mathbf{r}) - \mathbf{u}(\mathbf{x})]^2 \rangle \sim r^{2/3}$. It was Obukhov (1941) who gave an argument in spectral space and first wrote down that $E(k) \sim k^{-5/3}$. Kolmogorov's argument is regarded as more general and hence the 5/3 spectrum is usually named for him, but sometimes it is called the 'Kolmogorov–Obukhov' 5/3 spectrum.
- 6 Results kindly provided by R. Zhao.
- 7 The first observations confirming the -5/3 predictions were from a tidal channel (Grant et al. 1962). These results were initially presented at a turbulence conference in Marseille in 1961, ironically at the same time as Kolmogorov presented a modification of his original theory that incorporated a local mean dissipation rate, to try to take intermittency into account,

recognizing that his first theory was incomplete (Kolmogorov 1962). It is said to have been L. D. Landau who pointed out the consequences of intermittency to Kolmogorov, soon after the K41 theory first appeared.

- 8 Two early papers on two-dimensional turbulence are those of Lee (1951) and Fjørtoft (1953), the former noting the incompatibility of the material conservation of vorticity with Kolmogorov's energy inertial range, and the latter recognizing the two-dimensional nature of large-scale atmospheric motion. Batchelor (1953b) noted the tendency of energy to concentrate in small wavenumbers as a consequence of energy and enstrophy conservation. The theory was more fully developed by Kraichnan (1967) (who predicted the spectral slopes of the two-dimensional cascades), Leith (1968) and Batchelor (1969). Lilly (1969) performed some early numerical integrations. For reviews of two-dimensional turbulence see Kraichnan & Montgomery (1980) and Danilov & Gurarie (2001).
- 9 This similarity argument is due to Batchelor (1969), and its validity was explored by Bartello & Warn (1996) using numerical simulations of decaying two-dimensional turbulence. They found that the similarity hypothesis is not quantitatively accurate, and in particular that higher-order moments of the vorticity do not obey the predictions of the theory. This failure may be ascribed to the fact that in two-dimensional flow vorticity is conserved on parcels, and in the presence of coherent vortices this is an effective constraint that is not included in the theory. In flow with a finite deformation radius coherent vortices are found to be less important and an analogous similarity hypothesis appears to work better (Iwayama *et al.* 2002).
- 10 Arbic *et al.* (2005), Colin de Verdière (1980). See also Scott (2001).
- 11 To read more about strain rate, see Lesieur (1997).
- 12 The realization that in two-dimensional turbulence energy is transferred to larger scales was arguably one of the most important developments in fluid mechanics in the second half of the twentieth century. Its manifestation in stratified and/or rotating three-dimensional fluids is described in, for example, Peltier & Stuhne (2002) and Smith & Waleffe (1999). One might characterize the upscale cascade in two-dimensional turbulence by:

*Smaller whorls beget bigger whorls, that feed upon their energy;
And bigger whorls beget larger whorls, and so on, with much synergy.*
- 13 Adapted from Maltrud & Vallis (1991).
- 14 For numerical simulations illustrating these and other properties of two-dimensional turbulence see, among others, McWilliams (1984), Maltrud & Vallis (1991), Oetzel & Vallis (1997), Lindborg & Alvelius (2000) and Smith *et al.* (2002). Also look at Jupiter through a telescope! A statistical-mechanical argument that there is tendency for a small number of vortices to form in two-dimensional turbulence was given by Onsager (1949).
- 15 The Lorenz equations were written down by Lorenz (1963), based on some earlier work of Saltzman (1962), and inspired a veritable industry of study. The field of chaos, or more generally nonlinear dynamics, has grown enormously since then, prompted also by work in mathematics occurring at the same time, and its development is sometimes regarded as one of the true revolutions of science in the twentieth century. Aubin & Dahan Dalmedico (2002) write an interesting history of it. The correspondence of the Lorenz equations to a real fluid system is tenuous, but the importance of the properties they demonstrate transcends this; we regard the equations simply as an example of a chaotic system with some fluid relevance. The equations and variations about them have reappeared in studies of, among other things, lasers, dynamos, chemical reactions, mechanical waterwheels and El Niño.
- 16 That a turbulent flow is, *inter alia*, chaotic and unpredictable follows from the work of Lorenz

(1963), Ruelle & Takens (1971) and others who showed that fluid turbulence was generically a consequence of a small number of bifurcations as some controlling parameter (such as the Reynolds number) is changed. Prior to this, turbulence was sometimes thought, following Landau (1944), to be a large collection of periodic motions with incommensurate frequencies that would have complex and non-repeating but presumably predictable behaviour. Notwithstanding the Landau picture, it seems to have been known, well before the development of nonlinear dynamical systems theory in the 1960s and 1970s, that the weather was inherently unpredictable. Poincaré was perhaps the first to properly understand this at the turn of the twentieth century, although in the review of a book by the thermodynamicist P. Duhem, W. S. Franklin wrote in 1898 that 'An infinitesimal cause produces a finite effect. Long range detailed weather prediction is therefore impossible... the flight of a grasshopper in Montana may turn a storm aside from Philadelphia to New York!' Weather forecasters themselves long seem to have intuited that the atmosphere was intrinsically, and not just practically, difficult to forecast: in the 1941 novel *Storm* by G. R. Stewart (1895–1980, a professor of English at UC Berkeley) we find a forecaster recalling his old professor's saying that 'A Chinaman sneezing in Shen-si may set men to shoveling snow in New York City'. (*Storm* is also notable because it used female names for intense storms, a practice that became common among forecasters in World War II and that was used by the US Weather Service from 1953–1978, after which gender equity pertained.) In a more academic setting, the predictability problem is mentioned in the book by Godske *et al.* (1957) (much of which was written in the 1930s and 1940s), and Thompson (1957) and Novikov (1959) studied the unpredictability of atmospheric flows from the perspective of turbulence, evidently unaware of, or at least uninfluenced by, either Poincaré or Landau. Phillip Thompson himself was in the Joint Numerical Weather Prediction Unit of the U.S. government in the 1950s, whose task was to numerically produce weather forecasts, and this practical experience undoubtedly confronted and guided his theoretical thinking. These various strands came together and were clarified by the dynamical systems viewpoint coupled with the view of the atmosphere as a geostrophically turbulent fluid, and this led to the viewpoint we describe here, and to estimates of the limit to predictability of the atmospheric weather of about two weeks, although at any given time the predictability may certainly be shorter or longer than that.

- 17 The more technical phrase 'sensitive dependence on initial conditions' is a paraphrase of one of Poincaré, but the catchier one 'butterfly effect' is more recent. It seems to have first appeared in a meteorological context in Smagorinsky (1969), where we find 'Would the flutter of a butterfly's wings ultimately amplify to the point where the numerical simulation departed from reality...? If not the flutter of the butterfly's wings, the disturbance might be the result of instrumental errors in the initial conditions....' The phrase became more well known following a lecture by Lorenz to the American Association for the Advancement of Science (AAAS) in 1972 entitled 'Predictability: does the flap of a butterfly's wings in Brazil set off a tornado in Texas?' The shape of the Lorenz attractor in phase space also resembles a butterfly (see Fig. 8.11), and the two (not wholly unrelated) phenomena are sometimes conflated. See also endnote 16 above and the brief history by Hilborn (2004).
- 18 A related point is that the predictability of a turbulent system is not well characterized by its spectrum of Lyapunov exponents: in a turbulent system in three dimensions the largest Lyapunov exponent is likely to be associated with very small scales of motion, and the error growth associated with this effectively saturates at small scales.
- 19 First derived by Obukhov (1949). See also Corrsin (1951).
- 20 Batchelor (1959), who also suggests that the constant \mathcal{K}_B in (8.106) should have the value 2. There is some observational support for the k^{-1} viscous-advection range in the temperature spectra of the ocean, one of the first measurements being that of Grant *et al.* (1968). Aside

from their intrinsic interest, the viscous and diffusive scales are used in microstructure theory and measurements that lead to estimates of the ocean's energy dissipation rate. See, for example, Gregg (1998) and Stips (2005).

- 21 Batchelor *et al.* (1959). There is some numerical support for the $-17/3$ spectrum using a large-eddy simulation (LES) model (Chasnov 1991). See also O'Gorman & Pullin (2005).

Further reading

There are a number of books on turbulence, including:

Tennekes, H. & Lumley, J., 1972. *A First Course in Turbulence*.

This remains a classic introduction to the subject.

McComb, W. D., 1990. *The Physics of Fluid Turbulence*.

Lesieur, M., 1997. *Turbulence in Fluids*.

The above two books deal with both the statistical and phenomenological aspects of turbulence, with somewhat different emphases and styles.

Frisch, U., 1995. *Turbulence: the Legacy of A. N. Kolmogorov*.

Concentrates, as its title might suggest, on the statistical aspects of the subject.

A set of review articles covering a very wide range of topics related to turbulence in marine systems, on both large and small scales, is collected in:

Baumert, H. Z., Simpson, J. H. & Sünderman, J., Eds., 2003. *Marine Turbulence: Theories, Observations and Models*.

Given its price, check it out from your library.

The more applied side of numerical weather prediction and predictability is discussed by:

Kalnay, E., 2003. *Atmospheric Modeling, Data Assimilation and Predictability*.

Problems

- 8.1 Different averaging processes (time, space, or ensemble means, projection operators, etc.) may lead to different rules for the products of averages. Let $a = \bar{a} + a'$, where \bar{a} is some kind of average, and similarly for the variable b . Thus

$$\bar{a}\bar{b} = \overline{\bar{a}\bar{b}} + \overline{\bar{a}b'} + \overline{a'\bar{b}} + \overline{a'b'}. \quad (P8.1)$$

Is this expression always equal to $\bar{a}\bar{b} + a'b'$? Give examples. Similarly, is it always true that $\overline{ab} = \bar{a}\bar{b}$?

- 8.2 Consider the upscale energy transfer of two-dimensional turbulence, obeying the equation of motion, $D\zeta/Dt = -r\zeta + F$, where F is a stirring that is non-zero only at large wavenumbers and that produces an upscale energy transfer at the rate ε_0 , and r is a constant drag coefficient.

(a) Show that drag is important only for wavenumbers smaller than $k_r = (r^3/\varepsilon_0)^{1/2}$.

(b) ♦ Suppose that, at all wavenumbers, the energy spectrum $\mathcal{E}(k)$ and the cascade rate ε are related by $\varepsilon = \mathcal{E}(k)k/\tau_k$ where $\tau_k = [\mathcal{E}(k)k^3]^{-1/2}$. However, at wavenumbers smaller than k_r , the energy cascade rate diminishes with wavenumber because of frictional effects. Obtain an expression for the energy spectrum in this frictional range.

- 8.3 ♦ Two-dimensional flow with a finite deformation radius obeys

$$\frac{\partial q}{\partial t} + J(\psi, q) = 0, \quad q = (\nabla^2 - k_d^2)\psi. \quad (8.2a,b)$$

Consider the asymptotic case with $k_d \rightarrow \infty$.

- (a) Show that the equation may be written in the form $\partial \hat{q} / \partial \hat{t} + J(\hat{\psi}, \hat{q}) = 0$ where $\hat{t} = t/k_d^2$, $\hat{q} = \psi$ and $\hat{\psi} = \nabla^2 \hat{q}$.
- (b) What are the inviscid invariants of such a model?
- (c) Consider the forced-dissipative case, with forcing confined to a small range of wavenumbers. Argue that we may expect forward and inverse cascades of appropriate inviscid invariants, and obtain expressions for the associated spectra.
- (d) In the unforced but viscous case, assume that energy is conserved and that the energy spectrum evolves in a self-similar way. Show that the energy spectrum then obeys

$$\mathcal{E}(k) = k_d^{-3/4} E^{9/8} t^{1/4} F(k k_d^{-3/4} E^{1/8} t^{1/4}), \quad (\text{P8.3})$$

where F is a function of universal form and E is the total energy.

[See Watanabe *et al.* (1998) and Smith *et al.* (2002).]

8.4 Assuming that a small error cascades upscale.

- (a) Using the observed spectrum of the atmosphere as a guide, estimate the predictability time of the atmosphere.
- (b) Suppose that the large-scales of the atmosphere behave two dimensionally and the small scales three dimensionally. Estimate the predictability time as a function of the cross-over scale, and plot this.
- (c) Estimate the error doubling time as a function of the scale of the error in the atmosphere, and plot it.

8.5 ♦ Consider predictability in an isotropic fluid that has a -4 energy spectrum, and a largest scale at wavenumber k_0 .

- (a) Suppose that a small error projects entirely on to a small scale characterized by a wavenumber k_1 ($k_1 \gg k_0$), such that scales $1/k_1$ and smaller are effectively unobserved, whereas scales larger than this are perfectly observed. Supposing error cascades via spectrally local interactions to large scales, estimate the predictability time.
- (b) Now suppose that the initial error projects almost entirely on to the largest scale, so that it is a small perturbation on the scale k_0 , but that its amplitude (the r.m.s. difference between the perturbed and unperturbed streamfunction) is the same as in the first case. Estimate the predictability time by supposing that the error growth is exponential, and compare it to the solution of the first part.

8.6 In the limit of zero viscosity (8.72) implies that enstrophy dissipation remains constant, and therefore that pininstrophy (mean-square curl of the vorticity) is infinite somewhere. However, it has been shown rigorously that the two-dimensional Euler equations have *no* singularities in finite time if the initial conditions are sufficiently smooth, and also that their enstrophy dissipation is zero. Resolve this paradox.

Heuristic solution. We are concerned with the zero viscosity *limit* in (8.72). But in this limit it takes the fluid an infinite time to come to equilibrium with an infinitely long inertial range, and only then is the enstrophy dissipation non-zero. However, the rigorous results only prohibit singularities forming in finite time, and so there is no contradiction.

Oh brave new world, That has such people in't!

William Shakespeare, *The Tempest*, c. 1611.

CHAPTER NINE

Geostrophic Turbulence and Baroclinic Eddies

GEOSTROPHIC TURBULENCE is turbulence in flows that are stably stratified and in near-geostrophic balance. Like any problem in turbulence it is difficult, and a real ‘solution’ — meaning an accurate, informative statement about average states, without computation of the detailed evolution — may be out of our reach, and may not exist. Nevertheless, and ironically, it is sometimes easier to say something interesting about geostrophic turbulence than about incompressible isotropic two- or three-dimensional turbulence. In the latter class of problems there is nothing else to understand other than the problem of turbulence itself; on the other hand, rotation and stratification give one something else to grasp, a nettle though it may be, and it becomes possible to address geophysically interesting phenomena without having to solve the whole turbulence problem. Furthermore, in inhomogeneous geostrophic turbulence, asking questions about the *mean fields* is meaningful and useful, whereas this is trivial in isotropic turbulence.

Geostrophic turbulence is not restricted to quasi-geostrophic flow; indeed, the large scale turbulence of the Earth’s ocean and atmosphere is sometimes simply called ‘macro-turbulence’. Nevertheless, the quasi-geostrophic equations retain advective nonlinearity in the vorticity equation, and they capture the constraining effects of rotation and stratification that are so important in geophysical flows in a simple and direct way; for these reasons the quasi-geostrophic equations will be our main tool. Let us consider the effects of rotation first, then stratification.

9.1 EFFECTS OF DIFFERENTIAL ROTATION IN TWO-DIMENSIONAL TURBULENCE

In the limit of motion of a scale much shorter than the deformation radius, and with no topography, the quasi-geostrophic potential vorticity equation, (5.118), reduces to the

two-dimensional equation,

$$\frac{Dq}{Dt} = 0, \quad (9.1)$$

where $q = \zeta + f$. This is the perhaps the simplest equation with which to study the effects of rotation on turbulence. Suppose first that the Coriolis parameter is constant, so that $f = f_0$. Then (9.1) becomes simply the two-dimensional vorticity equation

$$\frac{D\zeta}{Dt} = 0. \quad (9.2)$$

Thus constant rotation has *no* effect on purely two-dimensional motion. Flow that is already two-dimensional — flow on a soap film, for example — is unaffected by rotation. (In the ocean and atmosphere, or in a rotating tank, it is of course the effects of rotation that lead to the flow being quasi-two dimensional in the first instance.)

Suppose, though, that the Coriolis parameter is variable, as in $f = f_0 + \beta y$. Then we have

$$\frac{D}{Dt}(\zeta + \beta y) = 0 \quad \text{or} \quad \frac{D\zeta}{Dt} + \beta v = 0. \quad (9.3a,b)$$

If the asymptotically dominant term in these equations is the one involving β , then we obtain $v = 0$. Put another way, if β is very large, then the meridional flow v must be correspondingly small to ensure that βv is bounded. Any flow must then be predominantly *zonal*. This constraint may be interpreted as a consequence of angular momentum and energy conservation as follows. A ring of fluid encircling the Earth at a velocity u has an angular momentum per unit mass $a \cos \theta (u + \Omega a \cos \theta)$, where θ is the latitude and a is the radius of the Earth. Moving this ring of air polewards (i.e., giving it a meridional velocity) while conserving its angular momentum requires that its zonal velocity and hence energy must increase. Unless there is a source for that energy the flow is constrained to remain zonal.

9.1.1 The wave-turbulence cross-over

Scaling

Let us now consider how turbulent flow might interact with Rossby waves. We write (9.1) in full as

$$\frac{\partial \zeta}{\partial t} + \mathbf{u} \cdot \nabla \zeta + \beta v = 0. \quad (9.4)$$

If $\zeta \sim U/L$ and if $t \sim T$ then the respective terms in this equation scale as

$$\frac{U}{LT} \quad \frac{U^2}{L^2} \quad \beta U. \quad (9.5)$$

How time scales (i.e., advectively or with a Rossby wave frequency scaling) is determined by which of the other two terms dominates, and this in turn is scale dependent. For large scales the β -term will be dominant, and at smaller scales the advective term is dominant. The cross-over scale, denoted L_R , is called the *Rhines scale* and is given by¹

$$L_R \sim \sqrt{\frac{U}{\beta}}.$$

(9.6)

The U in (9.6) should be interpreted as the root-mean-square velocity at the energy containing scales, not a mean or translational velocity. (We will refer to the specific scale $\sqrt{U/\beta}$ as the Rhines scale, and more general scales involving a balance between nonlinearity and β , discussed below, as the β -scale, and denote them L_β .)

This is not a unique way to arrive at a β -scale, since we have chosen the length scale that connects vorticity to velocity to also be the β -scale, and it is by no means clear a priori that this should be so. If the two scales are different, the three terms in (9.4) scale as

$$\frac{\mathcal{Z}}{T} : \frac{U\mathcal{Z}}{L} : \beta U, \quad (9.7)$$

respectively, where \mathcal{Z} is the scaling for vorticity [i.e., $\zeta = \mathcal{O}(\mathcal{Z})$]. Equating the second and third terms gives the scale

$$L_{\beta\mathcal{Z}} = \frac{\mathcal{Z}}{\beta}. \quad (9.8)$$

Nevertheless, (9.6) and (9.8) both indicate that at some *large* scale Rossby waves are likely to dominate whereas at small scales advection, and turbulence, dominates.

Another heuristic way to derive (9.6) is by a direct consideration of time scales. Ignoring anisotropy, Rossby wave frequency is β/k and an inverse advective time scale is Uk , where k is the wavenumber. Equating these two gives an equation for the Rhines wavenumber

$$k_R \sim \sqrt{\frac{\beta}{U}}. \quad (9.9)$$

This equation is the inverse of (9.6), but note that factors of order unity cannot be revealed by simple scaling arguments such as these. The cross-over between waves and turbulence is reasonably sharp, as indicated in Fig. 9.1.

Turbulent phenomenology

Let us now try to go beyond elementary scaling arguments and examine wave-turbulence cross-over using the phenomenology of two-dimensional turbulence. We will suppose that the fluid is stirred at some well-defined scale k_f , producing an energy input ε . Then (assuming no energy is lost to smaller scales) energy cascades to large scales at that same rate. At some scale, the β -term in the vorticity equation will start to make its presence felt. By analogy with the procedure for finding the viscous dissipation scale in turbulence, we can find the scale at which linear Rossby waves dominate by equating the inverse of the turbulent eddy turnover time to the Rossby wave frequency. The eddy-turnover time is

$$\tau_k = \varepsilon^{-1/3} k^{-2/3}, \quad (9.10)$$

and equating this to the inverse Rossby wave frequency k/β gives the β -wavenumber and its inverse, the β -scale:

$$k_\beta \sim \left(\frac{\beta^3}{\varepsilon} \right)^{1/5}, \quad L_\beta \sim \left(\frac{\varepsilon}{\beta^3} \right)^{1/5}.$$

(9.11a,b)

In a real fluid these expressions are harder to evaluate than (9.9), since it is generally much easier to measure velocities than energy transfer rates, or even vorticity. On the other hand, (9.11) is perhaps more satisfactory from the point of view of turbulence, and ε

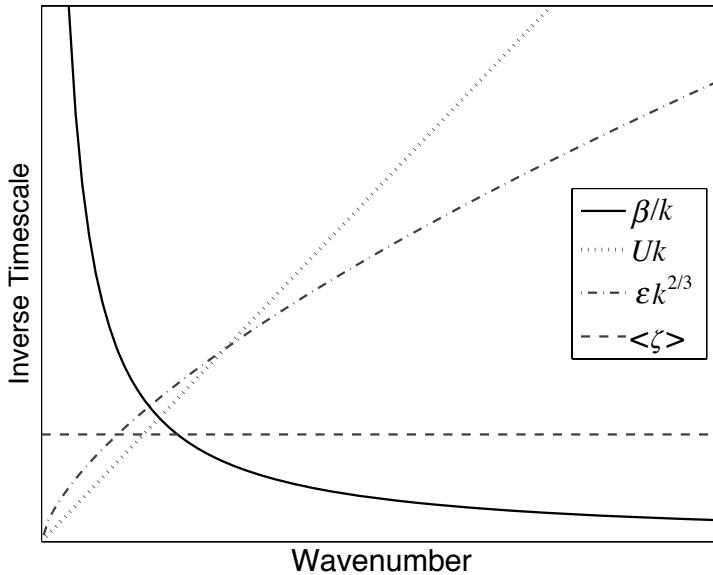


Fig. 9.1 Three estimates of the wave-turbulence cross-over, in wavenumber space. The solid curve is the frequency of Rossby waves, proportional to β/k . The other three curves are various estimates of the inverse turbulence time scale, or ‘turbulence frequency’. These are the turbulent eddy transfer rate, proportional to $\epsilon^{1/3}k^{2/3}$ in a $k^{-5/3}$ spectrum; the simple estimate Uk where U is a root-mean-squared velocity; and the mean vorticity, which is constant. Where the Rossby wave frequency is larger (smaller) than the turbulent frequency, i.e., at large (small) scales, Rossby waves (turbulence) dominate the dynamics.

may be determined by processes largely independent of β , whereas the magnitude of the eddies at the energy containing scales is likely to be a function of β . We also note that the scale given by (9.11b) is not necessarily the energy-containing scale, and may in principle differ considerably from the scale given by (9.9). This is because the inverse cascade is not necessarily *halted* at the scale (9.11b) — this is just the scale at which Rossby waves become important. Energy may continue to cascade to larger scales, albeit anisotropically as discussed below, and so the energy containing scale may be larger.

9.1.2 Generation of zonal flows and jets

None of the effects discussed so far takes into account the anisotropy inherent in Rossby waves, and such anisotropy can give rise to predominantly zonal flows and jets. To understand this, let us first note that energy transfer will be relatively inefficient at those scales where linear Rossby waves dominate the dynamics. But the wave-turbulence boundary is not isotropic; the Rossby wave frequency is quite anisotropic, being given by

$$\omega = -\frac{\beta k^x}{k^{x^2} + k^{y^2}}. \quad (9.12)$$

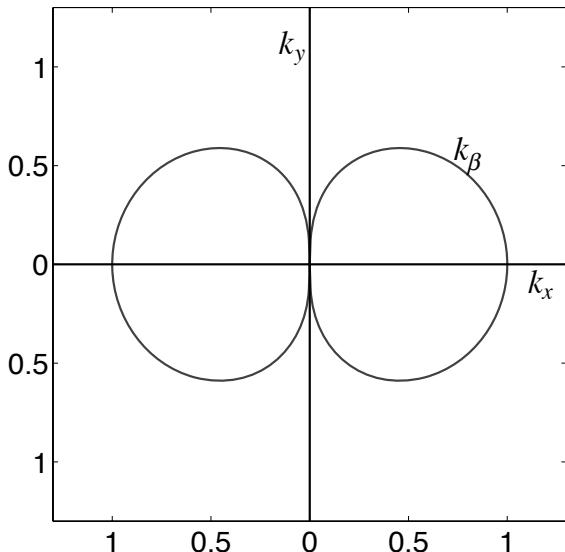


Fig. 9.2 The anisotropic wave-turbulence boundary k_β , in wave-vector space calculated by equating the turbulent eddy transfer rate, proportional to $k^{2/3}$ in a $k^{-5/3}$ spectrum, to the Rossby wave frequency $\beta k^x/k^2$, as in (9.14). Within the dumbbell Rossby waves dominate and energy transfer is inhibited. The inverse cascade plus Rossby waves thus leads to a generation of zonal flow. [The wavenumbers are scaled such that $(\beta^3/\varepsilon) = 1$.]

If, albeit a little crudely, we suppose that the turbulent part of the flow remains isotropic, the wave-turbulence boundary is then given by equating the inverse of (9.10) with (9.12); that is solution of

$$\varepsilon^{1/3} k^{2/3} = \frac{\beta k^x}{k^2}, \quad (9.13)$$

where k is the isotropic wavenumber. Solving this gives expressions for the x - and y -wavenumber components of the wave-turbulence boundary, namely

$$k_\beta^x = \left(\frac{\beta^3}{\varepsilon} \right)^{1/5} \cos^{8/5} \theta, \quad k_\beta^y = \left(\frac{\beta^3}{\varepsilon} \right)^{1/5} \sin \theta \cos^{3/5} \theta, \quad (9.14)$$

where the polar coordinate is parameterized by the angle $\theta = \tan^{-1}(k^y/k^x)$. This rather uninformative-looking formula is illustrated in Fig. 9.2. (Slight variations on this theme are produced by using different expressions for the turbulence time scale; see problem 9.1.)

What occurs physically? The region inside the dumbbell shapes in Fig. 9.2 is dominated by Rossby waves, where the natural frequency of the oscillation is *higher* than the turbulent frequency. If the flow is stirred at a wavenumber higher than this the energy will cascade to larger scales, but because of the frequency mismatch the turbulent flow will be unable to efficiently excite modes within the dumbbell. Nevertheless, there is still a natural tendency of the energy to seek the gravest mode, and it will do this by cascading toward the $k^x = 0$ axis; that is, toward zonal flow. Thus, the combination of Rossby waves and turbulence will lead to the formation of zonal flow and, potentially, zonal jets.³

Figure 9.3 illustrates this mechanism; it shows the freely evolving (unforced, inviscid) energy spectrum in a simulation on a β -plane, with an initially isotropic spectrum. The energy implodes, cascading to larger scales but avoiding the region inside the dumbbell and piling up at $k^x = 0$. In physical space this mechanism manifests itself by the formation of zonally elongated structures and jets, in both freely-decaying and forced-dissipative simulations (Fig. 9.4 and Fig. 9.5).

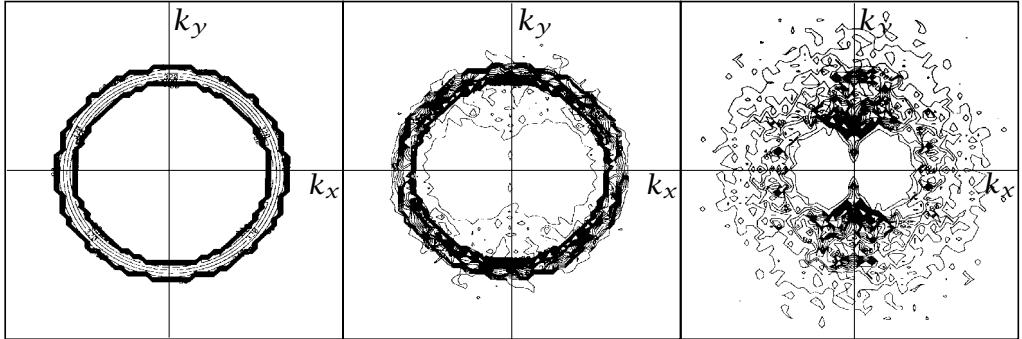


Fig. 9.3 Evolution of the energy spectrum in a freely evolving two-dimensional simulation on the β -plane. The panels show contours of energy in wavenumber (k_x, k_y) space at successive times. The initial spectrum is isotropic. The energy ‘implodes’, but its passage to large scales is impeded by the β -effect, and second and third panels show the spectrum at later times, illustrating the dumbbell predicted by (9.14) and Fig. 9.2.²

9.1.3 † Joint effect of β and friction

The β term does not remove energy from a fluid. Thus, if energy is being added to a fluid at some small scales, and the energy is cascading to larger scales, then the β -effect does not of itself halt the inverse cascade, it merely deflects the cascade such that the flow becomes more zonal. Suppose that the fluid obeys the barotropic vorticity equation,

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) + \beta \frac{\partial \psi}{\partial x} = F - r\zeta + \nu \nabla^2 \zeta, \quad (9.15)$$

where the viscosity, ν , is small and acts only to remove enstrophy, and not energy, at very small scales. The forcing, F , supplies energy at a rate ε and this is cascaded upscale and removed by the linear drag term $-r\zeta$, where the drag coefficient r is a constant. If the friction is sufficiently large, then the energy is removed before it feels the effect of β . A scaling of (9.15) suggests that the relative importance of the β -effect relative to friction is parameterized by the non-dimensional number $\beta L/r$, where L is the length scale of the energy containing modes. This length scale is not known a priori, and a more predictive measure of the importance of friction and β is obtained by comparing the frictional wavenumber (8.74), and the β -wavenumber, (9.11). The ratio of these scales, γ , is given by

$$\gamma = \left(\frac{\beta \varepsilon^{1/2}}{r^{5/2}} \right)^{3/5}. \quad (9.16)$$

When γ is large the β -effect will be felt, but when γ is small frictional effects will dominate at large scales.

Even if γ is large, then frictional effects must still be important somewhere, in order that energy may be removed. Forming an energy equation from (9.15) by multiplying by $-\psi$ and spatially integrating (and neglecting viscosity) we find the energy balance

$$\varepsilon = -\frac{1}{A} \int_A \psi F \, dA = \frac{r}{A} \int_A (\nabla \psi)^2 \, dA = 2r\bar{E}, \quad (9.17)$$

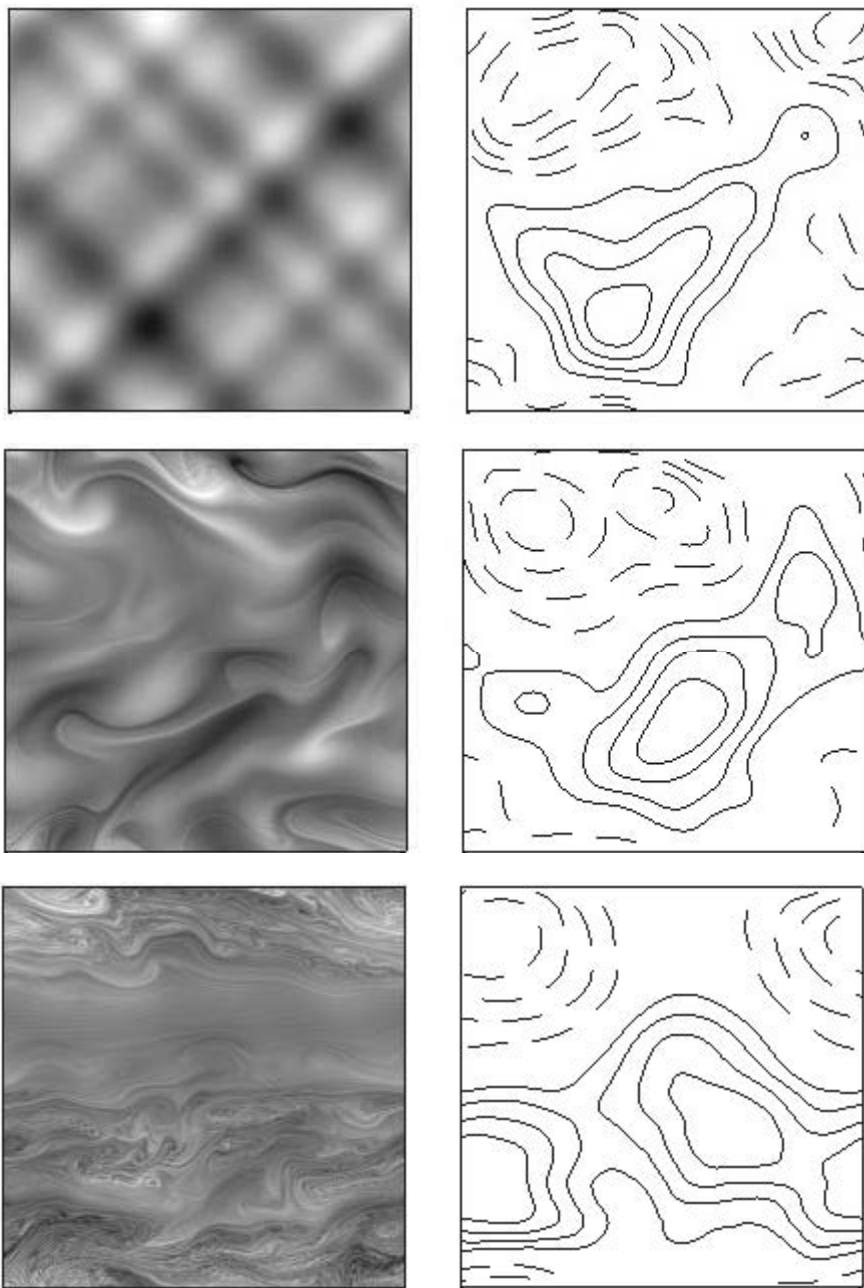


Fig. 9.4 Evolution of vorticity (greyscale, left column) and streamfunction (contour plots, right column) in a doubly periodic square domain (of length 2π) at times $t = 0, 50$ and 260 (in units of inverse vorticity), obeying (9.4) with the addition of a weak viscous term on the right-hand side. The initial conditions are the same as for Fig. 8.8, with a maximum value of vorticity of about 3. As $\beta = 3$, the β -Rossby number, $|\zeta|/\beta L$ is about unity. Compared to Fig. 8.8, vortex formation is inhibited and there is a tendency toward zonal flow.

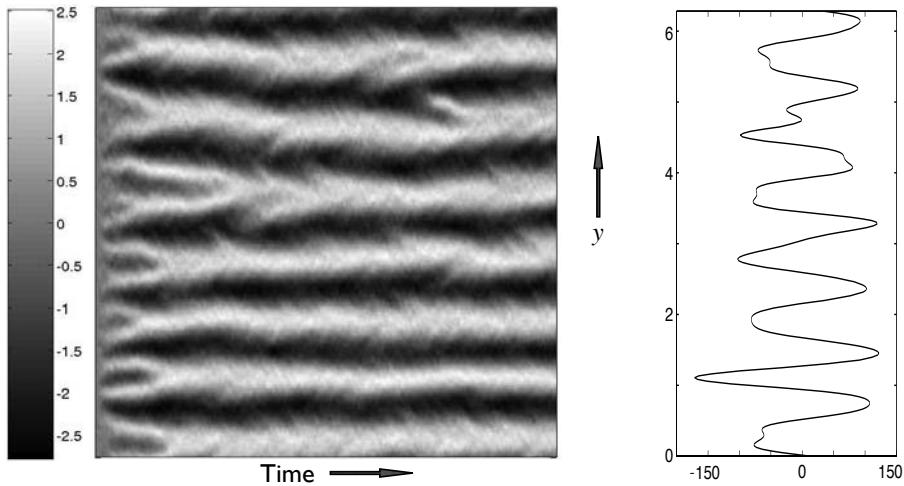


Fig. 9.5 Left: Grey-scale image of zonally averaged zonal velocity (\bar{u}) as a function of time and latitude (Y), produced in a simulation forced around wavenumber 80 and with $k_\beta = \sqrt{\beta/U} \approx 10$ (in a square domain of side 2π). Right: Values of $\partial^2 \bar{u} / \partial y^2$ as a function of latitude, late in the integration. Jets form very quickly from the random initial conditions, and are subsequently quite steady.⁴

where \bar{E} is the average energy of the fluid per unit mass. Using (9.6) with $U = \sqrt{2\bar{E}}$, where \bar{E} is obtained from (9.17), we obtain the Rhines scale

$$L_R = \left(\frac{\varepsilon}{r\beta^2} \right)^{1/4}. \quad (9.18)$$

What do the two scales, (9.11b) and (9.18), represent? The first is the scale at which the β -effect is first felt by the inverse cascade, and this parameterizes the overall size of the dumbbell of Fig. 9.2. It is this scale that is most relevant for large-scale meridional mixing, and so for the meridional heat transport in the atmosphere. But if the inverse cascade continues past this scale, especially in the zonal direction, then (9.18) may characterize the largest scale reached by the inverse cascade, and so the meridional scale of the jets.⁵

9.2 STRATIFIED GEOSTROPHIC TURBULENCE

9.2.1 An analogue to two-dimensional flow

Now let us consider stratified effects in a simple setting, using the quasi-geostrophic equations with constant Coriolis parameter and constant stratification.⁶ The (dimensional) unforced and inviscid governing equation may then be written as

$$\frac{Dq}{Dt} = 0, \quad q = \nabla^2 \psi + Pr^2 \frac{\partial^2 \psi}{\partial z^2}, \quad (9.19a)$$

where $Pr = f_0/N$ is the *Prandtl ratio* (and Pr/H is the inverse of the deformation radius) and $D/Dt = \partial/\partial t + \mathbf{u} \cdot \nabla$ is the two-dimensional material derivative. The vertical boundary

conditions are

$$\frac{D}{Dt} \left(\frac{\partial \psi}{\partial z} \right) = 0, \quad \text{at } z = 0, H. \quad (9.19b)$$

These equations are analogous to the equations of motion for purely two-dimensional flow. In particular, with periodic lateral boundary conditions, or conditions of no-normal flow, there are two quadratic invariants of the motion, the energy and the enstrophy, which are obtained by multiplying (9.19a) by $-\psi$ and q and integrating over the domain, as in chapter 5. The conserved energy is

$$\frac{d\hat{E}}{dt} = 0, \quad \hat{E} = \frac{1}{2} \int_V \left[(\nabla \psi)^2 + Pr^2 \left(\frac{\partial \psi}{\partial z} \right)^2 \right] dV, \quad (9.20)$$

where the integral is over a *three-dimensional* domain. The enstrophy is conserved at each vertical level, and of course the volume integral is also conserved, namely

$$\frac{d\hat{Z}}{dt} = 0, \quad \hat{Z} = \frac{1}{2} \int_V q^2 dV = \frac{1}{2} \int_V \left[\nabla^2 \psi + Pr^2 \left(\frac{\partial^2 \psi}{\partial z^2} \right) \right]^2 dV. \quad (9.21)$$

The analogy with two-dimensional flow is even more transparent if we further rescale the vertical coordinate by $1/Pr$, and so let $z' = z/Pr$. Then the energy and enstrophy invariants are:

$$\hat{E} = \int (\nabla_3 \psi)^2 dV, \quad \hat{Z} = \int q^2 dV = \int (\nabla_3^2 \psi)^2 dV, \quad (9.22)$$

where $\nabla_3 = \mathbf{i} \partial/\partial x + \mathbf{j} \partial/\partial y + \mathbf{k} \partial/\partial z'$. The invariants then have almost the same form as the two-dimensional invariants, but with a three-dimensional Laplacian operator instead of a two-dimensional one.

Given these invariants, we should expect that any dynamical behaviour that occurs in the two-dimensional equations *that depends solely on the energy/enstrophy constraints* should have an analogue in quasi-geostrophic flow. In particular, the transfer of energy to large-scales and enstrophy to small scales will also occur in quasi-geostrophic flow with, in so far as these transfers are effected by a local cascade, corresponding spectra of $k^{-5/3}$ and k^{-3} . However, in the quasi-geostrophic case, it is the *three-dimensional* wavenumber that is relevant, with the vertical component scaled by the Prandtl ratio. As a consequence, the energy cascade to larger horizontal scales is generally accompanied by a cascade to larger vertical scales — a *barotropization* of the flow, and this is an important and robust process in geostrophic turbulence. Still, the analogy between two-dimensional and quasi-geostrophic cascades should not be taken too far, because in the latter the potential vorticity is advected only by the horizontal flow. Thus, the dynamics of quasi-geostrophic turbulence will *not* in general be isotropic in three-dimensional wavenumber. To examine these dynamics more fully we first turn to a simpler model, that of two-layer flow.

9.2.2 Two-layer geostrophic turbulence

Let us now consider flow in two layers of equal depth, governed by the quasi-geostrophic equations with (for now) $\beta = 0$, namely

$$\frac{\partial q_i}{\partial t} + J(\psi_i, q_i) = 0, \quad i = 1, 2, \quad (9.23)$$

where

$$q_1 = \nabla^2 \psi_1 + \frac{1}{2} k_d^2 (\psi_2 - \psi_1), \quad q_2 = \nabla^2 \psi_2 + \frac{1}{2} k_d^2 (\psi_1 - \psi_2), \quad (9.24a)$$

$$J(a, b) = \frac{\partial a}{\partial x} \frac{\partial b}{\partial y} - \frac{\partial a}{\partial y} \frac{\partial b}{\partial x}, \quad \frac{1}{2} k_d^2 = \frac{2f_0^2}{g'H} \equiv \frac{4f_0^2}{N^2 H^2}. \quad (9.24b)$$

The wavenumber k_d is inversely proportional to the baroclinic radius of deformation, and the two equivalent expressions given are appropriate in a layered model and a level model, respectively. The equations conserve the total energy,

$$\frac{d\hat{E}}{dt} = 0, \quad \hat{E} = \frac{1}{2} \int_A \left[(\nabla \psi_1)^2 + (\nabla \psi_2)^2 + \frac{1}{2} k_d^2 (\psi_1 - \psi_2)^2 \right] dA, \quad (9.25)$$

and the enstrophy in each layer

$$\frac{d\hat{Z}_1}{dt} = 0, \quad \hat{Z}_1 = \int_A q_1^2 dA, \quad (9.26a)$$

$$\frac{d\hat{Z}_2}{dt} = 0, \quad \hat{Z}_2 = \int_A q_2^2 dA. \quad (9.26b)$$

The first two terms in the energy expression, (9.25), represent the kinetic energy, and the last term is the available potential energy, proportional to the variance of temperature.

Baroclinic and barotropic decomposition

Define the barotropic and baroclinic streamfunctions by

$$\psi \equiv \frac{1}{2} (\psi_1 + \psi_2), \quad \tau \equiv \frac{1}{2} (\psi_1 - \psi_2). \quad (9.27)$$

Then the potential vorticities for each layer may be written as

$$q_1 = \nabla^2 \psi + (\nabla^2 - k_d^2) \tau, \quad q_2 = \nabla^2 \psi - (\nabla^2 - k_d^2) \tau \quad (9.28a,b)$$

and the equations of motion may be rewritten as evolution equations for ψ and τ as follows:

$$\frac{\partial}{\partial t} \nabla^2 \psi + J(\psi, \nabla^2 \psi) + J(\tau, (\nabla^2 - k_d^2) \tau) = 0 \quad (9.29a)$$

$$\frac{\partial}{\partial t} (\nabla^2 - k_d^2) \tau + J(\tau, \nabla^2 \psi) + J(\psi, (\nabla^2 - k_d^2) \tau) = 0. \quad (9.29b)$$

We note the following about the nonlinear interactions in (9.29).

- (i) ψ and τ are like vertical modes. That is, ψ is the barotropic mode with a ‘vertical wavenumber’, k^z , of zero, and τ is a baroclinic mode with an effective vertical wavenumber of one.
- (ii) Just as purely two-dimensional turbulence can be considered to be a plethora of interacting triads, whose two-dimensional vector wavenumbers sum to zero, it is clear from (9.29b) that geostrophic turbulence may be considered to be similarly composed of a sum of interacting triads, although now there are two distinct kinds, namely

$$(i) \quad (\psi, \psi) \rightarrow \psi, \quad (ii) \quad (\tau, \tau) \rightarrow \psi \quad \text{or} \quad (\psi, \tau) \rightarrow \tau. \quad (9.30)$$

The first kind is a *barotropic triad*, for it involves only the barotropic mode. The other two are examples of a *baroclinic triad*. If a barotropic mode has a vertical wavenumber of zero, and a baroclinic mode has a vertical wavenumber of plus or minus one, then the three vertical wavenumbers of the triad interactions must sum to zero. There is no triad that involves only the baroclinic mode, as we may see from the form of (9.29). (If the layers are of unequal depths, then purely baroclinic triads do exist.)

- (iii) Wherever the Laplacian operator acts on τ , it is accompanied by $-k_d^2$. That is, it is *as if* the effective horizontal wavenumber (squared) of τ is shifted, so that $k^2 \rightarrow k^2 + k_d^2$.

Conservation properties

Multiplying (9.29a) by ψ and (9.29b) by τ and horizontally integrating over the domain of area A , assuming once again that this is either periodic or has solid walls, gives

$$\hat{T} = \int_A (\nabla \psi)^2 dA, \quad \frac{d\hat{T}}{dt} = \int_A \psi J(\tau, (\nabla^2 - k_d^2)\tau) dA \quad (9.31a)$$

$$\hat{C} = \int_A [(\nabla \tau)^2 + k_d^2 \tau^2] dA, \quad \frac{d\hat{C}}{dt} = \int_A \tau J(\psi, (\nabla^2 - k_d^2)\tau) dA. \quad (9.31b)$$

Here, \hat{T} is the energy associated with the barotropic flow and \hat{C} is the energy of the baroclinic flow. An integration by parts shows that

$$\int_A \psi J(\tau, (\nabla^2 - k_d^2)\tau) dA = - \int_A \tau J(\psi, (\nabla^2 - k_d^2)\tau) dA, \quad (9.32)$$

and therefore

$$\frac{d\hat{E}}{dt} = \frac{d}{dt}(\hat{T} + \hat{C}) = 0. \quad (9.33)$$

That is, total energy is conserved.

An enstrophy invariant is obtained by multiplying (9.29a) by $\nabla^2 \psi$ and (9.29b) by $(\nabla^2 - k_d^2)\tau$ and integrating over the domain and adding the two expressions. The result is

$$\frac{d\hat{Z}}{dt} = 0, \quad \hat{Z} = \int_A (\nabla^2 \psi)^2 + [(\nabla^2 - k_d^2)\tau]^2 dA. \quad (9.34)$$

This also follows from (9.26).

Just as for two-dimensional turbulence, we may define the spectra of the energy and enstrophy. Then, with obvious notation, for the energy we have

$$\frac{1}{A} \hat{T} = \int \mathcal{T}(k) dk \quad \text{and} \quad \frac{1}{A} \hat{C} = \int C(k) dk, \quad (9.35)$$

The enstrophy spectrum $\mathcal{Z}(k)$ is related to the energy spectra by

$$\frac{1}{A} \hat{Z} = \int \mathcal{Z}(k) dk = \int [k^2 \mathcal{T}(k) + (k^2 + k_d^2) C(k)] dk. \quad (9.36)$$

which is analogous to the relationship between energy and enstrophy in two-dimensional flow. We thus begin to suspect that the phenomenology of two-layer turbulence is closely related to, but richer than, that of two-dimensional turbulence.

9.2.3 Phenomenology of two-layer turbulence

In this section we explore the phenomenology of two-layer geostrophic turbulence, by examining how the equations of motion simplify if we suppose that the nonlinear interactions occur at particular scales, and by looking at the forms of the triad interactions.

Triad interactions and scales of interactions

Two types of triad interactions are, in general, possible.

Barotropic triads. An interaction that is purely barotropic (i.e., as if $\tau = 0$) conserves \hat{T} , the barotropic energy, and the associated enstrophy $\int k^2 \mathcal{T}(k) dk$, and a barotropic triad behaves as purely two-dimensional flow. Explicitly, the conserved quantities are

$$\text{energy:} \quad \frac{d}{dt} [\mathcal{T}(k) + \mathcal{T}(p) + \mathcal{T}(q)] = 0, \quad (9.37a)$$

$$\text{enstrophy:} \quad \frac{d}{dt} [k^2 \mathcal{T}(k) + p^2 \mathcal{T}(p) + q^2 \mathcal{T}(q)] = 0. \quad (9.37b)$$

Baroclinic triads. Baroclinic triads involve two baroclinic wavenumbers (say p, q) interacting with a barotropic wavenumber (say k). The energy and enstrophy conservation laws for this triad are

$$\text{energy:} \quad \frac{d}{dt} [\mathcal{T}(k) + C(p) + C(q)] = 0, \quad (9.38a)$$

$$\text{enstrophy:} \quad \frac{d}{dt} [k^2 \mathcal{T}(k) + (p^2 + k_d^2)C(p) + (q^2 + k_d^2)C(q)] = 0. \quad (9.38b)$$

The turbulent interactions may be divided into three general types:

- (i) interactions only involving scales much smaller than the deformation radius;
- (ii) interactions only involving scales much larger than the deformation radius;
- (iii) interactions involving scales comparable to the deformation radius.

We consider each in turn.

(i) Interactions at small scales

For small scales the potential vorticity of each layer is given by

$$q_i = \nabla^2 \psi_i + \frac{1}{2} k_d^2 (\psi_j - \psi_i) \approx \nabla^2 \psi_i \quad (9.39)$$

where $i = 1, 2$ and $j = 3 - i$. Thus, each layer is decoupled from the other and obeys the equations of purely two-dimensional turbulence. Enstrophy is cascaded to small scales and, were there to be an energy source at small scales, energy would be transferred upscale until it reached a scale comparable to the deformation scale. As regards triad interactions, interactions at small scales correspond to the case $(p, q) \gg k_d$ in (9.38). If we neglect k_d^2 then we see that a baroclinic triad behaves like a barotropic triad, for (9.38) has the same form as (9.37). In reality, the small scales of a continuously stratified flow may not be representable by a two-layer model, because in a continuously stratified quasi-geostrophic model the enstrophy cascade occurs in *three-dimensional* wavenumber space. Thus, as the horizontal scales become smaller, so does the vertical scale and higher deformation radii [e.g., $n > 1$ in (5.208)] may play a role in reality.

(ii) Interactions at large scales

At scales much larger than the deformation radius we take $|\nabla^2| \sim k^2 \ll k_d^2$, and neglect terms in (9.29) that involve ∇^2 if they appear alongside terms involving k_d^2 . Noting that $J(\tau, k_d^2 \tau) = 0$, and assuming that $|J(\tau, \nabla^2 \psi)| \sim |J(\psi, \nabla^2 \tau)|$, we obtain

$$\frac{\partial}{\partial t} \nabla^2 \psi + J(\psi, \nabla^2 \psi) = -J(\tau, \nabla^2 \tau), \quad (9.40a)$$

$$\frac{\partial \tau}{\partial t} + J(\psi, \tau) = 0. \quad (9.40b)$$

Thus, the baroclinic streamfunction obeys the equation of a passive tracer, although because of the term on the right-hand side of (9.40a) it is not wholly passive. Nevertheless, and given the discussion of passive tracers in section 8.5, these equations suggest that the variance of the baroclinic streamfunction, and thus the energy in the baroclinic mode, will be transferred to smaller scales.

In terms of triad interactions, this case corresponds to $(p, q, k) \ll k_d$. The energy and enstrophy conservation triad interaction laws, (9.38), collapse to

$$\frac{d}{dt} [C(p) + C(q)] = 0. \quad (9.41)$$

That is to say, energy is conserved among the baroclinic modes alone, with the barotropic mode k mediating the interaction. Consistent with the analysis in physical space, there is no constraint preventing the transfer of baroclinic energy to smaller scales, and no production of barotropic energy at $k \ll k_d$.

(iii) Interactions at scales comparable to the deformation radius

In this case there is, in general, no simplification of the equations in physical space, or of the conservation laws governing triad interactions, and both baroclinic and barotropic modes are important. However, we can write the triad interaction conservation laws in an instructive form if we define the pseudowavenumber k' by $k'^2 \equiv k^2 + k_d^2$ for a baroclinic mode and $k'^2 \equiv k^2$ for a barotropic mode, and similarly for p' and q' . Then (9.37) and (9.38) can both be written as

$$\frac{d}{dt} [\mathcal{E}(k) + \mathcal{E}(p) + \mathcal{E}(q)] = 0, \quad (9.42a)$$

$$\frac{d}{dt} [k'^2 \mathcal{E}(k) + p'^2 \mathcal{E}(p) + q'^2 \mathcal{E}(q)] = 0, \quad (9.42b)$$

where $\mathcal{E}(k)$ is the energy (barotropic or baroclinic) of the particular mode. These are formally identical with the conservation laws for purely two-dimensional flow and so we expect energy to seek the gravest mode (i.e., smallest pseudowavenumber). Since the gravest mode has $k_d = 0$ this implies a tendency of energy to be transferred into the barotropic mode — a *barotropization* of the flow.

Baroclinic instability in the classic two-layer problem concerns the instability of a flow with vertical but no horizontal shear. This is like a triad interaction for which $p \ll (k, q, k_d)$, where p is the wavenumber of a baroclinic mode, and thus $k^2 \sim q^2$. The triad conservation

laws then become

$$\begin{aligned}\frac{d}{dt} [\mathcal{T}(k) + C(p) + C(q)] &= 0, \\ \frac{d}{dt} [k^2 \mathcal{T}(k) + k_d^2 C(p) + (q^2 + k_d^2) C(q)] &= 0.\end{aligned}\tag{9.43}$$

From these two equations, and with $k^2 \approx q^2$, we derive

$$q^2 \dot{C}(q) = (k_d^2 - k^2) \dot{\mathcal{T}}(k).\tag{9.44}$$

Baroclinic instability requires that both $\dot{C}(q)$ and $\dot{\mathcal{T}}(k)$ be positive. This can only occur if

$$k^2 < k_d^2.\tag{9.45}$$

Thus, there is a *high-wavenumber cut-off* for baroclinic instability, as we previously found in section 6.6 by direct calculation. We can see that the cut-off arises by virtue of energy and enstrophy conservation, and is not dependent on linearizing the equations and looking for exponentially growing normal-mode instabilities.

For small scales, we previously noted that each layer is decoupled from the other and that enstrophy, but not energy, may cascade to smaller scales. Now, baroclinic instability (of the large-scale flow) occurs at scales comparable to or larger than the deformation radius. Thus, the energy that is extracted from the mean flow is essentially trapped at scales larger than the deformation scale: it cannot be transferred to smaller scales in a cascade; rather, it is transferred to barotropic flow at scales comparable to the deformation radius, from which it cascades upwards to larger barotropic scales.

Summary of two-layer phenomenology

Putting together the considerations above leads to the following picture of geostrophic turbulence in a two-layer system (Fig. 9.6). At large horizontal scales we imagine some source of baroclinic energy, which in the atmosphere might be the differential heating between pole and equator, or in the ocean might be the wind and surface heat fluxes. Baroclinic instability effects a non-local transfer of energy to the deformation scale, where both baroclinic and barotropic modes are excited. From here there is an enstrophy cascade in each layer to smaller and smaller scales, until eventually the wavenumber is large enough (denoted by k_{3D} in Fig. 9.6) that non-geostrophic effects become important and enstrophy is scattered by three-dimensional effects, and dissipated. At scales larger than the deformation radius, there is an inverse barotropic cascade of energy to larger scales, where it is modified by the β -effect and halted by the effects of friction, with the latter being responsible for the dissipation of the large-scale barotropic energy.

There are two aspects of this phenomenology that prevent it from being a complete theory of baroclinic eddies in the atmosphere or ocean.

- (i) Even as a model of two-layer quasi-geostrophic turbulence the model is not ideal, because our assumptions are questionable. For example, we have largely neglected the effects of friction and we have assumed the baroclinic mode to be passive at large scales. Numerical simulations tend to support the model qualitatively, but they also show a significant transfer of energy from the baroclinic mode to the barotropic mode at scales larger than the deformation radius.

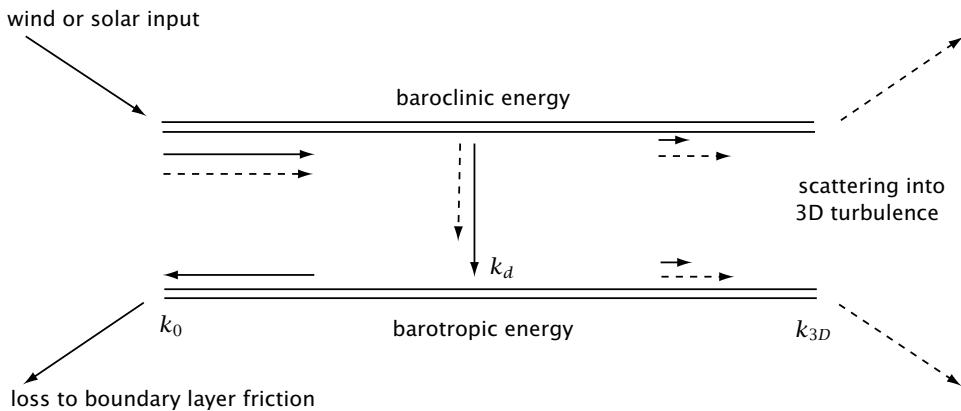


Fig. 9.6 Schema of idealized two-layer baroclinic turbulence.⁷ The horizontal axis represents horizontal wavenumber, and the vertical variation is decomposed into two vertical modes — the barotropic and first baroclinic. Energy transfer is shown by solid arrows and enstrophy transfer by dashed arrows. Large-scale forcing maintains the available potential energy, and so provides energy to the baroclinic mode at very large scales. Energy is transferred to smaller baroclinic scales, and then into barotropic energy at horizontal scales comparable to and larger than the deformation radius (this is baroclinic instability), and then transferred to larger barotropic scales in an upscale cascade. The entire process of baroclinic instability and energy transfer may be thought of as a generalized inverse cascade in which the energy passes to smaller pseudowavenumber $k'^2 \equiv k^2 + k_d^2$.

- (ii) The ideas do not perfectly apply to either the atmosphere or ocean. In the latter, the turbulence is quite inhomogeneous except perhaps in the Antarctic Circumpolar Current. In the atmosphere, observations indicate that the deformation radius is almost as large as the Rhines scale, and is only a factor of a few smaller than the pole-equator scale, leaving little room for isotropic inverse cascade to fully develop. On the other hand, the atmosphere does display k^{-3} spectra at scales similar to and somewhat smaller than the deformation radius (Fig. 9.7), and analysis of this indicates that it may indeed be associated with a forward cascade of enstrophy.⁸

9.3 † A SCALING THEORY FOR GEOSTROPHIC TURBULENCE

Let us now build on the phenomenological model of the previous section and construct a quantitative theory of two-layer forced-dissipative geostrophic turbulence.¹⁰ Although, as we have noted, the underlying model is imperfect, some of its qualitative properties transcend its limitations and are quite revealing about the real system. We will consider a system in which the basic state is a purely zonal flow, with constant vertical shear and no horizontal variation; our goal is to predict the amplitude and the scale of the eddies that result from the baroclinic instability of this flow. The assumptions that the mean flow and the stratification are constants are quite severe and thus the model, even if correct within its own terms of reference, is certainly not a complete theory of the baroclinic eddies in mid-latitude flow.

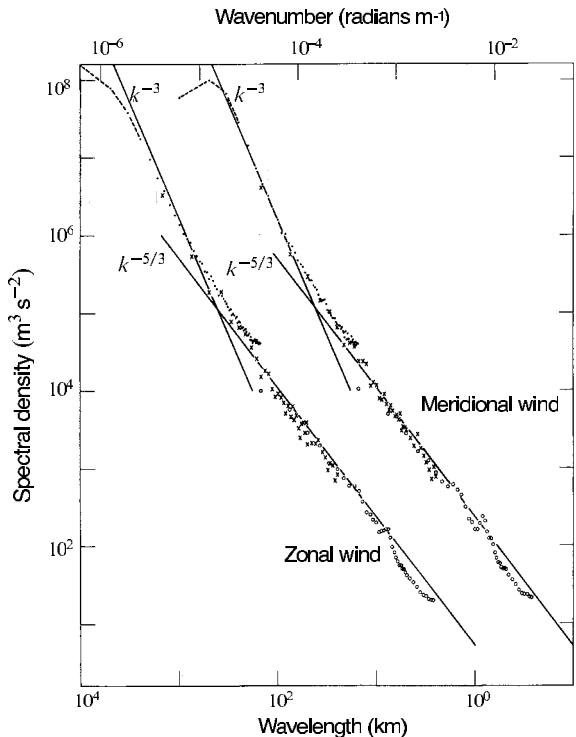


Fig. 9.7 Energy spectra of the zonal and meridional wind near the tropopause, from thousands of commercial aircraft measurements between 1975 and 1979. The meridional spectrum is shifted one decade to the right. The -3 spectrum may well be associated with a forward enstrophy cascade, but the origin of the $-5/3$ spectrum at smaller scales is not definitely known.⁹

9.3.1 Preliminaries

We shall suppose that, as in Fig. 9.6, baroclinic instability at large scales leads to a transfer of energy to the barotropic mode at a scale comparable to the deformation radius, followed by an inverse cascade of energy within the barotropic mode, and the energy is finally dissipated in the Ekman layer. At scales smaller than the deformation radius the layers are largely uncoupled, and in each layer there is an enstrophy cascade to small scales. The equations of motion describing all of this are (9.29), but we will explicitly recognize the effect of a shear flow by replacing τ by $\tau - Uy$, where U is constant, this being equivalent to supposing there is a constant shear in the flow. The equations of motion, (9.29), become

$$\frac{\partial}{\partial t} \nabla^2 \psi + J(\psi, \nabla^2 \psi) + J(\tau, (\nabla^2 - k_d^2) \tau) + U \frac{\partial}{\partial x} \nabla^2 \tau = D[\psi], \quad (9.46a)$$

$$\frac{\partial}{\partial t} (\nabla^2 - k_d^2) \tau + J(\tau, \nabla^2 \psi) + J(\psi, (\nabla^2 - k_d^2) \tau) + U \frac{\partial}{\partial x} (\nabla^2 \psi + k_d^2 \psi) = D[\tau]. \quad (9.46b)$$

These are similar to the equations used for studying two-layer baroclinic instability in chapter 6, but we now retain the nonlinear terms and include dissipation, represented by $D[\cdot]$.

For scales much larger than the deformation radius (9.46) become, analogously to (9.40),

$$\frac{\partial}{\partial t} \nabla^2 \psi + J(\psi, \nabla^2 \psi) = -J(\tau, \nabla^2 \tau) - U \frac{\partial}{\partial x} \nabla^2 \tau + D[\psi], \quad (9.47)$$

$$\frac{\partial \tau}{\partial t} + J(\psi, \tau) = U \frac{\partial \psi}{\partial x} - k_d^{-2} D[\tau]. \quad (9.48)$$

The equation for ψ is just the barotropic vorticity equation, ‘forced’ via its interaction with the baroclinic mode, namely the terms on the right-hand side of (9.47). The equation for the baroclinic streamfunction is the same as the equation for a passive scalar, except for the forcing term $U\partial\psi/\partial x$.

9.3.2 Scaling properties

In the barotropic equation, we may argue that in the energy-containing scales, $k^2 \ll k_d^2$, the magnitude of the barotropic streamfunction is much larger than that of the baroclinic streamfunction; that is, $|\psi| \gg |\tau|$, as follows. We may reasonably suppose that the forcing of the barotropic vorticity equation occurs at wavenumbers close to k_d , as in baroclinic instability. At larger scales the barotropic streamfunction obeys the two-dimensional vorticity equation, and we may expect an energy cascade to large scales with energy spectrum given by

$$\mathcal{E}_\psi(k) = \mathcal{K}_1 \varepsilon^{2/3} k^{-5/3}, \quad (9.49)$$

where \mathcal{K}_1 is a constant and ε is the as yet undetermined energy flux through the system. We may suppose that this cascade holds for wavenumbers $k_0 < k \ll k_d$, where the wavenumber k_0 is the halting scale of the inverse cascade, determined by one or more of: friction, the β -effect, and the domain size. Now, in this same wavenumber regime the baroclinic streamfunction is being advected as a passive tracer — it is being stirred by ψ . Thus, any baroclinic energy that is put in at large scales by the interaction with the mean flow (via the term proportional to $U\psi_x$) will be cascaded to smaller scales. Thus, we expect the baroclinic energy spectrum to be that of a passive tracer whose variance is cascading to smaller scales in a forward cascade, with a spectrum given by [cf. (8.98)]

$$\mathcal{E}_\tau(k) = \mathcal{K}_2 \varepsilon_\tau \varepsilon^{-1/3} k^{-5/3}, \quad (9.50)$$

where \mathcal{K}_2 is a constant, ε_τ is the transfer rate of baroclinic energy and ε is the same quantity appearing in (9.49). Now, because energy is not lost to small scales we have $\varepsilon_\tau = \varepsilon$, both being equal to the energy flux in the system. Thus, using (9.49) and (9.50), the energy in the barotropic and baroclinic modes are comparable at sufficiently large scales. Since the energy density in the former is $(\nabla\psi)^2$ and in the latter $(\nabla\tau)^2 + k_d^2\tau^2 \sim k_d^2\tau^2$, the magnitude of ψ must then be much larger than that of τ , and specifically

$$|\psi| \sim \frac{k_d |\tau|}{k_0} \gg |\tau|. \quad (9.51)$$

Let us write the baroclinic equation (9.48) in the form

$$\frac{\partial \tau}{\partial t} + J(\psi, \tau - U\gamma) = 0, \quad (9.52)$$

which is the equation for a passive tracer (τ) in a mean gradient (U), stirred by the flow (ψ), and we omit dissipation. Because there is a large scale separation (in fact, an infinite one) between the scale of the mean gradient of τ (i.e., the scale of variations of U) and the scale of its fluctuations, we can write

$$\tau \sim l' \frac{\partial \bar{\tau}}{\partial \gamma} = -l' U, \quad (9.53)$$

where l' is the scale of the fluctuation. Thus, at the scale k_0^{-1} , the magnitude of τ , and the associated baroclinic velocity v_τ , are given by

$$\tau \sim \frac{U}{k_0}, \quad v_\tau \sim U. \quad (9.54)$$

At this scale, and using (9.51), the magnitude of the barotropic streamfunction and its associated velocity are given by

$$\psi \sim \frac{k_d U}{k_0^2}, \quad v_\psi \sim \frac{k_d U}{k_0}. \quad (9.55a,b)$$

This is an important result, regardless of the other details in the calculation. In particular, (9.55b) implies that the barotropic velocity, v_ψ , scales like the mean velocity U multiplied by the ratio of the length scale of the eddies to the deformation radius. In the Earth's atmosphere, this ratio is an $\mathcal{O}(1)$ number; in the ocean it is somewhat larger.

How much energy flows through the system? The mean shear is the ultimate source of energy here, and for simplicity this shear is kept constant in time, analogous to an infinite heat bath supplying energy to a smaller system without its own temperature changing. The conversion of energy from the mean shear to the eddy flow is given by multiplying (9.48) by $k_d^2 \tau$ and integrating over the domain. This gives an expression for the rate of increase of the available potential energy of the system, namely

$$\frac{d}{dt} APE = \frac{d}{dt} \int_A k_d^2 \tau^2 dA = \int_A 2U k_d^2 \tau \frac{\partial \psi}{\partial x} dA. \quad (9.56)$$

(Note that the energy input to the system equals the poleward heat flux.) From this we estimate the average energy flux as

$$\varepsilon = 2U k_d^2 \overline{\psi_x \tau} \sim \frac{U^3 k_d^3}{k_0^2}. \quad (9.57)$$

The correlation between ψ_x and τ cannot be determined by this argument. This aside, we have produced a physically based closure for the flux of energy through the system in terms only of the mean shear, the halting scale k_0 (discussed below) and the deformation scale k_d .

Finally, we calculate the eddy diffusivity defined by

$$\kappa \equiv -\frac{\overline{v_\psi \tau}}{\partial_y \overline{\tau}}. \quad (9.58)$$

(Eddy diffusion is considered at greater length in chapter 10.) Using (9.54) and (9.55) gives

$$\kappa \sim \frac{k_d U}{k_0^2} \quad (9.59)$$

which, if the mixing velocity is the barotropic stirring velocity, implies a mixing length of k_0^{-1} . (Note also that the eddy diffusivity is just the magnitude of the barotropic streamfunction at the energy-containing scales.)

9.3.3 The halting scale and the β -effect

Let us suppose that, as discussed in section 9.1, the β -effect provides a barrier for the inverse cascade at the scale (9.11), so that $k_0 = k_\beta \sim (\beta^3/\varepsilon)^{1/5}$. Using (9.57) we find that the stopping scale is given by

$$k_0 = k_\beta \sim \frac{\beta}{U k_d}, \quad (9.60)$$

and using (9.59) and (9.60) we obtain for the energy flux and the eddy diffusivity,

$$\varepsilon \sim \frac{U^5 k_d^5}{\beta^2}, \quad \kappa \sim \frac{U^3 k_d^3}{\beta^2}. \quad (9.61)$$

The magnitudes of the eddies themselves are easily given using (9.55) and (9.54), whence

$$\tau \sim \frac{U^2 k_d}{\beta}, \quad v_\tau \sim U, \quad \psi \sim \frac{U^3 k_d^3}{\beta^2}, \quad v_\psi \sim \frac{U^2 k_d^2}{\beta}. \quad (9.62)$$

Evidently, in this model (in which the mean shear and deformation radius are fixed), the eddies become more energetic with decreasing β , and the eddy amplitudes and poleward heat flux increase very rapidly with the mean shear, more so than in a model in which the energy-containing scale is fixed. This is because as β decreases, the inverse cascade can extend to larger scales, thereby increasing the overall energy of the flow. Similarly, as U increases not only does the eddy amplitude increase as a direct consequence [as in (9.54) and (9.55)] but also k_β falls [see (9.60)], and these effects combine to give a rapid increase of the eddy magnitudes with U .

Frictional effects

Whether β is present or not, friction is necessary to ultimately remove the energy flowing through the system, as well as to remove enstrophy at small scales. Friction provides another mechanism for halting the inverse cascade, and the simplest case is that of a linear drag representing Ekman friction, in which case we write

$$D[\psi] = -r \nabla^2 \psi. \quad (9.63)$$

and the stopping wavenumber for a given ε is given, as in (8.74), by $k_r = (r^3/\varepsilon)^{1/2}$. However, the use of this in (9.57) fails to give a result for ε (problem 9.2). From a physical perspective, a linear drag that is weak enough to allow an inverse cascade to form is, *ipso facto*, too weak to equilibrate the flow. A friction that becomes larger at larger scales, for example an ‘inverse Laplacian’, has no such problems. More physically, a nonlinear drag, proportional to the square of the amplitude of the flow, also leads to a well-posed problem with the cascade halting at a well-defined scale. Finally, we should point out that in neither the atmosphere nor ocean is there an extended inverse cascade, because the deformation scale and the beta scale are not asymptotically well separated (although the β -effect does not prevent an inverse cascade to large zonal scales).

9.4 † PHENOMENOLOGY OF BAROCLINIC EDDIES IN THE ATMOSPHERE AND OCEAN

In the remaining sections of this chapter we take a more informal approach, illustrated by numerical experiments, to the problem of baroclinic eddies in the atmosphere and ocean.

We draw from our treatment of geostrophic turbulence but by being a little less formal we are able to travel farther, for we spend less time looking at the map (but with a concomitant danger that we lose our way).

9.4.1 The magnitude and scale of baroclinic eddies

How big, in both amplitude and scale, do baroclinic eddies become? Suppose that the time-mean flow is given, and that it is baroclinically unstable. Eddies will grow, initially according to the linear theory of chapter 6, but they cannot and do not continue to amplify: they ultimately equilibrate, and this by way of nonlinear mechanisms. The eddies will extract energy from the mean flow, but at the same time the available energy of the mean flow is being replenished by external forcing (i.e., the maintenance of an equator-pole temperature gradient by radiative forcing in the atmosphere, and wind and buoyancy forcing at the surface in the ocean). Thus, we cannot *a priori* determine the amplitude of baroclinic eddies by simply assuming that all of the available potential energy in the mean flow is converted to eddying motion. To close the problem we find we need to make three, not necessarily independent, assumptions:

- (i) an assumption about the magnitude of the baroclinic eddies;
- (ii) an assumption relating eddy kinetic energy to eddy available potential energy;
- (iii) an assumption about the horizontal scale of the eddies.

Baroclinic eddies extract available potential energy (APE) from the mean flow, and it is reasonable to suppose that an eddy of horizontal scale L_e can extract, as an upper bound, the APE of the mean flow contained within that scale. The APE is proportional to the variation of the buoyancy field so that

$$(\Delta b')^2 \sim |\Delta \bar{b}|^2 \sim L_e^2 |\nabla \bar{b}|^2, \quad (9.64)$$

where Δb is the variation in the buoyancy over the horizontal scale L_e . (For simplicity we stay with the Boussinesq equations, and $b = -g\delta\rho/\rho_0$. However, we could extend the arguments to an ideal-gas atmosphere with $b = g\delta\theta/\theta_0$.) Equivalently, we might simply write

$$b' \sim L_e |\nabla \bar{b}|, \quad (9.65)$$

which arises from a mixing-length approach. Supposing that the temperature gradient is mainly in the y -direction then, using thermal wind, we have

$$b' \sim L_e f \frac{\partial \bar{u}}{\partial z} \quad \text{and} \quad v'_\tau \sim \bar{u}, \quad (9.66a,b)$$

where v'_τ is an estimate of the shear (multiplied by the depth scale) of the eddying flow. [These estimates are the same as (9.54), with \bar{u} replacing U .]

Our second assumption is to relate the barotropic eddy kinetic energy to the eddy available potential energy, and the most straightforward one to make is that there is a rough equipartition between the two. This assumption is reasonable because in the baroclinic lifecycle (or baroclinic inverse cascade) energy is continuously transferred from eddy available potential energy to eddy kinetic energy, and the assumption is then equivalent to supposing that the relevant eddy magnitude is always proportional to this rate of transfer.

Thus we assume $v_\psi^2 \sim (b'/N)^2$ or

$$v'_\psi \sim \frac{b'}{N}. \quad (9.67)$$

Finally, the scale of the eddies is determined by the extent to which the eddies might grow through nonlinear interactions. As we discussed earlier, possibilities for this scale include the deformation radius itself (if the inverse cascade is weak) or the Rhines scale (if the inverse cascade is slowed by the β -effect), or even the domain scale if neither of these apply.

Some consequences

The simple manipulations above have some very interesting consequences. Using (9.66) and (9.67) we find

$$v'_\psi \sim \frac{f L_e}{N H} \bar{u} = \frac{L_e}{L_d} \bar{u} \quad (9.68)$$

where $L_d = NH/f_0$ is the deformation radius and \bar{u} is the amplitude of the mean baroclinic velocity, that is the mean shear multiplied by the height scale. This important relationship relates the magnitude of the eddy kinetic energy to that of the mean. In the atmosphere the scale of the motion is not much larger than the deformation radius (which is about 1000 km) and the eddy and mean kinetic energies are, consistently, comparable to each other. In the ocean the deformation radius (about 50 km over large areas) is significantly smaller than the scale of mesoscale eddies (which is more like 200 km or more), and observations consistently reveal that the eddy kinetic energy is an order of magnitude larger than the mean kinetic energy.¹¹

One other important and slightly counter-intuitive result concerns the timescale of eddies. From (9.68) we have

$$T_e \sim \frac{L_e}{v'_\psi} \sim \frac{L_d}{\bar{u}} \equiv T_E, \quad (9.69)$$

where T_E is the Eady time scale. That is, the eddy time scale (at the scale of the largest eddies) is independent of the process that ultimately determines the spatial scale of those eddies; if the eddy length scale increases somehow, perhaps because friction or β are decreased, the velocity scale increases in proportion.

9.4.2 Baroclinic eddies and their lifecycle in the atmosphere

Amplitude and Scale

We saw in section 6.9.2 that baroclinic instability in the atmosphere occurs predominantly in the troposphere, i.e., in the lowest 10 km or so of the atmosphere, with the higher stratification of the eponymous stratosphere inhibiting instability. In the mid-latitude troposphere the vertical shear and the stratification are also relatively uniform which is why fairly simple models, such as the two-layer model or the Eady model (with the addition of the β -effect) are reasonable first-order models.

The mean pole-equator temperature gradient is about 40 K and the deformation radius NH/f is about 1000 km. The Rhines scale, $\sqrt{U/\beta}$ is a little larger than the deformation radius, being perhaps 2000 km, and is similar to the width of the main mid-latitude baroclinic

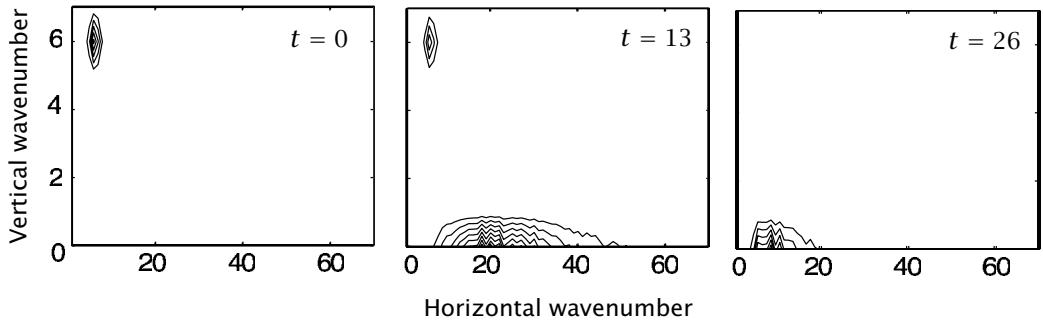


Fig. 9.8 A numerical simulation of a very idealized baroclinic lifecycle, showing contours of energy in spectral space at successive times. Initially, there is baroclinic energy at low horizontal wavenumber, as in a large-scale shear. Baroclinic instability transfers this energy to barotropic flow at the scale of the deformation radius, and this is followed by a barotropic inverse cascade to large scales. Most of the transfer to the barotropic mode in fact occurs quite quickly, between times 11 and 14, but the ensuing barotropic inverse cascade is slower. The entire process may be thought of as a generalized inverse cascade. The stratification (N^2) is uniform, and the first deformation radius is at about wavenumber 15. Times are in units of the eddy turnover time.¹³

zone which lies between about 40° and 65° , in either hemisphere. Given these, and especially given that the maximum wavelength for instability occurs at scales somewhat larger than the deformation radius, there is little prospect of an extended upscale cascade, and for this reason the Earth's atmosphere has comparable eddy kinetic and mean kinetic energies.¹²

The baroclinic lifecycle

The baroclinic lifecycle of geostrophic turbulence, sketched schematically in Fig. 9.6, can be nicely illustrated by way of numerical initial value problems, and we describe two such. The first is extremely idealized: take a doubly-periodic quasi-geostrophic model on the f -plane, initialize it with baroclinic energy at large horizontal scales, and then let the flow freely evolve. Figure 9.8 shows the results. The flow, initially concentrated in high vertical wavenumbers to best illustrate the energy transfer, is baroclinically unstable, and energy is transferred to barotropic flow at wavenumbers close to the first radius of deformation, here at about wavenumber 15. Energy then slowly cascades back to large scales in a predominantly barotropic inverse cascade, piling up at the largest scales much as in decaying, two-dimensional turbulence. Nearly all of the initial baroclinic energy is converted to barotropic, eddy kinetic energy and, even without any surface friction, the flow evolves to a baroclinically stable state. Couched in these terms, it is easy to see that the baroclinic lifecycle is a form of baroclinic inverse cascade, with an energy transfer to large total wavenumber, K_{tot} , that is made up of contributions from both horizontal and vertical wavenumbers:

$$K_{tot}^2 = K_h^2 + k_d^2 m^2, \quad (9.70)$$

where m is the vertical wavenumber and K_h is the horizontal wavenumber. As we noted earlier, the twin constraints of energy and enstrophy conservation prevent the excitation

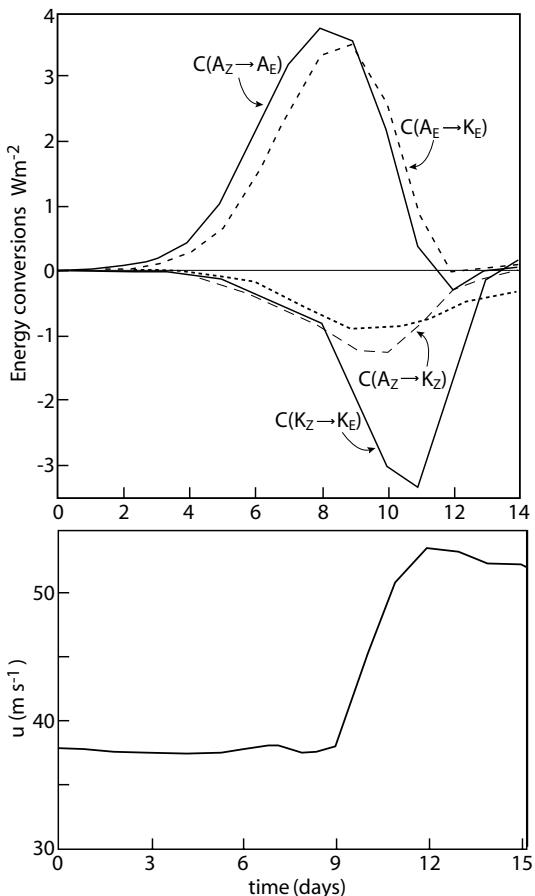


Fig. 9.9 Top: energy conversion and dissipation processes in a numerical simulation of an idealized atmospheric baroclinic lifecycle, simulated with a GCM. Bottom: evolution of the maximum zonal-mean velocity. Top: energy conversions. A_Z and A_E are zonal and eddy available potential energies (APEs), and K_Z and K_E are the corresponding kinetic energies, and $C(\cdot \rightarrow \cdot)$ represents the corresponding conversion. Initially baroclinic processes dominate, with conversions from zonal to eddy APE, $C(A_Z \rightarrow A_E)$, and then eddy APE to eddy kinetic energy, $C(A_E \rightarrow K_E)$, followed by the barotropic conversion of eddy kinetic to zonal kinetic energy [a negative value of $C(K_Z \rightarrow K_E)$]. The latter process is reflected in the *increase* of the maximum zonal-mean velocity at about day 10, shown in the lower panel.¹⁴

of horizontal scales with very large horizontal wavenumbers, and so the lifecycle proceeds through wavenumbers at the deformation scale.

The results of the second, and more realistic, initial value problem are illustrated in Fig. 9.9. Here, the atmospheric primitive equations on a sphere are integrated forward, beginning from a baroclinically unstable zonal flow, plus a small-amplitude disturbance at zonal wavenumber 6. The disturbance grows rapidly through baroclinic instability, accompanied by a conversion of energy initially from the zonal mean potential energy to eddy available potential energy (EAPE), and then from EAPE to eddy kinetic energy (EKE), and finally from EKE to zonal kinetic energy (ZKE). The last stage of this roughly corresponds to the barotropic inverse cascade of quasi-geostrophic theory, and because of the presence of a β -effect the flow becomes organized into a zonal jet. The parameters in the Earth's atmosphere are such that there is only one such jet, and in the lower panel of Fig. 9.9 we see its amplitude increase quickly from days 10 to 12, associated with the conversion of EKE to ZKE. (In other lifecycle experiments, the end state is found to be more akin to a barotropic vortex or, in meteorological parlance, a 'cut-off cyclone'.¹⁵) A commonality between all the lifecycles is the evolution toward a generally barotropic flow, with variable degrees of zonality in the final state, depending on the importance of the β -effect.

Of course, the real atmosphere is never in the zonally uniform state that is used in idealized baroclinic instability or lifecycle studies. Rather, at any given time, finite-amplitude eddies exist and these provide a finite amplitude perturbation to the baroclinically unstable zonal flow. For this reason we rarely, if ever, see an exponentially growing normal mode. Furthermore, given any instantaneous atmospheric state, zonally symmetric or otherwise, the fastest growing (linear) instability is not necessarily exponential but may be ‘non-modal’, with a secular or linear growth that, over some finite time period and in some given norm, is much more rapid than exponential.¹⁶

9.4.3 Baroclinic eddies and their lifecycle in the ocean

Basic ideas

Baroclinic instability was first developed as a theory for mid-latitude synoptic-scale instabilities in the atmosphere and the original problems were accordingly set in a zonally re-entrant channel. The ocean, apart from the Antarctic Circumpolar Current (ACC), is not zonally re-entrant. However, the ocean is driven by buoyancy and wind-forcing at the surface, and these combine to produce a region of enhanced stratification and associated shear in the ocean in the upper 500–1000 m or so — that is, in the ‘thermocline’ — as discussed more fully in chapter 16. The sloping isopycnals indicate that there is a pool of available potential energy that might be converted to kinetic energy, and so the ocean is potentially baroclinically unstable. Satellite observations indicate that baroclinic eddies are in fact almost ubiquitous in the mid- and high-latitude oceans, two particularly eddy-rich regions being the areas in and surrounding intense western boundary currents, such as the Gulf Stream, and the ACC. The ocean is, literally, a sea of eddies.¹⁷

In addition to the geometry, the main differences between the oceanic and atmospheric problems are twofold.

- (i) In the ocean, the shear and the stratification are not uniform between two rigid lids, nor even uniform between one rigid lid and a structure like the tropopause. Instead, both stratification and shear are largest in the upper ocean, decaying into a quiescent and nearly unstratified abyss.
- (ii) In the ocean, the first radius of deformation is much smaller than the scale of the large-scale flow; that is, of the gyres or the large-scale overturning circulation. On dimensional grounds we have, using a height scale and stratification representative of the upper ocean, $L_d \sim NH/f \sim 10^{-2} \times 10^3/10^{-4} = 100 \text{ km}$.

A consequence of the enhanced shear in the upper ocean is that the amplitude of the growing waves is also largely concentrated in the upper ocean, as we saw in Fig. 6.22. Regarding the stratification, in quasi-geostrophic theory we may, as in section 5.8.2, define the deformation radii by solution of the eigenvalue problem

$$\frac{\partial}{\partial z} \frac{f^2}{N^2} \frac{\partial \phi_n}{\partial z} + \Gamma_n \phi_n = 0, \quad \text{with} \quad \frac{\partial \phi_n}{\partial z} = 0 \quad \text{at } z = 0, H. \quad (9.71\text{a,b})$$

The successive eigenvalues, Γ_n , are related to the successive deformation radii, L_n , by $L_n^2 = 1/\Gamma_n$. If the stratification (N^2) is uniform the resulting eigenfunctions are cosines with corresponding eigenvalues and deformation radii given by

$$\Gamma_n = n^2 \frac{f^2 \pi^2}{N^2 H^2}, \quad L_n = \frac{1}{\sqrt{\Gamma_n}} = \frac{NH}{n \pi f_0}. \quad (9.72\text{a,b})$$

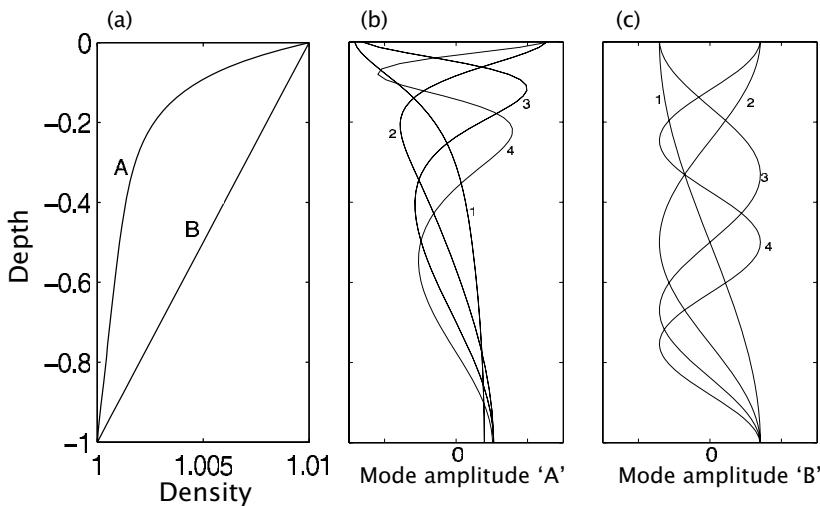


Fig. 9.10 (a) Two density profiles $[1 + (\rho_0 - \rho)/\rho_0]$, a fairly realistic oceanic case with enhanced stratification in the upper ocean (profile A), and uniform stratification (profile B). (b) and (c) The first four baroclinic modes [eigenfunctions of (9.71)]. With profile B the eigenmodes are cosines, whereas in profile A they have a larger amplitude in the upper ocean. The number of zero crossings is equal to the mode number.

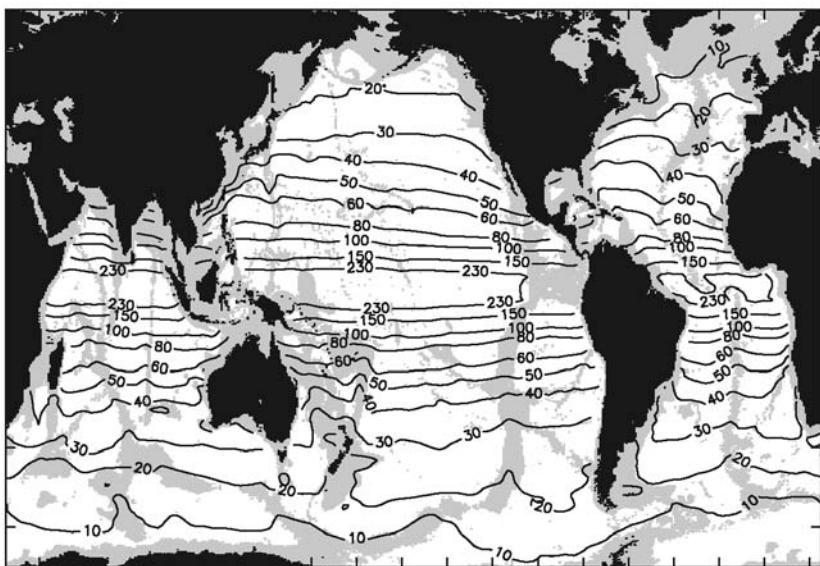


Fig. 9.11 The oceanic first deformation radius L_d , calculated using the observed stratification and an equation similar to (9.71). Near equatorial regions are excluded, and regions of ocean shallower than 3500 m are shaded. Variations in Coriolis parameter are responsible for much of large-scale variability, although weak stratification also reduces the deformation radius at high latitudes.²⁰

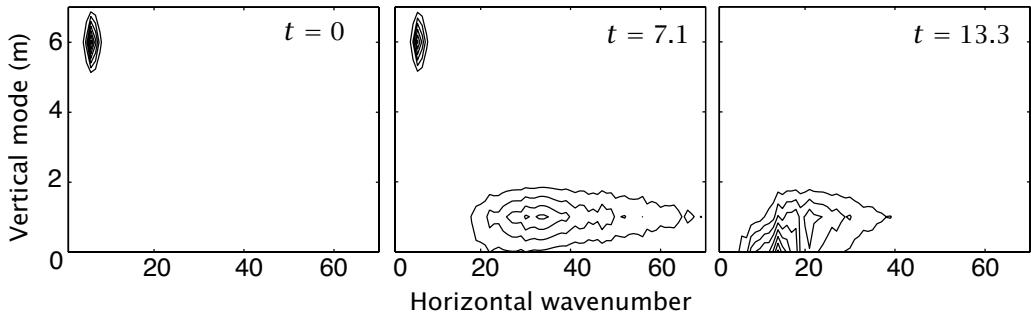


Fig. 9.12 Idealized baroclinic lifecycle, similar to that in Fig. 9.8, but with enhanced stratification of the basic state in the upper domain, representing the oceanic thermocline.

If the stratification is non-uniform, we must in most cases solve the eigenproblem numerically, and the results of one such calculation are given in Fig. 9.10. The case with uniform stratification reproduces cosine modes, whereas in the more realistic case the modes tend to have highest amplitude in the upper ocean, where the stratification is strongest — a result that is typical of oceanic profiles.¹⁸ The results of a calculation of the first deformation radius using observed oceanic profiles are given in Fig. 9.11, and values of 50–100 km emerge in mid-latitudes.¹⁹ We may therefore expect oceanic baroclinic instability to occur on a scale much smaller than that in the atmosphere, and much smaller than the scale of an ocean basin. (We should, however, note that the scale of baroclinic instability will typically be larger than the first deformation radius, L_1 , shown in Fig. 9.11, because of two compounding effects. First, in uniform stratification the first deformation radius, L_1 , as given by (9.72b) is a factor of π smaller than the simple definition NH/f . Second, in simple baroclinic instability problems like the Eady problem the wavelength of maximum instability is a few times NH/f . Thus, the wavelength of maximum instability will be an order of magnitude larger than L_1 .)

Eddy amplitudes and scales

The consequences of this small deformation radius on the lifecycle and finite-amplitude equilibration of oceanic baroclinic eddies are far-reaching, one being is that there is more scope for an inverse cascade than in the atmosphere, and indeed observations indicate that the horizontal scale of the eddies is typically a few to several times larger than the local deformation radius itself. The situation is not clear cut, however, as some observations²¹ indicate that the eddy size nevertheless scales with the local deformation radius, suggesting that the eddy scale may be set by the instability scale and not an inverse cascade.

In any case, suppose then that an ocean eddy has a horizontal scale 200 km, and that it sits in the subtropical gyre where the mean temperature gradient is 10^{-5} K m^{-1} , that the mean shear and ensuing baroclinic activity is mainly confined to the upper 1000 m of the ocean, and that the deformation radius is 50 km. The temperature gradient corresponds to a temperature difference of about 20 K across 2000 km, a horizontal buoyancy gradient of about $2 \times 10^{-9} \text{ s}^{-2}$ [using the simple equation of state $\rho = \rho(1 - \beta_T \Delta T)$ where $\beta_T = 2 \times 10^{-4} \text{ K}^{-1}$] and a shear of about 2 cm s^{-1} over the upper 1 km of ocean. Then, using (9.68),

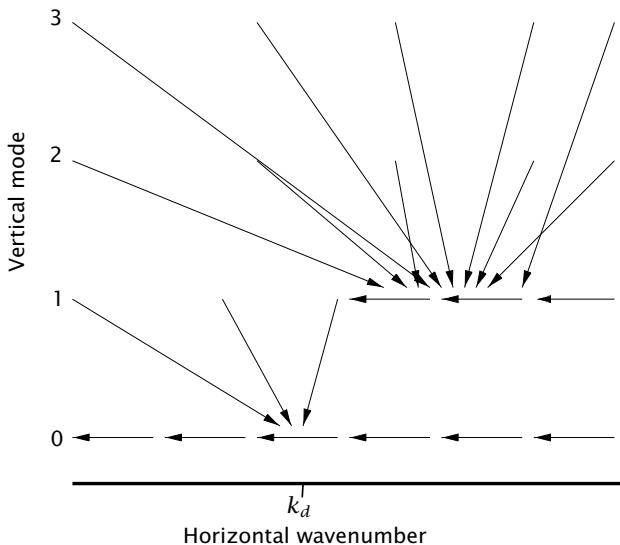


Fig. 9.13 Energy transfer paths as a function of vertical mode and horizontal wavenumber, in a fluid with an oceanic stratification; i.e., with a thermocline. Vertical mode 0 is the barotropic mode.

we can estimate a typical eddy velocity scale as

$$v'_\psi \sim \frac{L_e}{L_d} \bar{u} \approx 4\bar{u} \approx 8 \text{ cm s}^{-1}, \quad (9.73)$$

implying, as we noted earlier, an EKE that is an order of magnitude larger than the mean kinetic energy. Associated with this are typical temperature perturbations whose magnitude we can estimate using (9.65) or (9.66) as being about 2 K. These estimates are comparable to those observed in mid-ocean, with more energetic eddies forming near intense western boundary currents where gradients are large and barotropic instability also provides a source of energy for the eddies. There is least a factor of a few uncertainty, but it is noteworthy that they are roughly comparable to the values observed.

Eddy lifecycles

The lifecycle of a mid-oceanic baroclinic eddy will differ from its atmospheric counterpart in two main respects:

- (i) baroclinic eddies may be advected by the mean flow into regions with quite different properties from where they initially formed;
- (ii) the non-uniformity of the stratification affects the passage to barotropic flow.

Both of these can best be studied by numerical means. Regarding the first, eddies will often form in or near intense western boundary currents, but then will be advected by that current into the potentially less unstable open ocean before completing their lifecycle. Regarding the second, an oceanic analogue of the lifecycle illustrated in Fig. 9.8 is shown in Fig. 9.12. The main difference between this case and the atmospheric one is that baroclinic instability initially leads to the transfer of energy to vertical mode one, followed by a transfer to larger horizontal scales in the barotropic mode, as illustrated schematically in Fig. 9.13.²² If the energy is initially solely in the first baroclinic mode the cycle is more similar to the atmospheric one, but higher baroclinic modes may be more readily excited in the ocean than the atmosphere.

Notes

- 1 Rhines (1975). See also Holloway & Hendershott (1977) and Vallis & Maltrud (1993).
- 2 Adapted from Vallis & Maltrud (1993).
- 3 The mechanism described here follows Vallis & Maltrud (1993). Interactions between Rossby waves will also give rise to zonal flow, as described by Newell (1969) and Rhines (1975). Williams (1978) was one of the first numerical simulation to show the production of jets.
- 4 Simulation kindly performed by K. S. Smith.
- 5 More discussion is given by Smith *et al.* (2002), Danilov & Gryanik (2002) and Galperin *et al.* (2006).
- 6 Quasi-geostrophic turbulence was introduced by Charney (1971). Salmon (1980) and Rhines (1977) provided much of the two-layer phenomenology. Some laboratory experiments are reviewed by Read (2001).
- 7 Adapted from Salmon (1980)
- 8 Lindborg (1999) concluded that the data were consistent with a forward cascade of enstrophy between wavelengths of a few thousand kilometres and a few hundred kilometres. Boer & Shepherd (1983) and Shepherd (1987) also found a k^{-3} spectrum at similar scales, with the latter noting the importance of interactions involving stationary waves. At scales smaller than than 100 km or so the spectrum is shallower than -3 , and more like $-5/3$. This may be due to non-geostrophic effects, for example a forward cascade of energy associated with gravity wave breaking, or it may be due to a two-dimensional inverse cascade of energy with an energy source at very small scales associated with convection, or it may be due to effects associated with 'surface quasi-geostrophic' dynamics.
- 9 Adapted from Gage & Nastrom (1986).
- 10 Larichev & Held (1995) and Held & Larichev (1996). See Spall (2000) for an oceanic extension, and Thompson & Young (2006) for evidence that the theory fails to account for coherent structures.
- 11 Related arguments concerning eddy magnitudes were given by Gill *et al.* (1974). Atmospheric energetics, and atmospheric observations in general, are described by Peixoto & Oort (1992). For the oceanic case, see Wyrtki *et al.* (1976), Richardson (1983), and Stammer (1997).
- 12 This does not address the issue as to *why* the Rhines scale and deformation radius are similar. See also chapter 12.
- 13 Modified from Smith & Vallis (2001).
- 14 Adapted from Simmons & Hoskins (1978).
- 15 See Thorncroft *et al.* (1993). These authors identify two classes of lifecycles, which they call LC1 and LC2, which have differing degrees of decay of eddy kinetic energy in the later parts of the lifecycle, and with LC2 producing cut-off cyclones. Initial conditions and spatial inhomogeneities, including the horizontal shear of the flow and the presence of critical layers, play an important role in guiding the location of the wave breaking, and hence the final state that is reached.
- 16 The theory of this has been developed by Farrell (1984), Farrell & Ioannou (1996) and other papers by these authors. These authors have rightly emphasized that exponential growth is the exception, not the rule, in the real world. Indeed, if the basic flow is oscillatory, baroclinic instability can arise even if the basic flow is never unstable via the CSP criterion.

- 17 This realization came to fruition as a result of the bilateral US-USSR POLYMODE project in the 1970s. See Robinson (1984).
- 18 For example, Kundu *et al.* (1975).
- 19 The eigenproblem actually solved was

$$\partial^2 \phi / \partial z^2 + (N^2(z)/c^2) \phi = 0, \quad \phi = 0 \quad \text{at } z = 0, H, \quad (9.74)$$

where H is the ocean depth and N is the observed buoyancy frequency. The deformation radius is given by $L_d = c/f$ where c is the first eigenvalue and f is the latitudinally varying Coriolis parameter.

- 20 From Chelton *et al.* (1998).

- 21 Stammer (1997).

- 22 Fu & Flierl (1980) and Smith & Vallis (2001) examined this issue in more detail, both analytically and numerically. Figure 9.13 is adapted from these papers. Barnier *et al.* (1991) and Hua & Haidvogel (1986) looked at the role of high baroclinic modes in oceanic eddies and geostrophic turbulence.

Problems

- 9.1 Obtain an expression for the wave-turbulence boundary, (k_β^x, k_β^y) , by equating a ‘turbulence frequency’, Uk with a Rossby wave frequency, $\beta k^x / k^2$. Plot it and compare to Fig. 9.2.

Solution:

$$k_\beta^x = \left(\frac{\beta}{U} \right)^{1/2} \cos^{3/2} \theta, \quad k_\beta^y = \left(\frac{\beta}{U} \right)^{1/2} \sin \theta \cos^{1/2} \theta. \quad (9.1a,b)$$

- 9.2 ♦ Use the phenomenological model of section 9.3.

- (a) Suppose that $\beta = 0$ and that the friction is parameterized by a linear drag on the barotropic flow, $-r\zeta$, with a corresponding inverse cascade halting scale, L_h , given by $L_h = (\varepsilon/r^3)^{1/2}$. Show that the model fails to give a prediction for the stopping scale in terms of the external parameters (i.e., the imposed shear, the deformation radius and the frictional parameter r). Will the flow equilibrate? Discuss what this may mean from a physical standpoint.
- (b) Suppose that $\beta \neq 0$ and that the halting scale is given by (9.18), namely $L_h = [\varepsilon/(r\beta^2)]^{1/4}$. Does the model give a prediction for ε , and does the flow equilibrate? If so, what is the halting scale in terms of the external parameters? Compare the other predictions of this model with those obtained using the halting scale $(\varepsilon/\beta^3)^{1/5}$.
- (c) Suppose that the inverse cascade is halted by a nonlinear drag, such that in the zonal momentum equation the drag term is $-A(u^2 + v^2)u$, where A is a constant. What is the halting scale?

- 9.3 (a) If the inverse cascade is halted by linear drag, show that an estimate of the horizontal eddy diffusivity is

$$\kappa \sim \left(\frac{\varepsilon}{r^2} \right), \quad (P9.2)$$

where r is a drag coefficient (such that $D\zeta/Dt = F - r\zeta$) and ε is the energy cascade rate. Estimate the halting scale and the eddy diffusivity for the Earth’s atmosphere, stating clearly any assumptions you may make. Compare it to the halting scale and associated eddy diffusivity obtained if one assumes that the eddies occur at the β -scale, estimating this both with and without friction.

- (b) Suppose that in a constant-density two-dimensional fluid friction is parameterized by a nonlinear drag, such the momentum equation takes the form

$$\frac{\partial \mathbf{u}}{\partial t} + \text{other terms} = -\nabla p - \lambda |\mathbf{u}| \mathbf{u} \quad (P9.3)$$

where λ is a constant. Use scaling arguments to estimate the halting scale of an inverse cascade when the flow is forced at a high wavenumber with a rate of energy input equal to ε . Compare it to the scale obtained using a linear drag. Estimate these scales for the atmosphere, stating clearly any assumptions you make. (See also Griannik *et al.* 2004.)

Damn the torpedoes! Full speed ahead!
Admiral David Farragut, at the battle of Mobile Bay, 1864.

CHAPTER TEN

Turbulent Diffusion and Eddy Transport

THE TRANSPORT OF FLUID PROPERTIES BY UNSTEADY MOTION, that is, the way in which the properties of a fluid may be carried from one location to another by waves and turbulence, is one of the most important topics in geophysical fluid dynamics. It may be the dominant transport in a fluid, greatly exceeding that of the mean flow — in the atmosphere, for example, heat is transferred polewards primarily by the action of unsteady weather systems, not by the much weaker time-mean flow. However, we are often not interested in the details of the turbulent eddies and hence we might seek to *parameterize* the turbulent transport in terms of the mean flow; unfortunately, no general theory exists for such transport, for indeed such a theory would amount to a theory of turbulence. In the absence of this, we focus our attention in this chapter on the theory (such as it is) and practice of *turbulent diffusion*. In models of turbulent diffusion, the turbulent transport is generally related to the gradient of the mean flow, and it is the simplicity of the resulting expressions that has led to their wide adoption in areas as different as turbulent pipe flow, atmospheric boundary layer transport and large-scale ocean modelling. These models are, by and large, rational, simple and tractable — a blend of heuristic reasoning and elementary mathematics, the latter needed to ensure that certain basic requirements (conservation laws, for example) of a physical process are captured by a parameterization. However, just like other turbulent closures, they rely on rather physical assumptions that cannot be rigorously justified. In the first part of the chapter we consider turbulent diffusion from a general standpoint, and then specialize our discussion to geofluids, and in particular to large-scale transport by baroclinic eddies. Those readers with some prior knowledge of turbulent diffusion may choose to skip ahead to section 10.6.

10.1 DIFFUSIVE TRANSPORT

We begin with a brief discussion of the diffusion equation itself, to wit

$$\frac{\partial \phi}{\partial t} = \kappa \nabla^2 \phi, \quad (10.1)$$

where κ is a constant, positive, scalar diffusivity and the tracer ϕ is a scalar field. We expect that an initially concentrated blob of tracer would spread out — it would diffuse — and thus small parcels of tracer are transported. How quickly does this occur, or, put another way, is there an effective diffusive transport velocity?

If the rate of spreading becomes independent of the initial conditions then, purely from dimensional considerations, the spreading can depend only on the diffusivity and time itself and we can write

$$\overline{X^2} = \alpha \kappa t, \quad (10.2)$$

where $\overline{X^2}$ is the mean-square displacement, t is time and α is a non-dimensional constant. Let us quantify this with an explicit calculation. If ϕ is interpreted as the density of markers of fluid parcels, then the mean-square displacement of the markers is given by (in three dimensions)

$$\overline{X^2} = \frac{\int_0^\infty r^2 \phi r^2 dr}{\int_0^\infty \phi r^2 dr}, \quad (10.3)$$

where the denominator, the total amount of tracer present, is a constant and we have assumed a spherically symmetric distribution of tracer. Using (10.1) we find

$$\frac{d}{dt} \int_0^\infty r^2 \phi r^2 dr = \kappa \int_0^\infty r^2 \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \phi}{\partial r} \right) r^2 dr = 6\kappa \int_0^\infty \phi r^2 dr \quad (10.4)$$

after a couple of integrations by parts. Thus

$$\frac{d}{dt} \overline{X^2} = 6\kappa \quad (10.5)$$

and because κ is a constant we have the important result that

$$\boxed{\overline{X^2} = 6\kappa t}. \quad (10.6)$$

In two dimensions the equivalent calculation begins with

$$\overline{X^2} = \frac{\int_0^\infty r^2 \phi r dr}{\int_0^\infty \phi r dr} \quad (10.7)$$

and using the diffusion equation we find

$$\frac{d}{dt} \int_0^\infty r^2 \phi r dr = \kappa \int_0^\infty r^2 \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \phi}{\partial r} \right) r dr = 4\kappa \int_0^\infty \phi r dr. \quad (10.8)$$

Thus we obtain

$$\overline{X^2} = 4\kappa t. \quad (10.9)$$

Finally, in one dimension (i.e., spreading along a line) it is easy to show that

$$\overline{X^2} = 2\kappa t. \quad (10.10)$$

10.1.1 An explicit example

We gain a little more intuition about what the above calculations mean by considering the case in which the initial tracer distribution is a delta function at the origin. If the total amount of tracer is unity, then in three dimensions at subsequent times the tracer is given by the distribution

$$\phi(r, t) = \frac{1}{8(\pi\kappa t)^{3/2}} \exp(-r^2/4\kappa t), \quad (10.11)$$

as may be checked by substitution back into the equation of motion. (See problem 10.1.) The distribution clearly broadens with time, and the mean-square distance from the origin is given by

$$\overline{X^2} = \int_0^\infty \frac{4\pi r^2}{8(\pi\kappa t)^{3/2}} \exp(-r^2/4\kappa t) dr = 6\kappa t, \quad (10.12)$$

as in (10.6). The important point is that the mean distance travelled by a particle during a time interval t is proportional to the square root of that time interval. This is, of course, redolent of a random walk (see the shaded box on the following page), which brings us to the subject of turbulent diffusion.

10.2 TURBULENT DIFFUSION

Fluids differ from solids in that they can transport properties by advection — thus, heat is primarily transferred polewards in the atmosphere by means of air movement and not by molecular diffusion. Turbulent fluid motion differs from laminar fluid motion in that such advective transport may be greatly enhanced by the seemingly random motion of the fluid, the net transport being much larger than that which would be effected by the time-mean fluid motion alone. Indeed, to continue the atmospheric example, away from the tropics the poleward transport of heat in the atmosphere is largely effected by way of the (large-scale) turbulent transfer of heat in mid-latitude weather systems. Of course, such transfer is simply by advection, and if we could explicitly calculate the motion of the fluid parcels we could explicitly calculate the transport. However, turbulent transport is both very complicated and very sensitive to the initial conditions, so that any hope of performing such a calculation in a real situation is often a forlorn one.

Turbulent transport is most important in *inhomogeneous* situations, because it is the divergence of the transport that is important and the mean divergence is non-zero only if there is inhomogeneity. The theories of chapters 8 and 9 do not lend themselves to an easy extension to inhomogeneous flow, and we turn to a slightly more empirical approach.¹

10.2.1 Simple theory

Let us consider how fluid markers are transported in a statistically steady, homogeneous, turbulent flow. The markers are introduced at the origin $x = y = z = 0$ at $t = 0$; we may create an ensemble of such markers by performing many such tracer release experiments on different realizations of the turbulent flow, but with each flow having the same statistical properties. The question is, what is the average rate of dispersion of a single particle of fluid?

A Random Walk

Here we give an elementary derivation of the most basic result in random walk theory, the relationship of the mean-square displacement to the number of steps taken.² A loose analogy is that of drunkards staggering randomly from here to there, with no correlation between their successive steps. After any number of steps, the *mean* displacement of the drunkards is zero, but we expect their *root-mean-square* (r.m.s.) displacement to increase: this is because drunkards independently thrown out of the same bar will generally wander off in different directions (which is why the mean displacement is zero), but after some time most of them will indeed end up some distance away.

For simplicity consider steps, \mathbf{s}_n , each with random orientation but equal magnitude, s . The displacement after n steps is related to the displacement after $n - 1$ steps by

$$\mathbf{D}_n = \mathbf{D}_{n-1} + \mathbf{s}_n, \quad (\text{R.1})$$

so that the amplitude of \mathbf{D}_n , namely D_n , is given by

$$\begin{aligned} D_n^2 &= (\mathbf{D}_{n-1} + \mathbf{s}_n) \cdot (\mathbf{D}_{n-1} + \mathbf{s}_n) \\ &= D_{n-1}^2 + s^2 + 2\mathbf{D}_{n-1} \cdot \mathbf{s}_n. \end{aligned} \quad (\text{R.2})$$

Taking an ensemble average over many realizations gives

$$\overline{D_n^2} = \overline{D_{n-1}^2} + s^2 \quad (\text{R.3})$$

having used $\overline{\mathbf{D}_{n-1} \cdot \mathbf{s}_n} = 0$, because each step is random.

Now, $D_0 = 0$, so that $\overline{D_1^2} = s^2$, $\overline{D_2^2} = 2s^2$ and so on. Thus, using (R.3) to proceed inductively, we have

$$\overline{D_n^2} = ns^2, \quad (\text{R.4})$$

or

$$\boxed{\overline{D_n^2}^{1/2} = \sqrt{ns}}. \quad (\text{R.5})$$

Thus, in a random walk the *root-mean-square displacement increases with the half-power of the number of steps taken*. More work is required to calculate the distribution of the random walkers, but it may be shown that in the limit of infinitesimally small steps the random walk becomes a Wiener process and the distribution becomes Gaussian, as in (10.11) (with the exact form depending on the dimensionality of the problem), indicating a diffusive process. Diffusion may be considered a continuous random walk, with the displacement proportional to the square root of time.

The displacement of a marker at a time t is given by

$$X(t) = \int_0^t \mathbf{V}(t') dt', \quad (10.13)$$

where \mathbf{V} is the velocity of the fluid parcel — a material velocity. [We will use uppercase variables to denote material ('Lagrangian') quantities.] The mean-square displacement is

$$\overline{X^2(t)} = \int_0^t dt_1 \int_0^t \overline{\mathbf{V}(t_1) \cdot \mathbf{V}(t_2)} dt_2, \quad (10.14)$$

where the overbar denotes an ensemble average, and thus $\overline{\mathbf{V}(t_1) \cdot \mathbf{V}(t_2)}$ is a measure of the velocity correlation between the velocities of the fluid parcels at times t_1 and t_2 . That is,

$$\overline{\mathbf{V}(t_1) \cdot \mathbf{V}(t_2)} = \overline{v^2} R(t_2 - t_1) = \overline{v^2} R(\tau), \quad (10.15)$$

where $R(t_2 - t_1)$ is the velocity correlation function and, because the turbulence is statistically steady, this depends only on the time difference $\tau = t_2 - t_1$. Furthermore, $R(-\tau) = R(\tau)$. Thus,

$$\overline{X^2(t)} = \int_0^t dt_1 \overline{v^2} \int_0^t R(t_2 - t_1) dt_2 = \int_0^t d\hat{t} \overline{v^2} \int_{-\hat{t}}^{t-\hat{t}} R(\tau) d\tau, \quad (10.16)$$

changing variables to τ and $\hat{t} = t_1$ (Fig. 10.1). We expect the velocity correlation function to fall monotonically from its initial value of unity to a value approaching zero as $\tau \rightarrow \infty$, as in Fig. 10.2, and typically, there will be some characteristic time τ_{corr} that parameterizes the behaviour of the function. In general, we cannot obtain explicit general solutions without detailed knowledge of this correlation function, but there are two interesting limits.

- (i) *The short-time limit.* For small times, i.e., for $t \ll \tau_{corr}$ (and so $t_1, t_2 \ll \tau_{corr}$) the correlation function will be approximately unity and so (10.16) becomes

$$\overline{X^2(t)} \approx \int_0^t d\hat{t} \overline{v^2} \int_0^t d\hat{t} = \overline{v^2} t^2. \quad (10.17)$$

Thus, the root-mean-square displacement increases linearly with time, and linearly with the root-mean-square velocity of the flow. For small times, the fluid parcel's behaviour is well correlated with that at the initial time, and so the displacement increases linearly in the direction it was initially going. Indeed, directly from (10.13) we have, for small times, $X(t) \approx Vt$, which leads directly to (10.17).

- (ii) *The long-time limit.* We are now concerned with the case $t \gg \tau_{corr}$. Because the correlation function falls with time, most of the contributions to the second integrand [involving $R(\tau)$] in (10.16) are from $\tau \leq \tau_{corr}$. Thus, without much loss in accuracy, we can replace the limits of integration by $-\infty$ and $+\infty$; that is

$$\overline{X^2(t)} \approx \int_0^t d\hat{t} \overline{v^2} \int_{-\infty}^{\infty} R(\tau) d\tau. \quad (10.18)$$

Assuming the second integral converges, it is just a number; in fact, noting that $R(\tau) = R(-\tau)$, we may use it to *define* the correlation time τ_{corr} by

$$\tau_{corr} \equiv \int_0^{\infty} R(\tau) d\tau. \quad (10.19)$$

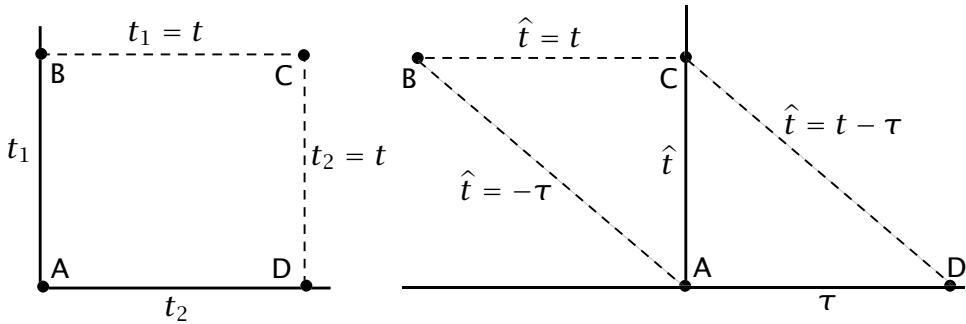


Fig. 10.1 Changes of time variables involved in (10.16) and (10.32). The original two-dimensional integral is over the rectangle ABCD. Defining $\tau = t_2 - t_1$ and $\hat{t} = t_1$, then the area is spanned by $[\hat{t} = (0, t), \tau = (-\hat{t}, t - \hat{t})]$ as in (10.16), or by $[\tau = (0, t), \hat{t} = (0, t - \tau)]$ (i.e., ACD) plus $[\tau = (-t, 0), \hat{t} = (-\tau, t)]$ (i.e., ABC) in (10.32).

We then have the important result that

$$\boxed{\overline{X^2(t)} \approx 2\overline{v^2}t \int_0^\infty R(\tau) d\tau = 2\overline{v^2}\tau_{corr}t.} \quad (10.20)$$

That is, for times that are long compared with the turbulence correlation time, the distance travelled by a fluid parcel in some time interval is proportional to the square-root of that time interval, just as for a diffusive process; this is because the fluid parcels are essentially undergoing random walks. The equation (10.20) connects two quite different fluid properties: the left-hand side tells us how tracers are dispersed in a turbulent flow, a material property, whereas the right-hand side can be evaluated from the Eulerian velocity field at different times. Both the left- and right-hand sides can be directly measured, by looking at the dispersion of a dye and by measuring the velocity at successive times.

We may define a *coefficient of turbulent diffusivity* by

$$\boxed{K_{turb} = \frac{1}{3}\overline{v^2}\tau_{corr},} \quad (10.21)$$

and then we have the result that

$$\overline{X^2(t)} = 6K_{turb}t. \quad (10.22)$$

Comparison of (10.22) with (10.6) indicates that the transport of turbulent flow, under these conditions, is like a diffusive transport, with a coefficient of diffusivity given by (10.21). [Sometimes, the numerical factors are neglected, and a diffusivity is defined by the expressions

$$K_{turb} \equiv \frac{d\overline{X^2(t)}}{dt} \quad \text{or} \quad K_{turb} \equiv \frac{1}{2} \frac{d\overline{X^2(t)}}{dt}. \quad (10.23)$$

These lose the exact connection with a true diffusion coefficient, but usually the turbulent diffusivity can only be estimated, anyway.]

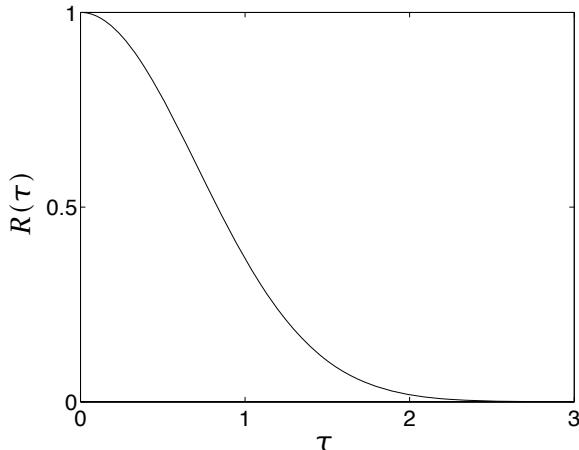


Fig. 10.2 Schematic of a velocity correlation function in turbulent flow with correlation time $\tau_{corr} = \mathcal{O}(1)$. For small times, $\tau \ll \tau_{corr}$, $R(\tau) \approx 1$. For large times, $\tau \gg \tau_{corr}$, $R(\tau) \ll 1$. We may define the correlation time by $\tau_{corr} = \int_0^\infty R(\tau) d\tau$.

We may define a correlation length scale to be the approximate distance that a parcel moves, on average, in a material (i.e., ‘Lagrangian’) correlation time. Thus

$$l_{corr} \equiv v_{rms} \tau_{corr}, \quad (10.24)$$

where $v_{rms} = (\bar{v^2})^{1/2}$, whence

$$K_{turb} = \frac{1}{3} v_{rms} l_{corr}. \quad (10.25)$$

In most situations, the numerical coefficient (1/3 here) cannot be trusted because a real turbulent flow is unlikely to satisfy the restrictions of stationarity and homogeneity that we have imposed. Nevertheless, a relationship similar to (10.25) — that a turbulent diffusivity is proportional to an r.m.s. turbulent velocity and a correlation length scale, is the foundation for semi-empirical *mixing length* theories that we discuss in section 10.4.

The simple relationships between the mean-square displacement, the Lagrangian time scale, the mean-square velocity and the eddy diffusivity allows the diffusivity to be computed from the statistics of particle trajectories. Thus, suppose that a cluster of floats is released into the ocean, or some balloons are released in the atmosphere. If neutrally buoyant, these instruments then essentially become labelled fluid particles, and one may compute K_{turb} directly from the dispersion of the cluster using (10.22). If it is possible to measure their root-mean-square velocity, then one may use (10.21) to estimate the diffusivity from this and the material correlation time scale.

10.2.2 * An anisotropic generalization

We now consider the correlation between the different components of the displacement in anisotropic, but still homogeneous, flow.³ The displacement of a fluid particle is given by (10.13), and this is a random vector. Thus, generalizing (10.14), we may define the fluid

particle displacement covariance tensor by

$$D_{ij}(t) = \overline{X_i(t)X_j(t)} = \int_0^t \int_0^t \overline{V_i(t_1)V_j(t_2)} dt_1 dt_2, \quad (10.26)$$

where the velocity denoted by V_i is the i th component of the velocity of a fluid element. For small times, $X_i(t) \approx v_i(\mathbf{a}, 0)t$, where $v_i(\mathbf{a}, 0)$ is the fluid velocity at the parcel's initial position, \mathbf{a} , and we obtain

$$D_{ij}(t) \approx \overline{v_i(\mathbf{a}, 0)v_j(\mathbf{a}, 0)}t^2. \quad (10.27)$$

If the flow is statistically steady and homogeneous the average of any quantity has no spatial or temporal dependence and so

$$D_{ij}(t) = A_{ij}t^2, \quad (10.28)$$

where the tensor $A_{ij} = \overline{v_i v_j}$ has constant entries, and this is a slight generalization of (10.17).

The velocity covariance of a fluid parcel at times t_1 and t_2 is, as before, a function only of the time difference $t_1 - t_2$ and so it must have the form

$$\overline{V_i(t_1)V_j(t_2)} = (\overline{v_i^2} \overline{v_j^2})^{1/2} R_{ij}(t_2 - t_1). \quad (10.29)$$

Except in the case of isotropic flow $R_{ij}(\tau) \neq R_{ij}(-\tau)$, but we do have, in general,

$$R_{ij}(\tau) = R_{ji}(-\tau). \quad (10.30)$$

Now, to obtain a generalization of (10.20), we first use (10.29) in (10.26) to obtain

$$D_{ij}(t) = (\overline{v_i^2} \overline{v_j^2})^{1/2} \int_0^t \int_0^t R_{ij}(t_2 - t_1) dt_1 dt_2. \quad (10.31)$$

If we change variables to $\tau = t_2 - t_1$ and $\hat{t} = t_1$ (see Fig. 10.1) we obtain⁴

$$D_{ij}(t) = (\overline{v_i^2} \overline{v_j^2})^{1/2} \left(\int_0^t d\tau \int_0^{t-\tau} d\hat{t} R_{ij}(\tau) + \int_{-t}^0 d\tau \int_{-\tau}^t d\hat{t} R_{ij}(\tau) \right), \quad (10.32)$$

and using (10.30) this becomes

$$D_{ij}(t) = 2 (\overline{v_i^2} \overline{v_j^2})^{1/2} \int_0^t d\tau \int_0^{t-\tau} d\hat{t} \hat{R}_{ij}(\tau), \quad (10.33)$$

where $\hat{R}_{ij} = (R_{ij} + R_{ji})/2$. This order of integration enables us to perform the integration over \hat{t} , giving

$$D_{ij}(t) = 2 (\overline{v_i^2} \overline{v_j^2})^{1/2} \int_0^t (t - \tau) \hat{R}_{ij}(\tau) d\tau. \quad (10.34)$$

For long times, i.e., for $t \gg \tau_{corr}$, the upper limit of the integration may be taken to be infinity, again because the contributions to the integrand from $\hat{R}_{ij}(\tau)$ all come from small τ . Furthermore, we expect that for large t ,

$$\int_0^\infty t \hat{R}_{ij}(\tau) d\tau \gg \int_0^\infty \tau \hat{R}_{ij}(\tau) d\tau, \quad (10.35)$$

because $R_{ij}(\tau)$ is only non-negligible for small τ , and $t \gg \tau$ in this range. Thus, we finally obtain a generalization of (10.20) for the displacement covariance of two components of the displacement, namely

$$D_{ij} = 2 \left(\overline{v_i^2} \overline{v_j^2} \right)^{1/2} t \int_0^\infty \hat{R}_{ij}(\tau) d\tau. \quad (10.36)$$

The integral is a tensor with constant entries, analogous to the turbulent decorrelation time scale of (10.19). Then, with $\tau_{ij} \equiv \int_0^\infty \hat{R}_{ij}(\tau) d\tau$, the corresponding turbulent diffusivity is

$$K_{ij} = \frac{1}{3} \left(\overline{v_i^2} \overline{v_j^2} \right)^{1/2} \tau_{ij}. \quad (10.37)$$

10.2.3 Discussion

We have shown that, for sufficiently long times, the distance travelled by a fluid parcel in some time is proportional to the square root of that time, just as for a diffusive process and just as for a random walk. The motion of our fluid parcel is analogous to that of a dust particle undergoing Brownian motion — both are continually buffeted and undergo random walks as a result. Still, it may appear that the usefulness of our results is limited by the assumptions of stationarity and homogeneity — it is well-nigh impossible in nature to produce a statistically stationary, homogeneous turbulent flow, because statistical stationarity implies there must be an energy source and this, as well as the presence of boundaries, militate against homogeneity. However, we should not be so pessimistic, on two counts.

- (i) Similar ideas may be directly applied to flows that are homogeneous in one direction.⁵
- (ii) Often, a flow will *not* be homogeneous in any direction. However, if the *statistics* of the eddy motion vary on a space scale that is longer than $v_{rms}\tau_{corr}$, then the eddy transport properties may be determined by a local theory. For example, the size of the eddy diffusivity is then determined by $D_t \sim v_{rms}l_{corr}$ where the parameters, and hence the diffusion coefficient, vary, but only on a scale longer than the energy-containing scale.

The essential results of this section then lie in (10.20), (10.21) and (10.25): that the dispersion of a fluid particle in a turbulent flow is *diffusive* in nature, and that the turbulent diffusivity is proportional to the product of the root-mean-square velocity and the correlation length.

10.3 TWO-PARTICLE DIFFUSIVITY

Let us now consider the problem of determining the mutual separation of two fluid parcels; the problem is relevant to geofluids because by tracking the separation of floats in the ocean, or balloons in the atmosphere, we can learn much about the nature of large-scale turbulence in those systems. The problem differs from the one-particle problem, because the separation of the particles itself will affect the rate of increase of the separation. In the one-particle problem in homogeneous flow, the position of the particular tagged fluid particle plays no direct role in determining its rate of spreading from its initial condition — any one position is the same as any other. But if two particles are close together, they may be swept away together by some large eddy, without affecting their mutual separation whereas two particles that are widely separated will undergo largely uncorrelated motion. Thus, we identify two regimes.

- (i) A regime in which the separation of the particles is greater than the scale of the largest eddies. In this case, each particle is undergoing a random walk that is effectively uncorrelated with that of the other particle.
- (ii) A regime in which the separation of the particles is less than the energy-containing scale of the motion. In this case, the eddies that contribute most to the two-particle separation are those that are comparable in scale to the separation itself.

If we attempt to apply the Taylor analysis ab initio we evidently have, by analogy to (10.13)

$$\mathbf{Y}(t) = \mathbf{X}_1(0) - \mathbf{X}_2(0) + \int_0^t [\mathbf{V}_1(t') - \mathbf{V}_2(t')] dt', \quad (10.38)$$

and a mean-square separation of

$$\overline{\mathbf{Y}^2(t)} = \overline{(\mathbf{X}_1(0) - \mathbf{X}_2(0))^2} + \int_0^t dt_1 \int_0^t [\mathbf{W}_1(t_1) \cdot \mathbf{W}_2(t_2)] dt_2, \quad (10.39)$$

where $\mathbf{W}(t) = \mathbf{V}_1(t) - \mathbf{V}_2(t)$. However, it is now difficult to proceed much further. The problem is that we cannot write

$$\overline{\mathbf{W}(t_1) \cdot \mathbf{W}(t_2)} = \overline{w^2} R(t_2 - t_1), \quad (10.40)$$

because the correlation will depend on the initial separation of the particles as well as the time since then. Thus, the diffusivity itself will depend on both time and the initial particle separation, and the results analogous to those of the single-particle diffusivity cannot easily be recovered. However, we can make some progress by separately considering the two above-mentioned regimes.

10.3.1 Large particle separation

This case is analogous to the single-particle case. The particle separation is given by

$$\mathbf{Y}(t) = \mathbf{X}_1(t) - \mathbf{X}_2(t), \quad (10.41)$$

so the mean-square separation is

$$\overline{\mathbf{Y}^2(t)} = \overline{\mathbf{X}_1(t)^2} + \overline{\mathbf{X}_2(t)^2} - 2\overline{\mathbf{X}_1(t) \cdot \mathbf{X}_2(t)}. \quad (10.42)$$

For long times, the last term is zero because the motion of the two particles is uncorrelated. Furthermore, each of the first two terms is given by (10.20) or (10.22), so that the mean separation varies as

$$\overline{\mathbf{Y}^2(t)} = 4\overline{v^2}\tau_{corr}t \quad (10.43)$$

and the rate of separation, for large t , is given by

$$\frac{d\overline{\mathbf{Y}^2(t)}}{dt} = 4\overline{v^2}\tau_{corr} = 12K_{turb}. \quad (10.44)$$

Thus, the relative diffusion is twice that of the single-particle process, in the limit that the particles are separated by an amount larger than the largest eddies.

10.3.2 Separation within the inertial range

How do fluid parcels whose separation is at inertial scales behave relative to each other?⁶ Suppose that two particles are tagged, and that their separation is greater than the viscous scale but smaller than the scales of the largest eddies — that is, the separation lies within the inertial range of the flow. Then, the rate of separation of the two particles can depend only on two quantities, the separation itself and properties of the inertial range, meaning (in three dimensions) the energy flux, ε , through the system. It cannot depend on the time, because this would imply that the subsequent rate of particle separation depends on the history of how the particles came to their current positions. Thus we can write

$$\frac{d\bar{L}^2}{dt} = g(\bar{L}, \varepsilon), \quad (10.45)$$

where $\bar{L} \equiv \overline{Y(t)^2}^{1/2}$. Dimensional analysis then gives

$$\frac{d\bar{L}^2}{dt} = A\varepsilon^{1/3}\bar{L}^{4/3}, \quad (10.46)$$

where A is a non-dimensional constant, and this is known as 'Richardson's four-thirds law'. We can integrate (10.46) to give

$$\bar{L}^2 \sim \varepsilon t^3. \quad (10.47)$$

Another way of deriving (10.46) is to suppose that the separation obeys the diffusive law

$$\frac{d\bar{L}^2}{dt} = K_{turb}, \quad (10.48)$$

where K_{turb} is a turbulent diffusivity that is *a function of the separation itself*. This is because the farther apart the eddies are, the larger the scale of the eddies that can move the two particles independently, rather than just sweeping them along together. An estimate of the diffusivity is then

$$K_{turb} \sim \nu l, \quad (10.49)$$

where ν is the characteristic velocity of an eddy of scale l , and $l \sim \bar{L} = \overline{Y^2}^{1/2}$. Using the inertial range scaling $\nu \sim (l\varepsilon)^{1/3}$ this is

$$K_{turb} \sim \varepsilon^{1/3}\bar{L}^{4/3}, \quad (10.50)$$

and so (10.48) becomes

$$\frac{d\bar{L}^2}{dt} \sim \varepsilon^{1/3}\bar{L}^{4/3}, \quad (10.51)$$

as before. Of course, dimensional consistency demands that we obtain the same result, but the derivation is intuitive and the estimate of the two-particle diffusivity (i.e., (10.50), that the eddy diffusivity governing the separation of two fluid parcels goes as the 4/3 power of their root-mean-square separation) is useful. If the particle separation is greater than the scale of the largest eddies in the system, l_{max} , then

$$K_{turb} \sim \nu(l_{max})l_{max} \sim \varepsilon^{1/3}l_{max}^{4/3} = \text{constant}. \quad (10.52)$$

The two-particle separation then proceeds as a conventional random walk or diffusive process, with the mean-square separation increasing linearly with time.

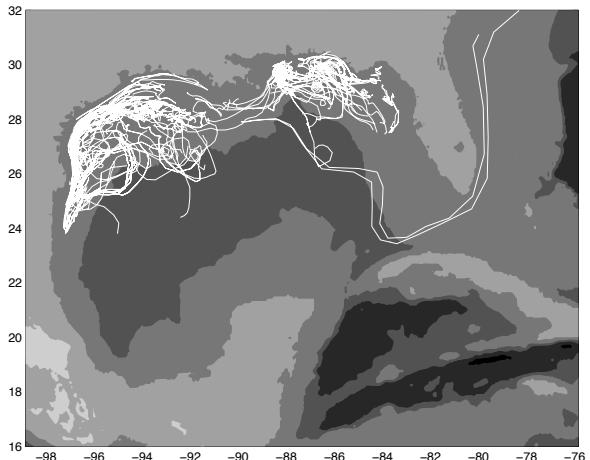


Fig. 10.3 Trajectories of surface drifters in the Gulf of Mexico, each truncated to produce paths of just 25 days. The drifters were released as part of 'SCULP' — the Surface Current and Lagrangian drift Program.⁷

Diffusion in two-dimensional flow

In two dimensions the turbulent diffusivity will differ depending on whether the two-particle separation is in the energy inertial range or in the enstrophy inertial range. In the energy inertial range the scaling is the same as in the three-dimensional case, but in the enstrophy inertial range the rate of separation will depend on the enstrophy cascade rate, η , and the separation itself. Dimensional analysis then leads to

$$\frac{d\bar{L}^2}{dt} = B\eta^{1/3}\bar{L}^2, \quad (10.53)$$

where B is a non-dimensional constant. This integrates to

$$\bar{L}^2 = \bar{L}(0)^2 \exp(B\eta^{1/3}t), \quad (10.54)$$

or $\overline{Y^2(t)} = \overline{Y^2(0)} \exp(B\eta^{1/3}t)$. Thus, the rate of separation is *exponential* in the enstrophy inertial range, a result unique to two-dimensional turbulence. Similarly, using $v \sim \eta^{1/3}l$, the turbulent diffusivity is given by

$$K_{turb} \sim \eta^{1/3}\bar{L}^2. \quad (10.55)$$

A geophysical example

The above ideas are well illustrated by analysing the trajectories of surface drifters in the Gulf of Mexico. The drifters are free-moving buoys which float about a half metre below the surface and which thus act as imperfect fluid markers — imperfect because they cannot follow the full three-dimensional motion of water parcels. Nevertheless, the motion at these scales can be expected to be quasi-geostrophic and nearly horizontal, so the associated error will be small. The drifters are tracked by satellite and their trajectories, proxies for the motion of fluid parcels, are shown in Fig. 10.3. The two-particle, or two-drifter, separation is illustrated in Fig. 10.4 and two regimes may be discerned. In the first, the pair separations grow approximately exponentially in time, with an e-folding time of 2 days, consistent with motion within an enstrophy inertial range using (10.54). The second regime is characterized by a power-law growth, proportional to $t^{2.2}$, somewhat slower than the t^3

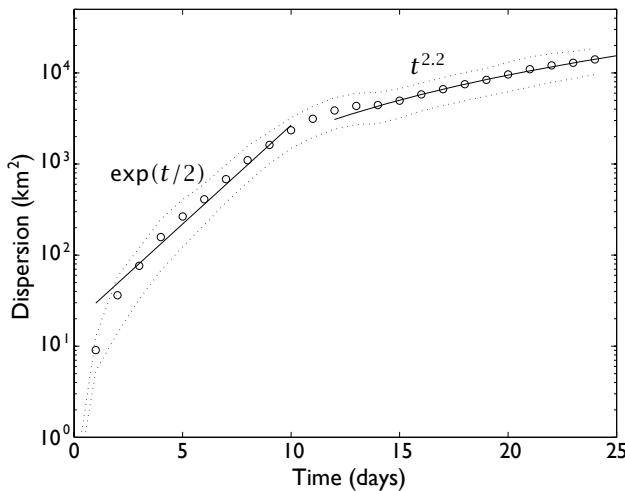


Fig. 10.4 Relative dispersion (the mean-square separation) for 140 drifter pairs as a function of time. The analysis utilizes all drifter pairs which come within 1 km of each other during their lifetimes.

separation expected for an energy inverse cascade using (10.47).⁸ The boundary for the two regimes occurs at about 75 km, which is similar to the first deformation radius. Note that no late-time diffusive regime (where the dispersion goes like t) is observed, suggesting that there exist long-time drifter correlations; these correlations arise because the separation of the drifters is never significantly larger than the energy-containing scale of the eddies themselves.

In the atmosphere similar exponential separation of pairs of drifting balloons in the stratosphere at scales of less than 1000 km has been seen, consistent with an enstrophy inertial range. Evidence of a t^3 separation at larger scales, consistent with an energy inverse cascade, has been less forthcoming.⁹

10.4 MIXING LENGTH THEORY

The discussion of the previous two sections deals with the dispersion of marked fluid parcels. However, both for practical and fundamental reasons, we would like to be able to represent the turbulent transport of a fluid property in an Eulerian form. Thus, consider the equation for a conserved quantity ϕ in an incompressible turbulent flow

$$\frac{D\bar{\phi}}{Dt} = -\nabla \cdot (\bar{\mathbf{v}'\phi'}) + \kappa \nabla^2 \bar{\phi}, \quad (10.56)$$

where κ is the molecular diffusivity and the overbar denotes some kind of averaging, perhaps an ensemble average over realizations with different initial conditions. (We also adopt the convention that, unless noted, whenever the material derivative written as D/Dt is applied to an averaged field, the advection is by the averaged velocity only.) We expect the transport of ϕ to be enhanced by the turbulent flow and, as we saw in the previous sections, in some circumstances this transport will have a diffusive nature, completely overwhelming the molecular diffusivity. Let us consider this from an Eulerian angle, and by analogy with molecular mixing.

Given the mean distribution $\bar{\phi}(x, y, z)$, let a fluid parcel be displaced from its mean

position by a turbulent fluctuation. Suppose that the displaced parcel of fluid is able to carry its initial properties a distance l' before mixing with its surroundings. Then just prior to mixing with the environment the fluctuation of ϕ is given by, in the one-dimensional case,

$$\phi' = -l' \frac{\partial \bar{\phi}}{\partial x} - \frac{1}{2} l'^2 \frac{\partial^2 \bar{\phi}}{\partial x^2} + \mathcal{O}(l'^3). \quad (10.57)$$

If the mean gradient is varying on a space-scale, L , that is larger than the mixing length l' , that is if

$$L \equiv \frac{|\partial \bar{\phi} / \partial x|}{|\partial^2 \bar{\phi} / \partial x^2|} \gg l', \quad (10.58)$$

then we can neglect terms in l'^2 and higher. The turbulent flux of ϕ -stuff is then given by

$$F = \overline{u' \phi'} = -\overline{u' l'} \frac{\partial \bar{\phi}}{\partial x}. \quad (10.59)$$

In more than one dimension, we have

$$F = F_i = -\overline{v'_i l'_j \partial_j \bar{\phi}} = -K_{ij} \partial_j \bar{\phi}, \quad (10.60)$$

with summation over repeated indices, where $K_{ij} \equiv \overline{v'_i l'_j}$. The quantity K_{ij} (which we also write as \mathbf{K}) is known as the eddy (or turbulent) diffusivity tensor. At high Reynolds number it is a property of the flow rather than the fluid itself but, supposing that it can somehow be determined, the equation for the mean value of ϕ becomes

$$\frac{D\bar{\phi}}{Dt} = \nabla \cdot (\mathbf{K} \nabla \bar{\phi}) = \partial_i (K_{ij} \partial_j \bar{\phi}), \quad (10.61)$$

neglecting molecular diffusion.

Suppose that there exists a coordinate system in which the displacements in one direction ($l'^{\hat{x}}$, displacements in the \hat{x} -direction) are not correlated with the fluctuating velocity in another, orthogonal, direction (v' , the velocity in the \hat{y} -direction) and for simplicity we restrict ourselves to two dimensions. Then, in that coordinate system \mathbf{K} is symmetric and

$$\mathbf{K} = \begin{pmatrix} \overline{u' l'^{\hat{x}}} & \overline{u' l'^{\hat{y}}} \\ \overline{v' l'^{\hat{x}}} & \overline{v' l'^{\hat{y}}} \end{pmatrix} = \begin{pmatrix} \overline{u' l'^{\hat{x}}} & 0 \\ 0 & \overline{v' l'^{\hat{y}}} \end{pmatrix}. \quad (10.62)$$

The tensor may then, if needs be, be rotated to some other Cartesian coordinate system, but it will remain a symmetric tensor. In isotropic flow the two diagonal entries are equal and the equation of motion is,

$$\frac{D\bar{\phi}}{Dt} = \nabla \cdot (K \nabla \bar{\phi}), \quad (10.63)$$

which is identical to the equation with molecular diffusion, save that the eddy diffusivity scalar, K , is different from the molecular diffusivity. To the extent, then, that K_{ij} is a symmetric tensor with constant entries, then the turbulence acts like an enhanced diffusion. If the flow is homogeneous, then K does not vary spatially.

10.4.1 Requirements for turbulent diffusion

Turbulent diffusion evidently is a tractable and rational approach for parameterizing the effects of turbulent transport.¹⁰ However, the underlying assumptions are not always satisfied and the derivation itself is rather heuristic, and turbulent diffusion is in no way a fundamental solution to the turbulence closure problem. Nonetheless, it can be an extremely useful parameterization in the appropriate circumstances, these being that:

- (i) there should be a scale separation between the mean gradient and the maximum mixing length, and the mixing length and decorrelation time scale should be well-defined.
- (ii) the diffused property ϕ should be a semi-materially conserved quantity; that is, materially conserved except for the effects of mixing (the next item).
- (iii) The diffused property ϕ should be able to *mix* with its environment.

These are all largely self-evident from the derivation, but items (ii) and (iii) deserve more discussion, as follows.

(ii) Material conservation of tracer

We assumed that a parcel of fluid carries its value of ϕ a distance, on average, equal to its mixing length before irreversibly mixing with its environment; this assumption is necessary in order that one may write $\phi' = -l' \partial \bar{\phi} / \partial x$. If ϕ is not materially-conserved over this scale other terms enter this formula. In particular, momentum is affected by the pressure force, and so is not normally a good candidate for turbulent diffusion. Potential vorticity is a better candidate, because it is a true material invariant, save for dissipative terms, and in large-scale geophysical flows potential vorticity also contains much of the information about the flow.

(iii) Tracer mixing and turbulent cascades

If a parcel cannot mix with its surroundings, then turbulent mixing cannot take place at all. Instead, we have what might be called turbulent stirring and if ϕ were, say, a dye then it would merely become threaded through the environment, producing streaks and swirls of colour rather than a truly mixed fluid. As another example, let ϕ be temperature and suppose that it has a mean gradient, so that temperature falls in the direction of increasing y . Now, we intuitively expect that turbulence can be a much more efficient transporter of heat than molecular diffusion. However, if a displaced parcel of fluid does not mix with or assume the value of its new environment at some stage, then there will be no correlation between the velocity producing the displacement and the value of the fluctuating quantity ϕ' . Suppose, for example, that an eddy causes parcels to be displaced from their mean positions. If a displaced parcel mixes with its surroundings, then a correlation will develop between v' and ϕ' , and we would have $\bar{v'}\bar{\phi'} \neq 0$. However, if no mixing occurs, then the eddy simply recirculates with eddies retaining their initial values, and $\bar{v'}\bar{\phi'}$ is zero because of a lack of correlation between the two quantities. Thus, it is essential that there be a degree of *irreversibility* to the flow in order for turbulent diffusion to be appropriate.

Molecular diffusion is not the only process that enables an eddy to assume the value of its surroundings — a Newtonian or other relaxation back to a specified temperature may have much the same effect. Indeed, in the atmosphere a displaced parcel will be subject to a radiation field that acts in qualitatively this way. For example, suppose that the temperature

equation is

$$\frac{DT}{Dt} = -\lambda(T - T^*(y, z)), \quad (10.64)$$

where the right-hand side crudely represents radiative effects via a relaxation back to a specified profile. Then a displaced parcel will be subject to a radiative damping that is different from that at its initial position, and this will allow the parcel to take on the value of its surroundings, and so potentially enable turbulent diffusion to occur (provided λ is small so that T is approximately materially conserved).

For molecular diffusion to be the mechanism whereby a parcel mixes with its surroundings, the turbulence must create scales that are small enough for diffusion to act. This means that turbulence must create a cascade of ϕ -stuff to small scales. This is quite consistent with the notion that ϕ is a materially conserved quantity, because a scalar field ϕ that satisfies

$$\frac{D\phi}{Dt} = F + \kappa \nabla^2 \phi, \quad (10.65)$$

where F might represent a spectrally local source of variance of ϕ , is certainly cascaded to smaller scales. The presence of a molecular diffusion does not substantially affect the requirement that ϕ be conserved on parcels, because on scales comparable to the eddy mixing length the effect of molecular diffusion is negligible. (And if it were not, perhaps because κ were extremely large or because the turbulence were anaemic, we would not be particularly interested in the turbulent transport.)

10.4.2 A macroscopic perspective

Consider turbulent diffusion from a more macroscopic point of view, and in particular consider the transport of a nearly materially conserved tracer obeying

$$\frac{D\phi}{Dt} = D, \quad (10.66)$$

where the advecting flow is incompressible and D is a dissipative process such that $\overline{D\phi} \leq 0$ (a conventional harmonic diffusion has this property). By decomposing the fields into mean and eddy components in the usual way an equation for the evolution of the tracer variance can be straightforwardly derived, namely

$$\frac{1}{2} \frac{\partial}{\partial t} \overline{\phi'^2} + \overline{\mathbf{v}'\phi'} \cdot \nabla \overline{\phi} + \frac{1}{2} \overline{\mathbf{v}} \cdot \nabla \overline{\phi'^2} + \frac{1}{2} \nabla \cdot \overline{\mathbf{v}'\phi'^2} = \overline{D'\phi'}, \quad (10.67)$$

and where we may assume $\overline{D'\phi'} < 0$. If the mean flow is small and if the third-order term may be neglected then in a statistically steady state we have

$$\overline{\mathbf{v}'\phi'} \cdot \nabla \overline{\phi} \approx \overline{D'\phi'} < 0. \quad (10.68)$$

Therefore, on average, the flux of ϕ is downgradient in regions of dissipation, implying a positive average eddy diffusivity, and a balance is maintained between the downgradient flux of ϕ (which increases the variance) and dissipation. However, it should also be clear from (10.67) that if the turbulence is not statistically stationary, or if there is a mean flow,

then downgradient transport cannot necessarily be expected. Indeed, the transport may be upgradient in regions where the eddy variance is falling, for then we may have the balance

$$\overline{\mathbf{v}'\phi'} \cdot \nabla \overline{\phi} \approx -\frac{1}{2} \frac{\partial}{\partial t} \overline{\phi'^2} > 0. \quad (10.69)$$

10.5 HOMOGENIZATION OF A SCALAR THAT IS ADVECTED AND DIFFUSED

Let us now assume that the effects of turbulence on a tracer are indeed diffusive. An important consequence of this is that, in the absence of additional forcing, there can be no extreme values of the tracer in the interior of the fluid and, in some circumstances, the diffusion will *homogenize* values of the tracer in broad regions. In this section we demonstrate and explore these properties.

10.5.1 Non-existence of extrema

Consider a tracer that obeys the equation

$$\frac{D\phi}{Dt} = \nabla \cdot (\kappa \nabla \phi) + S, \quad (10.70)$$

where $\kappa > 0$ and the advecting velocity is divergence-free. We now show that in regions where the source term, S , is zero there can be no interior extrema of ϕ if the flow is steady. The proof is in the form of a *reductio ad absurdum* argument — we first suppose there *is* an extrema of ϕ in the fluid, and prove a contradiction.

Given an extremum, there will then be a surrounding surface (in three dimensions), or a surrounding contour (in two), connecting constant values of ϕ . For definiteness consider two-dimensional incompressible flow for which the steady flow satisfies

$$\nabla \cdot (\mathbf{u}\phi) = \nabla \cdot (\kappa \nabla \phi). \quad (10.71)$$

Integrating the left-hand side over the area, A , enclosed by the iso-line of ϕ , and applying the divergence theorem, gives

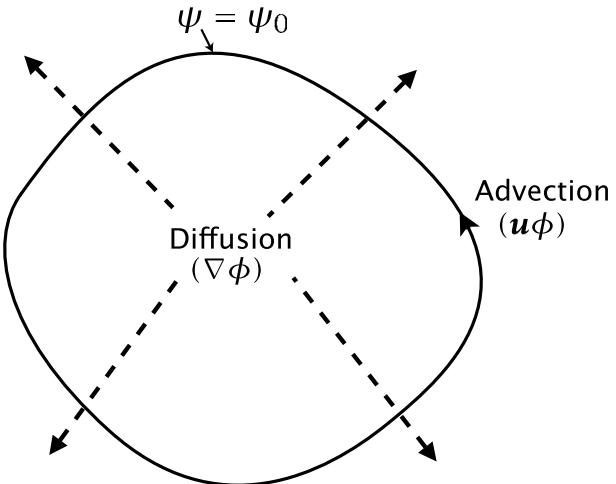
$$\iint_A \nabla \cdot (\mathbf{u}\phi) dA = \oint (\mathbf{u}\phi) \cdot \mathbf{n} dl = \phi \oint \mathbf{u} \cdot \mathbf{n} dl = \phi \iint_A \nabla \cdot \mathbf{u} dA = 0. \quad (10.72)$$

where \mathbf{n} is a unit vector normal to the contour. (If we were to integrate over an area bounded by a velocity contour, then $\mathbf{u} \cdot \mathbf{n} = 0$ and the integral would similarly vanish.) But the integral of the right-hand side of (10.71) over the same area is non-zero; that is

$$\iint_A \nabla \cdot (\kappa \nabla \phi) dA = \oint \kappa \nabla \phi \cdot \mathbf{n} dl \neq 0, \quad (10.73)$$

if the integral surrounds an extremum. This is a contradiction for steady flow. Hence, there can be no isolated extrema of a conserved quantity in the interior of a fluid, if there is any diffusion at all. The result (which applies in two or three dimensions) is kinematic, in that ϕ can be any tracer at all, active or passive. The physical essence of the result is the integrated effects of diffusion are non-zero surrounding an extremum, and cannot be balanced by advection. Thus, if the initial conditions contain an extremum, diffusion will smooth away the extremum until it no longer exists. This process, and the homogenization discussed below, are illustrated in Fig. 10.5.

Fig. 10.5 If an extremum of a tracer ϕ exists in the fluid interior, then diffusion will provide a down-gradient tracer flux. Over an area bounded by a streamline, or by an isoline of ϕ , the net advective flux is zero. Thus, the diffusion cannot be balanced by advection, and in a steady state no extrema can exist.



10.5.2 Homogenization in two-dimensional flow

For two-dimensional flow we can obtain a still stronger result if we allow ourselves to make more assumptions about the strength and nature of the diffusion. The steady distribution of a scalar quantity being advected by an incompressible flow is governed by

$$J(\psi, \phi) = \nabla \cdot (\kappa \nabla \phi) + S, \quad (10.74)$$

where the terms on the right-hand side represent diffusion and source terms. Suppose that these terms are small, in the sense that the individual terms on the left-hand side nearly balance each other, so that

$$|J(\psi, \phi)| \ll \frac{U\Phi}{L}. \quad (10.75)$$

This means we are in the high Peclet number limit ($P = UL/\kappa \gg 1$), and the dominance of advection suggests that any steady solution to (10.74) is of the form

$$\phi = G(\psi) + \mathcal{O}(P^{-1}), \quad (10.76)$$

where G is (for the moment) any function of its argument. Thus, isolines of ϕ are nearly coincident with streamlines, and

$$\nabla \phi \approx \nabla \psi \frac{d\phi}{d\psi}. \quad (10.77)$$

On integrating (10.74) over the area, A , bounded by some closed streamline, $\psi = \psi_0$ say, the left-hand side vanishes and we obtain

$$0 = \iint_A S \, dA + \oint_{\psi_0} \kappa \nabla \phi \cdot \mathbf{n} \, dl. \quad (10.78)$$

Using (10.77) then gives

$$\iint_A S \, dA = - \oint_{\psi_0} \kappa \frac{d\phi}{d\psi} \nabla \psi \cdot \mathbf{n} \, dl. \quad (10.79)$$

Since $d\phi/d\psi$ is constant along streamlines, and using $\mathbf{u} = \nabla^\perp \psi$, we have

$$\frac{d\phi}{d\psi} = - \frac{\iint S \, dA}{\oint_{\psi_0} \kappa \mathbf{u} \cdot d\mathbf{l}}. \quad (10.80)$$

This relationship determines ϕ as a function of ψ — that is, it determines $G(\psi)$ — in terms of the forcing and dissipation acting on the fluid. If the fluid is both unforced and inviscid, then a steady solution pertains when ϕ is an *arbitrary* function of ψ . If the source term S is zero, but dissipation is non-zero then the denominator of (10.80) is non-zero and therefore ϕ must be uniform: ϕ has been *homogenized*. The homogenization result also follows if we choose to integrate over an area surrounded by an isoline of ϕ , ϕ_0 say, but we leave this as a problem for the reader.

Interpretation

The homogenization result applies to a statistically steady flow in which the eddy transport of ϕ -stuff by the eddying motion may be parameterized diffusively, and in which there is an approximate functional relationship between mean ϕ and mean ψ . The first of these assumptions we have discussed at length in previous sections. The second requires that the diffusion must not be too strong, so that locally the tracer is conserved on fluid parcels. In the steady state the tracer is then a function of the streamfunction, the same function everywhere within the closed region.

Given these assumptions, the dynamics giving rise to homogenization is transparent: integrating round a contour of ψ or ϕ the effect of the advective terms vanish; the source (S) and the diffusion must balance each other, and if there is no source term there can be no tracer gradient. Put another way, the flow will circulate endlessly and steadily around the contours of ψ , which nearly coincide with contours of ϕ . Advection cannot alter the mean value of ϕ , so diffusion smooths out gradients within the closed contours, effectively *expelling* gradients of ϕ to the boundaries and forming a plateau of ϕ -values. Because extrema of ϕ are forbidden, the value of ϕ on the plateau cannot be a maximum or minimum: at the edge of the plateau the values of ϕ must fall somewhere, and rise somewhere else. The plateau can be a flat region etched out of a hillside, but a plateau on top of a butte is forbidden, for in that case diffusion would erode the butte down to the level of the surrounding land. Our derivation makes no distinction between a passive scalar like a dye and an active scalar, like potential vorticity. In reality, in the latter case the dynamics will further constrain the flow because the scalar distribution must be consistent with the velocity field that advects it, and this is particularly important in the dynamics of ocean gyres.

10.6 † TRANSPORT BY BAROCLINIC EDDIES

In the next few sections we discuss the transport of fluid properties by large-scale eddies typically generated by baroclinic instability — mesoscale eddies in the ocean, and weather systems in the atmosphere. Our motivation is twofold.

- (i) Mesoscale eddies in the ocean cannot be easily resolved in numerical models of its large-scale circulation, especially those used for climate simulations involving integrations of the global ocean over decades and centuries. In such models, the effects of eddies must be *parameterized* in terms of properties of the mean flow.

- (ii) Eddy-mean flow interaction is simply a key part of the dynamics of the atmosphere and the ocean.

The first of these will, before too long, be solved for us by the increasing power of computers, and of course in the atmosphere numerical models of the general circulation already resolve most of the effects of baroclinic eddies. However, as computer power increases and we resolve mesoscale eddies, item (i) is replaced with the hardly less difficult problem of understanding those massive, turbulent, numerical integrations, and for that task a theory of turbulent transport is a *sine qua non*.

10.6.1 Symmetric and antisymmetric diffusivity tensors

A tracer evolving freely save for the effects of molecular diffusivity obeys the equation

$$\frac{D\phi}{Dt} = \nabla \cdot (\kappa_m \nabla \phi), \quad (10.81)$$

where κ_m is the molecular diffusivity, a positive scalar quantity. In the more general case we have

$$\frac{D\phi}{Dt} = -\nabla \cdot \mathbf{F} = \nabla \cdot \mathbf{K} \nabla \phi, \quad (10.82)$$

where \mathbf{K} is (if ϕ is a scalar) a second-rank tensor and $\mathbf{F} = -\mathbf{K} \nabla \phi$ is the diffusive flux of ϕ . (We will omit the identifying field in square brackets if no ambiguity is so caused.) The flux has a component across the isosurfaces of ϕ , called the diffusive flux, and a component along the iso-surfaces, called the skew flux [cf. (7.91) and (7.92)]. We will see below that these fluxes are associated with the symmetric and antisymmetric components of the diffusivity tensor, respectively, where

$$\mathbf{K} = \mathbf{S} + \mathbf{A}, \quad (10.83)$$

and, using component notation,

$$S_{mn} = \frac{1}{2}(K_{mn} + K_{nm}), \quad A_{mn} = \frac{1}{2}(K_{mn} - K_{nm}). \quad (10.84)$$

The diagonal elements of the antisymmetric tensor are zero. The transport that is effected by these two tensors has different physical characteristics, as we now discuss.

10.6.2 * Diffusion with the symmetric tensor

In the simplest case of all, with an isotropic medium, \mathbf{K} is diagonal with equal entries,

$$\mathbf{K} = \mathbf{S} = \begin{pmatrix} \kappa & 0 & 0 \\ 0 & \kappa & 0 \\ 0 & 0 & \kappa \end{pmatrix}, \quad (10.85)$$

and we have the familiar $\mathbf{F} = -\kappa \nabla \phi$, and (10.82) has the same form as (10.81). If κ is positive, then the flux is *downgradient*, meaning that

$$\mathbf{F} \cdot \nabla \phi < 0, \quad (10.86)$$

even if κ is spatially non-uniform. Furthermore, such a diffusion is variance-dissipating; to see this, suppose we have the equation of motion

$$\frac{D\phi}{Dt} = \nabla \cdot (\kappa \nabla \phi). \quad (10.87)$$

Multiplying by ϕ and integrating over the domain V gives

$$\frac{1}{2} \frac{d}{dt} \int_V \phi^2 dV = \int_V \mathbf{F} \cdot \nabla \phi dV = - \int_V \kappa (\nabla \phi)^2 dV \leq 0, \quad (10.88)$$

after an integration by parts and assuming that the normal derivative of ϕ vanishes at the boundaries; that is, there is no flux of ϕ -stuff through the boundary. However, diffusion does preserve the first moment of the field; that is

$$\frac{d}{dt} \int_V \phi dV = \int_V \nabla \cdot (\kappa \nabla \phi) dV = 0, \quad (10.89)$$

again assuming no flux through the boundaries.

The transport that is effected by the symmetric diffusion tensor is the diffusive flux, \mathbf{F}_d , where

$$\mathbf{F}_d = -\mathbf{S} \nabla \phi = -S_{mn} \partial_n \phi, \quad (10.90)$$

where we employ the common convention that repeated indices are summed. In general, the flux has a component that is parallel to the tracer gradient; that is, $\mathbf{F}_d \cdot \nabla \phi \neq 0$. Suppose, for simplicity, we have the equation of motion

$$\frac{\partial \phi}{\partial t} = -\nabla \cdot \mathbf{F}_d = \nabla \cdot (\mathbf{S} \nabla \phi). \quad (10.91)$$

This equation preserves the first moment of ϕ , provided there is no flux through the boundary. Tracer variance evolves according to

$$\frac{1}{2} \frac{\partial}{\partial t} \int_V \phi^2 dV = \int_V \phi \nabla \cdot (\mathbf{S} \nabla \phi) = - \int_V (\mathbf{S} \nabla \phi) \cdot \nabla \phi dV. \quad (10.92)$$

This can be shown to be negative or zero, provided that \mathbf{S} is positive semi-definite, meaning that

$$\nabla \phi \mathbf{S} \nabla \phi = \partial_m \phi S_{mn} \partial_n \phi \geq 0. \quad (10.93)$$

The flux effected by such a diffusivity is then downgradient in the sense that

$$\mathbf{F}_d \cdot \nabla \phi = -\mathbf{S} \nabla \phi \cdot \nabla \phi \leq 0. \quad (10.94)$$

10.6.3 * The skew flux

The transport associated with the antisymmetric transport tensor is perpendicular to the gradient of ϕ , and so is neither upgradient nor downgradient. The flux is

$$\mathbf{F}_{sk} = -\mathbf{A} \nabla \phi = -A_{mn} \partial_n \phi \quad (10.95)$$

and thus

$$\mathbf{F}_{sk} \cdot \nabla \phi = -\mathbf{A} \nabla \phi \cdot \nabla \phi = -A_{mn} \partial_n \phi \partial_m \phi = 0, \quad (10.96)$$

where the final result follows because of the antisymmetry of \mathbf{A} — the contraction of a symmetric tensor and an antisymmetric tensor is zero.¹¹ For this reason, the associated transport is known as a *skew flux* (a term applying in general to fluxes that are perpendicular to the tracer gradient) or a *skew diffusion* (when those fluxes are parameterized using an antisymmetric diffusivity). It follows from this that if a tracer obeys

$$\frac{\partial \phi}{\partial t} = \nabla \cdot (\mathbf{A} \nabla \phi), \quad (10.97)$$

then the tracer variance is conserved. This may be verified by multiplying this equation by ϕ and integrating by parts, assuming that the flux vanishes at the boundaries. That is, a *skew diffusion has no effect on the variance of the skew diffused variable*. One other familiar physical process shares these properties, and that is advection by a divergence-free flow. A skew diffusion is physically equivalent to such an advection in that the divergence of a skew diffusive flux is the same as the divergence of an appropriately chosen advective flux. We proved this in section 7.3.4, but let us look at it again from the point of view of diffusion.

We define an advective flux of a tracer ϕ to be a flux of the form

$$\mathbf{F}_{ad} \equiv \tilde{\mathbf{v}} \phi, \quad (10.98)$$

where $\tilde{\mathbf{v}}$ is a divergence-free vector field. The divergence of the flux is just

$$\nabla \cdot \mathbf{F}_{ad} = \nabla \cdot (\tilde{\mathbf{v}} \phi) = \tilde{\mathbf{v}} \cdot \nabla \phi. \quad (10.99)$$

The field $\tilde{\mathbf{v}}$ might be called a *pseudovelocity* or a *quasi-velocity* — it acts like a velocity but is not necessarily the velocity of any fluid particle. Because $\tilde{\mathbf{v}}$ is divergence-free, we may, as in section 7.3.4, define a vector streamfunction $\boldsymbol{\psi}$ such that

$$\tilde{\mathbf{v}} = \nabla \times \boldsymbol{\psi}. \quad (10.100)$$

The field $\boldsymbol{\psi}$ is not unique: the gradient of an arbitrary function may be added to it, this gradient vanishing upon taking the curl, resulting in the same velocity field. That is, if $\boldsymbol{\psi}' = \boldsymbol{\psi} + \nabla \gamma$, then $\tilde{\mathbf{v}} = \nabla \times \boldsymbol{\psi} = \nabla \times \boldsymbol{\psi}'$. The scalar field γ is known as the *gauge*, and the freedom to choose it is the gauge freedom.

The advective flux \mathbf{F}_{ad} is related to the skew flux \mathbf{F}_{sk} by

$$\phi \tilde{\mathbf{v}} = \phi \nabla \times \boldsymbol{\psi} = \nabla \times (\phi \boldsymbol{\psi}) - \nabla \phi \times \boldsymbol{\psi}, \quad (10.101)$$

or

$$\mathbf{F}_{ad} = \mathbf{F}_r + \mathbf{F}_{sk}, \quad (10.102)$$

where $\mathbf{F}_r = \nabla \times (\phi \boldsymbol{\psi})$ is a rotational flux with no divergence, and

$$\mathbf{F}_{sk} = -\nabla \phi \times \boldsymbol{\psi} \quad (10.103)$$

is a skew flux — ‘skewed’ because it is manifestly orthogonal to the gradient of ϕ , i.e., $\nabla \phi \cdot \mathbf{F}_{sk} = 0$. Because $\nabla \cdot \mathbf{F}_r = 0$ we have

$$\nabla \cdot \mathbf{F}_{ad} = \nabla \cdot \mathbf{F}_{sk}. \quad (10.104)$$

However, the skew flux, $-\nabla \phi \times \boldsymbol{\psi}$, and the advective flux, $\phi \nabla \times \boldsymbol{\psi}$, may have, and in

general do have, different magnitudes and directions; only their divergences are equal. If the divergence of the skew fluxes given by (10.95) and (10.103) are to be the same then ψ must be related to the antisymmetric tensor A . Using (10.95) we have

$$\begin{aligned}\nabla \cdot \mathbf{F}_{sk} &= -\partial_m(A_{mn}\partial_n\phi) \\ &= -(\partial_n\phi)(\partial_mA_{mn}) - [A_{mn}\partial_n\partial_mA\phi] \\ &= -\partial_n(\phi\partial_mA_{mn}) + [\phi\partial_n\partial_mA_{mn}],\end{aligned}\quad (10.105)$$

where the quantities in square brackets are zero as a consequence of the antisymmetry of A — a symmetric operator acting on an antisymmetric tensor is zero. But the skew flux divergence is equal to the advective flux divergence

$$\nabla \cdot \mathbf{F}_{sk} = \nabla \cdot \mathbf{F}_{ad} = \partial_n(\phi\tilde{v}_n), \quad (10.106)$$

so the associated skew velocity is given by

$$\tilde{v}_n = -\partial_mA_{mn}, \quad (10.107)$$

and this is divergence-free because $\partial_n\partial_mA_{mn} = 0$. The streamfunction and the antisymmetric tensor are thus related, because

$$\tilde{v}_n = \epsilon_{lmn}\partial_l\psi_m. \quad (10.108)$$

This expression is equivalent to $\tilde{\mathbf{v}} = \nabla \times \psi$, as can be verified by expanding the expression out in Cartesian coordinates. The Levi-Civita symbol ϵ_{mnp} is such that $\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1$, $\epsilon_{132} = \epsilon_{321} = \epsilon_{213} = -1$, and $\epsilon_{mnp} = 0$ for other combinations.

Using (10.107) and (10.108), and just a little algebra, gives

$$A_{mn} = \epsilon_{mnp}\psi_p = \begin{pmatrix} 0 & \psi_3 & -\psi_2 \\ -\psi_3 & 0 & \psi_1 \\ \psi_2 & -\psi_1 & 0 \end{pmatrix}. \quad (10.109)$$

This expression provides a connection between the antisymmetric tensor A_{mn} and the streamfunction for the skew velocity $\tilde{\mathbf{v}}$.

10.6.4 The story so far

In this chapter we are making a rather large assumption — that we can parameterize turbulent fluxes via some form of diffusive parameterization, but given that let us be reasonably careful how we do it. Thus far, we have established the following.

- ★ Any flux can be decomposed into a component across iso-surfaces of a scalar (the along-gradient or diffusive flux) and a component along isosurfaces and so perpendicular to the gradient (the skew flux).
- ★ The along-gradient (usually downgradient) flux is effected by a diffusion using a symmetric diffusivity tensor.
- ★ The skew flux is effected by a diffusion using an antisymmetric diffusivity tensor, and this is equivalent to an advection by some divergence-free velocity.
- ★ The diffusive flux reduces tracer variance if the diffusivity is positive (in which case the diffusion is downgradient), whereas the skew diffusion has no effect on variance.

Let us now consider how all of this is relevant to the large-scale flow in the atmosphere and the ocean.

10.7 † EDDY DIFFUSION IN THE ATMOSPHERE AND OCEAN

Ah, but a man's reach should exceed his grasp, Or what's a heaven for?

Robert Browning, *Andrea del Sarto*, 1855.

10.7.1 Preliminaries

Consider a tracer that obeys the equation

$$\frac{D\phi}{Dt} = \nabla \cdot (\kappa_m \nabla \phi), \quad (10.110)$$

where κ_m is the molecular diffusivity. If the advecting flow is divergence-free then the ensemble average flow obeys, neglecting the molecular diffusion,

$$\frac{D\bar{\phi}}{Dt} = -\nabla \cdot \overline{\mathbf{v}'\phi'}. \quad (10.111)$$

We can prepare the ensemble by, for example, varying the initial conditions slightly; if the flow is chaotic, the realizations will differ from each other. If we parameterize the eddy transport by a diffusion then

$$\overline{\mathbf{v}'\phi'} = -\mathbf{K}\nabla\bar{\phi}, \quad (10.112)$$

where \mathbf{K} is, in general, a second-rank tensor. If the average is a zonal average then

$$\frac{D\bar{\phi}}{Dt} = -\frac{\partial \overline{v'\phi'}}{\partial y} - \frac{\partial \overline{w'\phi'}}{\partial z}. \quad (10.113)$$

If we are to employ a diffusive parameterization for the eddy terms in these equations, the issues that then arise fall into two general camps:

- (i) the overall *magnitude* of the eddy diffusivity, possibly as a function of the mean flow;
- (ii) the *structure* of the diffusivity tensor, and in particular the separate structure of its symmetric and antisymmetric parts.

10.7.2 Magnitude of the eddy diffusivity

The magnitude and scale of eddies were previously considered in sections 9.3 and 9.4. Here we see how these give rise to corresponding estimates of the magnitude of an eddy diffusivity. If we restrict attention for the moment to the meridional transfer of tracer properties, then we might write

$$\overline{v'\phi'} = -\kappa^{yy} \frac{\partial \bar{\phi}}{\partial y} - \kappa^{yz} \frac{\partial \bar{\phi}}{\partial z}, \quad (10.114)$$

where κ^{yy} and κ^{yz} are components of the eddy diffusivity tensor. These components have the dimensions of a length times a velocity and, to the extent that the diffusion represents the eddying motion we expect that κ^{yy} has an approximate magnitude of

$$\kappa^{yy} \sim v'l', \quad (10.115)$$

where v' is a typical magnitude of the horizontal eddy velocity, and l' is *mixing length* of the eddies, generally taken to be a typical length scale of the eddies. Larger and more energetic

eddies thus have a larger effect on the mean flow. We can estimate ν' and l' in a number of reasonable ways depending on the flow conditions, and we consider a few such below.¹² The magnitude of the component κ^{yy} may then be estimated by making choices about the plane of parcel displacements, and this is considered in the next section.

Perhaps the simplest assumption to make follows from the fact that the eddies are a consequence of baroclinic instability, and so one might suppose that the eddy length scale is the scale of the instability — the first deformation radius. One might also suppose that the eddy velocity is of the same approximate magnitude as the mean flow, \bar{u} , thus giving

$$\kappa^{yy} \sim L_d \bar{u} = \frac{NH\bar{u}}{f}. \quad (10.116)$$

Another way of deriving this result is by noting that $\kappa^{yy} \sim l'^2/T_e$, where T_e is a characteristic eddy time scale. If T_e is the Eady time scale, L_d/\bar{u} , and if $l' \sim L_d$, we reproduce (10.116). Note also that (10.116) may be written as

$$\kappa^{yy} \sim L_d \bar{u} \sim \frac{L_d^2 f}{\sqrt{Ri}} \sim L_d^2 Fr f, \quad (10.117)$$

where $Ri \equiv N^2/\Lambda^2 = N^2 H^2/\bar{u}^2$ and $Fr \equiv U/(NH)$ are the Richardson and Froude numbers for this problem, respectively.

A little more generally, if there is a cascade to larger scales then the eddy scale, L_e say, may be larger than the deformation scale. Depending on circumstances, L_e might be the domain scale (if eddies grow to the size of the domain), the β -scale (if the β -effect halts the cascade), or some scale determined by frictional effects (possibly in conjunction with β). However, the arguments of section 9.3 suggest that the eddy time scale is the Eady time scale in all cases. We therefore have:

$$\text{eddy length scale} \sim L_e \quad (10.118a)$$

$$\text{eddy time scale, } T_e \sim L_d/\bar{u} \quad (10.118b)$$

$$\text{eddy velocity scale, } U_e \sim \bar{u}(L_e/L_d). \quad (10.118c)$$

These give the general estimate for the horizontal diffusivity of

$$\kappa^{yy} \sim \bar{u} \left(\frac{L_e^2}{L_d} \right). \quad (10.119)$$

The estimate (10.116) is a special case of this, with $L_e = L_d$; the two estimates will thus differ if the eddy scale is much larger than the deformation radius.

In the case in which the inverse cascade is modified by the Rossby waves we might (and neglecting friction effects; see section 9.1.3) suppose that the eddy scale is the β -scale, (9.6), and we have

$$L_e \sim L_\beta = \left(\frac{U_e}{\beta} \right)^{1/2} = \frac{\bar{u}}{\beta L_d}, \quad (10.120)$$

using (10.118c). The eddy velocity scale is, using (10.118b),

$$U_e \sim \bar{u} \frac{L_e}{L_d} = \frac{\bar{u}^2}{\beta L_d^2}, \quad (10.121)$$

and combining (10.120) and (10.121) gives the estimate for the eddy diffusivity,

$$\kappa^{yy} \sim \frac{\bar{u}^3}{\beta^2 L_d^3}. \quad (10.122)$$

A similar estimate be written in terms of the inverse energy cascade rate, ε , giving

$$\kappa^{yy} \sim \left(\frac{\varepsilon^3}{\beta^4} \right)^{1/5}. \quad (10.123)$$

This expression may be obtained purely by dimensional analysis, if it is assumed that the only factors determining κ are ε and β . The estimate may be useful if ε is known independently, for example by calculating the energy throughput in the atmosphere.

To summarize: the magnitude of any eddy diffusion may be estimated as the product of the velocity scale and the energy-containing length scale of the eddies. If we assume that the time scale is the Eady time scale we obtain (10.119), where L_e is undetermined. If the eddy scale is the β -scale, then (10.119) becomes (10.122). However, in neither the atmosphere nor the ocean is the β -scale significantly (e.g., an order of magnitude) larger than the deformation scale, but nor, complicating matters, does the inverse cascade necessarily halt at the β -scale (see section 9.1.3). From an observational standpoint, the atmosphere has no $-5/3$ inverse cascade, although there is some evidence for one in some regions of ocean.¹³ In other oceanic regions the eddies may be advected away from each other and away from the unstable zone, or dispersed by Rossby waves, before an inverse cascade can be organized, and energy will remain at the deformation scale. These arguments suggest that although we can make sensible estimates, we cannot determine with certainty what the magnitude of an eddy diffusivity should be, in either the atmosphere or ocean.

10.7.3 * Structure: the symmetric transport tensor

Along-gradient transport, or diffusion, is transport by a symmetric transport tensor, as in (10.90). To explore its structure, let us consider transfer in a re-entrant channel with zonally homogeneous eddy statistics so that the averaging operator is the zonal average; the meridional and upward transport of tracer are then given by

$$\bar{v}'\bar{\phi}' = -\kappa^{yy}\frac{\partial\bar{\phi}}{\partial y} - \kappa^{yz}\frac{\partial\bar{\phi}}{\partial z} \quad (10.124)$$

$$\bar{w}'\bar{\phi}' = -\kappa^{wy}\frac{\partial\bar{\phi}}{\partial y} - \kappa^{wz}\frac{\partial\bar{\phi}}{\partial z}, \quad (10.125)$$

where $\kappa^{wy} = \kappa^{yz}$ by the posited symmetry. The relationship between the various transfer coefficients will be determined by the trajectories of the fluid parcels in the eddying motion. In the Cartesian y - z frame the transport tensor is not necessarily diagonal (i.e., κ^{yz} and κ^{wy} may be non-zero) but locally there is always a natural coordinate system in which the diffusivity tensor is diagonal. (A symmetric matrix may always be diagonalized by a suitable rotation of axes.) In that diagonal frame we can write

$$\mathbf{S}' = \kappa_s \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix}, \quad (10.126)$$

where κ_s determines the overall size of the transfer coefficients (as estimated in the previous section), and α is the ratio of sizes of the components in the two orthogonal directions. Now, fluid displacements in large-scale baroclinic eddies are nearly, but not exactly, horizontal — they may be along isopycnals, for example, or at an angle between the horizontal and the isopycnals. We may argue that the coordinate system in which the tensor is diagonal is the coordinate system defined by the plane along which fluid displacements occur. This is sensible because the transfers along and orthogonal to the fluid paths are each a consequence of different physical phenomena, and so we may expect the transfer tensor to be diagonal in these coordinates.

Because eddy displacements are predominantly horizontal, the diagonal coordinate system has a small slope, s , at an angle θ with respect to the horizontal, where $s = \tan \theta \approx \theta \ll 1$. Furthermore, we expect the parameter α to be small (i.e., $\alpha \ll 1$), because this represents transfer in a direction orthogonal to the eddy fluid motion. We rotate the tensor \mathbf{S}' through an angle θ to move into the usual y - z frame; that is

$$\mathbf{S} = \kappa_s \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad (10.127a)$$

$$\approx \kappa_s \begin{pmatrix} 1 + s^2 \alpha & s(1 - \alpha) \\ s(1 - \alpha) & s^2 + \alpha \end{pmatrix} \quad (\text{for small } s) \quad (10.127b)$$

$$\approx \kappa_s \begin{pmatrix} 1 & s \\ s & s^2 + \alpha \end{pmatrix} \quad (\text{for small } s \text{ and small } \alpha). \quad (10.127c)$$

We can follow the same procedure in three dimensions. Then, if the eddy transport is isotropic in the plane of eddy displacements, the three-dimensional transport tensor is

$$\mathbf{S}' = \kappa_s \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \alpha \end{pmatrix}, \quad (10.128)$$

and the slope of the motion is a two-dimensional vector $\mathbf{s} = (s^x, s^y)$, with the superscripts denoting components. If we rotate the transport tensor into physical space then we obtain, analogously to (10.127),

$$\mathbf{S} = \kappa_s \begin{pmatrix} 1 + s^{y^2} + \alpha s^{x^2} & (\alpha - 1)s^x s^y & (1 - \alpha)s^x \\ (\alpha - 1)s^x s^y & 1 + s^{x^2} + \alpha s^{y^2} & (1 - \alpha)s^y \\ (1 - \alpha)s^x & (1 - \alpha)s^y & \alpha + s^2 \end{pmatrix} \quad (10.129a)$$

$$\approx \kappa_s \begin{pmatrix} 1 & 0 & s^x \\ 0 & 1 & s^y \\ s^x & s^y & \alpha + s^2 \end{pmatrix}, \quad (10.129b)$$

for small s and small α , where $s^2 = s^{x^2} + s^{y^2}$.

The plane of eddy displacements

We are now in a position to make heuristic choices about the transfer coefficients, and we will consider two bases for this.

1. *Using linear baroclinic instability theory.*¹⁴ In a simple model of a growing baroclinic

(Eady) wave, parcel trajectories that are along half the slope of the mean isopycnals are able to release the most potential energy. We thus suppose that $s = s_\rho/2$, where s_ρ is the isopycnal slope, and that $\alpha = 0$ in (10.127c) or (10.129b). In two dimensions this gives

$$\mathbf{S} = \kappa_s \begin{pmatrix} 1 & s_\rho/2 \\ s_\rho/2 & s_\rho^2/4 \end{pmatrix} \quad (10.130)$$

and so

$$\overline{v' \phi'} = -\kappa_s \left(\frac{\partial \overline{\phi}}{\partial y} + \frac{1}{2} s_\rho \frac{\partial \overline{\phi}}{\partial z} \right) \quad (10.131a)$$

$$\overline{w' \phi'} = -\frac{1}{2} \kappa_s s_\rho \left(\frac{\partial \overline{\phi}}{\partial y} + \frac{1}{2} s_\rho \frac{\partial \overline{\phi}}{\partial z} \right). \quad (10.131b)$$

If the tracer ϕ is potential temperature (and not just a passive tracer) then (10.131), along with one of the estimates for the size of κ_s given in section 10.7.2, constitutes a parameterization for the diffusive poleward and upward heat flux in the atmosphere.

II. Flow along neutral surfaces. If the fluid interior is adiabatic and steady, then fluid trajectories are along neutral surfaces; that is, along surfaces of potential density or potential temperature. One might therefore be inclined to assume that the eddy fluxes are aligned along the mean neutral surfaces and choose $s = s_\rho$. However, even in the adiabatic case, this is not always a good choice. From the adiabatic thermodynamic equation $Db/Dt = 0$ we may derive the equation for the eddy buoyancy variance, namely

$$\begin{aligned} \frac{1}{2} \frac{\partial \overline{b'^2}}{\partial t} + \frac{1}{2} \overline{\mathbf{u} \cdot \nabla_z b'^2} + \frac{1}{2} \overline{w \frac{\partial b'^2}{\partial z}} + \overline{\mathbf{u}' b'} \cdot \nabla_z \overline{b} + \overline{w' b'} \frac{\partial \overline{b}}{\partial z} \\ + \frac{1}{2} \nabla_z \cdot \overline{\mathbf{u}' b'^2} + \frac{1}{2} \frac{\partial}{\partial z} \overline{w' b'^2} = 0, \end{aligned} \quad (10.132)$$

and specialize to the case of a zonally uniform basic state and small-amplitude wave. In that case

$$\frac{1}{2} \frac{\partial \overline{b'^2}}{\partial t} = -\overline{v' b'} \frac{\partial \overline{b}}{\partial y} - \overline{w' b'} \frac{\partial \overline{b}}{\partial z}. \quad (10.133)$$

If the wave is statistically steady then the left-hand side is zero and

$$\overline{\mathbf{v}'_m b'} \cdot \nabla_x \overline{b} = 0, \quad (10.134)$$

where the subscript x indicates that the vectors are in the meridional plane, with no variation in x . In this case there is indeed no along-gradient flux. However, if the wave is growing then $\overline{\mathbf{v}'_m b'} \cdot \nabla_x \overline{b} < 0$ and if $\partial \overline{b} / \partial y < 0$ and $\overline{v' b'} > 0$, as in the Northern Hemisphere, then

$$\frac{\overline{w' b'}}{\overline{v' b'}} > -\frac{\partial \overline{b} / \partial y}{\partial \overline{b} / \partial z}, \quad (10.135)$$

and so the mixing slope is *less* steep than the mean isopycnal slope, even though the flow may be adiabatic. (The mixing slope result also holds if $\overline{v' b'} < 0$, although the sign of the inequality is different.) Similarly, if the wave is decaying the mixing slope is steeper than that of the mean isopycnals. In an inhomogeneous flow, the advection

by the mean flow in (10.132) plays a similar role to time dependence: the advection of eddy variance by the mean flow into a region of larger variance will give rise to a mixing slope that is less steep than the isopycnal slope, and conversely for a flow entering a region of less variance. Only for a statistically steady, adiabatic, linear wave field is the mixing slope guaranteed to be along the isopycnals.

Having said all this, let us proceed by assuming that the fluid trajectories are indeed along neutral surfaces. If there is no diffusion orthogonal to this $\alpha = 0$, and the transport tensor is

$$\mathbf{S} = \kappa_s \begin{pmatrix} 1 & 0 & s_\rho^x \\ 0 & 1 & s_\rho^y \\ s_\rho^x & s_\rho^y & |s_\rho|^2 \end{pmatrix}, \quad (10.136)$$

or in two dimensions

$$\mathbf{S} = \kappa_s \begin{pmatrix} 1 & s_\rho \\ s_\rho & s_\rho^2 \end{pmatrix}, \quad (10.137)$$

and in this case

$$\overline{v' \phi'} = -\kappa_s \left(\frac{\partial \overline{\phi}}{\partial y} + s_\rho \frac{\partial \overline{\phi}}{\partial z} \right), \quad \overline{w' \phi'} = -\kappa_s s_\rho \left(\frac{\partial \overline{\phi}}{\partial y} + s_\rho \frac{\partial \overline{\phi}}{\partial z} \right). \quad (10.138a,b)$$

Suppose that ϕ is potential temperature θ , and that surfaces of potential temperature do define neutral surfaces. Then plainly eddy motion along potential temperature surfaces does not transfer potential temperature, and the diffusion defined by (10.138) should have no effect. The equations themselves respect this, for then

$$s_\rho = -\frac{\partial_y \overline{\theta}}{\partial_z \overline{\theta}} \quad (10.139)$$

and using this in (10.138) gives

$$\overline{v' \theta'} = 0, \quad \overline{w' \theta'} = 0. \quad (10.140a,b)$$

In the real ocean the presence of salinity means that the potential temperature, potential density and salinity surfaces are, in general, not parallel and there will be eddy diffusion of θ and S (salinity) along neutral surfaces. However, this does *not* provide a parameterization for the heat flux by baroclinic eddies, because we cannot expect such a flux to depend for its existence on the presence of a second tracer, salinity. For such a parameterization, we turn to the antisymmetric transport tensor.

10.7.4 * Structure: the antisymmetric transport tensor

The antisymmetric transport tensor gives rise to the skew flux, or the pseudoadvection. In two dimensions (one horizontal, one vertical) we can immediately write down its form, namely

$$\mathbf{A} = \begin{pmatrix} 0 & -\kappa'_a \\ \kappa'_a & 0 \end{pmatrix}, \quad (10.141)$$

where the transfer coefficient κ'_a , which may vary in space and time depending on the flow itself, determines the overall strength of the transport. In three dimensions then by inspection we can write

$$\mathbf{A} = A_{ij} = \begin{pmatrix} 0 & 0 & -\kappa'_a x \\ 0 & 0 & -\kappa'_a y \\ \kappa'_a x & \kappa'_a y & 0 \end{pmatrix}, \quad (10.142)$$

where we have used our gauge freedom to choose that $A_{21} = -A_{12} = 0$. Note that (10.142) preserves the form of (10.141) if one of the horizontal dimensions is absent — that is, if either row one and column one, or row two and column two, is eliminated. Our remaining choice is to determine the sign and magnitude of the transport coefficients.

An adiabatic, potential-energy diminishing, eddy transport scheme

A very useful parameterization for the transport of tracers in ocean models by baroclinic eddy fluxes, commonly known as the Gent–McWilliams or GM scheme,¹⁵ can be constructed using the antisymmetric transport tensor. The maintenance of two properties is the foundation of the scheme.

- (i) Moments of the tracer should be preserved; in particular, the amount of fluid between two isopycnal surfaces should be preserved. This suggests the scheme should not diffuse buoyancy across constant-buoyancy surfaces.
- (ii) The amount of available potential energy in the flow should be reduced, so mimicking the effects of baroclinic instability, which transfers available potential energy to kinetic energy.

The first of these is automatically satisfied by using an antisymmetric diffusivity tensor. The second property can be satisfied by choosing the transfer coefficients to be proportional to the slope of the isopycnals, in which case we may write (10.142) as

$$\mathbf{A} = \kappa_a \begin{pmatrix} 0 & 0 & -s^x \\ 0 & 0 & -s^y \\ s^x & s^y & 0 \end{pmatrix}, \quad (10.143)$$

where $\mathbf{s} = (s^x, s^y) = \nabla_\rho z = -\nabla_z \rho / (\partial \rho / \partial z)$ is the isopycnal slope (recall that s^x denotes a component of a vector, and s_x a derivative) and κ_a determines the overall magnitude of the diffusivity. In an ocean model separately carrying temperature and salinity fields, then (10.143) would be applied to each of these, with the isopycnal slope being determined using the equation of state. To more easily see what properties are implied by the transport, let us specialize to the salt-free case, with buoyancy, b , the only thermodynamic variable. The isopycnal slope is then $\mathbf{s} = -(b_x / b_z, b_y / b_z)$ and the horizontal eddy buoyancy transfer $\mathbf{F}_h = (F^x, F^y)$ is given by

$$\mathbf{F}_h = - \left(-\kappa_a \mathbf{s} \frac{\partial b}{\partial z} \right) = -\kappa_a \left(\frac{\partial b}{\partial x}, \frac{\partial b}{\partial y} \right) = -\kappa_a \nabla_z b, \quad (10.144a)$$

which for positive κ_a is the same as conventional downgradient diffusion.

The vertical transfer is given by

$$F^z = -\kappa_a \left(s^x \frac{\partial b}{\partial x} + s^y \frac{\partial b}{\partial y} \right) = \kappa_a s^2 \frac{\partial b}{\partial z}, \quad (10.144b)$$

where $s^2 = \mathbf{s} \cdot \mathbf{s}$. This flux is *up* the vertical gradient; however, by construction, the total skew flux is neither upgradient nor downgradient.

The combination of the downgradient horizontal flux and the upgradient vertical flux acts to reduce the potential energy of the flow at the same time as preserving the volume of fluid within each density interval. The upgradient flux in the vertical is a consequence of the need to reduce the available potential energy: suppose warm light fluid overlays cold dense fluid in a statically stable configuration, then a downgradient vertical diffusion would raise the centre of gravity of the fluid, increasing its potential energy — just the opposite of the action of baroclinic instability. Thus, the sign on the vertical diffusivity must be negative and this, in combination with the structure of (10.143) (and so a positive horizontal diffusivity) allows both properties (i) and (ii) above to be satisfied. The parameterization does not preserve total energy; the loss of potential energy is not balanced by a corresponding gain of kinetic energy, rather it is assumed to be lost to dissipation. Finally, to determine the magnitude of the (skew) eddy diffusivity we may turn again to the phenomenological estimates of section 10.7.2.

The eddy transport velocity

Applying (10.107) to (10.143) gives the eddy transport velocities,

$$\boxed{\begin{aligned}\tilde{\mathbf{u}} &= -\frac{\partial}{\partial z}(\kappa_a \mathbf{s}), \\ \tilde{w} &= \nabla_z \cdot (\kappa_a \mathbf{s})\end{aligned}}. \quad (10.145)$$

The streamfunction associated with \mathbf{A} is found using (10.109) and (10.143) giving

$$\boldsymbol{\psi} = (-\kappa_a s^y, \kappa s^x, 0) = \mathbf{k} \times \kappa_a \mathbf{s}. \quad (10.146)$$

Two equivalent ways of implementing the GM parameterization are thus as a skew flux, as in (10.144), or as an advection by the pseudovelocities (10.145). The vanishing of the normal component of the velocity is equivalent to the vanishing of the normal component of the flux at the boundary, and ensures that the scheme conserves tracer moments. The advective flux of buoyancy is just

$$\mathbf{F}_{ad} = b \tilde{\mathbf{v}} = b \nabla \times \boldsymbol{\psi} = b \nabla \times (\mathbf{k} \times \kappa_a \mathbf{s}), \quad (10.147)$$

whereas using (10.103) the skew flux is given by

$$\mathbf{F}_{sk} = -\nabla b \times \boldsymbol{\psi} = -\nabla b \times (\mathbf{k} \times \kappa_a \mathbf{s}). \quad (10.148)$$

Vector manipulation readily shows that the divergences of these two fluxes are equal.

10.7.5 Examples

Consider a situation with sloping isotherms (and with the density determined solely by temperature) as illustrated in Fig. 10.6. The vertical flux attempts to tighten the temperature distribution, whereas the horizontal flux, being downgradient, attempts to smooth out horizontal inhomogeneities. Taken together, their net effect is to preserve the amount of

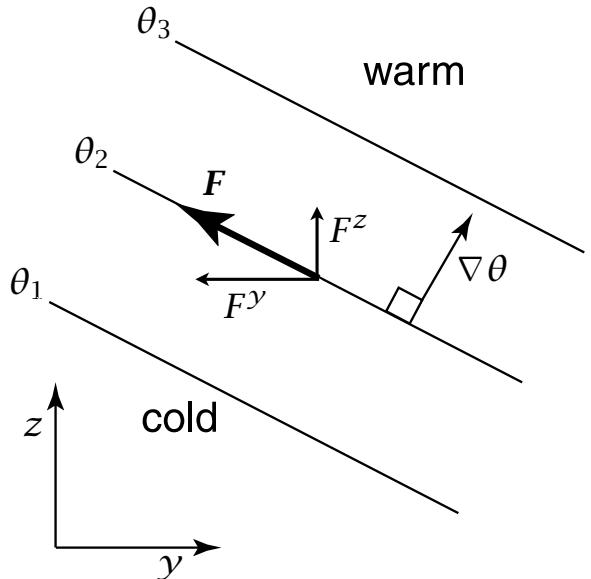


Fig. 10.6 The GM skew fluxes arising from sloping isotherms. The flux itself, F , is parallel to the isotherms, with the horizontal flux being directed down the horizontal gradient but the vertical flux being upgradient. The effect of the vertical flux is to lower the centre of gravity of the fluid, and reduce the potential energy. The horizontal flux tries to make the temperature more uniform in the horizontal direction. The net effect of the skew flux is to *flatten* the isotherms.

fluid between any two isotherms, but at the same time to rotate and flatten the isotherms, so reducing the available potential energy of the flow. This is different from a conventional downgradient diffusion. A purely horizontal diffusion would, in principle, act to equalize values at each level, and a three-dimensional downgradient diffusion would try to equalize all values. Thus, a skew flux behaves quite differently from the usual downgradient diffusion, which merely acts to reduce gradients without caring much about other fluid properties.

To illustrate this consider a very simple example, that of a two-dimensional (y - z) fluid in which the initial density field is a 3×3 grid, with initial conditions

$$\rho_{init} = \begin{bmatrix} 2 & 1 & 1 \\ 3 & 2 & 1 \\ 3 & 3 & 2 \end{bmatrix}. \quad (10.149)$$

The isopycnals are sloping, much as in Fig. 10.6, and the flow is statically stable everywhere.

A purely horizontal diffusion would lead to, in the absence of other processes and with zero normal flux at the boundaries, a final state of

$$\rho_{hd} = \begin{bmatrix} 1.33 & 1.33 & 1.33 \\ 2 & 2 & 2 \\ 2.66 & 2.66 & 2.66 \end{bmatrix}, \quad (10.150)$$

and a full (vertical and horizontal) diffusion would give

$$\rho_{hvd} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}. \quad (10.151)$$

Neither of the above two final states preserves the density census (i.e., its distribution) and

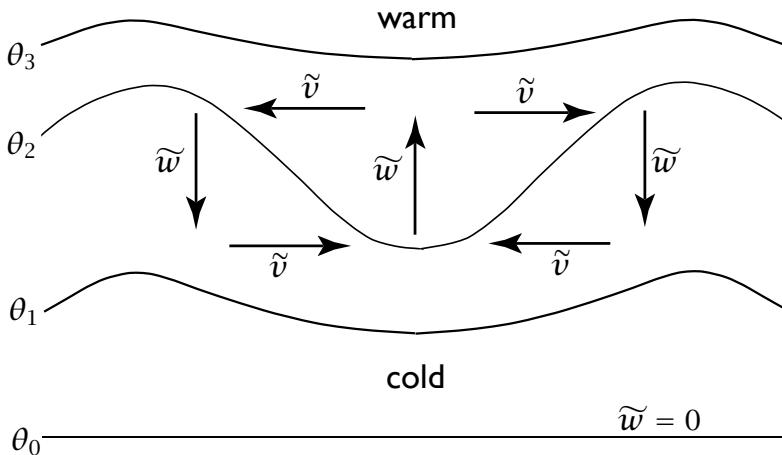


Fig. 10.7 The eddy-induced velocities in the Gent-McWilliams parameterization. The induced circulation attempts to flatten the sloping isopycnals. The induced vertical velocity, \tilde{w} , is zero on flat isopycnals.

both imply strong diabatic effects — the fluid has been *mixed*, and the density variance has been reduced.

In contrast, a skew diffusion or eddy-transport advection will rotate the density surfaces until the isopycnal slope is zero, at which point the value of the transfer coefficients becomes zero and the process stops. The final state is then

$$\rho_{GM} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix}. \quad (10.152)$$

This both preserves the density census and reduces the available potential energy.

We can equally well interpret these effects in terms of eddy-transport velocities, so emphasizing that it is not the eddy flux itself that is important; rather, it is the flux divergence. If the slopes of Fig. 10.6 extended uniformly everywhere, then the associated fluxes would have zero divergence, and the eddy-induced velocities, given by (10.145), would be zero. On the other hand, consider the case illustrated in Fig. 10.7, with variously sloping isotherms. For a constant value of the eddy diffusivity κ the slope of the isopycnals, $s = -(\partial \rho / \partial y) / (\partial \rho / \partial z)$, provides the streamfunction for the eddy-induced velocity:

$$\tilde{v} = -\frac{\partial \psi}{\partial z} = -\frac{\partial}{\partial z}(\kappa_a s), \quad \tilde{w} = \frac{\partial \psi}{\partial y} = \frac{\partial}{\partial y}(\kappa_a s). \quad (10.153)$$

This induces the velocities illustrated in Fig. 10.7, which evidently serve to flatten the isopycnals. (If the isopycnals steepen as they approach the surface, the effects of the scheme may be harder to interpret.)

Finally, we note that downgradient (symmetric) diffusion and skew diffusion would normally be used together, and the transport tensor will then have both symmetric and antisymmetric components. In ocean models the mixing of temperature and salinity is often chosen be along isopycnal surfaces. On the other hand, in the atmosphere (especially in

the troposphere) diabatic effects are quite important and the symmetric tensor needs to be chosen appropriately.

10.8 † THICKNESS DIFFUSION

In the previous section, we considered the structure of the diffusivity tensor, and then chose the entries by physical reasoning to mimic the effects of baroclinic instability. An alternative approach is to choose, *a priori*, a quantity to be diffused downgradient (i.e., not skew diffused), and to then transform the equations appropriately to represent the effect in the equations of motion as commonly used. In this section we explore the use of thickness as a ‘diffusee’. Thickness is the vertical distance between two isotherms or isopycnals, and it is a candidate for diffusion because a downgradient thickness transfer within an isopycnal layer satisfies two conditions:

- (i) the total mass contained between two isopycnals is preserved, provided there are no boundary fluxes, so the effect is adiabatic;
- (ii) if there is no orography a thickness flux serves to flatten isopycnals, and hence to reduce the available potential energy of the flow, mimicking baroclinic instability.

These are similar to the two properties listed on page 436, and indeed we will find that thickness diffusion is very similar to the GM scheme. However, thickness is *not* a materially conserved quantity; thus, the arguments of section 10.2 do not apply and so turbulent diffusion of thickness is somewhat ad hoc. Let us explore all these issues further.

10.8.1 Equations of motion

Thickness and its variance

Recalling the results of section 3.9, in a Boussinesq fluid the distance between two surfaces of constant buoyancy is given by

$$\text{thickness} = \Sigma = \int_{z(b_1)}^{z(b_2)} dz = \int_{b_1}^{b_2} \frac{\partial z}{\partial b} db, \quad (10.154)$$

and thus we may define the thickness field (strictly a ‘thickness density’ field), $\sigma \equiv \partial z / \partial b$. The volume of fluid between two isopycnal surfaces is proportional to $\sigma \Delta A$, where ΔA is an infinitesimal area, and in the absence of diabatic processes this is conserved. Thus, we have $D(\sigma \Delta A) / Dt = 0$ and using $D\Delta A / Dt = \Delta A \nabla_b \cdot \mathbf{u}$ we obtain the equation of motion for thickness [cf. (3.166)]

$$\frac{D\sigma}{Dt} + \sigma \nabla_b \cdot \mathbf{u} = D_\sigma \quad \text{or} \quad \frac{\partial \sigma}{\partial t} + \nabla_b \cdot (\mathbf{u} \sigma) = D_\sigma, \quad (10.155)$$

now including a term D_σ to represents any diabatic terms. From (10.155) we obtain, after a little algebra, the variance equation

$$\frac{1}{2} \frac{\partial}{\partial t} \overline{\sigma'^2} + \overline{\mathbf{u}' \sigma'} \cdot \nabla_b \overline{\sigma} + \frac{1}{2} \overline{\mathbf{u}} \cdot \nabla_b \overline{\sigma'^2} + \frac{1}{2} \overline{\mathbf{u}' \cdot \nabla_b \sigma'^2} = -\overline{\mathbf{w}' \sigma'} + \overline{D_\sigma' \sigma'}, \quad (10.156)$$

where we have written $\mathbf{w}' \equiv (\sigma \nabla_b \cdot \mathbf{u})'$. This equation is to be compared with the corresponding equation for a conserved tracer, (10.67). If the mean flow is small and the

third-order correlations may be neglected then (10.156) becomes, in a statistically steady state,

$$\overline{\mathbf{u}'\sigma'} \cdot \nabla_b \overline{\sigma} \approx -\overline{w'\sigma'} + \overline{D_\sigma \sigma'}. \quad (10.157)$$

Unlike the case for a materially conserved tracer [cf. (10.68)], the transport of thickness is not necessarily downgradient in regions of dissipation. However, in regions of baroclinic instability where there may be conversion of available potential energy to kinetic energy $\overline{w'\sigma'}$ is positive and we do then expect thickness to be transferred downgradient, suggesting a diffusive parameterization.

The eddy-induced and residual velocities

Now, let us decompose these variables in the usual manner into a mean component (say an ensemble mean, denoted with an overbar), and an eddy component, denoted with a prime. Then the averaged thickness equation is

$$\frac{\partial \overline{\sigma}}{\partial t} + \nabla_b \cdot (\overline{\sigma} \overline{\mathbf{u}} + \overline{\sigma'} \overline{\mathbf{u}'}) = 0, \quad (10.158)$$

where $\overline{\sigma'} \overline{\mathbf{u}'}$ is the eddy thickness flux. This equation may be written as

$$\frac{\partial \overline{\sigma}}{\partial t} + \nabla_b \cdot [(\overline{\mathbf{u}} + \tilde{\mathbf{u}}) \overline{\sigma}] = 0, \quad (10.159)$$

where

$$\tilde{\mathbf{u}} \equiv \frac{\overline{\sigma'} \overline{\mathbf{u}'}}{\overline{\sigma}}$$

(10.160)

is the ‘eddy-induced velocity’, sometimes referred to as the ‘bolus velocity’, so-called because the thickness flux is said to be evocative of a peristaltic transfer along a passage bounded by impermeable but elastic walls. The quantity

$$\overline{\mathbf{u}}^* = \overline{\mathbf{u}} + \tilde{\mathbf{u}} \quad (10.161)$$

is the residual velocity we encountered in chapter 7, and it accounts for the total transport of thickness, including both eddy and Eulerian means.

In adiabatic flow, the evolution of a materially conserved tracer τ is given by

$$\frac{D}{Dt}(\tau \sigma \Delta A) = 0, \quad (10.162)$$

whence, because $\sigma \Delta A$ is a constant,

$$\frac{\partial \tau}{\partial t} + \mathbf{u} \cdot \nabla_b \tau = 0. \quad (10.163)$$

Combining this with the thickness equation, (10.155) with $D_\sigma = 0$, gives

$$\frac{\partial}{\partial t}(\sigma \tau) + \nabla_b \cdot (\sigma \mathbf{u} \tau) = 0, \quad (10.164)$$

which in turn leads to

$$\frac{\partial}{\partial t}(\overline{\sigma} \overline{\tau} + \overline{\sigma'} \overline{\tau'}) + \nabla_b \cdot (\overline{\sigma} \overline{\mathbf{u}} \overline{\tau}) + \nabla_b \cdot \overline{\sigma'} \overline{\mathbf{u}' \tau} + \nabla_b \cdot [(\overline{\sigma \mathbf{u}})' \overline{\tau'}] = 0, \quad (10.165)$$

or, using (10.158),

$$\frac{\partial \bar{\tau}}{\partial t} + \frac{1}{\bar{\sigma}} \frac{\partial}{\partial t} (\bar{\sigma}' \bar{\tau}') + \left[\bar{\mathbf{u}} + \frac{\bar{\sigma}' \mathbf{u}'}{\bar{\sigma}} \right] \cdot \nabla_b \bar{\tau} = -\frac{1}{\bar{\sigma}} \nabla_b \cdot [\bar{(\sigma \mathbf{u})}' \bar{\tau}']. \quad (10.166)$$

(To derive these, first let $\bar{\sigma} \bar{\tau} \bar{\mathbf{u}} = \bar{(\sigma \mathbf{u})}' \bar{\tau}' + \bar{\sigma} \bar{\mathbf{u}} \bar{\tau}$.) If we neglect the correlation between σ' and τ' , then (10.166) has the form

$$\frac{\partial \bar{\tau}}{\partial t} + (\bar{\mathbf{u}} + \tilde{\mathbf{u}}) \cdot \nabla_b \bar{\tau} = -\frac{1}{\bar{\sigma}} \nabla_b \cdot [\bar{(\sigma \mathbf{u})}' \bar{\tau}']. \quad (10.167)$$

Thus, the averaged tracer evolves as if it were advected by two velocity fields: the large-scale field itself, $\bar{\mathbf{u}}$, and the eddy-induced velocity $\tilde{\mathbf{u}}$, their sum being the residual velocity. The term on the right-hand side of (10.167) is the divergence of the transport of the tracer along the isopycnals by the eddy transport $(\sigma \mathbf{u})'$.

10.8.2 Diffusive thickness transport

A downgradient diffusion of thickness parameterizes the eddy transport velocity by

$$\tilde{\mathbf{u}} \equiv \frac{\bar{\sigma}' \mathbf{u}'}{\bar{\sigma}} = -\frac{1}{\bar{\sigma}} \kappa \nabla_b \bar{\sigma}, \quad (10.168)$$

where κ is an eddy diffusivity. Similarly, we might parameterize the right-hand side of (10.166) by

$$-\frac{1}{\bar{\sigma}} \nabla_b \cdot [\bar{(\sigma \mathbf{u})}' \bar{\tau}'] = \frac{1}{\bar{\sigma}} \nabla_b \cdot (\kappa \bar{\sigma} \nabla_b \bar{\tau}), \quad (10.169)$$

that is, as a diffusion of the tracer along isopycnals.

In height coordinates the eddy transport velocity will be a three-dimensional field, obtained by appropriately transforming $\tilde{\mathbf{u}}$. We have

$$\tilde{\mathbf{u}} \approx -\frac{1}{\sigma} \kappa \nabla_b \bar{\sigma} = -\kappa \frac{\partial b}{\partial z} \nabla_b \left(\frac{\partial \bar{z}}{\partial b} \right) = -\kappa \frac{\partial b}{\partial z} \frac{\partial \bar{s}}{\partial b} = -\kappa \frac{\partial \bar{s}}{\partial z}, \quad (10.170)$$

where the third equality uses $\nabla_b z = \mathbf{s}$, the isopycnal slope. The final result is not quite the same as (10.145), because the diffusivity is now outside the z -derivative. It is a subtle but important distinction, because it means that if κ varies the vertical velocity can no longer be obtained easily as a local function. That is to say, given (10.170), we no longer have $\tilde{w} = \nabla_z \cdot \kappa \bar{s}$ as in (10.145). Rather, \tilde{w} must be evaluated by a non-local integration of the mass conservation requirement, so that

$$\tilde{w} = \int \nabla_z \cdot \left(\kappa \frac{\partial \bar{s}}{\partial z} \right) dz. \quad (10.171)$$

This should not be disconcerting from a physical standpoint, because the baroclinic activity of eddies certainly involves vertical communication — recall the tendency toward barotropic flow in baroclinic lifecycles. From a computational standpoint, it is a little less convenient. A less satisfactory feature of thickness diffusion arises when the ocean floor is not flat; a strict thickness diffusion might then increase the slope of the isopycnals and increase the available potential energy, which would be an unwanted effect. Nevertheless, overall the GM scheme is evidently similar to a thickness diffusion.

10.9 † EDDY TRANSPORT AND THE TRANSFORMED EULERIAN MEAN

A natural framework to discuss how eddy fluxes interact with the mean flow is the transformed Eulerian mean (TEM), discussed in chapter 7. It provides a simplification of the formalism (at least if one is already familiar with TEM theory!) and, more importantly, a means to consider more general cases, such as the flux of potential vorticity and other variables.

From section 7.3.4 the zonally averaged thermodynamic equation may be written as

$$\frac{\partial \bar{b}}{\partial t} + (\bar{\mathbf{u}} + \tilde{\mathbf{u}}) \cdot \nabla \bar{b} + (\bar{w} + \tilde{w}) \frac{\partial \bar{b}}{\partial z} = \frac{\partial G}{\partial z}, \quad (10.172)$$

where

$$G = \frac{\bar{\mathbf{v}}' \bar{b}' \cdot \nabla \bar{b}}{\partial_z \bar{b}} = \frac{1}{\partial_z \bar{b}} \left(\bar{\mathbf{u}}' \bar{b}' \cdot \nabla_z \bar{b} + \bar{w}' \bar{b}' \frac{\partial \bar{b}}{\partial z} \right), \quad (10.173)$$

and

$$\tilde{\mathbf{u}} = -\frac{\partial \tilde{\boldsymbol{\phi}}}{\partial z}, \quad \tilde{w} = \nabla \cdot \tilde{\boldsymbol{\phi}}, \quad \tilde{\boldsymbol{\phi}} = \frac{\mathbf{F}[b]}{\partial_z \bar{b}} = \frac{\bar{\mathbf{u}}' \bar{b}'}{\partial_z \bar{b}}. \quad (10.174)$$

If the eddy fluxes are parallel to the isopycnals, then $G = 0$ and the sole effect of the eddies on the mean flow is via the eddy-induced velocity. Using the results of section 7.3.3 to connect buoyancy transport to thickness transport, and then diffusing thickness, we can express these velocities as

$$\tilde{\mathbf{u}} = -\frac{\partial}{\partial z} \left(\frac{\bar{\mathbf{u}}' \bar{b}'}{\partial_z \bar{b}} \right) \approx \left(\frac{\bar{\mathbf{u}}' \sigma'}{\sigma} \right) \approx -\frac{1}{\sigma} \kappa \nabla_b \bar{\sigma} \approx -\kappa_a \frac{\partial \mathbf{s}}{\partial z}, \quad (10.175)$$

using (10.170), so recovering, for constant κ , the GM scheme. Note that if we write the residual velocity in standard TEM fashion as $\bar{\mathbf{v}}^* = \bar{\mathbf{v}} + \nabla \times \boldsymbol{\psi}$ then with the GM parameterization the TEM streamfunction may be written as

$$\boldsymbol{\psi} = -\kappa \frac{\mathbf{k} \times \nabla_z \bar{b}}{\partial_z \bar{b}}, \quad (10.176)$$

which reprises (10.146).

10.9.1 Potential vorticity diffusion

Preliminaries

From a more fundamental perspective, potential vorticity, Q , is a better candidate for diffusion than thickness because it is a materially conserved quantity.¹⁶ It is not the only variable that is materially conserved — potential temperature is also. However, potential temperature is advected by the *three-dimensional* velocity field, and the vertical advection complicates matters, since any diffusion tensor certainly cannot be isotropic and is probably not symmetric. On the other hand, in isentropic coordinates the adiabatic potential vorticity advection occurs in the isentropic plane (and in quasi-geostrophic flow the advection is purely horizontal). Thus, only the two-dimensional diffusion need be considered, and the diffusion tensor will be much simplified. Indeed, if the effects of differential rotation are unimportant, the turbulence may be nearly isotropic and the diffusivity becomes a simple scalar.

However, near the upper and lower boundaries buoyancy may in fact be the appropriate field to diffuse. This is because (for flat boundaries) $w = 0$ and buoyancy is conserved on parcels when advected by the horizontal flow:

$$\frac{\partial b}{\partial t} + \mathbf{u} \cdot \nabla_z b = 0. \quad (10.177)$$

Potential vorticity diffusion, on the other hand, may be difficult at boundaries where outcropping occurs: the thickness between two isentropic surfaces goes to zero and $(\zeta + f)/\sigma$ is ill-defined. This has an analogue in quasi-geostrophic dynamics, where the equations of motion are horizontal advection of potential vorticity in the fluid interior with horizontal buoyancy advection at the boundaries providing the boundary conditions. Horizontal diffusion of buoyancy is, of course, not an adiabatic parameterization, but diabatic effects do occur at the surface. These considerations suggest that downgradient potential vorticity diffusion on isentropic surfaces in the fluid interior, combined with downgradient buoyancy diffusion at the upper and lower boundaries, is as rational a parameterization of eddy transfer effects as any simple diffusion scheme can be.

Implementations and approximations

When the equations of motion are written in conventional form the potential vorticity flux does not directly appear, and the equations must be manipulated in order to implement a potential vorticity scheme. Suppose that the large-scale fields obey planetary-geostrophic scaling, and for simplicity we remain on the β -plane. Then, in isentropic coordinates the potential vorticity is

$$Q = \frac{f + \zeta}{\sigma} \approx \frac{f}{\sigma}, \quad (10.178)$$

which gives

$$\nabla_b \bar{Q} = \frac{1}{\sigma} \nabla_b f - \frac{f}{\sigma^2} \nabla_b \bar{\sigma} = \frac{\beta}{\sigma} \mathbf{j} - \frac{f}{\sigma^2} \nabla_b \bar{\sigma}. \quad (10.179)$$

Potential vorticity diffusion is then just

$$\bar{\mathbf{u}}' \bar{Q}' = -\mathbf{K} \nabla_b \bar{Q} = -\mathbf{K} \left(\frac{\beta}{\sigma} \mathbf{j} - \frac{f}{\sigma^2} \nabla_b \bar{\sigma} \right), \quad (10.180)$$

where \mathbf{K} is the diffusivity (a two-dimensional symmetric tensor in the isentropic plane, or a scalar). If the eddy flux of potential vorticity also satisfies planetary-geostrophic scaling then

$$\bar{\mathbf{u}}' \bar{Q}' \approx -\frac{f}{\sigma^2} \bar{\mathbf{u}}' \bar{\sigma}'. \quad (10.181)$$

Using (10.160) and (10.181) gives

$$\tilde{\mathbf{u}} = \frac{1}{\sigma} \bar{\mathbf{u}}' \bar{\sigma}' \approx -\frac{\bar{\sigma}}{f} \bar{\mathbf{u}}' \bar{Q}', \quad (10.182)$$

and using (10.180) yields

$$\tilde{\mathbf{u}} = \mathbf{K} \left(\frac{\beta}{f} \mathbf{j} - \frac{1}{\sigma} \nabla_b \bar{\sigma} \right). \quad (10.183)$$

This differs from (10.168) mainly in the existence of the term involving β on the right-hand side. The expression is singular at the equator, a consequence of ignoring the relative

vorticity term in the expression for potential vorticity. For use in z -coordinates, (10.183) can be transformed in the same way as (10.170), giving

$$\tilde{\mathbf{u}} \approx \kappa \left(\frac{\beta}{f} \mathbf{j} - \frac{\partial \bar{s}}{\partial z} \right). \quad (10.184)$$

Other approaches involving the diffusion of potential vorticity may be possible; for example, one could in principle choose the transformed Eulerian mean to be taken with respect to potential vorticity rather than potential temperature. One might also write the entire system of equations in TEM form, whence potential vorticity fluxes appear directly in the momentum equation.

Notes

- 1 Much of the theory of turbulent diffusion stems from G.I. (Geoffrey Ingram) Taylor (1886–1975), who made important contributions to both fluid and solid mechanics, in the former to meteorology, oceanography, and aerodynamics. In addition to his work in turbulence, Taylor is known for his work on the theory of rotating fluids (the ‘Taylor–Proudman’ effect, for example) and on hydrodynamic stability (analysis of stability of Couette flow, for example), and for his clear and simple laboratory experiments. The main results of this section were first derived by Taylor (1921a). Ludwig Prandtl (1875–1953) was the other great pioneer of turbulent diffusion; he is also famous for his work in boundary-layer theory and aerodynamics.
- 2 A number of textbooks in both fluid dynamics and stochastic processes give more detail on this topic. Gardiner (1985) is one.
- 3 Batchelor & Townsend (1956).
- 4 The way variables are changed between (10.31) and (10.32) could also have been used to derive (10.20), but in that case a simpler transformation sufficed.
- 5 Batchelor & Townsend (1956).
- 6 This topic was first addressed empirically by Richardson (1926), although it was Obukhov (1941) who first theoretically obtained the ‘4/3 power law’ describing how the eddy diffusivity varies with separation for parcels in the inertial range. Our treatment takes advantage of Kolmogorov scaling.
- 7 This figure and Fig. 10.4 were kindly provided by Joe LaCasce; see LaCasce & Ohlmann (2003).
- 8 In the open ocean, Ollitrault *et al.* (2005) do find float separation that increases with t^3 , consistent with a $-5/3$ inverse cascade range.
- 9 Morel & Larcheveque (1974) and Er-El & Peskin (1981). Earlier dispersion calculations were made by Richardson (1926) who measured smoke spreading from chimneys, finding results that are consistent with a three-dimensional energy inertial range at small scales.
- 10 Turbulent diffusion is both widely used and widely criticized. If there is a scale-separation between a well-defined mean flow and the eddies then turbulent diffusion can be a very useful parameterization. However, this condition is often *not* satisfied, because it is unusual in fluid mechanics for the turbulent eddies to be significantly smaller than the mean flow. Baroclinic turbulence is something of an exception because there is a natural scale of the turbulence — the deformation radius — that is in general different from the scale of the mean flow, although even this scale separation may be lost if there is an inverse cascade

or if the deformation scale is sufficiently large, as in the Earth's atmosphere. Furthermore, properly choosing what variable is to be diffused, and ensuring that various fluid conservation properties remain respected by the diffusion, remain difficult problems.

- 11 If S_{ij} and A_{ij} are symmetric and antisymmetric tensors respectively, then, summing over repeated indices, their contraction is $A_{ij}S_{ij} = -A_{ji}S_{ij} = -A_{ji}S_{ji} = -A_{ij}S_{ij}$, where the last equality follows because the indices are dummy. Thus, the contraction must equal zero.
- 12 Green (1970) and Stone (1972), in the context of the meridional transport of heat in the Earth's atmosphere, suggested that the magnitude of the turbulent diffusivity coefficients could be obtained by dynamical arguments using such things as baroclinic instability theory and the amount of available potential energy in the atmosphere, although their suggestions differ in such important details as the eddy mixing length. Other efforts have drawn on geostrophic turbulence theory, for example Larichev & Held (1995) and Smith & Vallis (2002).
- 13 Ollitrault *et al.* (2005).
- 14 Green (1970).
- 15 The Gent-McWilliams (GM) scheme originated in Gent & McWilliams (1990) and was clarified by Gent *et al.* (1995). Previously, Plumb (1979) had noted the connection between symmetric and antisymmetric diffusivities and diffusive and advective fluxes, and Griffies (1998) explicitly showed how the GM bolus velocities are related to a skew flux and can be calculated using an antisymmetric diffusivity tensor. See McDougall (1998) and (Griffies 2004) for reviews and more discussion. Visbeck *et al.* (1997) suggested that the values of eddy diffusivities in the GM scheme might be determined by dynamical arguments similar to those of Green (1970) and Stone (1972).
- 16 Potential vorticity diffusion was suggested by Green (1970) as a parameterization for large-scale eddies in the atmosphere, and used in ocean contexts by Welander (1973), Marshall (1981) and Rhines & Young (1982a). Tréguier *et al.* (1997) and Greatbatch (1998) explored the use of potential vorticity based parameterizations for ocean models. Lee *et al.* (1997), Marshall *et al.* (1999), Drijfhout & Hazeleger (2001), and others, have explored numerically whether the eddy transfer of tracers in the ocean is in fact diffusive, and whether potential vorticity or thickness is a better quantity to diffuse. Numerical results can be difficult to interpret because if a property is diffused downgradient then that property tends to become homogeneous, and the resulting diagnostics may be inaccurate. The construction of parameterizations for ocean models remains an active area of research. See for example Killworth (1997), Smith & Vallis (2002) and Ferrari & McWilliams (2006).

Further reading

Aris, R., 1962. *Vectors, Tensors and the Basic Equations of Fluid Mechanics*.

A straightforward introduction to tensors and Cartesian tensors and their application to fluid mechanics.

Monin, A. S. & Yaglom, A. M., 1971. *Statistical Fluid Mechanics: Mechanics of Turbulence*.

This two-volume encyclopaedic reference contains a wealth of information on all aspects of turbulence, and it is particularly strong on turbulent diffusion.

Problems

- 10.1 (a) On the line the Dirac delta function is given (non-uniquely) by

$\delta(x) = \lim_{\epsilon \rightarrow 0} (\pi\epsilon)^{-1/2} \exp(-x^2/\epsilon)$. What are the two- and three-dimensional analogues, given circular and spherical symmetry?

(b) In two dimensions, show that the solution to the diffusion equation

$$\frac{\partial \phi}{\partial t} = D \nabla^2 \phi \quad (\text{P10.1})$$

with initial conditions $\phi(r, t = 0) = \delta(r)$ is given by

$$\phi(r, t) = \frac{1}{4\pi Dt} \exp(-r^2/4Dt). \quad (\text{P10.2})$$

Show that the mean-square distance of a particle that spreads from the origin is given by $\bar{r}^2 = 4Dt$. You may assume a circularly symmetric solution.

(c) Solve this problem in one dimension.

- 10.2 ♦ Repeat problem 10.1, but in a space of arbitrary dimension m , which if needs be may be integer.
- 10.3 ♦ Solve the two-dimensional advection-diffusion equation, $\partial \phi / \partial t + \mathbf{u} \cdot \nabla \phi = \kappa \nabla^2 \phi$, exactly for a passive tracer ϕ in a flow $(u, v) = (\Gamma y, 0)$, with the initial condition $\phi = A \cos kx$. Provide a physical explanation of the long-time behaviour of the solution. *Hint:* Use an integrating factor, and look for a solution like $\phi \sim A e^{-B[x - u(y)t]}$.
- 10.4 Review the derivation of (10.80). Show that if instead of integrating over an area bounded by a streamline, we integrate over an area bounded by a line of constant ϕ , we obtain

$$\iint S \, dA = - \oint_{\phi_0} \frac{\nabla \phi \cdot \nabla \phi}{|\nabla \phi|} \, dl \quad (\text{P10.3})$$

and therefore that if the source term is zero ϕ must be uniform.

Part III

**LARGE-SCALE ATMOSPHERIC
CIRCULATION**

I think the causes of the general trade-winds have not been fully explained by any of those who have wrote on that subject... That the action of the Sun is the original cause of these Winds, I think all are agreed.

George Hadley, *Concerning the Cause of the General Trade Winds*, 1735.

CHAPTER ELEVEN

The Overturning Circulation: Hadley and Ferrel Cells

IN THIS CHAPTER AND THE TWO FOLLOWING we discuss the large-scale circulation, and in particular the *general circulation*, of the atmosphere, this being the mean flow on scales from the synoptic eddy scale — about 1000 km — to the global scale. In this chapter we focus on the dynamics of the Hadley Cell and then, rather descriptively, on the mid-latitude overturning cell or the Ferrel Cell. The latter provides a starting point for chapter 12 which discusses the dynamics of the extratropical zonally averaged circulation. Finally, in chapter 13, we consider the deviations from zonal symmetry, or more specifically the stationary wave pattern, and the stratosphere. In these three chapters we will use many of the tools developed in the previous chapters, but those readers who already have some acquaintance with geophysical fluid dynamics may simply wish to jump in here.

The atmosphere is a terribly complex system, and we cannot hope to fully explain its motion as the analytic solution to a small set of equations. Rather, a full understanding of the atmosphere requires describing it in a consistent way on many levels simultaneously. One of these levels involves simulating the flow by numerically solving the governing equations of motion as completely as possible, for example by using a comprehensive General Circulation Model (GCM). However, such a simulation brings problems of its own, including the problem of understanding the simulation, and discerning whether it is a good representation of reality. Thus, in this chapter and the two following we concentrate on simpler, more conceptual models. We begin this chapter with a brief observational overview of some of the pre-eminent large-scale features of the atmosphere, concentrating on the zonally averaged fields.¹

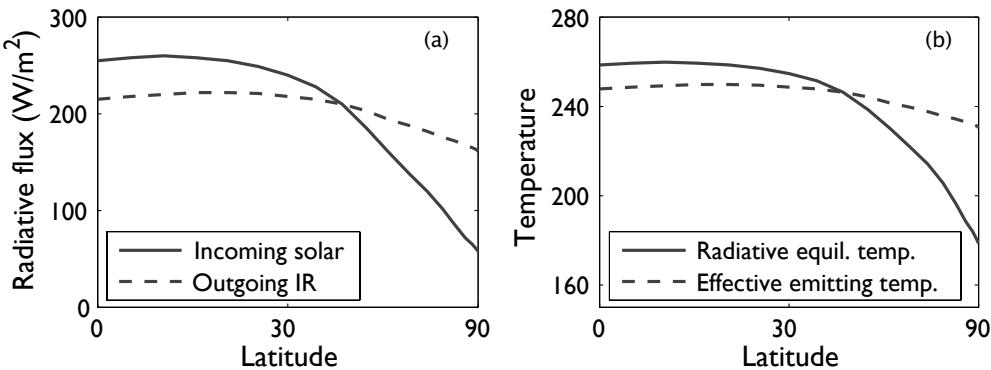


Fig. 11.1 (a) The (approximate) observed net average incoming solar radiation and outgoing infrared radiation at the top of the atmosphere, as a function of latitude (plotted on a sine scale). (b) The temperatures associated with these fluxes, calculated using $T = (R/\sigma)^{1/4}$, where R is the solar flux for the radiative equilibrium temperature and R is the infrared flux for the effective emitting temperature. Thus, the solid line is an approximate radiative equilibrium temperature

11.1 BASIC FEATURES OF THE ATMOSPHERE

11.1.1 The radiative equilibrium distribution

A gross but informative measure characterizing the atmosphere, and the effects that dynamics have on it, is the pole-to-equator temperature distribution. The *radiative equilibrium* temperature is the hypothetical, three-dimensional, temperature field that would obtain if there were no atmospheric or oceanic motion, given the composition and radiative properties of the atmosphere and surface. The field is a function only of the incoming solar radiation at the top of the atmosphere, although to evaluate it entails a complicated calculation, especially as the radiative properties of the atmosphere depend on the amount of water vapour and cloudiness in the atmosphere. (The distribution of absorbers is usually taken to be that which obtains in the observed, moving, atmosphere, in order that the differences between the calculated radiative equilibrium temperature and the observed temperature are due to fluid motion.)

A much simpler calculation that illustrates the essence of the situation is to first note that at the top of the atmosphere the globally averaged incoming solar radiation is balanced by the outgoing infrared radiation. If there is no lateral transport of energy in the atmosphere or ocean then *at each latitude* the incoming solar radiation will be balanced by the outgoing infrared radiation, and if we parameterize the latter using a single latitudinally-dependent temperature we will obtain a crude radiative-equilibrium temperature for the atmospheric column at each latitude. Specifically, a black body subject to a net incoming radiation of S (watts per square metre) has a radiative-equilibrium temperature T_{rad} given by $\sigma T_{\text{rad}}^4 = S$, this being Stefan's law with Stefan-Boltzmann constant $\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$. Thus, for the Earth, we have, at each latitude,

$$\sigma T_{\text{rad}}^4 = S(\vartheta)(1 - \alpha), \quad (11.1)$$

where α is the albedo of the Earth and $S(\vartheta)$ is the incoming solar radiation at the top of the

atmosphere, and its solution is shown in Fig. 11.1. The solid lines in the two panels show the net solar radiation and the solution to (11.1), T_{rad} ; the dashed lines show the observed outgoing infrared radiative flux, I , and the effective emitting temperature associated with it, $(I/\sigma)^{1/4}$. The emitting temperature does not quantitatively characterize that temperature at the Earth's surface, nor at any single level in the atmosphere, because the atmosphere is not a black body and the outgoing radiation originates from multiple levels. Nevertheless, the qualitative point is evident: the radiative equilibrium temperature has a much stronger pole-to-equator gradient than does the effective emitting temperature, indicating that there is a poleward transport of heat in the atmosphere-ocean system. More detailed calculations indicate that the atmosphere is further from its radiative equilibrium in winter than summer, indicating a larger heat transport. The transport occurs because polewards moving air tends to have a higher static energy ($c_p T + gz$ for dry air; in addition there is some energy transport associated with water vapour evaporation and condensation) than the equatorwards moving air, most of this movement being associated with the large-scale circulation. The radiative forcing thus seeks to maintain a pole-to-equator temperature gradient, and the ensuing circulation seeks to reduce this gradient.

11.1.2 Observed wind and temperature fields

The observed zonally averaged temperature and zonal wind fields are illustrated in Fig. 11.2. The vertical coordinate is log pressure, multiplied by a constant factor $H = RT_0/g = 7.5$ km, so that the ordinate is similar to height in kilometres. [In an isothermal hydrostatic atmosphere $(RT_0/g)d\ln p = -dz$, and the value of H chosen corresponds to $T_0 = 256$ K.] To a good approximation temperature and zonal wind are related by thermal wind balance, which in pressure coordinates is

$$f \frac{\partial u}{\partial p} = \frac{R}{p} \frac{\partial T}{\partial y}. \quad (11.2)$$

In the lowest several kilometres of the atmosphere temperature falls almost monotonically with latitude and height, and this region is called the *troposphere*. The temperature in the lower troposphere in fact varies more rapidly with latitude than does the effective emitting temperature, T_E , the latter being more characteristic of the temperature in the mid-to-upper troposphere. The meridional temperature gradient is much larger in winter than summer, because in winter high latitudes receive virtually no direct heating from the Sun. It is also strongest at the edge of the subtropics, and here it is associated with a zonal jet, particularly strong in winter. There is no need to 'drive' this wind with any kind of convergent momentum fluxes: given the temperature, the flow is a consequence of thermal wind balance, and to the extent that the upper troposphere is relatively frictionless there is no need to maintain it against dissipation. Of course just as the radiative-equilibrium temperature gradient is much larger than that observed, so the zonal wind shear associated with it is much larger than that observed. Thus, the overall effect of the atmospheric and oceanic circulation, and in particular of the turbulent circulation of the mid-latitude atmosphere, is to *reduce* the amplitude of the vertical shear of the eastward flow by way of a poleward heat transport. Observations indicate that about two-thirds of this transport is effected by the atmosphere, and about a third by the ocean, more in low latitudes.²

Above the troposphere is the *stratosphere*, and here temperature typically increases with height. The boundary between the two regions is called the *tropopause*, and this varies in

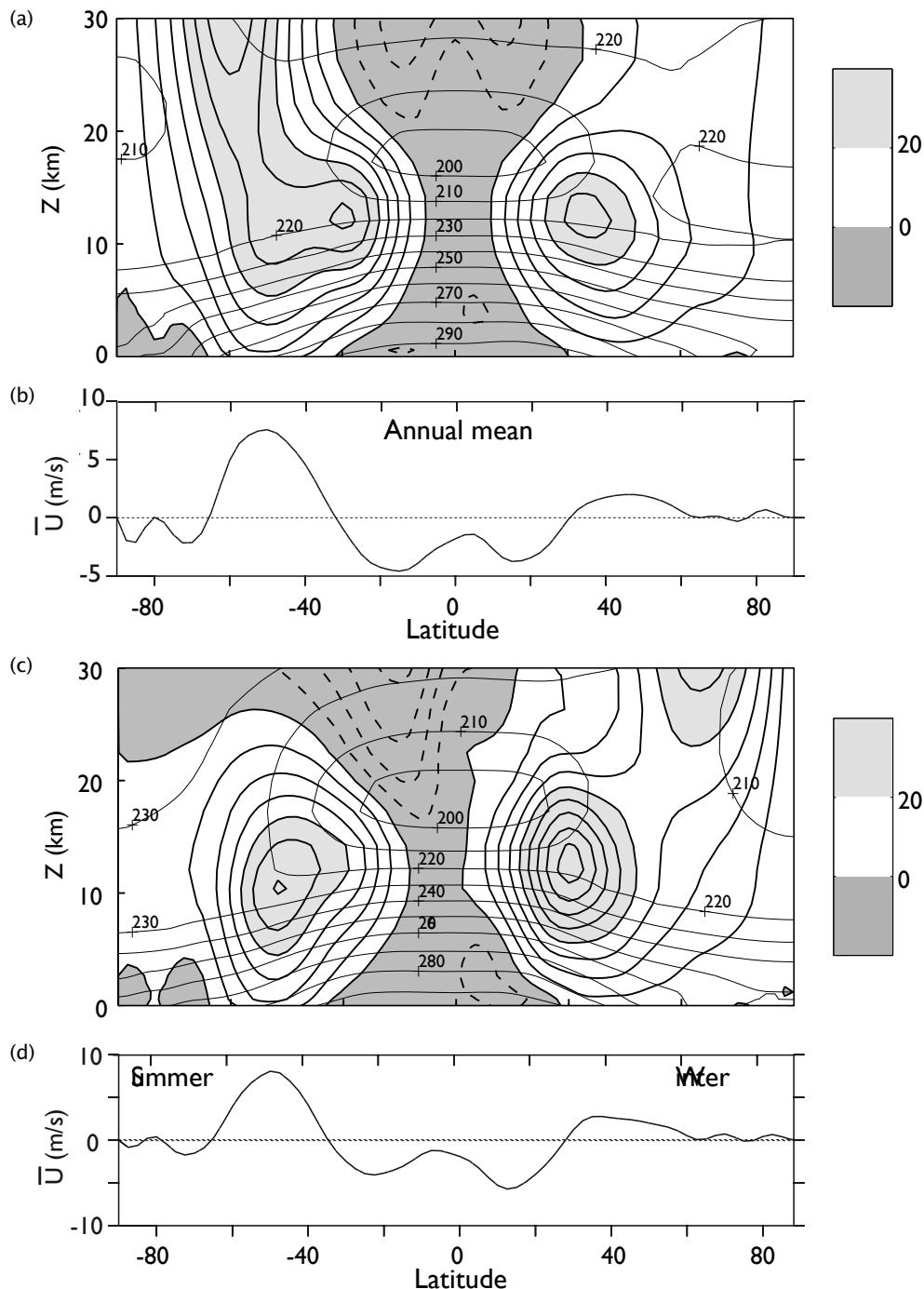


Fig. 11.2 (a) Annual mean, zonally averaged zonal wind (heavy contours and shading) and the zonally averaged temperature (lighter contours). (b) Annual mean, zonally averaged zonal winds at the surface. (c) and (d) Same as (a) and (b), except for northern hemisphere winter (DJF). The wind contours are at intervals of 5 m s^{-1} with shading for eastward winds above 20 m s^{-1} and for all westward winds, and the temperature contours are labelled. The ordinate of (a) and (c) is $Z = -H \log(p/p_R)$, where p_R is a constant, with scale height $H = 7.5 \text{ km}$.

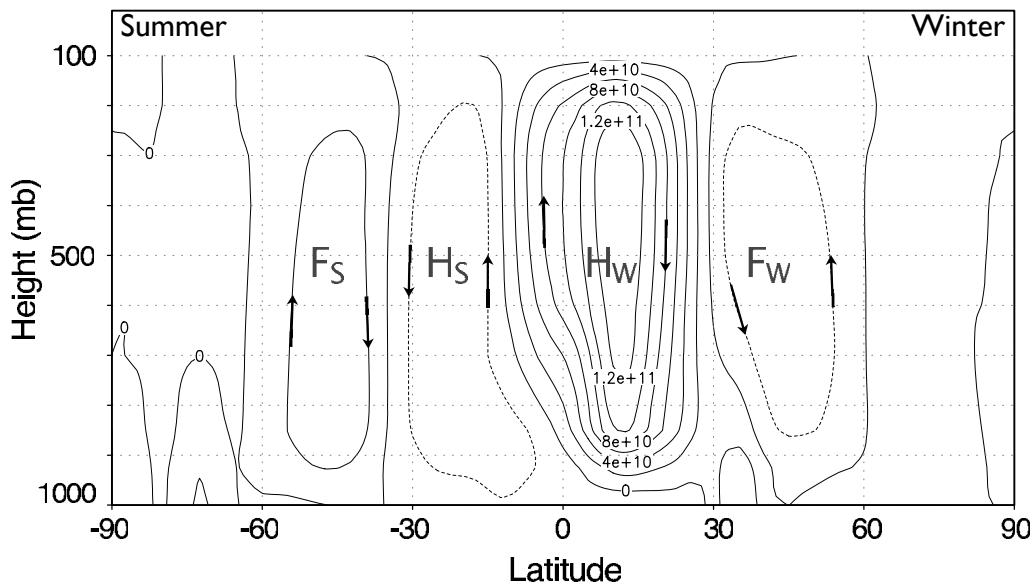


Fig. 11.3 The observed, zonally averaged, meridional overturning circulation of the atmosphere, in units of kg s^{-1} , averaged over December–January–February (DJF). In each hemisphere note the presence of a direct Hadley Cell (H_w and H_s in winter and summer) with rising motion near the equator, descending motion in the subtropics, and an indirect Ferrel Cell (F_w and F_s) at mid-latitudes. There are also hints of a weak direct cell at high latitudes. The winter Hadley Cell is far stronger than the summer one.

height from about 16 km in the tropics to about 8 km in polar regions. We consider the maintenance of this stratification in section 12.5.

The surface winds typically have, going from the equator to the pole, an E-W-E (easterly–westerly–easterly) pattern, although the polar easterlies are weak and barely present in the Northern Hemisphere. (Meteorologists use ‘westerly’ to denote winds from the west, that is eastward winds; similarly ‘easterlies’ are westward winds. We will use both ‘westerly’ and ‘eastward’, and both ‘easterly’ and ‘westward’, and the reader should be comfortable with all these terms.) In a given hemisphere, the surface winds are stronger in winter than summer, and they are also consistently stronger in the Southern Hemisphere than in the Northern Hemisphere, because in the former the surface drag is weaker because of the relative lack of continental land masses and topography. The surface winds are *not* explained by thermal wind balance. Indeed, unlike the upper level winds, they must be maintained against the dissipating effects of friction, and this implies a momentum convergence into regions of surface westerlies and a divergence into regions of surface easterlies. Typically, the maxima in the eastward surface winds are in mid-latitudes and somewhat polewards of the subtropical maxima in the upper-level westerlies and at latitudes where the zonal flow is a little more constant with height. The mechanisms of the momentum transport in the mid-latitudes and the maintenance of the surface westerly winds are the topics of section 12.1.

Some Features of the Large-scale Atmospheric Circulation

From Figures 11.1–11.3 we see or infer the following.

- ★ A pole-equator temperature gradient that is much smaller than the radiative equilibrium gradient.
- ★ A troposphere, in which temperature generally falls with height, above which lies the stratosphere, in which temperature increases with height. The two regions are separated by a tropopause, which varies in height from about 16 km at the equator to about 6 km at the pole.
- ★ A monotonically decreasing temperature from equator to pole in the troposphere, but a weakening and sometimes reversal of this above the tropopause.
- ★ A westerly (i.e., eastward) tropospheric jet. The time and zonally averaged jet is a maximum at the edge or just polewards of the subtropics, where it is associated with a strong meridional temperature gradient. In mid-latitudes the jet has a stronger barotropic component.
- ★ An E-W-E (easterlies–westerlies–easterlies) surface wind distribution. The latitude of the maximum in the surface westerlies is in mid-latitudes, where the zonally averaged flow is more barotropic.

11.1.3 Meridional overturning circulation

The observed (Eulerian) zonally averaged meridional overturning circulation is illustrated in Fig. 11.3. The figure shows a streamfunction, Ψ for the vertical and meridional velocities such that, in the pressure coordinates used in the figure,

$$\frac{\partial \Psi}{\partial y} = \bar{\omega}, \quad \frac{\partial \Psi}{\partial p} = -\bar{v}. \quad (11.3)$$

where the overbar indicates a zonal average. In each hemisphere there is rising motion near the equator and sinking in the subtropics, and this circulation is known as the *Hadley Cell*.³ The Hadley Cell is a thermally direct cell (i.e., the warmer fluid rises, the colder fluid sinks), is much stronger in the winter hemisphere, and extends to about 30°. In mid-latitudes the sense of the overturning circulation is apparently reversed, with rising motion in the high-mid-latitudes, at around 60° and sinking in the subtropics, and this is known as the *Ferrel Cell*. However, as with most pictures of averaged streamlines in unsteady flow, this gives a misleading impression as to the actual material flow of parcels of air because of the presence of eddying motion, and we discuss this in the next chapter. At low latitudes the circulation is more nearly zonally symmetric and the picture does give a qualitatively correct representation of the actual flow. At high latitudes there is again a thermally direct cell (although it is weak and not always present), and thus the atmosphere is often referred to as having a three-celled structure.

11.1.4 Summary

Some of the main features of the zonally averaged circulation are summarized in the shaded box on the preceding page. We emphasize that the zonally averaged circulation is not synonymous with a zonally symmetric circulation, and the mid-latitude circulation is highly asymmetric. Any model of the mid-latitudes that did not take into account the zonal asymmetries in the circulation — of which the weather is the main manifestation — would be seriously in error. This was first explicitly realized in the 1920s, and taking into account such asymmetries is the main task of the dynamical meteorology of the mid-latitudes, and is the subject of the next chapter. On the other hand, the large-scale tropical circulation of the atmosphere is to a large degree zonally symmetric or nearly so, and although monsoonal circulations and the Walker circulation (a cell with rising air in the Eastern Pacific and descending motion in the Western Pacific) are zonally asymmetric, they are also relatively weaker than typical mid-latitude weather systems. Indeed the boundary between the tropics and mid-latitude may be usefully defined by the latitude at which such zonal asymmetries become dynamically important on the large scale and this boundary, at about 30° on average, roughly coincides with the latitude at which the mean meridional overturning circulation vanishes. We begin our dynamical description with a study of the low-latitude zonally symmetric atmospheric circulation.

11.2 A STEADY MODEL OF THE HADLEY CELL

Ceci n'est pas une pipe.

René Magritte (1898–1967), title of painting.

11.2.1 Assumptions

Let us try to construct a zonally symmetric model of the Hadley Cell,⁴ recognizing that such a model is likely applicable mainly to the tropical atmosphere, this being more zonally symmetric than the mid-latitudes. We will suppose that heating is maximum at the equator, and our intuitive picture, drawing on the observed flow of Fig. 11.3, is of air rising at the equator and moving polewards at some height H , descending at some latitude ϑ_H , and returning equatorwards near the surface. We will make three major assumptions:

- (i) that the circulation is steady;
- (ii) that the polewards moving air conserves its axial angular momentum, whereas the zonal flow associated with the near-surface, equatorwards moving flow is frictionally retarded and is weak;
- (iii) that the circulation is in thermal wind balance.

We also assume the model is symmetric about the equator (an assumption we relax in section 11.4). These are all reasonable assumptions, but they cannot be rigorously justified; in other words, we are constructing a *model* of the Hadley Cell, schematically illustrated in Fig. 11.4. The model defines a limiting case — steady, inviscid, zonally-symmetric flow — that cannot be expected to describe the atmosphere quantitatively, but that can be analysed fairly completely. Another limiting case, in which eddies play a significant role, is described in section 11.5. The real atmosphere may defy such simple characterizations, but the two limits provide invaluable benchmarks of understanding.

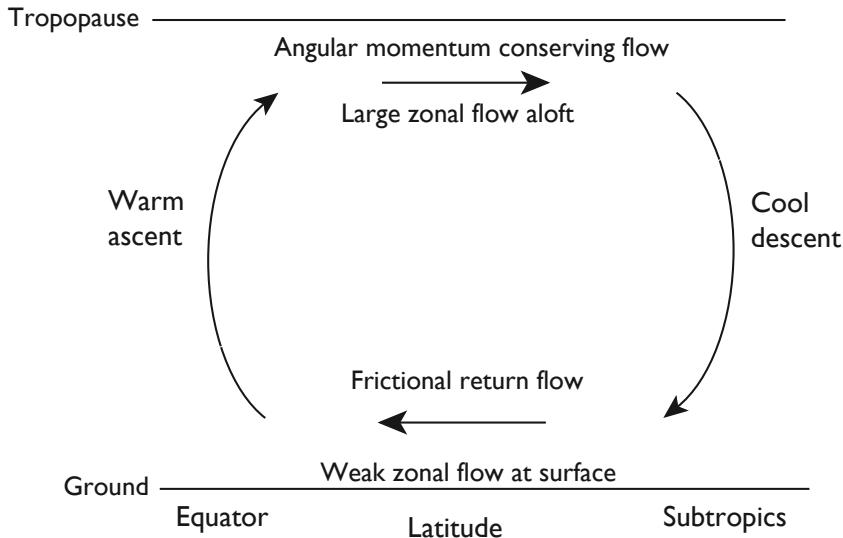


Fig. 11.4 A simple model of the Hadley Cell. Rising air near the equator moves polewards near the tropopause, descending in the subtropics and returning near the surface. The polewards moving air conserves its axial angular momentum, leading to a zonal flow that increases away from the equator. By the thermal wind relation the temperature of the air falls as it moves polewards, and to satisfy the thermodynamic budget it sinks in the subtropics. The return flow at the surface is frictionally retarded and small.

11.2.2 Dynamics

We now try to determine the strength and poleward extent of the Hadley circulation in our steady model. For simplicity we will work with a Boussinesq atmosphere, but this is not an essential aspect. We will first derive the conditions under which conservation on angular momentum will hold, and then determine the consequences of that.

The zonally averaged zonal momentum equation may be easily derived from (2.50a) and/or (2.62) and in the absence of friction it is

$$\frac{\partial \bar{u}}{\partial t} - (f + \bar{\zeta})\bar{v} + \bar{w}\frac{\partial \bar{u}}{\partial z} = -\frac{1}{a \cos^2 \theta} \frac{\partial}{\partial \theta} (\cos^2 \theta \bar{u}' \bar{v}') - \frac{\partial \bar{u}' \bar{w}'}{\partial z}, \quad (11.4)$$

where $\bar{\zeta} = -(\alpha \cos \theta)^{-1} \partial_{\theta} (\bar{u} \cos \theta)$ and the overbars represent zonal averages. If we neglect the vertical advection and the eddy terms on the right-hand side, then a steady solution, if it exists, obeys

$$(f + \bar{\zeta})\bar{v} = 0. \quad (11.5)$$

Presuming that the meridional flow \bar{v} is non-zero (an issue we address in section 11.2.8) then $f + \bar{\zeta} = 0$, or equivalently

$$2\Omega \sin \theta = \frac{1}{a} \frac{\partial \bar{u}}{\partial \theta} - \frac{\bar{u} \tan \theta}{a}. \quad (11.6)$$

At the equator we shall assume that $\bar{u} = 0$, because here parcels have risen from the surface

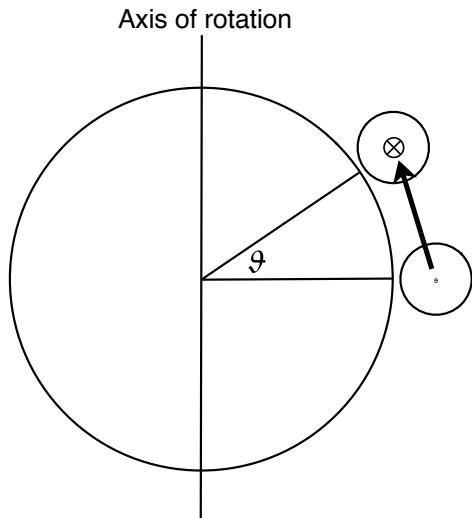


Fig. 11.5 If a ring of air at the equator moves polewards it moves closer to the axis of rotation. If the parcels in the ring conserve their angular momentum their zonal velocity must increase; thus, if $m = (\bar{u} + \Omega a \cos \vartheta) a \cos \vartheta$ is preserved and $\bar{u} = 0$ at $\vartheta = 0$ we recover (11.7).

where, by assumption, the flow is weak. Equation (11.6) then has a solution of

$$\bar{u} = \Omega a \frac{\sin^2 \vartheta}{\cos \vartheta} \equiv U_M. \quad (11.7)$$

This gives the zonal velocity of the polewards moving air in the upper branch of the (model) Hadley Cell, above the frictional boundary layer. We can derive (11.7) directly from the conservation of axial angular momentum, m , of a parcel of air at a latitude ϑ . In the shallow atmosphere approximation we have [cf. (2.64) and equations following]

$$\bar{m} = (\bar{u} + \Omega a \cos \vartheta) a \cos \vartheta, \quad (11.8)$$

and if $\bar{u} = 0$ at $\vartheta = 0$ and if \bar{m} is conserved on a polewards moving parcel, then (11.8) leads to (11.7). It also may be directly checked that

$$f + \bar{\zeta} = -\frac{1}{a^2 \cos \vartheta} \frac{\partial \bar{m}}{\partial \vartheta}. \quad (11.9)$$

We have thus shown that, if eddy fluxes and frictional effects are negligible, the poleward flow will conserve its angular momentum, the result of which, by (11.7), is that magnitude of the zonal flow in the Earth's rotating frame will increase with latitude (see Fig. 11.5). (Also, given the absence of eddies our model is zonally symmetric and we shall drop the overbars over the variables.)

If (11.7) gives the zonal velocity in the upper branch of the Hadley Cell, and that in the lower branch is close to zero, then the thermal wind equation can be used to infer the vertically averaged temperature. Although the geostrophic wind relation is not valid at the equator (a more accurate balance is the so-called cyclostrophic balance, $fu + u^2 \tan \vartheta/a = -a^{-1} \partial \phi / \partial \vartheta$) the zonal wind is in fact geostrophically balanced until very close to the equator, and at the equator itself the horizontal temperature gradient in our model vanishes, because of the assumed interhemispheric symmetry. Thus, conventional thermal wind

balance suffices for our purposes, and this is

$$2\Omega \sin \vartheta \frac{\partial u}{\partial z} = -\frac{1}{a} \frac{\partial b}{\partial \vartheta}, \quad (11.10)$$

where $b = g \delta\theta / \theta_0$ is the buoyancy and $\delta\theta$ is the deviation of potential temperature from a constant reference value θ_0 . (Be reminded that θ is potential temperature, whereas ϑ is latitude.) Vertically integrating from the ground to the height H where the outflow occurs and substituting (11.7) for u yields

$$\frac{1}{a\theta_0} \frac{\partial \theta}{\partial \vartheta} = -\frac{2\Omega^2 a}{gH} \frac{\sin^3 \vartheta}{\cos \vartheta}, \quad (11.11)$$

where $\theta = H^{-1} \int_0^H \delta\theta \, dz$ is the vertically averaged potential temperature. If the latitudinal extent of the Hadley Cell is not too great we can make the small-angle approximation, and replace $\sin \vartheta$ by ϑ and $\cos \vartheta$ by one, then integrating (11.11) gives

$$\theta = \theta(0) - \frac{\theta_0 \Omega^2 \vartheta^4}{2gH a^2}, \quad (11.12)$$

where $\vartheta = a\vartheta$ and $\theta(0)$ is the potential temperature at the equator, as yet unknown. Away from the equator, the zonal velocity given by (11.7) increases rapidly polewards and the temperature correspondingly drops. How far polewards is this solution valid? And what determines the value of the integration constant $\theta(0)$? To answer these questions we turn to thermodynamics.

11.2.3 Thermodynamics

In the above discussion, the temperature field is slaved to the momentum field in that it seems to follow passively from the dynamics of the momentum equation. Nevertheless, the thermodynamic equation must still be satisfied. Let us assume that the thermodynamic forcing can be represented by a Newtonian cooling to some specified radiative equilibrium temperature, θ_E ; this is a severe simplification, especially in equatorial regions where the release of heat by condensation is important. The thermodynamic equation is then

$$\frac{D\theta}{Dt} = \frac{\theta_E - \theta}{\tau}, \quad (11.13)$$

where τ is a relaxation time scale, perhaps a few weeks. Let us suppose that θ_E falls monotonically from the equator to the pole, and that it increases linearly with height, and a simple representation of this is

$$\frac{\theta_E(\vartheta, z)}{\theta_0} = 1 - \frac{2}{3} \Delta_H P_2(\sin \vartheta) + \Delta_V \left(\frac{z}{H} - \frac{1}{2} \right), \quad (11.14)$$

where Δ_H and Δ_V are non-dimensional constants that determine the fractional temperature difference between the equator and the pole, and the ground and the top of the fluid, respectively. P_2 is the second Legendre polynomial, and it is usually the leading term in the Taylor expansion of symmetric functions (symmetric around the equator) that decrease from

pole to equator; it also integrates to zero over the sphere. $P_2(y) = (3y^2 - 1)/2$, so that in the small-angle approximation and at $z = H/2$, or for the vertically averaged field, we have

$$\frac{\theta_E}{\theta_0} = 1 + \frac{1}{3}\Delta_H - \Delta_H \left(\frac{y}{a}\right)^2 \quad (11.15)$$

or

$$\theta_E = \theta_{E0} - \Delta\theta \left(\frac{y}{a}\right)^2, \quad (11.16)$$

where θ_{E0} is the equilibrium temperature at the equator, $\Delta\theta$ determines the equator-pole radiative-equilibrium temperature difference, and

$$\theta_{E0} = \theta_0(1 + \Delta_H/3), \quad \Delta\theta = \theta_0\Delta_H. \quad (11.17)$$

Now, let us suppose that the solution (11.12) is valid between the equator and a latitude ϑ_H where $v = 0$, so that within this region the system is essentially closed. Conservation of potential temperature then requires that the solution (11.12) must satisfy

$$\int_0^{Y_H} \theta \, dy = \int_0^{Y_H} \theta_E \, dy, \quad (11.18)$$

where $Y_H = a\vartheta_H$ is as yet undetermined. Polewards of this, the solution is just $\theta = \theta_E$. Now, we may demand that the solution be continuous at $y = Y_H$ (without temperature continuity the thermal wind would be infinite) and so

$$\theta(Y_H) = \theta_E(Y_H). \quad (11.19)$$

The constraints (11.18) and (11.19) determine the values of the unknowns $\theta(0)$ and Y_H . A little algebra (problem 11.1) gives

$$Y_H = \left(\frac{5\Delta\theta g H}{3\Omega^2 \theta_0}\right)^{1/2}, \quad (11.20)$$

and

$$\theta(0) = \theta_{E0} - \left(\frac{5\Delta\theta^2 g H}{18a^2 \Omega^2 \theta_0}\right). \quad (11.21)$$

A useful non-dimensional number that parameterizes these solutions is

$$R \equiv \frac{g H \Delta\theta}{\theta_0 \Omega^2 a^2} = \frac{g H \Delta_H}{\Omega^2 a^2}, \quad (11.22)$$

which is the square of the ratio of the speed of shallow water waves to the rotational velocity of the Earth, multiplied by the fractional temperature difference from equator to pole. Typical values for the Earth's atmosphere are a little less than 0.1. In terms of R we have

$$Y_H = a \left(\frac{5}{3}R\right)^{1/2}, \quad (11.23)$$

and

$$\theta(0) = \theta_{E0} - \left(\frac{5}{18}R\right) \Delta\theta. \quad (11.24)$$

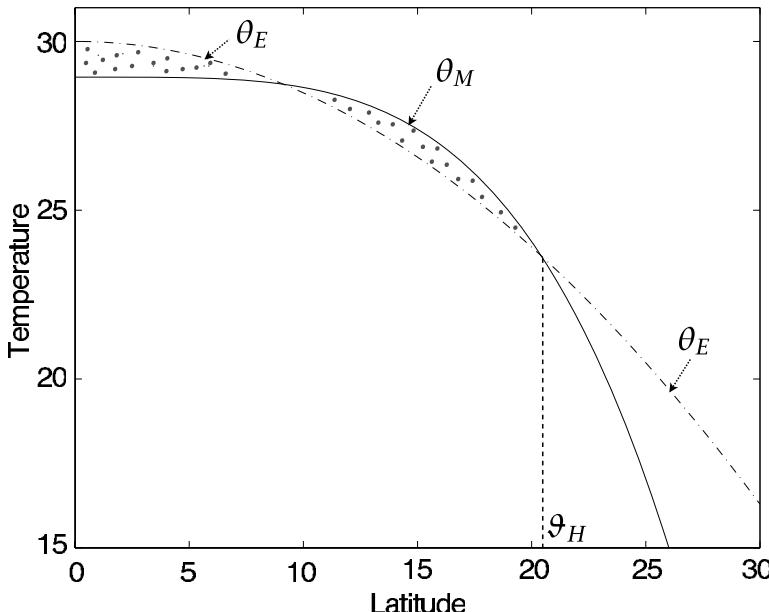


Fig. 11.6 The radiative equilibrium temperature (θ_E , dashed line) and the angular-momentum-conserving solution (θ_M , solid line) as a function of latitude. The two dotted regions have equal areas. The parameters are: $\theta_{EO} = 303\text{ K}$, $\Delta\theta = 50\text{ K}$, $\theta_0 = 300\text{ K}$, $\Omega = 7.272 \times 10^{-5}\text{ s}^{-1}$, $g = 9.81\text{ m s}^{-2}$, $H = 10\text{ km}$. These give $R = 0.076$ and $Y_H/a = 0.356$, corresponding to $\theta_H = 20.4^\circ$.

The solution, (11.12) with $\theta(0)$ given by (11.24) is plotted in Fig. 11.6. Perhaps the single most important aspect of the model is that it predicts that the Hadley Cell has a *finite* meridional extent, *even for an atmosphere that is completely zonally symmetric*. The baroclinic instability that does occur in mid-latitudes is not necessary for the Hadley Cell to terminate in the subtropics, although it may be an important factor, or even the determining factor, in the real world.

11.2.4 Zonal wind

The angular-momentum-conserving zonal wind is given by (11.7), which in the small-angle approximation becomes

$$U_M = \Omega \frac{y^2}{a}. \quad (11.25)$$

This relation holds for $y < Y_H$. The zonal wind corresponding to the radiative-equilibrium solution is given using thermal wind balance and (11.16), which leads to

$$U_E = \Omega a R. \quad (11.26)$$

That the radiative-equilibrium zonal wind is a constant follows from our choice of the second Legendre function for the radiative equilibrium temperature and is not a fundamental result; nonetheless, for most reasonable choices of θ_E the corresponding zonal wind will vary

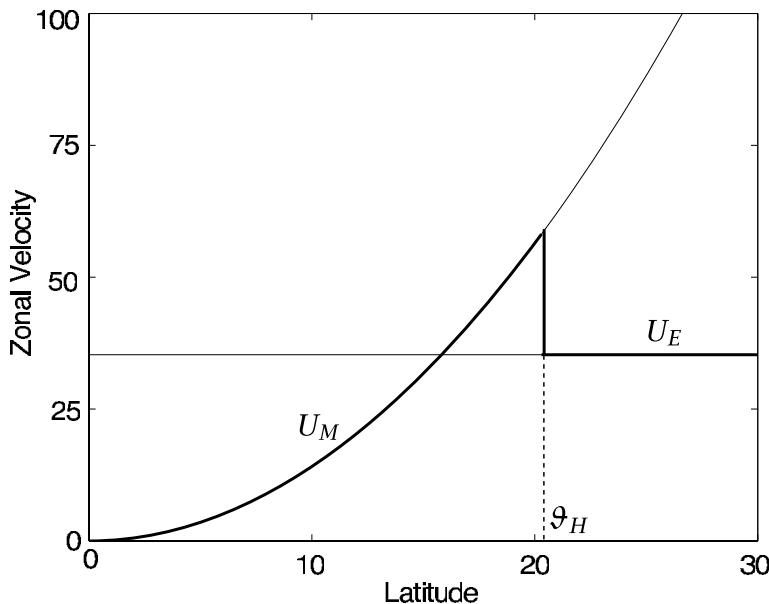


Fig. 11.7 The zonal wind corresponding to the radiative equilibrium temperature (U_E), and the angular-momentum-conserving solution (U_M) as a function of latitude, given (11.25) and (11.26) respectively. The parameters are the same as those of Fig. 11.6, and the radiative equilibrium wind, U_E is a constant, ΩaR . The actual zonal wind (in the model) follows the thick solid line: $u = U_M$ for $\vartheta < \vartheta_H$ ($\gamma < Y_H$), and $u = U_E$ for $\vartheta > \vartheta_H$ ($\gamma > Y_H$).

much less than the angular-momentum-conserving wind (11.25). The winds are illustrated in Fig. 11.7. There is a discontinuity in the zonal wind at the edge of the Hadley Cell, and of the meridional temperature gradient, but not of the temperature itself.

11.2.5 Properties of solution

From (11.23) we can see that the model predicts that the latitudinal extent of the Hadley Cell is:

- * proportional to the square root of the meridional radiative equilibrium temperature gradient: the stronger the gradient, the farther the circulation must extend to achieve thermodynamic balance via the equal-area construction in Fig. 11.6;
- * proportional to the square root of the height of the outward flowing branch: the higher the outward flowing branch, the weaker the ensuing temperature gradient of the solution (via thermal wind balance), and so the further polewards the circulation must go;
- * inversely proportional to the rotation rate Ω : the stronger the rotation rate, the stronger the angular-momentum-conserving wind, the stronger the ensuing temperature gradient and so the more compact the circulation.

These precise dependencies on particular powers of parameters are not especially significant in themselves, nor are they robust to changes in parameters. For example, were we to choose a meridional distribution of radiative equilibrium temperature different from (11.14) we might find different exponents in some of the solutions, although we would expect the same qualitative dependencies. However, the dependencies do provide predictions that may be tested with a numerical model. Also, as we have already noted, a key property of the model is that it predicts that the Hadley Cell has a finite meridional extent, even in the absence of mid-latitude baroclinic instability.

Another interesting property of the solutions is a discontinuity in the zonal wind. For tropical latitudes (i.e., $y < Y_H$), then $\bar{u} = U_M$ (the constant angular momentum solution), whereas for $y > Y_H$, $\bar{u} = U_E$ (the thermal wind associated with radiative equilibrium temperature θ_E). There is therefore a discontinuity of \bar{u} at $y = Y_H$, because u is related to the meridional gradient of θ which changes discontinuously, even though θ itself is continuous. No such discontinuity is observed in the real world, although one may observe a baroclinic jet at the edge of the Hadley Cell.

11.2.6 Strength of the circulation

We can make an estimate of the strength of the Hadley Cell by consideration of the thermodynamic equation at the equator, namely

$$w \frac{\partial \theta}{\partial z} \approx \frac{\theta_{E0} - \theta}{\tau}, \quad (11.27)$$

this being a balance between adiabatic cooling and radiative heating. If the static stability is determined largely by the forcing, and not by the meridional circulation itself, then

$$\frac{1}{\theta_0} \frac{\partial \theta}{\partial z} \approx \frac{\Delta_V}{H}, \quad (11.28)$$

and (11.27) gives

$$w \approx \frac{H}{\theta_0 \Delta_V} \frac{\theta_{E0} - \theta}{\tau}. \quad (11.29)$$

Thus, the strength of the circulation is proportional to the distance of the solution from the radiative equilibrium temperature. The right-hand side of (11.27) can be evaluated from the solution itself, and from (11.24) we have

$$\frac{\theta_{E0} - \theta}{\tau} = \frac{5R\Delta\theta}{18\tau}. \quad (11.30)$$

The vertical velocity is then given by

$$w \approx \frac{5R\Delta\theta H}{18\tau\Delta_V\theta_0} = \frac{5R\Delta_H H}{18\tau\Delta_V}. \quad (11.31)$$

Using mass continuity we can transform this into an estimate for the meridional velocity. Thus, if we let $(v/Y_H) \sim (w/H)$ and use (11.23), we obtain

$$v \sim \frac{R^{3/2} a \Delta_H}{\tau \Delta_V} \propto \frac{\Delta_H^{5/2}}{\Delta_V}, \quad (11.32)$$

and the mass flux, or the meridional overturning stream function Ψ , of the circulation scales as

$$\Psi \sim vH \sim \frac{R^{3/2} a H \Delta_H}{\tau \Delta_V} \propto (\Delta\theta)^{5/2}. \quad (11.33)$$

This evidently increases fairly rapidly as the gradient of the radiative equilibrium temperature increases. The characteristic overturning time of the circulation, τ_d is then

$$\tau_d = \frac{H}{w} \sim \frac{\tau \Delta_V}{R \Delta_H}. \quad (11.34)$$

We require $\tau_d/\tau \gg 1$ for the effects of the circulation on the static stability to be small and therefore $\Delta_V/(R\Delta_H) \gg 1$, or equivalently, using (11.17),

$$\theta_0 \Delta_V \gg R(\theta_{E0} - \theta_0). \quad (11.35)$$

If the converse were true, and $\tau \gg \tau_d$, then the potential temperature would be nearly conserved as a parcel ascended in the rising branch of the Hadley Cell, and the static stability would be nearly neutral.

11.2.7 † Effects of moisture

Suppose now that moisture is present, but that the Hadley Cell remains a self-contained system; that is, it neither imports nor exports moisture. We envision that water vapour joins the circulation by way of evaporation from a saturated surface into the equatorward, lower branch of the Hadley Cell, and that this water vapour then condenses in and near the upward branch of the cell. The latent heat released by condensation is exactly equal to the heat required to evaporate moisture from the surface, and no heat is lost or gained to the system. However, the heating *distribution* is changed from the dry case, becoming a strong function of the solution itself and likely to have a sharp maximum near the equator. Even if we were to try to parameterize the latent heat release by simply choosing a flow dependent radiative equilibrium temperature, the resulting problem would still be quite nonlinear and a general analytic solution seems out of our reach.⁵

Nevertheless, we may see quite easily the qualitative features of moisture, at least within the context of this model. The meridional distribution of temperature is still given by way of thermal wind balance with an angular-momentum-conserving zonal wind, and so is still given by (11.12). We may also assume that the meridional extent of the Hadley Cell is unaltered; that is, a solution exists with circulation confined to $\vartheta < \vartheta_H$ (although it may not be the unique solution). Then, if θ_E^* is the effective radiative equilibrium temperature of the moist solution, we have that $\theta_E^*(Y_H) = \theta_E(Y_H)$ and, in the small-angle approximation,

$$\int_0^{Y_H} \theta \, dy = \int_0^{Y_H} \theta_E^* \, dy = \int_0^{Y_H} \theta_E \, dy, \quad (11.36)$$

where the first equality holds because it defines the solution, and the second equality holds because moisture provides no net energy source. Because condensation will occur mainly in the upward branch of the Hadley Cell, θ_E^* will be peaked near the equator, as schematically sketched in Fig. 11.8. This construction makes it clear that the main difference between the dry and moist solutions is that the latter has a more intense overturning circulation,

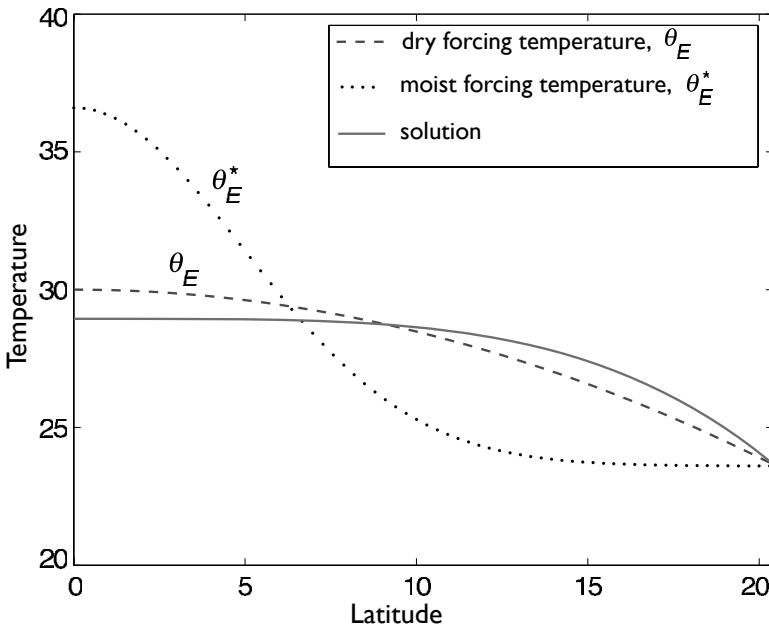


Fig. 11.8 Schema of the effects of moisture on a model of the Hadley Cell. The temperature of the solution (solid line) is the same as that of a dry model, because this is determined from the angular-momentum-conserving wind. The heating distribution (as parameterized by a forcing temperature) is peaked near the equator in the moist case, leading to a more vigorous overturning circulation.

because, from (11.27), the circulation increases with the temperature difference between the solution and the forcing temperature. Concomitantly, our intuition suggests that the upward branch of the moist Hadley circulation will become much narrower and more intense than the downward branch because of the enhanced efficiency of moist convection, and these expectations are generally confirmed by numerical integrations of the moist equations of motion.

11.2.8 The radiative equilibrium solution

Instead of a solution given by (11.12), could the temperature not simply be in radiative equilibrium everywhere? Such a state would have no meridional overturning circulation and the zonal velocity would be determined by thermal wind balance; that is,

$$v = 0, \quad \theta = \theta_E, \quad f \frac{u}{H} = -g \frac{\partial}{\partial y} \left(\frac{\theta_E}{\theta_0} \right). \quad (11.37)$$

To answer this question we consider the steady zonally symmetric zonal angular momentum equation with viscosity; that is, the zonally averaged, viscous, steady, shallow atmosphere version of (2.68), namely

$$\frac{1}{a \cos \theta} \frac{\partial}{\partial \theta} (v m \cos \theta) + \frac{\partial (m w)}{\partial z} = \frac{v}{a \cos \theta} \frac{\partial}{\partial \theta} \left(\cos^2 \theta \frac{\partial}{\partial \theta} \frac{u}{\cos \theta} \right) + v a \cos \theta \frac{\partial^2 u}{\partial z^2}, \quad (11.38)$$

where the variables vary only in the ϑ - z , or γ - z , plane. The viscous term on the right-hand side arises from the expansion in spherical coordinates of the Laplacian. Note that it is angular velocity, not the angular momentum, that is diffused, because there is no diffusion of the angular momentum due to the Earth's rotation. However, to a very good approximation, the viscous term will be dominated by vertical derivatives and we may then write (11.38) as

$$\nabla_x \cdot (\mathbf{v} m) = \nu \frac{\partial^2 m}{\partial z^2}. \quad (11.39)$$

where $\nabla_x \cdot$ is the divergence in the meridional plane. The right-hand side now has a diffusive form, and in section 10.5.1 we showed that variables obeying equations like this can have no extrema within the fluid. Thus, there can be no maximum or minimum of angular momentum in the interior of the fluid, a result sometimes called Hide's theorem.⁶ In effect, diffusion always acts to smooth away an isolated extremum, and this cannot be counterbalanced by advection. The result also implies that there cannot be any interior extrema in a statistically steady state if there is any zonally asymmetric eddy motion that transports angular momentum downgradient.

If the viscosity were so large that the viscous term were dominant in (11.38), then the fluid would evolve toward a state of solid body rotation, this being the fluid state with no internal stresses. In that case, there would be a maximum of angular momentum at the equator — a state of 'super-rotation'. (Related mechanisms have been proposed for the maintenance of super-rotation on Venus.⁷)

A maxima of m can, however, occur at the surface, even with a viscous term like that of (11.39). Suppose we add a surface stress to the right-hand side of (11.39), and that this stress acts in the opposite direction to that of the zonal wind. Then, in a region of surface easterlies the surface stress contribution would be positive, acting as a source of positive angular momentum; there then exists the possibility that this will exactly balance the diffusion term allowing a maximum of m to occur. However, such a surface stress does not allow this maximum to be a region of surface westerlies at the equator, because the stress would then act in the same way as the diffusion and reduce the angular momentum.

Returning now to the question posed at the head of this section, suppose that the radiative equilibrium solution does hold. Then a radiative equilibrium temperature decreasing away from the equator more rapidly than the angular-momentum-conserving solution θ_M implies, using thermal wind balance, a maximum of m at the equator and above the surface, in violation of the no-extremum principle. Of course, we have derived the angular-momentum-conserving solution in the inviscid limit, in which the no-extrema principle does not apply. But any small viscosity will make the radiative equilibrium solution completely invalid, but potentially have only a small effect on the angular-momentum-conserving solution; that is, in the *limit of small viscosity* the angular-momentum-conserving solution can conceivably hold approximately, at least in the absence of boundary layers, whereas the radiative equilibrium solution cannot.

However, if the radiative equilibrium temperature varies more slowly with latitude than the temperature corresponding to the angular momentum conserving solution then a radiative equilibrium solution *can* pertain, without violating Hide's theorem. In particular, this is the case if $\theta_E \propto P_4(\sin \vartheta)$, where P_4 is the fourth Legendre polynomial, and so the possibility exists of two equilibrium solutions for the same forcing; however, P_4 is an unrealistically flat radiative equilibrium temperature for the Earth's atmosphere.

11.3 A SHALLOW WATER MODEL OF THE HADLEY CELL

Although expressed in the notation of the primitive equations, the model described above takes no account of any vertical structure in its stratification and is, *de facto*, a shallow water model. Furthermore, the geometric aspects of sphericity play no essential role. Thus, we may transparently express the essence of the model by:

- (i) explicitly using the shallow water equations instead of the stratified equations;
- (ii) using the equatorial β -plane, with $f = f_0 + \beta y$ and $f_0 = 0$.

Let us therefore, as an exercise, construct a reduced-gravity model with an active upper layer overlying a stationary lower layer.

11.3.1 Momentum balance

The inviscid zonal momentum equation of the upper layer is

$$\frac{Du}{Dt} - \beta y v = 0 \quad (11.40)$$

or

$$\frac{D}{Dt} \left(u - \frac{\beta y^2}{2} \right) = 0, \quad (11.41)$$

which is the β -plane analogue of the conservation of axial angular momentum. (In this section, all variables are zonally averaged, but we omit any explicit notation denoting this.) From (11.41) we obtain the zonal wind as a function of latitude,

$$u = \frac{1}{2} \beta y^2 + A, \quad (11.42)$$

where A is a constant, which is zero if $u = 0$ at the ‘equator’, $y = 0$. The flow given by (11.42) is then analogous to the angular momentum conserving flow in the spherical model, (11.7). Because the lower layer is stationary, the analogue of thermal wind balance in the stratified model is just geostrophic balance, namely

$$f u = -g' \frac{\partial h}{\partial y}, \quad (11.43)$$

where h is the thickness of the active upper layer. Using (11.43) and $f = \beta y$ we obtain

$$g' \frac{\partial h}{\partial y} = -\frac{1}{2} \beta^2 y^3 \quad (11.44)$$

giving

$$h = -\frac{1}{8g'} \beta^2 y^4 + h(0) \quad (11.45)$$

where $h(0)$ is the value of h at $y = 0$.

11.3.2 Thermodynamic balance

The thermodynamic equation in the shallow water equations is just the mass conservation equation, which we write as

$$\frac{Dh}{Dt} = -\frac{1}{\tau} (h - h^*), \quad (11.46)$$

where the right-hand side represents heating — h^* is the field to which the height relaxes on a time scale τ . For illustrative purposes we will choose

$$h^* = h_0(1 - \alpha|\gamma|). \quad (11.47)$$

(If we chose the more realistic quadratic dependence on γ , the model would be more similar to that of the previous section.) To be in thermodynamic equilibrium we require that the right-hand side integrates to zero over the Hadley Cell; that is

$$\int_0^Y (h - h^*) \, d\gamma = 0 \quad (11.48)$$

where Y is the latitude of the poleward extent of the Hadley Cell, thus far unknown. Polewards of this, the height field is simply in equilibrium with the forcing — there is no meridional motion and $h = h^*$. Since the height field must be continuous, we require that

$$h(Y) = h^*(Y). \quad (11.49)$$

The two constraints (11.48) and (11.49) provide values of the unknowns $h(0)$ and Y , and give

$$Y = \left(\frac{5h_0\alpha g'}{\beta^2} \right)^{1/3}, \quad (11.50)$$

which is analogous to (11.20), as well as an expression for $h(0)$ that we leave as a problem for the reader. The qualitative dependence on the parameters is similar to that of the full model, although the latitudinal extent of the Hadley Cell is proportional to the cube root of the meridional thickness gradient α .

11.4 † ASYMMETRY AROUND THE EQUATOR

The Sun is overhead at the equator but two days out of the year, and in this section we investigate the effects that asymmetric heating has on the Hadley circulation. Observations indicate except for the brief periods around the equinoxes, the circulation is dominated by a single cell with rising motion centred in the summer hemisphere, but extending well into the winter hemisphere. That is, as seen in Fig. 11.3, the ‘winter cell’ is broader and stronger than the ‘summer cell’, and it behoves us to try to explain this. We will stay in the framework of the inviscid angular-momentum model of section 11.2, changing only the forcing field to represent the asymmetry and being a little more attentive to the details of spherical geometry.⁸

To represent an asymmetric heating we may choose a radiative equilibrium temperature of the form

$$\begin{aligned} \frac{\theta_E(\vartheta, z)}{\theta_0} &= 1 - \frac{2}{3} \Delta_H P_2(\sin \vartheta - \sin \vartheta_0) + \Delta_V \left(\frac{z}{H} - \frac{1}{2} \right) \\ &= 1 + \frac{\Delta_H}{3} \left[1 - 3(\sin \vartheta - \sin \vartheta_0)^2 \right] + \Delta_V \left(\frac{z}{H} - \frac{1}{2} \right). \end{aligned} \quad (11.51)$$

This is similar to (11.14), but now the forcing temperature falls monotonically from a specified latitude ϑ_0 . If $\vartheta_0 = 0$ the model is identical to the earlier one, but if not we envision

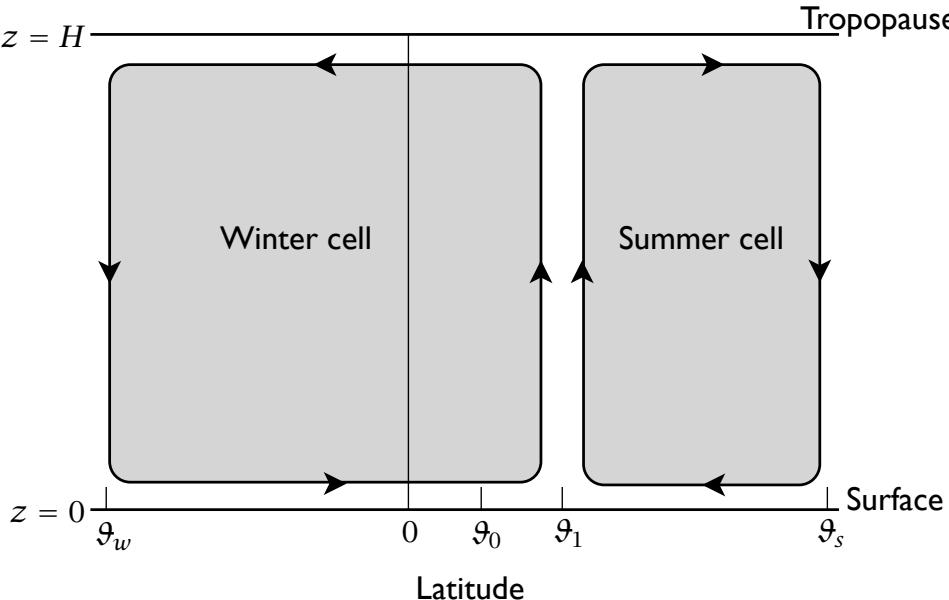


Fig. 11.9 Schematic of a Hadley circulation model when the heating is centred off the equator, at a latitude ϑ_0 . The lower level convergence occurs at a latitude ϑ_1 that is not in general equal to ϑ_0 . The resulting winter Hadley Cell is stronger and wider than the summer cell.

a circulation as qualitatively sketched in Fig. 11.9, with rising motion off the equator at some latitude ϑ_1 , extending into the winter hemisphere to a latitude ϑ_w , and into the summer hemisphere to ϑ_s . We will discover that, in general, $\vartheta_1 \neq \vartheta_0$ except when $\vartheta_0 = 0$. Following our procedure we used in the symmetric case as closely as possible, we then make the following assumptions.

- (i) The flow is quasi-steady. That is, for any given time of year the flow adjusts to a steady circulation on a time scale more rapid than that on which the solar zenith angle appreciably changes. Then, even though the forcing is time-dependent, we neglect local time derivatives in the momentum and thermodynamic equations.
- (ii) The flows in the upper branches conserve angular momentum, m . Further assuming that $u = 0$ at $\vartheta = \vartheta_1$ so that $m = \Omega a^2 \cos^2 \vartheta_1$ we obtain

$$u(\vartheta) = \frac{\Omega a (\cos^2 \vartheta_1 - \cos^2 \vartheta)}{\cos \vartheta}. \quad (11.52)$$

Thus, we expect to see westward (negative) winds aloft at the equator. In the lower branches the zonal flow is assumed to be approximately zero, i.e., $u(0) \approx 0$.

- (iii) The flow satisfies cyclostrophic and hydrostatic balance. Cyclostrophic balance in the meridional momentum equation is

$$fu + \frac{u^2 \tan \vartheta}{a} = -\frac{1}{a} \frac{\partial \phi}{\partial \vartheta}, \quad (11.53)$$

and because the flow crosses the equator we cannot neglect the second term on the left-hand side. Combining this with hydrostatic balance ($\partial\phi/\partial z = g\delta\theta/\theta_0$) leads to a generalized thermal wind balance, which may be written as

$$m \frac{\partial m}{\partial z} = -\frac{ga^2 \cos^2 \vartheta}{2\theta_0 \tan \vartheta} \frac{\partial \theta}{\partial \vartheta}, \quad (11.54)$$

where here and henceforth we write θ in place of $\delta\theta$. If the undifferentiated m is approximated by $\Omega a^2 \cos^2 \vartheta$, this reduces to conventional thermal wind balance, (11.10). The form of (11.54) is useful because we are assuming that m is conserved, and so from it we can immediately infer the temperature distribution.

- (iv) Potential temperature in each cell is conserved when integrated over the extent of the cell. Thus,

$$\int_{\vartheta_1}^{\vartheta_s} (\theta - \theta_E) \cos \vartheta \, d\vartheta = 0, \quad \int_{\vartheta_1}^{\vartheta_w} (\theta - \theta_E) \cos \vartheta \, d\vartheta = 0, \quad (11.55)$$

for the summer and winter cells, respectively, where θ is the vertically averaged potential temperature.

- (v) Potential temperature is continuous at the edge of each cell, so that

$$\theta(\vartheta_s) = \theta_E(\vartheta_s), \quad \theta(\vartheta_w) = \theta_E(\vartheta_w), \quad (11.56)$$

and is also continuous at ϑ_1 . This last condition must be explicitly imposed in the asymmetric model, whereas in the symmetric model it holds by symmetry. Now, recall from the symmetric model that the value of the temperature at the equator was determined by the integral constraint (11.18) and the continuity constraint (11.19). We have analogues of these in each hemisphere [(11.55) and (11.56)] and thus, if ϑ_1 is set equal to ϑ_0 we cannot expect that they each would give the same temperature at ϑ_0 . Thus, ϑ_1 must be a free parameter to be determined.

Given these assumptions, the solution may be calculated. Using thermal wind balance, (11.54), with $m(H) = \Omega a^2 \cos^2 \vartheta_1$ and $m(0) = \Omega a^2 \cos^2 \vartheta$ we find

$$-\frac{1}{\theta_0} \frac{\partial \theta}{\partial \vartheta} = \frac{\Omega^2 a^2}{gH} \left(\frac{\sin \vartheta}{\cos^3 \vartheta} \cos^4 \vartheta_1 - \sin \vartheta \cos \vartheta \right), \quad (11.57)$$

which integrates to

$$\theta(\vartheta) - \theta(\vartheta_1) = -\frac{\theta_0 \Omega^2 a^2}{2gH} \frac{(\sin^2 \vartheta - \sin^2 \vartheta_1)^2}{\cos^2 \vartheta}. \quad (11.58)$$

The value of ϑ_1 , and the value of $\theta(\vartheta_1)$, are determined by the constraints (11.55) and (11.56). It is not in general possible to obtain a solution analytically, but one may be found numerically by an iterative procedure and one such is illustrated in Fig. 11.10. The zonal wind of the solution is always symmetric around the equator, because it is determined solely by angular momentum conservation. The temperature is therefore also symmetric, as (11.58) explicitly shows. However, the width of the solution in each hemisphere will, in general, be different. Furthermore, because the strength of the circulation increases with difference between the temperature of the solution and the radiative equilibrium temperature, the

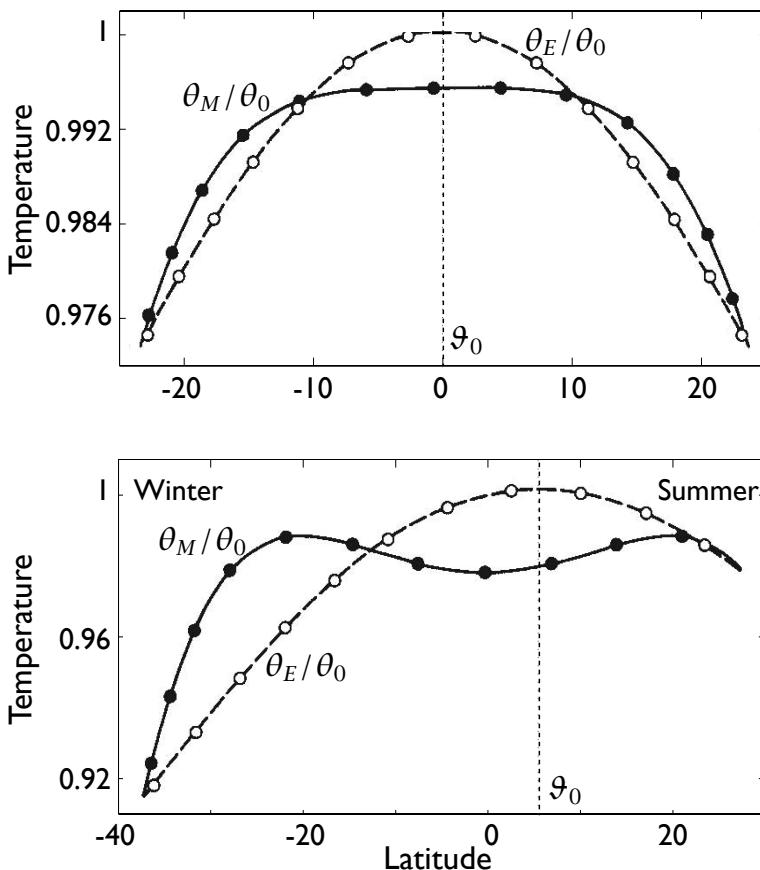


Fig. 11.10 Solutions of the Hadley Cell model with heating centred at the equator ($\vartheta_0 = 0^\circ$, top) and off the equator ($\vartheta_0 = +6^\circ N$, bottom), with $\Delta_H = 1/6$. The dashed line is the radiative equilibrium temperature and the solid line is the angular-momentum-conserving solution. In the lower panel, $\vartheta_1 \approx +18^\circ$, and the circulation is dominated by the cell extending from $+18^\circ$ to -36° .⁹

circulation in the winter hemisphere will also be much stronger than that in the summer, a prediction that is qualitatively consistent with the observations (see Fig. 11.3). More detailed calculations show that, because the strength of the model Hadley Cell increases nonlinearly with ϑ_0 , the time-average strength of the Hadley Cell with seasonal forcing is stronger than that produced by annually averaged forcing. However, this does not appear to be a feature of either the observations or more complete numerical simulations, suggesting that an angular-momentum-conserving model has, at the least, quantitative deficiencies, as follows.¹⁰

- (i) The lack of consideration of zonal asymmetries, such as monsoonal circulations.
- (ii) The quasi-steady assumption, given the presence of a temporally progressing seasonal cycle (even in presence of zonally symmetric boundary conditions). Because the latitude of the upward branch of the Hadley Cell varies with season, the value of the angular

momentum entering the system also varies, and so a homogenized value of angular momentum is hard to achieve. For quasi-steadiness to hold we require

$$\tau_s \gg \tau_e, \quad (11.59)$$

where τ_e is some dynamical equilibration time scale, similar to the dynamical timescale $\tau_d = H/w$ (section 11.2.6), and τ_s is the seasonal time scale

- (iii) The lack of angular momentum conservation in reality (a criticism that would also apply to a steady model with zonally symmetric boundary conditions). Such non-conservation will arise if either diffusion of momentum caused by small-scale turbulence, or the angular momentum transport by baroclinic eddies, are significant.

Nonetheless, the overall picture that the model paints, and its qualitative explanation of the strengthened and extended winter Hadley Cell, are invaluable aids to our understanding of the circulation.

11.5 EDDIES, VISCOSITY AND THE HADLEY CELL

So far, we have ignored the effects of baroclinic eddies on the Hadley circulation — ‘ignored’ rather than ‘neglected’, because we have no a priori or observational reason to believe that their effects will be negligible. If their effects are strong, then none of the models we discussed above will be quantitatively valid. With this in mind, in this section we look at the Hadley circulation from a quite different perspective, by supposing that the zonal momentum equation is linear, except for the effects of eddy fluxes on the right-hand side. Our approach is illustrative, not quantitative, and we again stay within the Boussinesq approximation.

We might expect eddy fluxes to be important because the angular momentum conserving solution will develop a large vertical shear and if this extends sufficiently far polewards it will become baroclinically unstable (compare Fig. 11.7 with the minimum shear needed for baroclinic instability sketched in Fig. 6.16). It is a quantitative issue as to whether the Hadley flow becomes strongly unstable before it reaches its poleward extent, and if it does not the angular-momentum-conserving solution might be expected to be a good one. But here let us assume that the flow is strongly unstable and that the ensuing instability transfers both heat and angular momentum polewards (the mechanisms of this are discussed in the next chapter). This transfer will lead to the nonconservation of angular momentum and, potentially, render invalid the models of the previous sections.

11.5.1 Qualitative considerations

The zonally averaged zonal momentum equation, (11.4), may be written as an equation for angular momentum, \bar{m} . Referring back to section 2.2 if needs be, the equation may be written as

$$\begin{aligned} \frac{\partial \bar{m}}{\partial t} + \frac{1}{\cos \vartheta} \frac{\partial}{\partial \gamma} (\bar{v} \bar{m} \cos \vartheta) + \frac{\partial}{\partial z} (\bar{w} \bar{m}) \\ = - \frac{1}{\cos \vartheta} \frac{\partial}{\partial \gamma} (\bar{m}' \bar{v}' \cos \vartheta) - \frac{\partial}{\partial z} (\bar{m}' \bar{w}') \\ = - \frac{1}{\cos \vartheta} \frac{\partial}{\partial \gamma} (\bar{u}' \bar{v}' a \cos^2 \vartheta) - \frac{\partial}{\partial z} (\bar{u}' \bar{w}' a \cos \vartheta), \end{aligned} \quad (11.60)$$

where $\bar{m} = (\bar{u} + \Omega a \cos \vartheta) a \cos \vartheta$, $m' = u' a \cos \vartheta$, $y = a\vartheta$, and the vertical and meridional velocities are related by the mass continuity relation

$$\frac{1}{\cos \vartheta} \frac{\partial}{\partial y} (\bar{v} \cos \vartheta) + \frac{\partial \bar{w}}{\partial z} = 0. \quad (11.61)$$

In the angular-momentum-conserving model the eddy fluxes were neglected and (11.60) was approximated by the simple expression $\partial \bar{m} / \partial y = 0$, and by construction the Rossby number is $\mathcal{O}(1)$, because $\zeta = -f$.

The observed eddy heat and momentum fluxes are shown in Fig. 11.11. The eddy momentum flux is generally polewards, converging in the region of the surface westerlies. Its magnitude, and more particularly its meridional gradient, is as large or larger than the momentum flux associated with the mean flow. Neglecting vertical advection and vertical eddy fluxes, and using (11.61), (11.60) may be written as

$$\frac{\partial \bar{m}}{\partial t} + \bar{v} \frac{\partial \bar{m}}{\partial y} = - \frac{1}{\cos \vartheta} \frac{\partial}{\partial y} (\bar{u}' \bar{v}' a \cos^2 \vartheta). \quad (11.62)$$

Thus, if $\bar{v} > 0$ (as in the upper branch of the Northern Hemisphere Hadley Cell) and the flow is steady, the observed eddy fluxes are such as to cause the angular momentum of the zonal flow to decrease as it moves polewards, and the zonal velocity is lower than it would be in the absence of eddies. (In the Southern Hemisphere the signs of v and the eddy momentum flux are reversed, but the dynamics are equivalent.) Note that we cannot a priori determine whether eddies are likely to be important by comparing the magnitudes of the eddy terms with the terms on the left-hand side of (11.62) in the angular-momentum-conserving solution, because in that solution these terms are individually zero. Rather, we should compare the eddy fluxes to $\bar{v} \partial m_e / \partial y$, where $m_e = \Omega a^2 \cos^2 \vartheta$ is the angular momentum of the solid Earth.

The eddy flux of heat will also affect the Hadley Cell, although in a different fashion. We see from Fig. 11.11 that the eddy flux of temperature is predominantly polewards, and therefore that eddies export heat from the subtropics to higher latitudes. Now, the zonally averaged thermodynamic equation may be written

$$\begin{aligned} \frac{\partial \bar{b}}{\partial t} + \frac{1}{\cos \vartheta} \frac{\partial}{\partial y} (\bar{v} \bar{b} \cos \vartheta) + \frac{\partial}{\partial z} (\bar{w} \bar{b}) \\ = - \frac{1}{\cos \vartheta} \frac{\partial}{\partial y} (\bar{v}' \bar{b}' \cos \vartheta) - \frac{\partial}{\partial z} (\bar{w}' \bar{b}') + Q[b]. \end{aligned} \quad (11.63)$$

where $Q[b]$ represents the heating. After vertical averaging, the vertical advection terms vanish and the resulting equation is the thermodynamic equation implicitly used in the angular-momentum-conserving model, with the addition of the meridional eddy flux on the right-hand side. A diverging eddy heat flux in the subtropics (as in Fig. 11.11) is plainly equivalent to increasing the meridional gradient of the radiative equilibrium temperature, and therefore will increase the intensity of the overturning circulation.

11.5.2 An idealized eddy-driven model

Consider now the extreme case of an ‘eddy-driven’ Hadley Cell. (The driving for the Hadley Cell, and the atmospheric circulation in general, ultimately comes from the differential

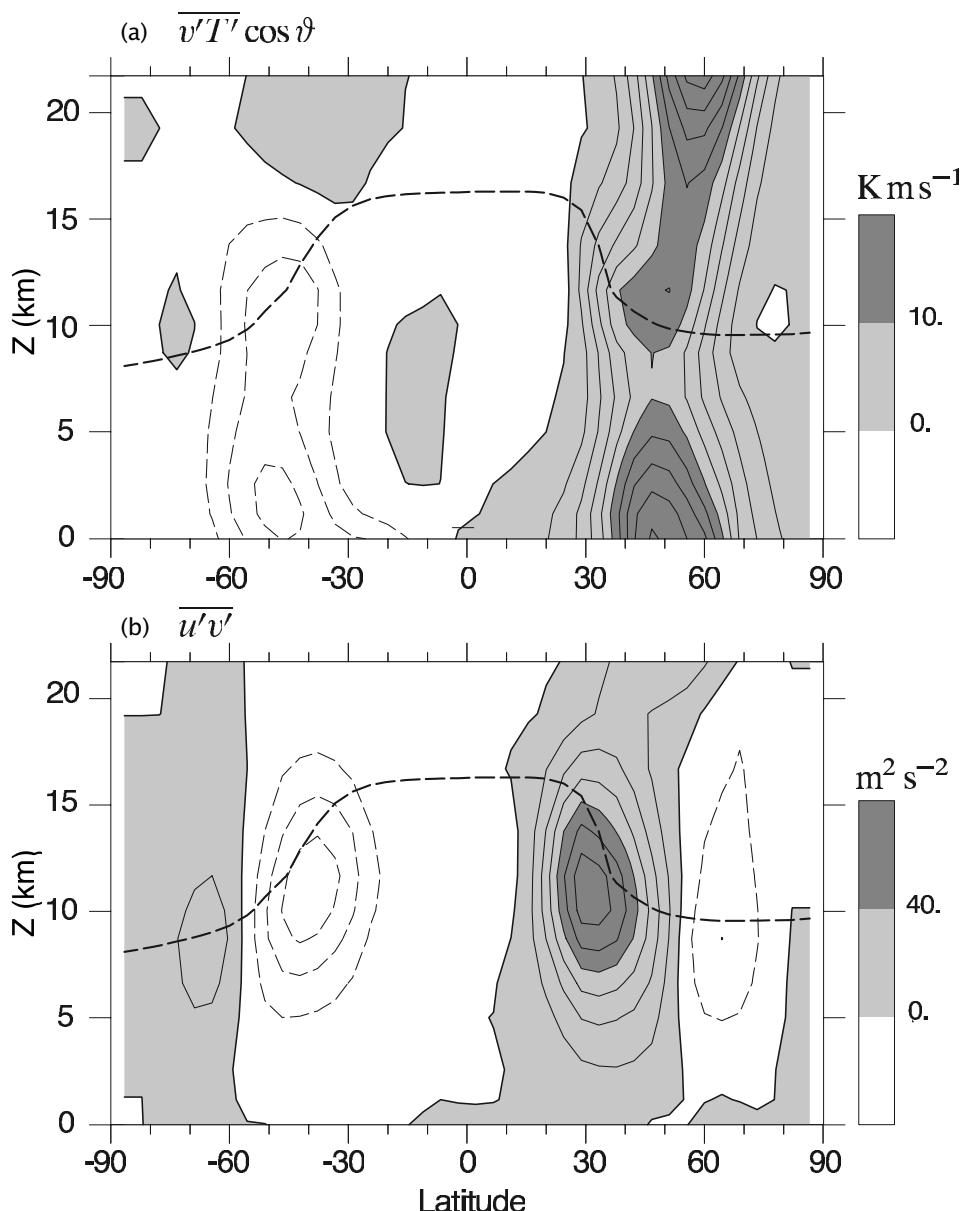


Fig. 11.11 (a) The average meridional eddy heat flux and (b) the eddy momentum flux in the northern hemisphere winter (DJF). The ordinate is log-pressure, with scale height $H = 7.5 \text{ km}$. Positive (northward) fluxes are shaded in both cases, and the dashed line marks the thermal tropopause. The eddy heat flux (contour interval 2 K m s^{-1}) is largely polewards, and down the temperature gradient, in both hemispheres. The eddy momentum flux (contour interval $10 \text{ m}^2 \text{s}^{-2}$) converges in mid-latitudes in the region of the mean jet, and must be upgradient there.¹¹

heating between equator and pole. Recognizing this, ‘eddy driving’ is a convenient way to refer to the mediating role of eddies in producing a zonally averaged circulation.) The model is over-simple, but revealing. The zonally averaged zonal momentum equation (11.4) may be written as

$$\frac{\partial \bar{u}}{\partial t} - (f + \bar{\zeta})\bar{v} = -\frac{1}{\cos^2 \vartheta} \frac{\partial}{\partial \vartheta} (\cos^2 \vartheta \bar{u}' \bar{v}'). \quad (11.64)$$

If the Rossby number is sufficiently low this becomes simply

$$\frac{\partial \bar{u}}{\partial t} - f\bar{v} = M, \quad (11.65)$$

where M represents the eddy terms. This approximation is not quantitatively accurate but it will highlight the role of the eddies. (Note the contrast between this model and the angular-momentum-conserving model. In the latter we assumed $f + \bar{\zeta} \approx 0$, and $Ro = \mathcal{O}(1)$; now we are neglecting $\bar{\zeta}$ and assuming the Rossby number is small.) At a similar level of approximation let us write the thermodynamic equation, (11.63), as

$$\frac{\partial \bar{b}}{\partial t} + N^2 \bar{w} = J, \quad (11.66)$$

where $J = Q[b] - (\cos \vartheta)^{-1} \partial_y (\bar{v}' \bar{b}') \cos \vartheta$ represents the diabatic terms and eddy forcing. We are assuming, as in quasi-geostrophic theory, that the mean stratification, N^2 is fixed, and now \bar{b} represents only the (zonally averaged) deviations from this. If we simplify further by using Cartesian geometry then the mass conservation is

$$\frac{\partial \bar{v}}{\partial y} + \frac{\partial \bar{w}}{\partial z} = 0, \quad (11.67)$$

and we may define a meridional streamfunction Ψ such that

$$\bar{w} = \frac{\partial \Psi}{\partial y}, \quad \bar{v} = -\frac{\partial \Psi}{\partial z}. \quad (11.68)$$

We may then use the thermal wind relation,

$$f \frac{\partial \bar{u}}{\partial z} = -\frac{\partial \bar{b}}{\partial y}, \quad (11.69)$$

to eliminate time derivatives in (11.65) and (11.66), giving

$$f^2 \frac{\partial^2 \Psi}{\partial z^2} + N^2 \frac{\partial^2 \Psi}{\partial y^2} = f \frac{\partial M}{\partial z} + \frac{\partial J}{\partial y}. \quad (11.70)$$

This is a linear equation for the overturning streamfunction, one that holds even if the flow is not in a steady state, and we see that the overturning circulation is forced by eddy fluxes of heat and momentum, as well as heating and other terms that might appear on the right-hand sides of (11.65) and (11.66). If we rescale the vertical coordinate by the Prandtl ratio (i.e., let $z = z' f / N$) then (11.70) is a Poisson equation for the streamfunction. A few other germane points are as follows.

- (i) The horizontal gradient of the thermodynamic forcing partially drives the circulation. At low latitudes, both the heating term and the horizontal eddy flux divergence act in the same sense. An overturning circulation that is forced by diabatic terms, and so with warm fluid rising and cold fluid sinking, is called a ‘direct cell’.

- (ii) The vertical gradient of the horizontal eddy momentum divergence also partially drives the circulation, and from Fig. 11.11 it is clear these fluxes will intensify the circulation. That is, the same terms that cause angular momentum non-conservation act to strengthen the overturning circulation. This balance is reflected in the momentum equation — the Coriolis term $f v$ is balanced by the eddy momentum flux convergence.
- (iii) If M contains frictional terms, such as $v \partial^2 u / \partial z^2$, then these may also act to strengthen the meridional circulation, and weaken angular momentum conservation.
- (iv) If N is small, then the circulation will become stronger if the other terms remain the same. That is, a dry atmosphere with a lapse rate close to that of a dry adiabat (i.e., $N = 0$) may have a stronger overturning circulation than otherwise, because the air can circulate without transporting any heat.
- (v) In winter, the increased strength of eddy momentum and buoyancy fluxes will drive a stronger Hadley Cell. This constitutes a different mechanism from that given in section 11.4 for the increased strength of the winter cell.

* *A slight generalization*

We can generalize (11.70) somewhat by replacing (11.65) by (11.62), namely

$$\frac{\partial \bar{m}}{\partial t} + \bar{v} \frac{\partial \bar{m}}{\partial y} = M. \quad (11.71)$$

Then, using thermal wind equation in the form

$$\frac{f}{a \cos \theta} \frac{\partial \bar{m}}{\partial z} = - \frac{\partial \bar{b}}{\partial y}, \quad (11.72)$$

an equation very similar to (11.70) may be derived. However, the coefficients on the left-hand side are functions of the solution, and $f^2 \partial_{zz} \Psi$ in (11.70) is replaced by a term like $(f \partial_y \bar{m}) (\partial_{zz} \Psi)$. Then, to the extent that $\partial \bar{m} / \partial y < f$ and the other terms are the same, the overturning will be stronger than that given using (11.70).¹²

11.6 THE HADLEY CELL: SUMMARY AND NUMERICAL SOLUTIONS

We have presented two models for the Hadley Cell: (i) an angular momentum conserving model and (ii) a largely eddy-driven model. The two models are opposite extremes, both being severe approximations to a more complete representation of the Hadley Cell that might comprise the zonal momentum equation with eddies (11.60), the thermodynamic equation (11.63) (with the effects of moisture included) and the meridional momentum equation, perhaps approximated as cyclostrophic wind balance. In reality, both the conservative effects of angular momentum advection and the effects of eddy fluxes likely play a role, and delineating the importance of their respective effects is a task that must be guided by observations and numerical simulations.

Illustrative results from two idealized GCM experiments are shown in Figs. 11.12 and 11.13. The GCM has no explicit representation of moisture, except that the lapse rate is adjusted to a value close to the moist adiabatic lapse rate if it exceeds that value. In one experiment the model is constrained to produce an axisymmetric solution (top panels of

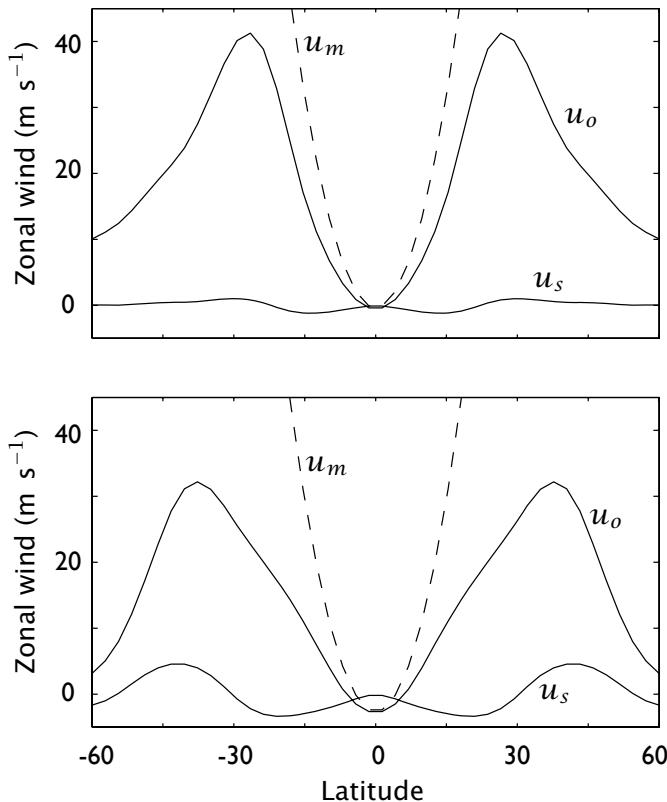


Fig. 11.12 The zonal wind in two numerical simulations. The lower panel is from an idealized dry, three-dimensional atmospheric GCM, and the upper panel is an axisymmetric version of the same model. Plotted are the zonal wind at the level of the Hadley Cell outflow, u_o ; the surface wind, u_s ; and the angular-momentum-conserving value, u_m .¹³

the figures), and the zonal wind produced by the model in the Hadley Cell outflow is fairly close to being angular-momentum-conserving. (The lack of perfect angular momentum conservation is due to the presence of a small but finite vertical viscosity that is necessary in order to reach a steady state; without it, the steady state becomes symmetrically unstable.) In a three-dimensional version of the model, in which baroclinic eddies are allowed to form, the zonal wind is significantly reduced from its angular-momentum-conserving value, and correspondingly the overturning circulation is much stronger. Indeed, the strength of the Hadley Cell increases roughly linearly with the strength of the eddies in a sequence of numerical integrations similar to those shown, as suggested by (11.70). Qualitatively similar results are found in a model with no convective parameterization. In this case, the lapse rate is closer to neutral, N^2 is small, and the overturning circulation is generally stronger, as also expected from (11.70).

Is the real Hadley circulation ‘eddy-driven’, as in section 11.5.2, or is it a largely zonally symmetric structure constrained by angular momentum conservation, as in section 11.2 and

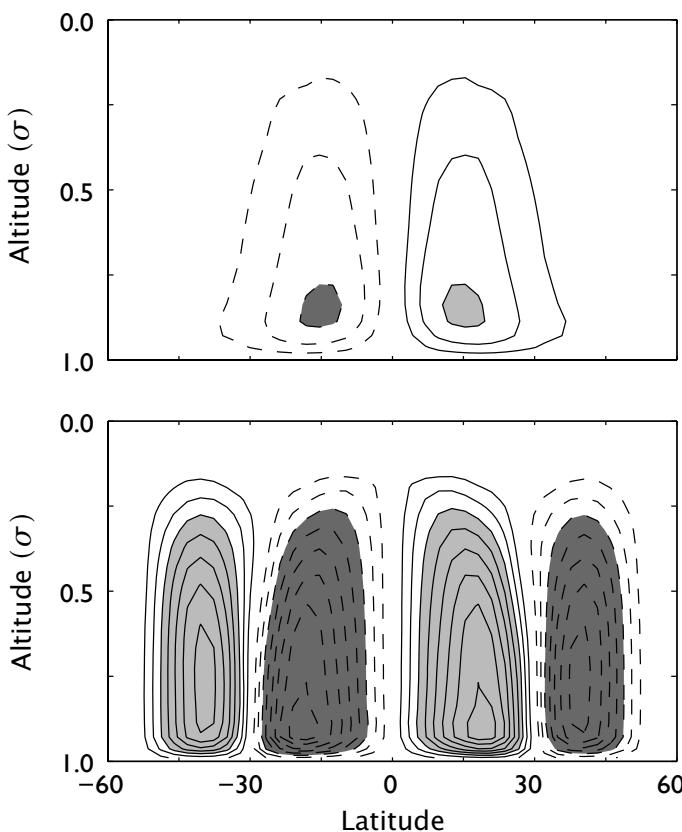
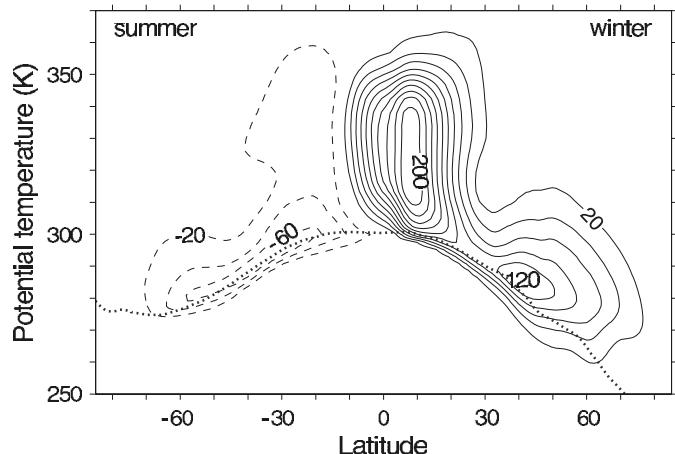


Fig. 11.13 As for Fig. 11.12, but now showing the streamfunction of the overturning circulation. ‘Altitude’ is $\sigma = p/p_s$, where p_s is surface pressure, and contour interval is 5 Sv (i.e., $5 \times 10^9 \text{ kg s}^{-1}$).

its hemispherically asymmetric extensions? Observations of the overturning flow in summer and winter provide a guide. Figure 11.14 shows the thickness weighted transport overturning circulation in isentropic coordinates, and (as discussed more in chapter 7) this circulation includes both the Eulerian mean transport and the transport due to eddies. In winter there is considerable recirculation within the Hadley Cell, most of it coming from the zonally symmetric flow, and it is a quite distinct structure from the mid-latitude circulation. This suggests that the poleward extent of the winter Hadley Cell is influenced by axisymmetric dynamics, for if it were solely a response to eddy heat fluxes one might expect it to join more smoothly with the mid-latitude Ferrel Cell. It may be that the effects of condensation and the concentration of the thermodynamic source act to give the axisymmetric circulation a significant role. However, a winter Hadley Cell strongly influenced by eddy momentum fluxes might still strongly recirculate, and eddy effects do almost certainly play an important role. In summer, there is virtually no recirculation within the Hadley Cell and it does not appear as a self-contained structure, and this is suggestive of eddy effects and/or a strong mid-latitude influence. However, the above remarks are very speculative.

Fig. 11.14 The observed mass transport streamfunction in isentropic coordinates in northern hemisphere winter (DJF). The dotted line is the median surface temperature. The return flow is nearly all in a layer near the surface, much of it at a lower temperature than the median surface temperature. Note the more vigorous circulation in the winter hemisphere.¹⁴



11.7 THE FERREL CELL

In this section we give a brief introduction to the Ferrel Cell, taking the eddy fluxes of heat and momentum to be given and viewing the circulation from a zonally averaged and Eulerian perspective. We investigate the associated dynamics in the next chapter.

The Ferrel Cell is an indirect meridional overturning circulation in mid-latitudes (see Fig. 11.3) that is apparent in the zonally averaged v and w fields, or the meridional overturning circulation defined by (11.3) or (11.68). It is ‘indirect’ because cool air apparently rises in high latitudes, moves equatorwards and sinks in the subtropics. Why should such a circulation exist? The answer, in short, is that it is there to balance the eddy momentum convergence of the mid-latitude eddies and it is effectively driven by those eddies. To see this, consider the zonally averaged zonal momentum equation in mid-latitudes; at low Rossby number, and for steady flow this is just

$$-f\bar{v} = -\frac{1}{\cos^2 \theta} \frac{\partial}{\partial \theta} (\cos^2 \theta \bar{u}' \bar{v}') + \frac{1}{\rho} \frac{\partial \tau}{\partial z}. \quad (11.73)$$

This is a steady version of (11.65) with the addition of a frictional term $\partial \tau / \partial z$ on the right-hand side. At the surface we might approximate the stress by a drag, $\tau = r\bar{u}_s$, where r is a constant, with the stress falling away with height so that it is important only in the lowest kilometre or so of the atmosphere, in the atmospheric Ekman layer. Above this layer, the eddy momentum flux convergence is balanced by the Coriolis force on the meridional flow. In mid-latitudes (from about 30° to 70°) the eddy momentum flux divergence is negative in both hemispheres (Fig. 11.11) and therefore, from (11.73), the averaged meridional flow must be equatorwards, as illustrated schematically in Fig. 11.15.

The flow cannot be equatorwards everywhere, simply by mass continuity, and the return flow occurs largely in the Ekman layer, of depth d say. Here the eddy balance is between the Coriolis term and the frictional term, and integrating over this layer gives

$$-fV \approx -r\bar{u}_s, \quad (11.74)$$

where $V = \int_0^d \rho \bar{v} dz$ is the meridional transport in the boundary layer, above which the stress vanishes. The return flow is polewards (i.e., $V > 0$ in the Northern Hemisphere) producing

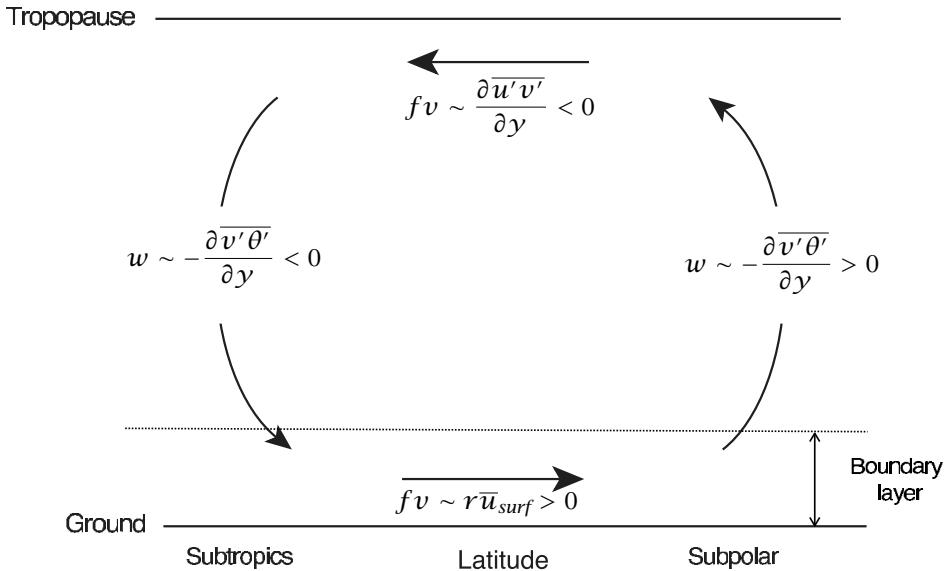


Fig. 11.15 The eddy-driven Ferrel Cell, from an Eulerian point of view. Above the planetary boundary layer the mean flow is largely in balance with the eddy heat and momentum fluxes, as shown. The lower branch of the Ferrel Cell is largely confined to the boundary layer, where it is in a frictional-geostrophic balance.

an eastward Coriolis force. This can be balanced by a westward frictional force provided that the surface flow has an eastward component. In this picture, then, the mid-latitude eastward zonal flow at the surface is a consequence of the polewards flowing surface branch of the Ferrel Cell, this poleward flow being required by mass continuity given the equatorward flow in the upper branch of the cell. In this way, the Ferrel Cell is responsible for bringing the mid-latitude eddy momentum flux convergence to the surface where it may be balanced by friction (refer again to Fig. 11.15).

A more direct way to see that the surface flow must be eastwards, given the eddy momentum flux convergence, is to vertically integrate (11.73) from the surface to the top of the atmosphere. By mass conservation, the Coriolis term vanishes (i.e., $\int_0^\infty f \rho \bar{v} dz = 0$) and we obtain

$$\int_0^\infty \frac{1}{\cos^2 \vartheta} \frac{\partial}{\partial \vartheta} (\cos^2 \vartheta \bar{u}' \bar{v}') \rho dz = [\tau]_0^\infty = -r \bar{u}_s. \quad (11.75)$$

That is, the surface wind is proportional to the vertically integrated eddy momentum flux convergence. Because there is a momentum flux convergence, the left-hand side is negative and the surface winds are eastwards.

The eddy heat flux also plays a role in the Ferrel Cell, for in a steady state we have, from (11.66)

$$w = \frac{1}{N^2} \left[Q[b] - \frac{1}{\cos \vartheta} \frac{\partial (\bar{v}' \bar{b}' \cos \vartheta)}{\partial y} \right] \quad (11.76)$$

and inspection of Fig. 11.11 shows that the observed eddy heat flux produces an overturning circulation in the same sense as the observed Ferrel Cell (again see Fig. 11.15).

Is the circulation produced by the heat fluxes *necessarily* the same as that produced by the momentum fluxes? In a non-steady state the effects of both heat and momentum fluxes on the Ferrel Cell are determined by (11.70) (an equation which in fact applies more accurately at mid-latitudes than at low ones because of the low-Rossby number assumption), and there is no particular need for the heat and momentum fluxes to act in the same way. But in a steady state they must act to produce a consistent circulation. To show this, for simplicity let us take f and N^2 to be constant, let us suppose the fluid is incompressible and work in Cartesian coordinates. Take the y -derivative of (11.73) and the z -derivative of (11.76) and use the mass continuity equation. Noting that $\bar{v}'\bar{\zeta}' = -\partial\bar{u}'\bar{v}'/\partial y$ we obtain

$$\frac{\partial}{\partial y} \left(\bar{v}'\bar{\zeta}' + \frac{f_0}{N^2} \frac{\partial \bar{v}'\bar{b}'}{\partial z} \right) = \frac{\partial}{\partial z} \left(\frac{f_0}{N^2} Q[b] \right) + \frac{\partial}{\partial y} \left(\frac{1}{\rho_0} \frac{\partial \tau}{\partial z} \right). \quad (11.77)$$

The expression on the left-hand side is the divergence of the eddy flux of quasi-geostrophic potential vorticity! That the heat and momentum fluxes act to produce a consistent overturning circulation is thus equivalent to requiring that the terms in the quasi-geostrophic potential vorticity equation are in a steady-state balance.

Notes

- 1 Many of the observations presented here are so-called 'reanalyses', prepared by the National Centers for Environmental Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMWF) (e.g., Kalnay 1996). Unless stated, we use the NCEP reanalysis with data from 1958–2003. Reanalysis products are syntheses of observations and model results and so are not wholly accurate representations of the atmosphere. However, especially in data-sparse regions of the globe and for poorly measured fields, they are likely to be more accurate representations of the atmosphere than could be achieved using only the raw data. Of course, this in turn means they contain biases introduced by the models.
- 2 For example, Trenberth & Caron (2001).
- 3 George Hadley (1685–1768) was a British meteorologist who formulated the first dynamical theory for the trade winds, presented in a paper (Hadley 1735) entitled 'Concerning the cause of the general trade winds.' (At that time, trade winds referred to any large-scale prevailing wind, and not just tropical winds. Some etymologists have associated the name with the commercial (i.e., trade) exploitation of the wind by mariners on long ocean journeys, but such an origin has been disputed: trade also means customary, and the winds customarily blow in one direction. Relatedly, in Middle English the word trade means path or track — hence the phrase 'the wind blows trade', meaning the wind is on track.) Hadley realized that in order to account for the zonal winds, the Earth's rotation makes it necessary for there also to be a meridional circulation. His vision was of air heated at low latitudes, cooled at high latitudes, giving rise to a single meridional cell between the equator and each pole. Although he thought of the cell as essentially filling the hemisphere, and he did not account for the instability of such a flow, it was nevertheless a foundational contribution to meteorology. The thermally direct cell in low latitudes is now named after him.

A three-celled circulation was proposed by William Ferrel (1817–1891), an American school teacher and meteorologist, and the middle of these cells is now named for him. His explanation of the cell (Ferrel 1856a) was not correct, but this is hardly surprising because the eddy motion producing the angular momentum convergence that drives the Ferrel Cell was not understood for another 100 years or so. Ferrel in fact abandoned his three-celled picture in

favour of a something more akin to a two-celled picture (Ferrel 1859), similar to that proposed by J. Thompson in 1857. The history of these ideas, and those of Hadley, is discussed by Thomson (1892). Ferrel did however give the first essentially correct description of the role of the Coriolis force and the geostrophic wind in the general circulation (Ferrel 1858, a paper with a quite modern style), a key development in the history of geophysical fluid dynamics. Ferrel also contributed to tidal theory [in Ferrel (1856b) he noted, for example, that the tidal force due to the moon would slow the Earth's rotation] and to ocean dynamics. (See <http://www.history.noaa.gov/giants/ferrel2.html>).

Although Hadley's single-celled viewpoint was in part superseded by the three-celled and two-celled structures, the modern view of the overturning circulation is, ironically, that of a single cell of 'residual circulation', which, although having distinct tropical and extratropical components, in some ways qualitatively resembles Hadley's original picture. See chapter 12.

- 4 Following Held & Hou (1980). Schneider (1977) and Schneider & Lindzen (1977) were particularly influential precursors.
- 5 Nevertheless, Fang & Tung (1996) do find some analytic solutions in the presence of moisture and convection.
- 6 After Hide (1969).
- 7 Gierasch (1975).
- 8 Largely following Lindzen & Hou (1988).
- 9 Solutions from Lindzen & Hou (1988).
- 10 See Dima & Wallace (2003) for some relevant observations. They noted that the asymmetry of the Hadley Cell is affected by monsoonal circulations (which are, of course, not accounted for in the model presented here). Fang & Tung (1999) investigated the effects of time dependence, and essentially noted that (11.59) is not well satisfied, although this alone was unable to limit the nonlinear amplification effect. Walker & Schneider (2005) showed that the effects of vertical momentum diffusion and of momentum transport by baroclinic eddies are both significant in a GCM, and these limit the nonlinear amplification.
- 11 Figure courtesy of M. Juckes, using an ECMWF reanalysis.
- 12 A still more general, usually elliptic equation for the overturning circulation may be derived from the zonally averaged primitive equations, assuming only that the zonally averaged zonal wind is in cyclostrophic balance with the pressure field (Vallis 1982).
- 13 Simulations kindly performed by C. Walker. See also Walker and Schneider (2005).
- 14 Figure courtesy of T. Schneider, using an ECMWF reanalysis.

Further reading

Lorenz, E. N., 1967. *The Nature and Theory of the General Circulation of the Atmosphere*.

A classic monograph on the atmospheric general circulation.

Peixoto, J. P. & Oort, A. H., 1992. *Physics of Climate*.

A descriptive but physically based discussion of the climate and the general circulation, with an emphasis on observations.

Problems

- 11.1 Explicitly derive equations (11.20) and (11.21).
- 11.2 Suppose that, in the vertically integrated Hadley Cell model considered in section 11.2 the radiative equilibrium temperature falls linearly from the equator to the pole. For example,

suppose that $\theta_E = \theta_{E0} - \Delta\theta(|y|/a)$, rather than the quadratic fall-off in (11.16). Obtain and discuss the solutions to the Held-Hou problem. Include an expression for the latitudinal extent of the Hadley Cell, and comment on any discontinuities at the edge of the Hadley Cell and at the equator.

- 11.3 By considering the form of the viscous term in spherical coordinates, show explicitly that an atmosphere in solid body rotation experiences no viscous stresses.
- 11.4 (a) Suppose the zonal wind is in thermal wind balance with the radiative equilibrium temperature, θ_E . Obtain a condition for the meridional variation of the radiative equilibrium temperature so that the wind does not violate Hide's theorem (that is, it does not produce an interior maximum of angular momentum). For example, show the radiative equilibrium temperature must not vary more rapidly with latitude faster than the latitude raised to some power. You may make the small-angle approximation. Alternatively, you could use thermal wind balance in the form (2.206) or (2.208). You may assume the surface wind is zero if needs be.
- (b) Suppose the radiative equilibrium temperature falls off with latitude as $P_4(\sin \vartheta)$, where P_4 is the fourth Legendre polynomial. Show that the zonal velocity that is in thermal wind balance with this does not violate Hide's theorem. Comment on the relevance of this to the issue of whether the radiative equilibrium solution is physically realizable.
- 11.5 In the angular-momentum-conserving model of the Hadley Cell, air that starts at rest at the equator develops a large zonal velocity, and hence a large kinetic energy, as it moves polewards. Explain carefully where this energy comes from. (Note that the Coriolis force itself does no work on a fluid parcel.)
- 11.6 A spinning ice skater with arms outstretched lowers his arms. Show that if the skater's angular momentum is conserved his kinetic energy increases. Where has this energy come from? Is it different if the skater raises his arms?
- 11.7 (a) Derive and plot the layer thickness as a function of latitude in the shallow water Hadley Cell model, and the corresponding zonal wind.
- (b) Suppose that the equilibrium thickness, h^* , falls quadratically with latitude, rather than linearly as we assume in (11.47). Obtain and plot expressions for the extent of the Hadley Cell, the thickness and the zonal wind.
- 11.8 ♦ The oceanic thermohaline circulation seems similar to Hadley's vision of the atmospheric circulation, with a large thermally driven cell between pole and equator. Discuss. Is conservation of angular momentum an important factor in the thermohaline circulation? If so, what are its manifestations? If not, why not?

Any theory of the atmospheric circulation must be based on a theory of (large-scale) atmospheric turbulence.

Eric Eady, *The cause of the general circulation of the atmosphere*, 1950.

CHAPTER TWELVE

Zonally Averaged Mid-latitude Atmospheric Circulation

OUR GOAL IN THIS CHAPTER is to gain an understanding of the zonally averaged circulation of the mid-latitude atmosphere. Because of the presence of strong zonal asymmetries — that is, eddies — this circulation differs considerably from the zonally symmetric circulation that would exist if eddies did not develop at all. (The idealized angular-momentum-conserving model of the Hadley Cell discussed in the previous chapter is an example of a zonally symmetric circulation.) Let us explain more.

When studying many aspects of the large-scale ocean circulation, or the low-latitude atmospheric circulation, we can make a great deal of progress by treating the large-scale flow as if it were absolutely steady. However, this approach fails miserably for the mid-latitude atmosphere: the large-scale mid-latitude circulation is intrinsically unsteady on the large-scale to the extent that the associated eddies essentially *are* the circulation. The eddies are also unpredictable and chaotic; that is to say, *the large-scale mid-latitude circulation of the atmosphere is a turbulent flow*. This turbulence involves the large-scale, geostrophically and hydrostatically balanced flow — that is, it is geostrophic turbulence — and so has different properties than the smaller-scale, more nearly three-dimensional turbulence that may occur in boundary layers and the like. In addition, the turbulence is neither fully-developed nor isotropic — it interacts with the Rossby waves that propagate on the large-scale background meridional vorticity gradient that is imparted by the planetary rotation, and it is this interaction that, we will find, produces the momentum convergence that gives rise to large-scale zonal flows and the surface winds.

Why is the large-scale flow turbulent? Why is it zonally asymmetric at all? There are two potential sources for zonal asymmetries.

- (i) The zonal asymmetries that exist in the underlying boundary conditions and forcing: mountains, land-sea contrasts, the diurnal cycle, and so on.
- (ii) Hydrodynamic instability: even if the surface and the forcing were exactly zonally symmetric, the corresponding zonally symmetric solutions of the equations of motion would have a large shear in the zonal wind and this might be baroclinically unstable to zonally asymmetric perturbations.

If the flow were not unstable, then we might expect that the zonal asymmetries of item (i) would give rise to corresponding, steady, zonal asymmetries in the resulting circulation, and this process is discussed in the next chapter. However, the flow *is* unstable, primarily via baroclinic instability (as discussed in chapter 6), and this leads to eddy growth and ultimately to geostrophic turbulence; this is, essentially, what gives rise to *weather*. The large-scale circulation is not, however, so turbulent that such things as Rossby waves cease to have meaning, and we find that we can make progress in understanding the circulation by drawing from the wave-mean flow theory discussed in chapter 7, as well as the theory of turbulence discussed in chapters 9 and 10. Of course, because the atmosphere is so complex — among other things it is inhomogeneous and anisotropic — applying ideas from such theories is by no means straightforward. We find that we must simplify and consider idealized situations, and of course in so doing we risk becoming divorced from reality; thus, we try to construct models that are robust in their predictions and that do not depend delicately on how we might parameterize some particular turbulent process.

In this chapter, then, we focus on the effects of item (ii), and try to understand the mid-latitude circulation of an atmosphere with zonally symmetric forcing and boundary conditions. We assume that the zonal asymmetries due to the boundary conditions do not *qualitatively* affect our arguments, and that the circulation of the atmosphere with a perfectly smooth surface would resemble that of the real circulation. One of the most basic features of the mid-latitude atmosphere is the westerly winds at the surface, and we begin with a discussion of the mechanisms that maintain these westerlies. We will present our arguments semi-independently of earlier chapters, and in particular we will develop some of the results of wave-mean flow interaction ab initio, so that a reader with just a little experience can start here and refer back to these chapters as needed.¹

12.1 SURFACE WESTERLIES AND THE MAINTENANCE OF A BAROTROPIC JET

12.1.1 Observations and motivation

The atmosphere above the surface has a generally eastward flow, with a broad maximum about 10 km above the surface at around 40° in either hemisphere. But if we look a little more at the zonally average wind in Fig. 11.2(a) we see hints of there being two jets — one (the subtropical jet) at around 30°, and another somewhat polewards of this, especially apparent in the Southern Hemisphere. Such a jet is particularly noticeable in certain regions of the globe, when a zonal average is not taken, as in Fig. 12.1. The subtropical jet is associated with a strong meridional temperature gradient at the edge of the Hadley Cell, and is quite baroclinic. On the other hand, the mid-latitude jet (sometimes called the subpolar jet) is more barotropic (it has little vertical structure, with less shear than the subtropical jet) and lies above an eastward surface flow. This flow feels the effect of friction and so there must be a momentum *convergence* into this region, as indeed is seen in Fig. 11.11. We will find

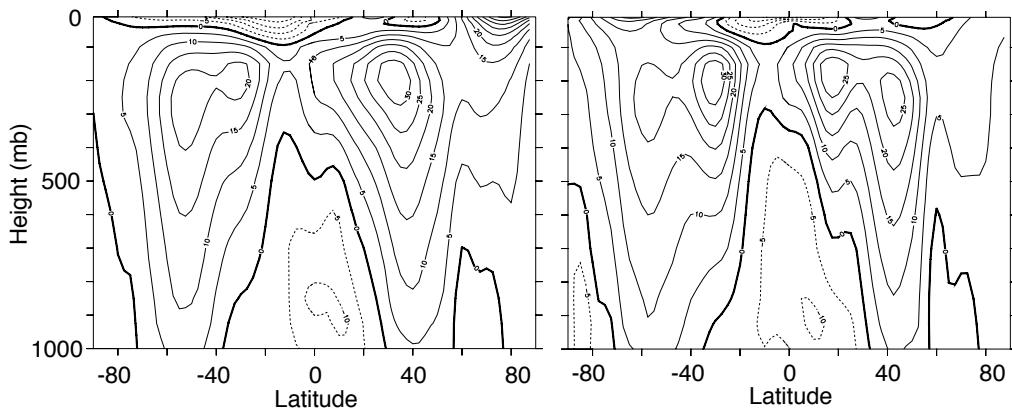


Fig. 12.1 The time averaged zonal wind at 150°W (in the mid-Pacific) in December–January–February (DJF, left), March–April–May (MAM, right). The contour interval is 5 m s⁻¹. Note the double jet in each hemisphere, one in the subtropics and one in mid-latitudes. The subtropical jets are associated with a strong meridional temperature gradient, whereas the mid-latitude jets have a stronger barotropic component and are associated with westerly winds at the surface.

that this momentum convergence occurs largely in transient eddies, and the jet is known as the *eddy-driven jet*. Although the eddies themselves are a product of baroclinic instability, the essential mechanism of jet production is present in barotropic dynamics, so we first consider how an eastward jet can be maintained in a turbulent two-dimensional flow on the surface of a rotating sphere.

In barotropic turbulence, alternating east–west jets can be maintained if β is non-zero, as described in section 9.1. However, that case was homogeneous, with no preferred latitude for a particular direction of jet, whereas in the atmosphere there appears to be but one mid-latitude jet, and although it meanders it certainly has a preferred average location. In the subsections that follow we give four explanations as to how the jet is maintained; the first has a different flavour from the others, but they are all really just different perspectives on the same mechanism.²

12.1.2 The mechanism of jet production

I. The vorticity budget

Suppose that the absolute vorticity normal to the surface (i.e., $\zeta + 2\Omega \sin \vartheta$) increases monotonically polewards. (A sufficient condition for this is that the fluid is at rest.) By Stokes' theorem, the circulation around a line of latitude circumscribing the polar cap, I , is equal to the integral of the absolute vorticity over the cap. That is,

$$I_i = \int_{\text{cap}} \boldsymbol{\omega}_{ia} \cdot d\mathbf{A} = \oint_C \mathbf{u}_{ia} \cdot d\mathbf{l} = \oint_C (u_i + 2\Omega a \cos \vartheta) dl, \quad (12.1)$$

where $\boldsymbol{\omega}_{ia}$ and \mathbf{u}_{ia} are the initial absolute vorticity and velocity, respectively, u_i is the initial zonal velocity in the Earth's frame of reference, and the line integrals are around the line of

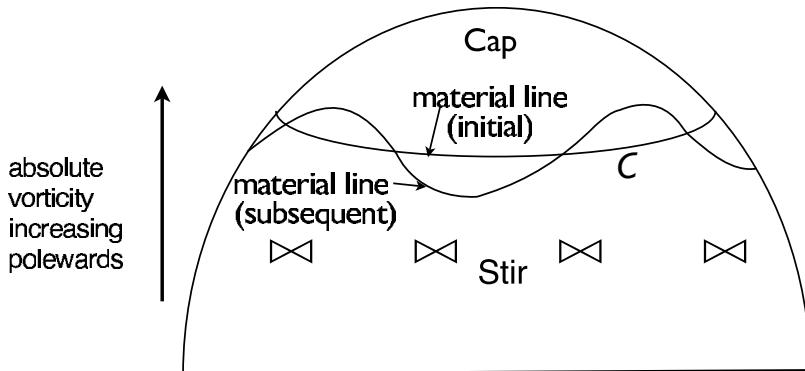


Fig. 12.2 The effects of a mid-latitude disturbance on the circulation around the latitude line C . If initially the absolute vorticity increases monotonically polewards, then the disturbance will bring fluid with lower absolute vorticity into the cap region. Then, using Stokes theorem, the velocity around the latitude line C will become more westwards.

latitude. For simplicity let us take $u_i = 0$ and suppose there is a disturbance equatorwards of the polar cap, and that this results in a distortion of the material line around the latitude circle C (Fig. 12.2). Since we are supposing the source of the disturbance to be distant from the latitude of interest, then if we neglect viscosity the circulation along the material line is conserved, by Kelvin's circulation theorem. Thus, vorticity with a lower value is brought into the region of the polar cap — that is, the region polewards of the latitude line C . Using Stokes' theorem again the circulation around the latitude circle C must therefore fall; that is, denoting values after the disturbance with a subscript f ,

$$I_f = \int_{\text{cap}} \boldsymbol{\omega}_{fa} \cdot d\mathbf{A} < I_i \quad (12.2)$$

so that

$$\oint_C (u_f + 2\Omega a \cos \vartheta) dl < \oint_C (u_i + 2\Omega a \cos \vartheta) dl \quad (12.3)$$

and

$$\bar{u}_f < \bar{u}_i \quad (12.4)$$

with the overbar indicating a zonal average. Thus, there is a tendency to produce *westward* flow polewards of the disturbance. By a similar argument westward flow is also produced equatorwards of the disturbance — to see this one might apply Kelvin's theorem over all of the globe south of the source of the disturbance (taking care to take the dot-product correctly between the direction of the vorticity vector and the direction normal to the surface). Finally, note that the overall situation is the same in the Southern Hemisphere. Thus, on the surface of a rotating sphere, external stirring will produce westward flow *away* from the region of the stirring.

Now suppose, furthermore, that the disturbance imparts no net angular momentum to the fluid. Then the integral of $ua \cos \vartheta$ over the entire hemisphere must be constant. But the fluid is accelerating westwards away from the disturbance. Therefore, the fluid in the region

of the disturbance must accelerate *eastwards*; that is, angular momentum must converge into the stirred region, producing an eastward flow. This simple mechanism is the essence of the production of eastward eddy-driven jets in the atmosphere, and of the eastward surface winds in mid-latitudes. The stirring that here we have externally imposed comes, in reality, from baroclinic instability.

If the stirring subsides then the flow may reversibly go back to its initial condition, with a concomitant reversal of the momentum convergence that caused the zonal flow. Thus, we must have some form of dissipation and irreversibility in order to produce permanent changes, and in particular we need to irreversibly mix vorticity. (This result is closely related to the non-acceleration results of chapter 7.) If the fluid is continuously mixed, then of course we also need a source that restores the absolute vorticity gradient, otherwise we will completely homogenize the vorticity over the hemisphere, so let us now set up a simple model that shows how a permanent jet structure can be maintained.

II. Rossby waves and momentum flux

We saw above that a mean gradient of vorticity is an essential ingredient in the mechanism whereby a mean flow is generated by stirring. Given such, we expect Rossby waves to be excited, and we now show how Rossby waves are intimately related to the momentum flux maintaining the mean flow.

If a stirring is present in mid-latitudes then we expect that Rossby waves will be generated there, propagate away and break and dissipate. To the extent that the waves are quasi-linear and do not interact, then just away from the source region each wave has the form

$$\psi = \operatorname{Re} C e^{i(kx+ly-\omega t)} = \operatorname{Re} C e^{i(kx+ly-kct)}, \quad (12.5)$$

where C is a constant, with dispersion relation

$$\omega = ck = \bar{u}k - \frac{\beta k}{k^2 + l^2} \equiv \omega_R, \quad (12.6)$$

provided that there is no meridional shear in the zonal flow. The meridional component of the group velocity is given by

$$c_g^y = \frac{\partial \omega}{\partial l} = \frac{2\beta kl}{(k^2 + l^2)^2}. \quad (12.7)$$

Now, the direction of the group velocity must be *away* from the source region; this is a radiation condition (discussed more in the next subsection), demanded by the requirement that Rossby waves transport energy *away* from the disturbance. Thus, northwards of the source kl is positive and southwards of the source kl is negative. That the product kl can be positive or negative arises because for each k there are two possible values of l that satisfy the dispersion relation (12.6), namely

$$l = \pm \left(\frac{\beta}{\bar{u} - c} - k^2 \right)^{1/2}, \quad (12.8)$$

assuming that the quantity in parentheses is positive.

The velocity variations associated with the Rossby waves are

$$u' = -\operatorname{Re} C il e^{i(kx+ly-\omega t)}, \quad v' = \operatorname{Re} C ik e^{i(kx+ly-\omega t)}, \quad (12.9a,b)$$

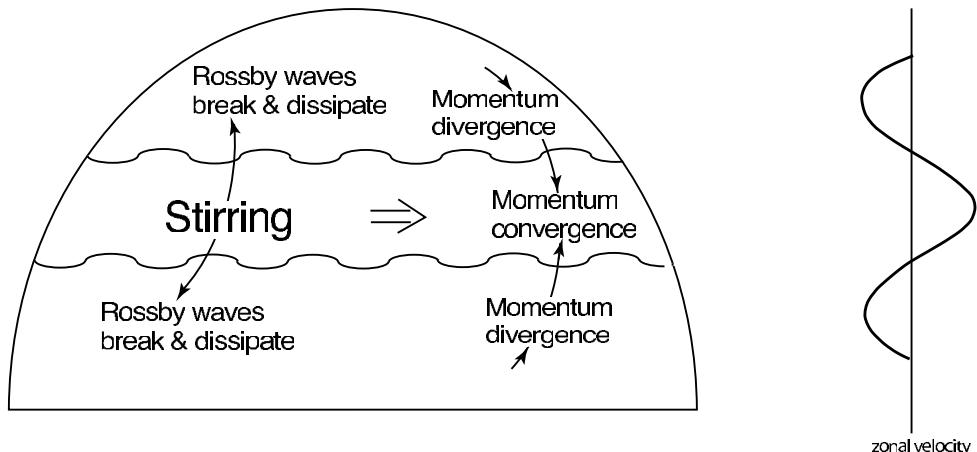


Fig. 12.3 Generation of zonal flow on a β -plane or on a rotating sphere. Stirring in mid-latitudes (by baroclinic eddies) generates Rossby waves that propagate away from the disturbance. Momentum converges in the region of stirring, producing eastward flow there and weaker westward flow on its flanks.

and the associated momentum flux is

$$\overline{u'v'} = -\frac{1}{2} C^2 kl. \quad (12.10)$$

Thus, given that the sign of kl is determined by the group velocity, northwards of the source the momentum flux associated with the Rossby waves is southward (i.e., $\overline{u'v'}$ is negative), and southwards of the source the momentum flux is northward (i.e., $\overline{u'v'}$ is positive). That is, the momentum flux associated with the Rossby waves is *toward* the source region. Momentum converges in the region of the stirring, producing net eastward flow there and westward flow to either side (Fig. 12.3).

Another way of describing the same effect is to note that if kl is positive then lines of constant phase ($kx + ly = \text{constant}$) are tilted north-west/south-east, and the momentum flux associated with such a disturbance is negative ($\overline{u'v'} < 0$). Similarly, if kl is negative then the constant-phase lines are tilted north-east/south-west and the associated momentum flux is positive ($\overline{u'v'} > 0$). The net result is a convergence of momentum flux into the source region. In physical space this is reflected by having eddies that are 'bow-shaped', as in Fig. 12.4.

* The radiation condition and Rayleigh friction

Why is the group velocity directed away from the source region? It is because the energy flux travels at the group velocity, and the energy flux must be directed away from the source region; the reader comfortable with that statement may stop here. (See section 7.2.2 for more on group velocity.) Another way to determine the direction of the group velocity is to employ a common trick in problems of wave propagation, that of adding a small amount of friction to the inviscid problem.³ The solution of the ensuing problem in the limit of small friction will often make clear which solution is physically meaningful in the inviscid problem, and

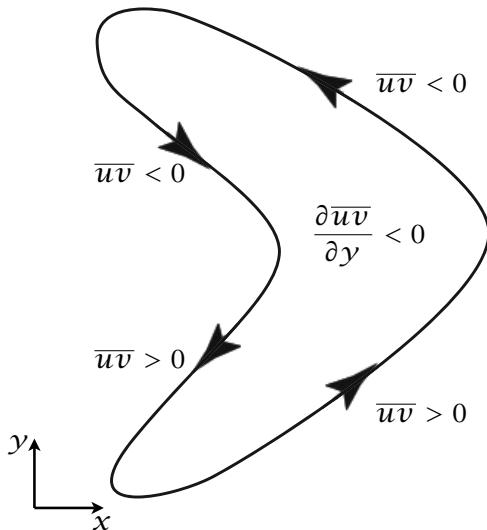


Fig. 12.4 The momentum transport in physical space, caused by the propagation of Rossby waves away from a source in mid-latitudes. The ensuing bow-shaped eddies are responsible for a convergence of momentum, as indicated in the idealization pictured.

therefore which solution nature chooses. Consider the linear barotropic vorticity equation with linear friction,

$$\frac{\partial \zeta}{\partial t} + \beta \frac{\partial \psi}{\partial x} = -r\zeta \quad (12.11)$$

where r is a small friction coefficient. The dispersion relation is

$$\omega = -\frac{\beta k}{K^2} - ir = \omega_R(k, l) - ir, \quad (12.12)$$

where ω_R is defined by (12.6), with $\bar{u} = 0$, and so the wave decays with time. Now suppose a wave is generated in some region, and that it propagates meridionally away, decaying as it moves away. Then, instead of an imaginary frequency, we may suppose that the frequency is real and the y -wavenumber is imaginary. Specifically, we take $l = l_0 + l'$, where $l_0 = \pm[\beta/(\bar{u} - c) - k^2]^{1/2}$ for some zonal wavenumber k , as in (12.8), and $\omega = \omega_R(k, l_0)$. For small friction, we obtain l' by Taylor-expanding the dispersion relation around its inviscid value, $\omega_R(k, l_0)$, giving

$$\omega + ir = \omega_R(k, l) \approx \omega_R(k, l_0) + \frac{\partial \omega_R(k, l)}{\partial l} \Big|_{l=l_0} l', \quad (12.13)$$

and therefore

$$l' = \frac{ir}{c_g^y}, \quad (12.14)$$

where $c_g^y = \partial_l \omega_R(k, l)|_{l=l_0}$ is the y -component of the group velocity. The wavenumber is imaginary, so that the wave either grows or decays in the y -direction, and the wave solution obeys

$$\psi \approx \operatorname{Re} C \exp[i(kx - \omega_R t)] \exp(il_0 y - ry/c_g^y). \quad (12.15)$$

We now demand that the solution decay away from the source, because any other choice is manifestly unphysical, even as we let r be as small as we please. Thus, with the source at

$\gamma = 0$, c_g^γ must be positive for positive γ and negative for negative γ . In other words, the group velocity must be directed *away* from the source region, and therefore momentum flux converges on the source region.

III. The pseudomomentum budget

The kinematic relation between vorticity flux and momentum flux for non-divergent two-dimensional flow is

$$v\zeta = \frac{1}{2} \frac{\partial}{\partial x} (v^2 - u^2) - \frac{\partial}{\partial y} (uv). \quad (12.16)$$

After zonal averaging this gives

$$\overline{v'\zeta'} = -\frac{\partial \overline{u'v'}}{\partial y}, \quad (12.17)$$

noting that $\overline{v} = 0$ for two-dimensional incompressible (or geostrophic) flow. For reference, in spherical coordinates this expression becomes

$$\overline{v'\zeta'} \cos \theta = -\frac{1}{a \cos \theta} \frac{\partial}{\partial \theta} (\cos^2 \theta \overline{u'v'}). \quad (12.18)$$

If (12.17) [or (12.18)] is integrated with respect to y between two quiescent latitudes then the right-hand side vanishes. That is the zonally averaged meridional vorticity flux vanishes when integrated over latitude.

Now, the barotropic zonal momentum equation is (for horizontally non-divergent flow)

$$\frac{\partial u}{\partial t} + \frac{\partial u^2}{\partial x} + \frac{\partial uv}{\partial y} - fv = -\frac{\partial \phi}{\partial x} + F_u - D_u, \quad (12.19)$$

where F_u and D_u represent the effects of any forcing and dissipation. Zonal averaging, with $\overline{v} = 0$, gives

$$\frac{\partial \overline{u}}{\partial t} = -\frac{\partial \overline{u'v'}}{\partial y} + \overline{F}_u - \overline{D}_u, \quad (12.20)$$

or, using (12.17),

$$\frac{\partial \overline{u}}{\partial t} = \overline{v'\zeta'} + \overline{F}_u - \overline{D}_u. \quad (12.21)$$

Thus, the zonally averaged wind is maintained by the zonally averaged vorticity flux. On average there is little if any direct forcing of horizontal momentum and we may set $\overline{F}_u = 0$, and if the dissipation is parameterized by a linear drag (12.21) becomes

$$\frac{\partial \overline{u}}{\partial t} = \overline{v'\zeta'} - r\overline{u}, \quad (12.22)$$

where the constant r is an inverse frictional time scale.

Now consider the maintenance of this vorticity flux. The barotropic vorticity equation is

$$\frac{\partial \zeta}{\partial t} + \mathbf{u} \cdot \nabla \zeta + \nu \beta = F_\zeta - D_\zeta, \quad (12.23)$$

where F_ζ and D_ζ are forcing and dissipation of vorticity. Linearize about a mean zonal flow to give

$$\frac{\partial \zeta'}{\partial t} + \overline{u} \frac{\partial \zeta'}{\partial x} + \gamma v' = F'_\zeta - D'_\zeta, \quad (12.24)$$

where

$$\gamma = \beta - \frac{\partial^2 \bar{u}}{\partial y^2} \quad (12.25)$$

is the meridional gradient of absolute vorticity. Multiply (12.24) by ζ' / γ and zonally average to form the pseudomomentum equation,

$$\frac{\partial \mathcal{A}}{\partial t} + \bar{v}' \bar{\zeta}' = \frac{1}{\gamma} (\bar{\zeta}' F'_\zeta - \bar{\zeta}' D'_\zeta), \quad (12.26)$$

where

$$\mathcal{A} = \frac{1}{2\gamma} \bar{\zeta}'^2 \quad (12.27)$$

is a wave activity density, equal to the (negative of) the pseudomomentum for this problem [see also (7.23b) on page 299, and section 7.2 more generally]. The parameter γ is positive if the average absolute vorticity increases monotonically northwards, and this is usually the case in both Northern and Southern Hemispheres.

In the absence of forcing and dissipation, (12.22) and (12.26) imply an important relationship between the change of the mean flow and the pseudomomentum, namely

$$\frac{\partial \bar{u}}{\partial t} + \frac{\partial \mathcal{A}}{\partial t} = 0. \quad (12.28)$$

Now if for some reason \mathcal{A} increases, perhaps because a wave enters an initially quiescent region because of stirring elsewhere, then mean flow must decrease. However, because the vorticity flux integrates to zero, the zonal flow cannot decrease everywhere. Thus, if the zonal flow decreases in regions away from the stirring, it must *increase* in the region of the stirring. In the presence of forcing and dissipation this mechanism can lead to the production of a statistically steady jet in the region of the forcing, since (12.22) and (12.26) combine to give

$$\frac{\partial \bar{u}}{\partial t} + \frac{\partial \mathcal{A}}{\partial t} = -r\bar{u} + \frac{1}{\gamma} (\bar{\zeta}' F'_\zeta - \bar{\zeta}' D'_\zeta), \quad (12.29)$$

and in a statistically steady state

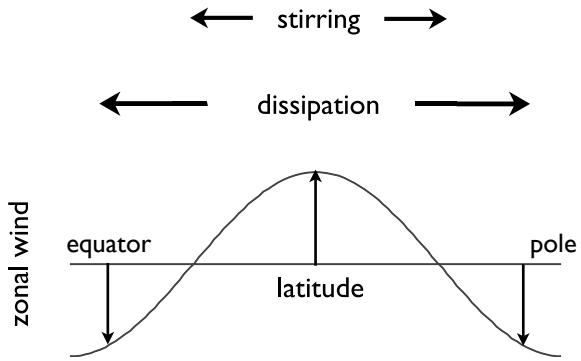
$$r\bar{u} = \frac{1}{\gamma} (\bar{\zeta}' F'_\zeta - \bar{\zeta}' D'_\zeta) .$$

(12.30)

The terms on the right-hand side represent the stirring and dissipation of vorticity, and integrated over latitude their sum will vanish, or otherwise the pseudomomentum budget cannot be in a steady state. However, let us suppose that forcing is confined to mid-latitudes. In the forcing region, the first term on the right-hand side of (12.30) will be larger than the second, and an eastward mean flow will be generated. Away from the direct influence of the forcing, the dissipation term will dominate and westward mean flows will be generated, as sketched in Fig. 12.5. Thus, *on a β -plane or on the surface of a rotating sphere an eastward mean zonal flow can be maintained by a vorticity stirring that imparts no net momentum to the fluid*. In general, stirring in the presence of a vorticity gradient will give rise to a mean flow, and on a spherical planet the vorticity gradient is provided by differential rotation.

It is crucial to the generation of a mean flow that the dissipation has a broader latitudinal distribution than the forcing: if all the dissipation occurred in the region of the forcing then

Fig. 12.5 Mean flow generation by a meridionally confined stirring. Because of Rossby wave propagation away from the source region, the distribution of pseudomomentum dissipation is broader than that of pseudomomentum forcing, and the sum of the two leads to the zonal wind distribution shown, with positive (eastward) values in the region of the stirring. See also Fig. 12.8.



from (12.30) no mean flow would be generated. However, Rossby waves are generated in the forcing region, and these propagate meridionally before dissipating thus broadening the dissipation distribution and allowing the generation of a mean flow.

IV. The Eliassen-Palm flux

The Eliassen-Palm (EP) flux (section 7.2) provides a convenient framework for determining how waves affect the mean flow, and the barotropic case is a particularly simple and instructive example. In the unforced case, the zonally averaged momentum equation may be written as

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v}^* = \nabla_x \cdot \mathcal{F}, \quad (12.31)$$

where \bar{v}^* is the residual meridional velocity and \mathcal{F} is the Eliassen-Palm flux, and $\nabla_x \cdot$ is the divergence in the meridional plane. In the barotropic case $\bar{v}^* = 0$ and

$$\mathcal{F} = -\mathbf{j} \bar{u}' \bar{v}' . \quad (12.32)$$

If the momentum flux is primarily the result of interacting nearly monochromatic waves, then the EP flux obeys the group velocity property (section 7.2.2), namely that the flux of wave activity density is equal to the group velocity times the wave activity density. Thus,

$$\mathcal{F}^y \equiv \mathbf{j} \cdot \mathcal{F} \approx c_g^y \mathcal{A}, \quad (12.33)$$

where \mathcal{A} is the wave activity density, or pseudomomentum,

$$\mathcal{A} = \frac{\overline{\zeta'^2}}{\overline{q}_y} = \frac{\overline{\zeta'^2}}{\gamma}, \quad (12.34)$$

and, if $\gamma > 0$, \mathcal{A} is a positive-definite quantity. The zonal momentum equation and the Eliassen-Palm relation, (7.23a), become respectively

$$\frac{\partial \bar{u}}{\partial t} = \frac{\partial}{\partial y} (c_g^y \mathcal{A}), \quad \frac{\partial \mathcal{A}}{\partial t} + \frac{\partial}{\partial y} (c_g^y \mathcal{A}) = 0, \quad (12.35a,b)$$

and so

$$\frac{\partial \bar{u}}{\partial t} = -\frac{\partial \mathcal{A}}{\partial t}, \quad (12.36)$$

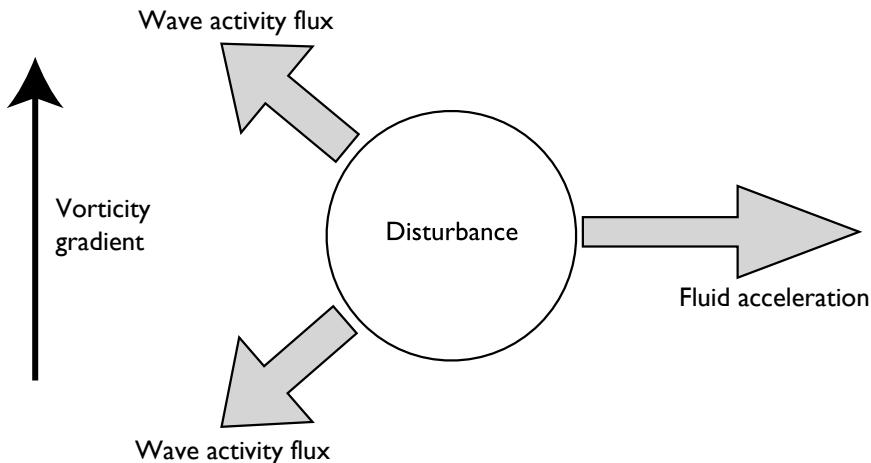


Fig. 12.6 If a region of fluid on the β -plane or on a rotating sphere is stirred, then Rossby waves propagate away from the disturbance, and this is the direction of the wave activity density flux vector. Thus, there is divergence of wave activity in the stirred region, and using (12.35) this produces an eastward acceleration.

as in (12.28).

Now suppose that we initiate a disturbance at some latitude, and then let the fluid evolve freely. The disturbance generates Rossby waves whose group velocity will be directed away from the region of disturbance, and the wave activity density \mathcal{A} will diminish in the region of the disturbance (and increase elsewhere). However, from (12.35a) the zonal velocity will *increase* in the region of the disturbance, and an eastward flow will be generated. That is, momentum converges in the region of the disturbance and an eastward jet is generated (Fig. 12.6).

This EP flux argument, the pseudomomentum argument and the Rossby wave argument are, in essence, just different expressions of the same physical process. Indeed, the result of (12.10) can be regarded as illustrating the group velocity property of the EP flux for barotropic Rossby waves. The vorticity budget argument is arguably more general than any of these, because it does not depend on linearization or small amplitude disturbances. At the same time, none of these arguments is really turbulent in nature, and although they are all fundamentally nonlinear, it is the presence of β -effect — a linear term in the vorticity equation — that is crucial in the development of a mean flow.

12.1.3 A numerical example

We conclude from the above arguments that momentum will converge into a rapidly rotating flow that is stirred in a meridionally localized region. To illustrate this, we numerically integrate the barotropic vorticity equation on the sphere, with a meridionally localized stirring term; explicitly, the equation that is integrated is

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) + \beta \frac{\partial \psi}{\partial x} = -r\zeta + \kappa \nabla^4 \zeta + F_\zeta. \quad (12.37)$$

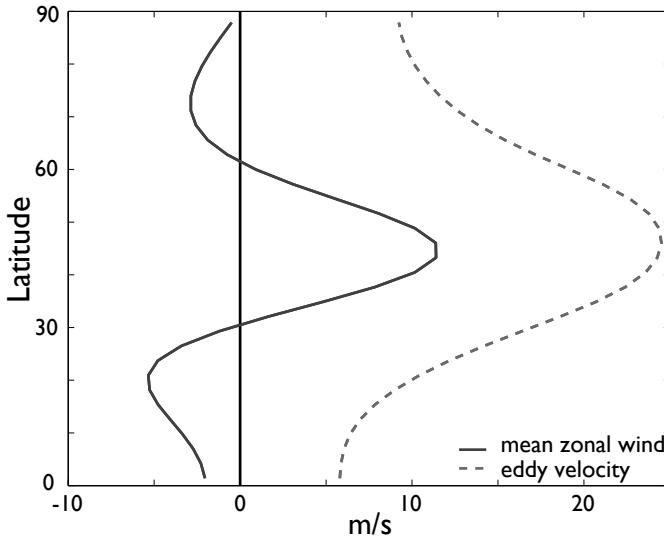


Fig. 12.7 The time and zonally averaged wind (solid line) obtained by an integration of the barotropic vorticity equation (12.37) on the sphere. The fluid is stirred in mid-latitudes by a random wavemaker that is statistically zonally uniform, acting around zonal wavenumber 8, and that supplies no net momentum. Momentum converges in the stirring region leading to an eastward jet with a westward flow to either side, and zero area-weighted spatially integrated velocity. The dashed line shows the r.m.s. (eddy) velocity created by the stirring.

The first term on the right-hand side is a linear drag, parameterizing momentum loss in an Ekman layer. The second term removes enstrophy that has cascaded to small scales; it has a negligible impact at large scales. The forcing term F_ζ is a wavemaker confined to a zonal strip of about 15° meridional extent, centred at about 45°N , that is statistically zonally uniform and that spatially integrates to zero. Within that region it is a random stirring with a temporal decorrelation scale of a few days and a spatial decorrelation scale corresponding to about wavenumber 8, thus mimicking weather scales. Thus, it provides no net source of vorticity or momentum, but it is a source of pseudomomentum because $\bar{F}_\zeta \bar{\zeta} > 0$.

The results of a numerical integration of (12.37) are illustrated in Figs. 12.7 and 12.8. An eastward jet forms in the vicinity of the forcing, with westward flow on either side. The pseudomomentum stirring and dissipation that produce this flow are shown in Fig. 12.8. As expected, the dissipation has a broader distribution than the forcing, and the sum of the two (the dot-dashed line in the figure) has the same meridional distribution as the zonal flow itself.

12.2 LAYERED MODELS OF THE MID-LATITUDE CIRCULATION

Let us now extend our barotropic model in the direction of increasing realism. So far we have shown that localized stirring can give rise to an eastward acceleration as Rossby waves propagate away from the disturbance. The source of the disturbance is of course baroclinic

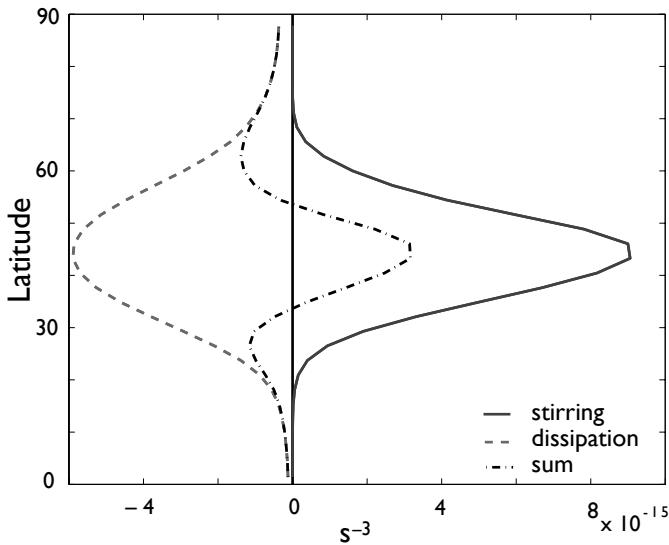


Fig. 12.8 The pseudomomentum stirring (solid line, $\overline{F'_\zeta \zeta'}$), dissipation (dashed line, $\overline{D'_\zeta \zeta'}$) and their sum (dot-dashed), for the same integration as Fig. 12.7. Because Rossby waves propagate away from the stirred region before breaking, the distribution of dissipation is broader than the forcing, resulting in an eastward jet where the stirring is centred, with westward flow on either side.

instability, and to incorporate these effects we will need to introduce some vertical structure into the problem; we do this by way of layered models of the circulation.⁴ That is, we will consider an atmosphere to consist of one or more *isentropic layers*, as described in section 3.9. The equations describing such layers are virtually isomorphic to the shallow water equations and, for the sake of familiarity and simplicity, and with no loss of essential dynamics, we will use the Boussinesq shallow water equations. We begin with a model of a single layer.

12.2.1 A single-layer model

We first consider a single layer obeying the shallow water equations. We further restrict the flow by supposing that it is constrained by two rigid surfaces: an upper flat lid and a lower, wavy (but stationary) surface (Fig. 12.9). We may imagine the fluid layer to crudely represent the upper troposphere, with the (given) lower wavy surface corresponding to the undulating mid-atmosphere interface of a two-layer model. (This section is in some ways an exercise, and too much realism should not be ascribed to the model.) Thus frictional effects are small in the momentum equation, and in particular there is no Ekman layer and no drag on the velocity field. However, there may be some dissipative effects in the vorticity equation, arising from the cascade of enstrophy to small scales. We also suppose that the Rossby number is small, that the variations in layer thickness are small compared to the mean layer thickness, and that variations in Coriolis parameter are small. Let the initial flow be a uniform zonal current, passing over the wavy lower boundary. The boundary is waviest

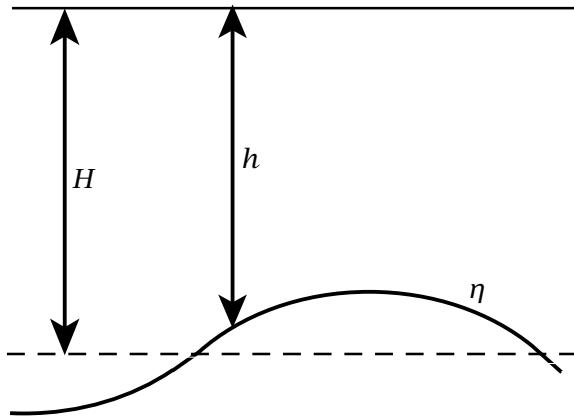


Fig. 12.9 A model atmosphere with an active layer of mean thickness H , local thickness h , and a variable lower surface of height displacement η , lying above a stationary layer with a slightly larger potential density.

in mid-latitudes, creating a disturbance from which Rossby waves emanate. Our questions are: (i) How does the wavy interface affect the mean zonal flow? (ii) What if any meridional circulation is induced?

Equations of motion

The zonal momentum equation for the layer may be written as

$$\frac{\partial u}{\partial t} - (f + \zeta)v = -\frac{\partial B}{\partial x}, \quad (12.38)$$

where $B = \phi + \mathbf{u}^2/2$ is the Bernoulli function and ϕ is the pressure in the layer. The zonal average of this is

$$\frac{\partial \bar{u}}{\partial t} - f\bar{v} = \bar{\zeta}\bar{v} + \bar{\zeta}'v'. \quad (12.39)$$

Note that \bar{v} is wholly ageostrophic ($\bar{v}_g = \bar{\partial}_x \psi = 0$). Now, using $\partial u/\partial x + \partial v/\partial y = 0$, the vorticity flux is related to the momentum flux by

$$v\zeta = -\frac{\partial}{\partial y}(uv) + \frac{1}{2}\frac{\partial}{\partial x}(v^2 - u^2), \quad (12.40)$$

so that, under quasi-geostrophic scaling, (12.39) simplifies to

$$\frac{\partial \bar{u}}{\partial t} - f_0\bar{v} = \bar{\zeta}'v' = -\frac{\partial}{\partial y}\bar{u}'v'. \quad (12.41)$$

Note that although \bar{v} is small and ageostrophic, mass conservation does not demand that it be zero, because the thickness of the layer is not constant [look ahead to (12.44)]. Thus, as \bar{v} is multiplied by the large term f_0 , the term $f_0\bar{v}$ term should be retained (whereas $\bar{\zeta}\bar{v}$ is dropped). If the flow is statistically steady and there are no sources or sinks of momentum (12.41) becomes

$$f_0\bar{v} = \frac{\partial}{\partial y}\bar{u}'v'. \quad (12.42)$$

To complete the model we use the zonally averaged mass conservation equation, namely

$$\frac{\partial \bar{h}}{\partial t} + \frac{\partial}{\partial y} \bar{v} \bar{h} = 0. \quad (12.43)$$

In the situation here $\partial \bar{h} / \partial t = 0$, because the flow is confined between two rigid surfaces, and so $\partial \bar{v} \bar{h} / \partial y = 0$. If the mass flux vanishes somewhere, for example at a meridional boundary, it therefore vanishes everywhere and we have

$$\bar{v} \bar{h} + \bar{v}' \bar{h}' = 0. \quad (12.44)$$

Using (12.42) and (12.44) gives

$$\frac{1}{f_0} \frac{\partial}{\partial y} \bar{u}' \bar{v}' + \frac{1}{\bar{h}} \bar{v}' \bar{h}' = 0, \quad (12.45)$$

or

$$\bar{v}' \bar{\zeta}' - f_0 \frac{1}{\bar{h}} \bar{v}' \bar{h}' = 0. \quad (12.46)$$

Because thickness variations are assumed to be small we may write this as

$$\bar{v}' \bar{\zeta}' - f_0 \frac{1}{H} \bar{v}' \bar{h}' = 0, \quad (12.47)$$

where H is the reference thickness of the layer, which may be taken as its mean thickness. [The left-hand side of (12.47) is just the potential vorticity flux for this problem — look ahead to (12.53).]

The potential vorticity equation for the layer is

$$\frac{DQ}{Dt} = \frac{D}{Dt} \left[\frac{\zeta + f}{h} \right] = 0, \quad (12.48)$$

where h is the fluid layer thickness. For small variations in layer thickness and Coriolis parameter this becomes

$$\frac{Dq}{Dt} = \frac{\partial q}{\partial t} + u \frac{\partial q}{\partial x} + v \frac{\partial q}{\partial y} = 0, \quad q = \zeta + \beta y + f_0 \frac{\eta}{H}, \quad (12.49a,b)$$

where $\eta = H - h$ is the height of the lower interface (Fig. 12.9) and this is a function of x and y but not, in this model, time. Using the horizontal non-divergence of the flow, the zonally averaged potential vorticity equation is

$$\frac{\partial \bar{q}}{\partial t} = - \frac{\partial \bar{v} \bar{q}}{\partial y} - \frac{\partial \bar{v}' \bar{q}'}{\partial y}. \quad (12.50)$$

The term involving \bar{v} is very small, and omitting it and using (12.49a) we obtain the perturbation potential vorticity equation

$$\frac{\partial q'}{\partial t} + \bar{u} \frac{\partial q'}{\partial x} + v' \frac{\partial \bar{q}}{\partial y} = -D' \quad (12.51)$$

where we include a term, D' , to represent dissipative processes. Multiplying by $q' / (\partial \bar{q} / \partial y)$ and zonally averaging we obtain the pseudomomentum equation for this system, namely

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial t} \left(\frac{\bar{q}^2}{2y} \right) = -\bar{v}' \bar{q}' - \frac{D' \bar{q}'}{\gamma}, \quad (12.52)$$

where $\gamma = \partial \bar{q} / \partial y$. This equation is the equivalent of (12.26), but now for the layered system. In a turbulent fluid we cannot, in general, demand that $D' = 0$, even as the viscosity goes to zero, because of the presence of an enstrophy flux to smaller scales and a concomitant dissipation. But in regions where D' is zero (where there is no wave breaking) then the potential vorticity flux must also be zero in a steady state. For our argument let us assume that, in fact, $D' = 0$.

Using (12.49b), the eddy potential vorticity flux is

$$\overline{v' q'} = \overline{v' \zeta'} + \frac{f_0}{H} \overline{v' \eta'} = \overline{v' \zeta'} - \frac{f_0}{H} \overline{v' h'}, \quad (12.53)$$

where η' is the topography and h' is the layer thickness perturbation. Using this in the zonal momentum equation (12.41) gives

$$\frac{\partial \bar{u}}{\partial t} = \overline{v' q'} + \frac{f_0}{H} \overline{v' h'} + f_0 \bar{v}. \quad (12.54)$$

But the last two terms on the right-hand side constitute the total mass flux so we finally write

$$\frac{\partial \bar{u}}{\partial t} = \overline{v' q'} + f_0 \bar{v}^*.$$

$$\bar{v}^* = \bar{v} + \frac{\overline{v' h'}}{H}.$$

(12.55a,b)

The quantity \bar{v}^* is the *residual circulation* for this problem; it is proportional to the sum of mass flux from the mean flow and the eddies (see section 7.3 for more discussion). Now, \bar{v}^* is proportional to the total meridional mass flux and therefore, because the flow is confined between rigid lids and if there are no sources or sinks of mass, $\bar{v}^* = 0$ everywhere [see (12.44) with $\bar{h} = H$].

Dynamics

When the flow passes over the wavy boundary, Rossby waves will, as in the barotropic case, cause momentum flux to converge in the generation region. If the flow is steady and dissipation-free then from the momentum equation

$$f_0 \bar{v} = \frac{\partial \overline{u' v'}}{\partial y}, \quad (12.56)$$

and, in regions of momentum flux convergence (i.e., where $\partial \overline{u' v'} / \partial y < 0$) *the mean meridional velocity is equatorwards*. Thus, whereas frictional forces balance the vorticity flux in a constant-thickness barotropic model (because in that case $\bar{v} = 0$) in the free atmosphere a meridional circulation may be generated, and this is basis of the equatorial flow in the upward branch of the Ferrel cell. However, this does not imply that the *total* mass flux is equatorwards; in fact, for this single-layer model it must be zero and therefore

$$\overline{v' h'} = -\bar{h} \bar{v} > 0. \quad (12.57)$$

That is, *the eddy mass flux is polewards*, balancing the equatorward mean flow. These balances are sketched in Fig. 12.10.

Informal Summary of the Single-Layer Arguments

The zonally averaged momentum equation is

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v} = \bar{v}' \zeta' = -\frac{\partial \bar{u}' v'}{\partial y}. \quad (\text{SL.1})$$

A region that is the source of Rossby waves will generally be a region where there is momentum flux convergence, where $\partial \bar{u}' v' / \partial y < 0$. In this region \bar{v} will be directed *equatorwards* if \bar{u} is steady, and this flow is the upper branch of the Ferrel Cell. To think about this in terms of potential vorticity, first define the residual meridional velocity by

$$\bar{v}^* = \frac{\bar{v}' h'}{h} + \bar{v}. \quad (\text{SL.2})$$

This is proportional to the total meridional mass flux and is zero in this one-layer model. The momentum equation is then

$$\frac{\partial \bar{u}}{\partial t} = f_0 \bar{v}^* - \frac{f_0}{h} \bar{v}' h' + \bar{v}' \zeta' \quad (\text{SL.3a})$$

$$= f_0 \bar{v}^* + \bar{v}' q'. \quad (\text{SL.3b})$$

using $\bar{v}' q' = \bar{v}' \zeta' - (f_0/h) \bar{v}' h'$, where q the is potential vorticity. The second term on the right-hand side of (SL.3a) is the *form drag* exerted by the topography on the flow, and in a steady state this balances the momentum flux convergence of the Rossby waves. Because of the presence of Rossby waves we expect $\bar{v}' \zeta' > 0$. If there is no dissipation then in steady flow $\bar{v}' q' = -f_0 \bar{v}^* = 0$ and the eddy mass flux is polewards (positive if $f_0 > 0$) and a meridional flow is generated as in Fig. 12.10.

We can infer the potential vorticity flux more directly using the pseudomomentum equation:

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial t} \left(\frac{q'^2}{2\gamma} \right) = -\bar{v}' q' - \frac{\bar{D}' q'}{\gamma}. \quad (\text{SL.4})$$

where $\gamma = \partial \bar{q} / \partial y$. If dissipation is identically zero, then the potential vorticity flux is zero if the waves are steady. Then, using (SL.3b), there is no acceleration of the zonal flow — an example of the *non-acceleration theorem*. More generally (and in the real atmosphere) there *will* be some dissipation away from the source region: Rossby waves will break in critical layers (where $\bar{u} = c$), or more generally Rossby waves will interact producing an enstrophy cascade, giving $\bar{D}' q' > 0$ and (for $\gamma > 0$) a negative potential vorticity flux, $\bar{v}' q' < 0$. In these regions, a balance in the momentum equation (SL.3b) can be achieved either by balancing the PV flux with a friction term, as in the barotropic model of section 12.1, or by a Coriolis force on a poleward residual meridional velocity. That is, $f_0 \bar{v}^* \approx -\bar{v}' q' > 0$, so generating a poleward residual flow.

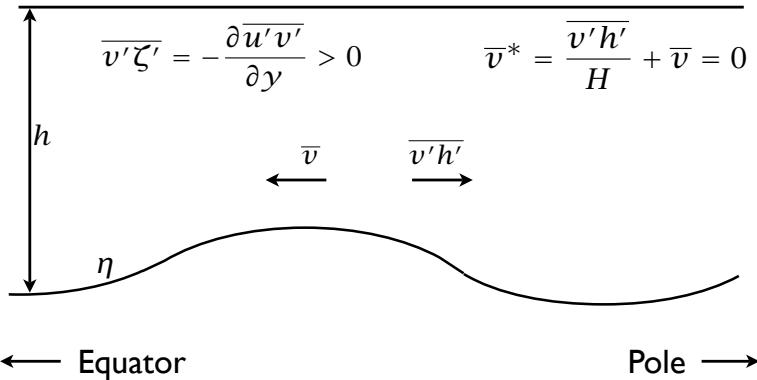


Fig. 12.10 Dynamics of a single layer, with no dissipation. The force on the active layer arises from the form drag exerted by the interface. Vorticity dynamics demands that this produce a converging eddy momentum flux ($\partial_y \bar{u}' \bar{v}' < 0$), which in turn produces a poleward eddy mass flux ($\bar{v}' \bar{h}' > 0$), and so an equatorward Eulerian flow.

Another way to arrive at this result is to utilize potential vorticity fluxes directly. For steady, dissipation-free flow the pseudomomentum equation (12.52) reveals that the potential vorticity flux vanishes. Then using (12.55) and noting that $\bar{v}^* = 0$ we have $\partial \bar{u} / \partial t = 0$ — an example of the non-acceleration theorem that steady non-dissipative waves do not induce a change in the zonal momentum. Then, using (12.53) we find

$$\bar{v}' \bar{\zeta}' = \frac{f_0}{H} \bar{v}' \bar{h}' \quad (12.58)$$

and using (12.44) we recover (12.56).

From the point of view of the momentum equation, the momentum flux convergence is balanced by the *form drag* caused by the flow over the wavy boundary. To see this we use (12.55b) to write the zonal momentum equation as

$$\frac{\partial \bar{u}}{\partial t} = f_0 \bar{v} + \bar{v}' \bar{\zeta}' = f_0 \bar{v}^* - \frac{f_0}{H} \bar{v}' \bar{h}' + \bar{v}' \bar{\zeta}', \quad (12.59)$$

where here \bar{v}^* (but not \bar{v}) is zero. The term $-(f_0/H) \bar{v}' \bar{h}'$ is the force on the fluid layer coming from the wavy boundary — the form drag, as described in section 3.5. Specifically, the average force per unit area exerted on the layer by the sloping surface is given by

$$F = -f_0 \rho_0 \bar{v}' \bar{\eta}' = f_0 \rho_0 \bar{v}' \bar{h}', \quad (12.60)$$

and dividing by $\rho_0 H$ provides the acceleration on the active fluid layer. The atmosphere also exerts an equal and opposite force on the wavy surface, an effect we consider in the next section. A steady state is achieved, without dissipation, when the form drag is balanced by the eddy momentum flux convergence. From (12.53) or (12.55), this is the same condition that the potential vorticity flux vanishes.

Final remarks on the one-layer model

A summary of the single-layer arguments is given in the shaded box on the previous page. In the single-layer model, as in the barotropic model, the zonal flow is proximately driven by

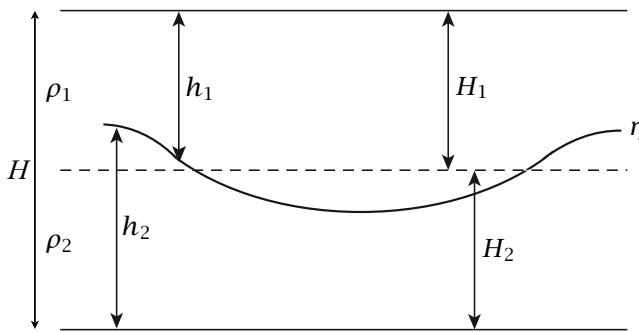


Fig. 12.11 An atmosphere with two homogeneous (or isentropic) layers of mean thickness H_1 and H_2 , local thickness h_1 and h_2 , and interface η , contained between two flat, rigid surfaces.

eddy fluxes of potential vorticity, and in the model the eddy fluxes must be zero if a steady state is to be achieved. *Vis-à-vis* the real atmosphere this is a little unrealistic, because from the pseudomomentum equation (12.52) we expect these fluxes to be negative, and there is then nothing to balance them in the momentum equation, (12.55), if $\bar{v}^* = 0$. In the real atmosphere, there are effectively sources and sinks in the mass conservation equation that arise from the thermodynamics that allow \bar{v}^* to be non-zero; we then expect $\bar{v}^* > 0$, but to explore this requires a two-layer model, in which the single layer of the one-layer model will correspond to the upper layer of the two-layer model.

12.2.2 A two-layer model

We now consider a model with two active layers, constructing what is probably the simplest model that can capture the dynamics of the mid-latitude tropospheric general circulation without undue approximation. Indeed virtually all of the phenomenology that we associate with the circulation — a thermal wind, mid-latitude surface westerly winds, the Ferrel cell, breaking Rossby waves — is present. A three-layer model introduces no new physics, although a continuously stratified model does lead to some differences of interpretation. The physical model we have in mind is one of two isentropic layers of a compressible ideal gas, virtually equivalent to a two-layer shallow water model illustrated in Fig. 12.11, and our presentation will be in terms of the latter. The upper layer may be thought of as being forced by an undulating interface between the lower and upper layers, a crude representation of stratification. We will continue to assume that quasi-geostrophic scaling holds; that is, the flow is in near geostrophic balance, variations in layer thickness are small compared to their mean thickness and variations in the Coriolis parameter are small. We also assume that the two fluid layers are held between two flat rigid lids — topography is an unnecessary complication at this stage.

Equations of motion

The equations of motion are those of a two-layer Boussinesq shallow water model confined between two rigid flat surfaces, and readers comfortable with these dynamics (see sections 3.3 and 3.4) may quickly skip through this section, glancing at the boxed equations. The momentum equations of each layer are

$$\frac{D\mathbf{u}_1}{Dt} + \mathbf{f} \times \mathbf{u}_1 = -\nabla \phi_1 \quad (12.61a)$$

$$\frac{D\mathbf{u}_2}{Dt} + \mathbf{f} \times \mathbf{u}_2 = -\nabla \phi_2 - r\mathbf{u}_2, \quad (12.61b)$$

where $\phi_1 = p_T/\rho_0$ and $\phi_2 = p_T/\rho_0 + g'\eta$, with p_T being the pressure at the lid at the top, η the interface displacement (see Fig. 12.11) and $g' = g(\rho_2 - \rho_1)/\rho_0$ the reduced gravity, and we may take $\rho_0 = \rho_1$. We have also included a simple representation of surface drag, $-r\mathbf{u}_2$, in the lowest layer, and r is a constant. We will use a constant value of the Coriolis parameter except where it is differentiated, and on zonal averaging the zonal components of (12.61) become

$$\frac{\partial \bar{u}_1}{\partial t} - f_0 \bar{v}_1 = \bar{v}'_1 \zeta'_1 \quad (12.62a)$$

$$\frac{\partial \bar{u}_2}{\partial t} - f_0 \bar{v}_2 = \bar{v}'_2 \zeta'_2 - r \bar{u}_2. \quad (12.62b)$$

Geostrophic balance in each layer implies

$$f_0 \mathbf{u}_{g1} = \mathbf{k} \times \nabla \phi_1, \quad f_0 \mathbf{u}_{g2} = \mathbf{k} \times \nabla \phi_1 + g' \mathbf{k} \times \nabla \eta, \quad (12.63a,b)$$

where the subscript g denotes geostrophic. Subtracting one equation from the other gives

$$\boxed{f_0(\mathbf{u}_1 - \mathbf{u}_2) = -g' \mathbf{k} \times \nabla \eta}, \quad (12.64)$$

dropping the subscripts g on \mathbf{u} . This is thermal wind balance (or the Margules relation) for this system. A temperature gradient thus corresponds to a slope of the interface height, with the interface sloping upwards toward lower temperatures, analogous to isentropes sloping up toward the pole in the real atmosphere.

The quasi-geostrophic potential vorticity for each layer is

$$q_i = \zeta_i + f - f_0 \frac{h_i}{H_i}, \quad (12.65)$$

where H_i is the reference thickness of each layer, which we take to be its mean thickness. The potential vorticity flux in each layer is then

$$\bar{v}'_i \bar{q}'_i = \bar{v}'_i \zeta'_i - \frac{f_0}{H_i} \bar{v}'_i h'_i. \quad (12.66)$$

Using this in (12.62) gives

$$\boxed{\begin{aligned} \frac{\partial \bar{u}_1}{\partial t} &= \bar{v}'_1 q'_1 + f_0 \bar{v}'_1 \\ \frac{\partial \bar{u}_2}{\partial t} &= \bar{v}'_2 q'_2 + f_0 \bar{v}'_2 - r \bar{u}_2 \end{aligned}}, \quad (12.67a,b)$$

where

$$\bar{v}'_i = \bar{v}_i + \frac{\bar{v}'_i h'_i}{H_i} \quad (12.68)$$

is the meridional component of the residual velocity in each layer, proportional to the *total*

meridional mass flux in each layer. These are the transformed Eulerian mean (TEM) forms of the equations, first encountered in section 7.3.

In the barotropic model of section 12.1 the mean meridional velocity vanished at every latitude, a consequence of mass conservation in a single layer between two rigid flat surfaces. In the single-layer model of section 12.2.1 the mean meridional velocity was in general non-zero, but the total meridional mass flux (i.e., the meridional component of the residual velocity) was zero if the domain is bounded laterally by solid walls. In the two-layer model we will allow a transformation of mass from one layer to another, which is the equivalent of heating: a conversion of mass from the lower layer to the upper layer is heating, and conversely for cooling. Thus, heating at low latitudes and cooling at high latitudes leads to the interface sloping upwards toward the pole. In the two-layer model the constraint that mass conservation supplies is that, assuming a statistically steady state, the total poleward mass flux summed over both layers must vanish.

The mass conservation equation for each layer is

$$\frac{\partial h_i}{\partial t} + \nabla \cdot (h_i \mathbf{u}_i) = S_i, \quad (12.69)$$

where S_i is the mass source term and we may suppose that $S_1 + S_2 = 0$ everywhere. A zonal average gives

$$\frac{\partial \bar{h}_i}{\partial t} + \frac{\partial \bar{h}_i v_i}{\partial y} = \bar{S}_i \quad (12.70)$$

or, setting $\bar{h}_i = H_i$ and using (12.68),

$$\frac{\partial \bar{h}_i}{\partial t} + H_i \frac{\partial \bar{v}_i^*}{\partial y} = \bar{S}_i \quad (12.71)$$

The mass source term in these equations is equivalent to heating, and let us suppose that this is such as to provide heating at low latitudes and cooling at high ones. This is equivalent to conversion of an upper-layer mass to a lower-layer mass at high latitudes, and the reverse at low latitudes; such a conversion can only be balanced by a poleward mass flux in the upper layer and an equatorward mass flux in the lower layer (Fig. 12.12). That is to say, an Earth-like radiative forcing between equator and pole implies that *the total mass flux in the upper layer will be polewards*. This is the opposite of the mean meridional circulation of the Ferrel cell shown in Fig. 11.3! What's going on? Before we can answer that, let us manipulate the equations of motion and obtain a couple of useful preliminary results.

Manipulating the equations

Because the total depth of the fluid is fixed, the mass conservation equations in each layer, (12.69), may each be written as an equation for the interface displacement, namely

$$\frac{\partial \eta}{\partial t} + \nabla \cdot (\eta \mathbf{u}_1) = -S_1, \quad \text{or} \quad \frac{\partial \eta}{\partial t} + \nabla \cdot (\eta \mathbf{u}_2) = S_2, \quad (12.72a,b)$$

where $\eta = H_1 - h_1 = h_2 - H_2$ and $h_1 + h_2 = H_1 + H_2$. Because of the thermal wind equation, (12.64), (12.72a) and (12.72b) are identical: $\mathbf{u}_1 \cdot \nabla \eta = \mathbf{u}_2 \cdot \nabla \eta$ and $S_1 = -S_2$. (If $S_1 \neq -S_2$

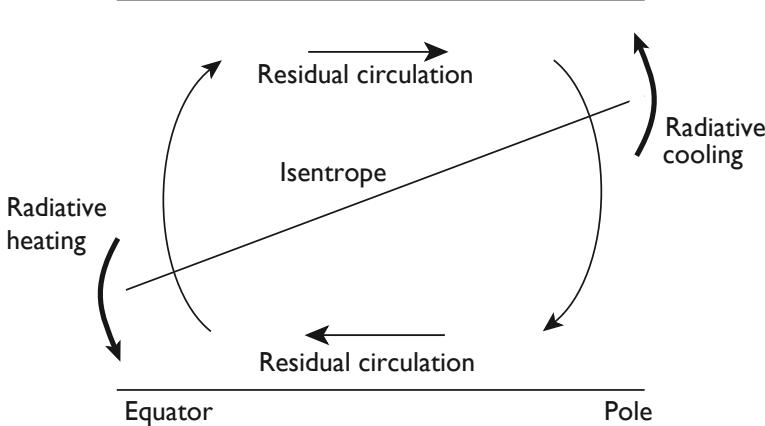


Fig. 12.12 Sketch of the zonally averaged thermodynamics of a two-layer model. Cooling at high latitudes and heating at low ones steepens the interface upward toward the pole (thicker arrows). Associated with this is a net mass flux — the residual flow, or the meridional overturning circulation (lighter arrows). In the tropics this circulation is accounted for by the Hadley Cell, and is nearly all in the mean flow. In mid-latitudes the circulation — the residual flow — is largely due to baroclinic eddies, and the smaller Eulerian mean flow is actually in the opposite sense.

the flow would not remain balanced and the thermal wind equation could not be satisfied.) The zonally averaged interface equation may be written as

$$\frac{\partial \bar{\eta}}{\partial t} - H_1 \frac{\partial \bar{v}_1^*}{\partial y} = \bar{S}, \quad \text{or} \quad \frac{\partial \bar{\eta}}{\partial t} + H_2 \frac{\partial \bar{v}_2^*}{\partial y} = \bar{S}, \quad (12.73)$$

where $\bar{S} = -\bar{S}_1 = +\bar{S}_2$, consistent with the mass conservation statement

$$H_1 \bar{v}_1^* + H_2 \bar{v}_2^* = 0, \quad (12.74)$$

which states that the vertically integrated total mass flux vanishes at each latitude.

Now, whereas (12.74) is a kinematic statement about the total mass flux, the dynamics provides a constraint on the *eddy* mass flux in each layer. Using the thermal wind relationship we have

$$f_0 (\bar{v}'_1 - \bar{v}'_2) \bar{\eta}' = g' \frac{\partial \bar{\eta}'}{\partial x} \bar{\eta}' = 0. \quad (12.75)$$

Hence, if the upper and lower surfaces are both flat, we have that

$$\bar{v}'_1 \bar{h}'_1 = -\bar{v}'_2 \bar{h}'_2, \quad (12.76)$$

and the eddy meridional mass fluxes in each layer are equal and opposite. If the bounding surfaces are not flat, we have

$$\bar{v}'_1 \bar{\eta}'_T - \bar{v}'_1 \bar{h}'_1 = \bar{v}'_2 \bar{\eta}'_B + \bar{v}'_2 \bar{h}'_2 \quad (12.77)$$

instead, where η_T and η_B are the topographies at the top and the bottom. Equations (12.76)

and (12.77) are *dynamical* results, and not just kinematic ones; they are equivalent to noting that the form drag on one layer due to the interface displacement is equal and opposite to that on the other, namely

$$\overline{v'_1 \eta'} = -[-\overline{v'_2 \eta'}], \quad (12.78)$$

where the minus sign inside the square brackets arises because the interface displacement is into layer one but out of layer two.

Using (12.66) and (12.77) the eddy potential vorticity fluxes in the two layers are related by

$$H_1 \overline{v'_1 q'_1} + H_2 \overline{v'_2 q'_2} = H_1 \overline{v'_1 \zeta'_1} + H_2 \overline{v'_2 \zeta'_2} - f_0 \overline{v'_1 \eta'_T} + f_0 \overline{v'_2 \eta'_B}, \quad (12.79)$$

which is the layered version of the continuous result [eq. (7.20) on page 298]

$$\int_B^T \overline{v' q'} dz = \int_B^T \overline{v' \zeta'} dz + \frac{f_0}{N^2} [\overline{v' b'}]_B^T. \quad (12.80)$$

For flat upper and lower surfaces, and using $\overline{v_i \zeta_i} = -\partial \overline{u_i v_i} / \partial y$, (12.79) becomes

$$H_1 \overline{v'_1 q'_1} + H_2 \overline{v'_2 q'_2} = -H_1 \frac{\partial}{\partial y} \overline{u'_1 v'_1} - H_2 \frac{\partial}{\partial y} \overline{u'_2 v'_2}, \quad (12.81)$$

and integrating with respect to y between quiescent latitudes gives

$$\int \left[H_1 \overline{v'_1 q'_1} + H_2 \overline{v'_2 q'_2} \right] dy = 0. \quad (12.82)$$

That is, the total meridional flux of potential vorticity must vanish. This is a consequence of the fact that the potential vorticity flux is the divergence of a vector field; in the continuous case

$$\overline{v' q'} = -\frac{\partial \overline{u' v'}}{\partial y} + f_0 \frac{\partial}{\partial z} \frac{\overline{v' b'}}{N^2}, \quad (12.83)$$

which similarly vanishes when integrated over a volume if there are no boundary contributions.

12.2.3 Dynamics of the two-layer model

We now consider the climate, or the time averaged statistics, of our two-layer model. The equations of motion are (12.62) or (12.67), and (12.70) or (12.71). These equations are not closed because of the presence of eddy fluxes, and in this section we make some phenomenological and rather general arguments about how these behave in order to get a sense of the general circulation. In the next section we use a specific closure to address the same problem.

Let us summarize the physical situation. The two layers of our model are confined in the vertical direction between two flat, rigid surfaces, and they are meridionally confined between slippery walls at high and low latitudes (the ‘pole’ and ‘equator’). The circulation is driven thermodynamically by heating at low latitudes and cooling at high ones, which translates to a conversion of layer 1 fluid to layer 2 fluid at high latitudes, and the converse at low latitudes (see Fig. 12.12). This sets up an interface that slopes upwards toward the

pole and, by thermal wind, a shear. This situation is baroclinically unstable, and this sets up a field of eddies, most vigorous in mid-latitudes where the temperature gradient (or interface slope) is largest. Three fields encapsulate the dynamics — the lower-layer wind field, the meridional circulation, and the meridional temperature gradient, and our goal is to understand their qualitative structure. We note from the outset that the residual circulation is polewards in the upper layer, equatorwards in the lower layer, and that this is a thermodynamic result, a consequence of heating at low latitudes and cooling at high latitudes.

From (12.67), the steady-state lower-layer wind is given by

$$rH_2\bar{u}_2 = H_1\bar{v}'_1\bar{q}'_1 + H_2\bar{v}'_2\bar{q}'_2 = H_1\bar{v}'_1\bar{\zeta}'_1 + H_2\bar{v}'_2\bar{\zeta}'_2 , \quad (12.84)$$

where the second equality uses (12.81). That is, *the lower-layer wind is determined by the vertical integral of either the vorticity flux or the potential vorticity flux.*

Neglecting contributions due to the mean horizontal shear (which are small if the *beta-Rossby number*, $U/\beta L^2$, is small) the potential vorticity gradient in each layer is given by

$$\frac{\partial \bar{q}_1}{\partial y} = \beta - \frac{f_0}{H_1} \frac{\partial \bar{h}_1}{\partial y} \gg 0 \quad \text{and} \quad \frac{\partial \bar{q}_2}{\partial y} = \beta - \frac{f_0}{H_2} \frac{\partial \bar{h}_2}{\partial y} \lesssim 0. \quad (12.85a,b)$$

In the upper layer $\partial \bar{h}_1 / \partial y$ is negative so that the total potential vorticity gradient is positive and larger than β itself. In the lower layer $\partial \bar{h}_2 / \partial y$ is positive and indeed if there is to be baroclinic instability it must be as large as β in order for $\partial \bar{q} / \partial y$ to change sign somewhere. Thus, although negative the potential vorticity gradient is much weaker in the lower layer. Thus, Rossby waves (meaning waves that exist because of a background gradient in potential vorticity) will propagate further in the upper layer, and this asymmetry is the key to the production of surface winds.

Now, the potential vorticity flux must be negative (and downgradient) in the upper layer, and there are various ways to see this. One is from the upper-layer momentum equation (12.67a) which in a steady state gives

$$\bar{v}'_1\bar{q}'_1 = -f_0\bar{v}_1^*. \quad (12.86)$$

Because \bar{v}_1^* is polewards, $f_0\bar{v}_1^*$ is positive and the potential vorticity flux is negative in both Northern and Southern Hemispheres. Equivalently, in the upper layer the radiative forcing is increasing the potential vorticity gradient between the equator and the pole, so there must be an equatorward potential vorticity flux to compensate. Finally, the perturbation enstrophy or pseudomomentum equations tell us that in a steady state the potential vorticity flux is downgradient (also see section 12.3.2). This is not an independent argument, since it merely says that the enstrophy budget may be balanced through a balance between production proportional to the potential vorticity gradient and the dissipation. For similar reasons we expect the potential vorticity flux to be positive (polewards) in the lower layer.

Now, (12.82) tells us that the latitudinally integrated potential vorticity flux is equal and opposite in the two layers. If the potential vorticity flux in the lower layer were everywhere equal and opposite to that in the upper layer, then using (12.84) there would be no surface wind, in contrast to the observations. In fact, the potential vorticity flux is more uniformly distributed in the upper layer, and this gives rise to the surface wind observed. Let us a

couple of perspectives (on the same argument) as to why this should be so. The argument centres around the fact that the potential vorticity gradient is stronger in the upper layer, as we can see from (12.85).

I. Rossby waves and the vorticity flux

The stronger potential vorticity gradient of the upper layer is better able to support linear Rossby waves than the lower layer. Thus, the vorticity flux in the region of Rossby-wave genesis in mid-latitudes will be large and positive in the upper layer, and small and negative in the lower layer. Thus, there will be more momentum convergence into the source region in the upper layer than in the lower layer, and the vertical integral of the vorticity flux will largely be dominated by that of the upper layer. This is positive in mid-latitudes and, to ensure that its latitudinal integral is zero, it is negative on either side. Using (12.84), a surface wind has the same pattern as the net vorticity flux, and so is eastwards in the mid-latitude source region and westwards on either side.

II. Potential vorticity flux

Rossby waves are generated in the region of baroclinic instability, at approximately the same latitude in both upper and lower layers. However, because the potential vorticity gradient is larger in the upper layer than in the lower layer, Rossby waves are able to propagate more efficiently and breaking and associated dissipation will tend to be further from the source region in the upper layer than in the lower layer. Now, the pseudomomentum equation for each layer is, similar to (12.52) for the one-layer case,

$$\frac{\partial P_i}{\partial t} = \frac{\partial}{\partial t} \left(\frac{\overline{q_i'^2}}{2\gamma_i} \right) = -\overline{v_i' q_i'} - \frac{\overline{D_i' q_i'}}{\gamma_i}, \quad i = 1, 2, \quad (12.87)$$

where γ_i , the potential vorticity gradient, has opposite signs in each layer. In a statistically steady state, the region of strongest dissipation is the region where the potential vorticity flux is largest. In the upper layer, Rossby-wave propagation allows the dissipation region to spread out from the source, whereas in the lower layer the dissipation region will be concentrated near the source. The distribution of the potential vorticity flux then becomes as illustrated in Fig. 12.13. The surface winds, being the vertical integral of the potential vorticity fluxes, are westerly in the baroclinic region and easterly to either side.

Momentum balance and the overturning circulation

From thermodynamic arguments we deduced that the residual circulation is ‘direct’, meaning that warm fluid rises at low latitudes, moves polewards aloft, and returns near the surface. At low latitudes where eddy effects are small the zonally averaged Eulerian circulation circulates in the same way, giving us the Hadley Cell. In mid-latitudes, we may determine the Eulerian circulation from the momentum equation, (12.62). In the upper layer the balance is between the vorticity flux and the Coriolis term, namely

$$f_0 \overline{v}_1 = -\overline{v_1' \zeta_1'} < 0. \quad (12.88)$$

That is, *the mean Eulerian flow is equatorwards*, and this is the upper branch of the Ferrel cell. Note that the Eulerian circulation is in the opposite sense to the residual circulation.

Phenomenology of a Two-layer Mid-latitude Atmosphere

A radiative forcing that heats low latitudes and cools high latitudes will lead to an interface that slopes upward with increasing latitude, and a poleward total mass flux in the upper layer and an equatorward flux in the lower layer. The interface implies a thermal wind shear between the two layers. Neglecting relative vorticity, the potential vorticity gradients in each layer are given by

$$\frac{\partial \bar{q}_1}{\partial y} = \beta - \frac{f_0}{H_1} \frac{\partial \bar{h}_1}{\partial y} > 0 \quad \text{and} \quad \frac{\partial \bar{q}_2}{\partial y} = \beta - \frac{f_0}{H_2} \frac{\partial \bar{h}_2}{\partial y} \lesssim 0. \quad (\text{TL.1})$$

The gradient is large and positive in the upper layer and small and negative in the lower layer — the gradient must change sign if there is to be baroclinic instability which we assume to be the case. This baroclinic instability generates eddy fluxes that largely determine the surface winds and the meridional overturning circulation. The zonal momentum equation in each layer is

$$\frac{\partial \bar{u}_1}{\partial t} = f_0 \bar{v}_1 + \bar{v}'_1 \bar{\zeta}'_1 = f_0 \bar{v}_1^* + \bar{v}'_1 \bar{q}'_1 \quad (\text{TL.2a})$$

$$\frac{\partial \bar{u}_2}{\partial t} = f_0 \bar{v}_2 + \bar{v}'_2 \bar{\zeta}'_2 - r \bar{u}_2 = f_0 \bar{v}_2^* + \bar{v}'_2 \bar{q}'_2 - r \bar{u}_2. \quad (\text{TL.2b})$$

In steady state the potential vorticity flux will be equatorwards in the upper layer and polewards in the lower layer. Because the mass flux in each layer is equal and opposite, the surface (i.e., lower-layer) wind is given by the vertical integral of the vorticity or potential vorticity fluxes, namely

$$r H_2 \bar{u}_2 = H_1 \bar{v}'_1 \bar{q}'_1 + H_2 \bar{v}'_2 \bar{q}'_2 = H_1 \bar{v}'_1 \bar{\zeta}'_1 + H_2 \bar{v}'_2 \bar{\zeta}'_2. \quad (\text{TL.3})$$

The vorticity flux is positive in the upper layer and negative in the lower layer. However, because the potential vorticity gradient in the upper layer is large, this layer is more linear than the lower layer and Rossby waves are better able to transport momentum. The magnitude of the vorticity flux is thus larger in the upper layer than in the lower layer and, using (TL.3), the surface winds are positive (eastwards) in the mid-latitude baroclinic zone (see Fig. 12.14). To balance the upper-layer mid-latitude momentum flux convergence a meridional overturning circulation (a Ferrel cell) is generated. In a steady state $f_0 \bar{v}_1 = -\bar{v}'_1 \bar{\zeta}'_1$ so that the zonally averaged upper level flow is equatorwards. However, the total mass flux in the upper level is polewards; thus, the equatorward meridional velocity in the upper branch of the Ferrel cell is the consequence of an Eulerian zonal average and does not correspond to a net equatorward mass transport.

The real atmosphere resembles the two-layer model in part. In the real atmosphere, the equatorward residual flow occurs close to the surface. Thus, for more realism, we might think of the lower layer as representing a near-surface layer, and $H_1 \gg H_2$.

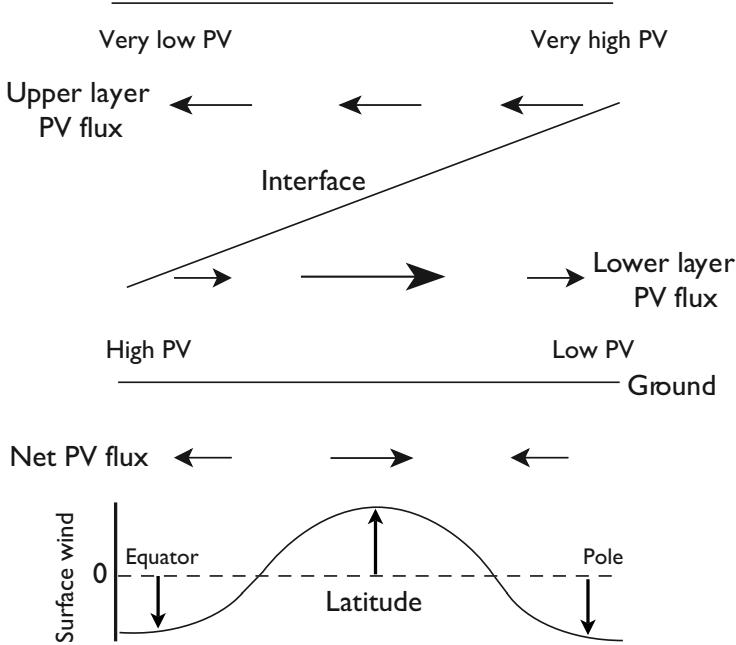


Fig. 12.13 Sketch of the potential vorticity fluxes in a two-layer model. The surface wind is proportional to their vertical integral. The PV fluxes are negative (positive) in the upper (lower) layer, but are more uniformly distributed at upper levels. The lower panel shows the net (vertically integrated) PV fluxes and the associated surface winds.

In the lower layer the vorticity fluxes are weak and the balance is largely between the Coriolis force on the meridional wind and the frictional force on the zonal wind (as in Fig. 11.15). If the upper-layer flow is equatorwards, the lower-layer flow must be polewards by mass conservation, and so the zonal wind is positive (eastwards); that is

$$r\bar{u}_2 \approx f_0\bar{v}_2 = -\frac{H_1}{H_2}f_0\bar{v}_1 > 0, \quad (12.89a,b)$$

where the second equality follows by mass conservation of the Eulerian flow.

In terms of the TEM form of the equations, (12.67), the corresponding balances in the centre of the domain are

$$f_0\bar{v}_1^* = -\bar{v}'_1\bar{q}'_1 > 0 \quad (12.90a)$$

and

$$r\bar{u}_2 = f_0\bar{v}_2^* + \bar{v}'_2\bar{q}'_2 = -f_0\frac{H_1}{H_2}\bar{v}_1^* + \bar{v}'_2\bar{q}'_2 = \frac{H_1}{H_2}\bar{v}'_1\bar{q}'_1 + \bar{v}'_2\bar{q}'_2 > 0, \quad (12.90b)$$

using mass conservation and the fact that the lower-layer potential vorticity fluxes are larger than those of the upper layer. Illustrations of the dynamical balances of the two-layer model are given in Figs. 12.13 and 12.14.

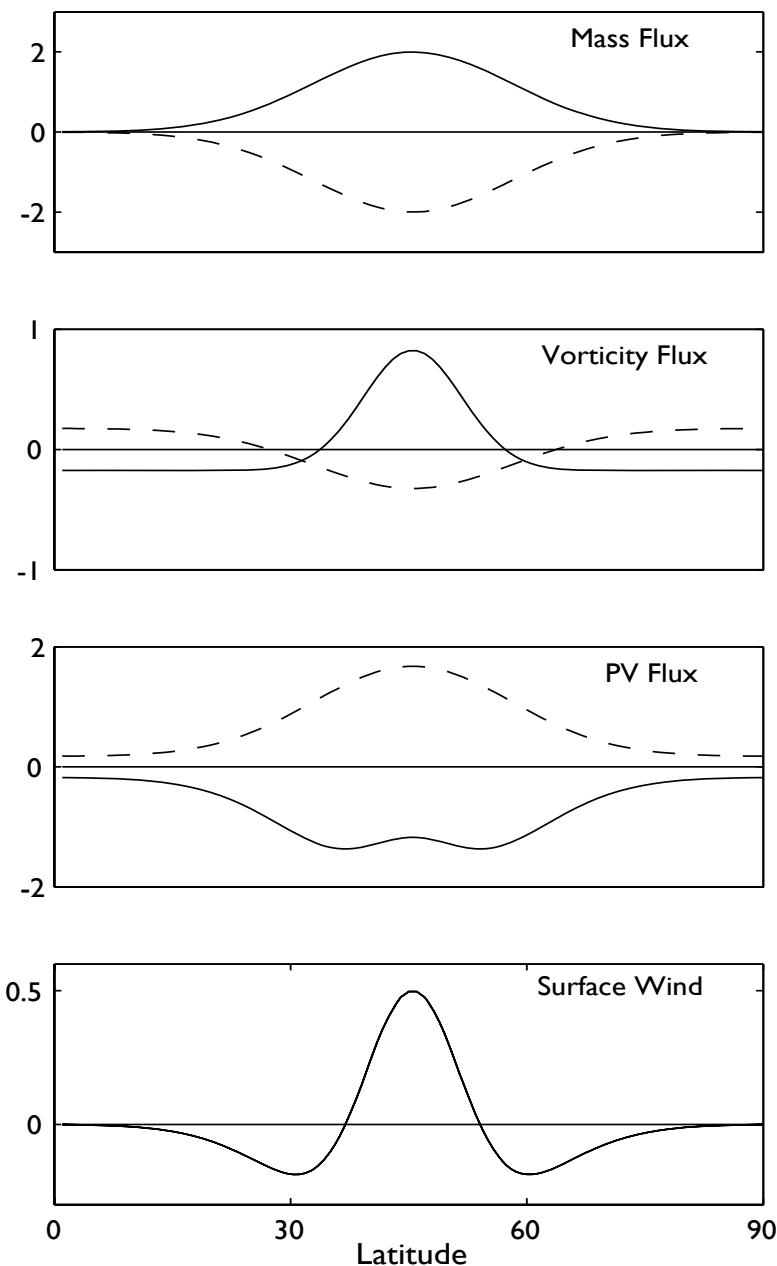


Fig. 12.14 Schema of the eddy fluxes in a two-layer model of an atmosphere with a single mid-latitude baroclinic zone. The upper-layer fluxes are solid lines and the lower-layer fluxes are dashed. The lowest panel shows the sum of the lower- and upper-layer vorticity fluxes (or, equivalently, the sum of the potential vorticity fluxes), which is proportional (when the surface friction is a linear drag) to the surface wind. The fluxes satisfy the various relationships and integral constraints of section 12.2.2 but are otherwise idealized.

12.3 † EDDY FLUXES AND AN EXAMPLE OF A CLOSED MODEL

The arguments above are heuristic and phenomenological and, although quite plausible they are not wholly systematic. In this section we give a more axiomatic calculation, and make certain closure assumptions that relate the eddy fluxes to the mean fields; specifically, we invoke the diffusion of potential vorticity, and then calculate the zonal winds and meridional circulation. The main purpose of this section is explicitly show that if we do invoke a potential vorticity closure then a complete solution of the flow follows. However, too much credence should not be given to the particular closure we do invoke, for it cannot be rigorously justified.

12.3.1 Equations for a closed model

With quasi-geostrophic scaling, the equations of motion are the momentum equations

$$\frac{\partial \bar{u}_1}{\partial t} = f_0 \bar{v}_1^* + \bar{v}_1' \bar{q}_1' \quad (12.91a)$$

$$\frac{\partial \bar{u}_2}{\partial t} = f_0 \bar{v}_2^* + \bar{v}_2' \bar{q}_2' - r \bar{u}_2 \quad (12.91b)$$

and the mass conservation equation for each layer which may be written as an equation for the interface height,

$$\frac{\partial \bar{\eta}}{\partial t} - H_1 \frac{\partial \bar{v}_1^*}{\partial y} = S, \quad (12.92)$$

where $\eta = H_1 - h_1 = h_2 - H_2$, with the notation of the previous sections. This can be written in terms of \bar{v}_2^* because

$$H_1 \bar{v}_1^* + H_2 \bar{v}_2^* = 0. \quad (12.93)$$

The velocities and thickness of the layers are related by the thermal wind relation

$$f_0(\bar{u}_1 - \bar{u}_2) = g' \frac{\partial \bar{\eta}}{\partial y} = -g' \frac{\partial \bar{h}_1}{\partial y}. \quad (12.94)$$

Using this to eliminate time derivatives between (12.91) and (12.92) reveals that the residual circulation satisfies

$$f_0^2 \frac{H}{H_2} \bar{v}_1^* - H_1 g' \frac{\partial^2 \bar{v}_1^*}{\partial y^2} = +g' \frac{\partial S}{\partial y} - f_0 (\bar{v}_1' \bar{q}_1' - \bar{v}_2' \bar{q}_2') - f_0 r \bar{u}_2, \quad (12.95)$$

where $H = H_1 + H_2$. Thus, *the residual circulation is driven by the potential vorticity fluxes, plus the diabatic terms*. We may derive a similar expression for the Eulerian mean meridional flow, namely

$$\begin{aligned} f_0^2 \frac{H}{H_2} \bar{v}_1 - H_1 g' \frac{\partial^2 \bar{v}_1}{\partial y^2} &= g' \frac{\partial S}{\partial y} + g' \frac{\partial^2}{\partial y^2} \bar{v}_1' \bar{h}_1' \\ &\quad - f_0 (\bar{v}_1' \bar{\zeta}_1' - \bar{v}_2' \bar{\zeta}_2') - f_0 r \bar{u}_2. \end{aligned} \quad (12.96)$$

However, the right-hand side now involves *both* the eddy vorticity fluxes and the eddy mass fluxes. The above equations illustrate the natural way in which the potential vorticity fluxes proximately ‘drive’ the extratropical atmosphere (see also the box on page 525).

Potential vorticity equation

A single prognostic equation for each layer is obtained by eliminating the residual circulation from (12.91) and (12.92), giving

$$\frac{\partial \bar{q}_1}{\partial t} = -\frac{\partial \bar{v}'_1 q'_1}{\partial y} + \frac{f_0}{H_1} S \quad (12.97a)$$

$$\frac{\partial \bar{q}_2}{\partial t} = -\frac{\partial \bar{v}'_2 q'_2}{\partial y} - \frac{f_0}{H_2} S + r \frac{\partial \bar{u}_2}{\partial y}, \quad (12.97b)$$

where q_i are the quasi-geostrophic potential vorticities of each layer given by

$$\bar{q}_1 = -\frac{\partial \bar{u}_1}{\partial y} + f_0 \frac{\bar{\eta}}{H_1}, \quad \bar{q}_2 = -\frac{\partial \bar{u}_2}{\partial y} - f_0 \frac{\bar{\eta}}{H_2}. \quad (12.98a,b)$$

Closure

If the potential vorticity fluxes can be expressed in terms of the mean fields then (12.97) is a closed set of equations. We can then solve for the potential vorticity in each layer and, using (12.95), for the residual circulation. One simple and rational closure is to assume that potential vorticity flux is transferred downgradient so that

$$\bar{v}'_i q'_i = -K_i \frac{\partial \bar{q}_i}{\partial y}, \quad (12.99)$$

where K_i is an eddy diffusivity, or transfer coefficient, which here is just a scalar quantity.⁵ Note that the model demands a closure of the potential vorticity flux — not momentum, vorticity or the mass flux — and potential vorticity, being a materially conserved variable, is also that field for which a diffusive closure is most applicable.

Such a closure has all of the features and problems associated with diffusive closures discussed in chapter 10, plus some of its own. One such is that a diffusive closure will not automatically respect the kinematic constraint that the volume integral of the potential vorticity flux must vanish, which for the two-layer model is expressed by (12.82). We may choose the vertical structure of the diffusivity in such a way that this constraint is satisfied, and in that case the model produces the results illustrated in Fig. 12.15. The diffusive closure does indeed then produce potential vorticity fluxes similar to the observed westward-eastward-westward surface wind pattern, and a residual circulation of the same sense as in Fig. 12.12, and constitutes perhaps the simplest closed model of the zonally averaged atmospheric circulation. Note that the surface wind is produced by the integral of the potential vorticity flux and, because the fluxes are quite different in the two layers, two layers are needed to produce a realistic pattern of surface wind without oversimplification, as well as to represent the meridional overturning and residual circulations. However, the model should not be regarded as being quantitatively valid, and the results depend on the structure of the transfer coefficients and the boundary conditions chosen (problem 12.1).

12.3.2 * Eddy fluxes and necessary conditions for instability

In linear baroclinic instability problems, a necessary condition for instability (the Charney-Stern-Pedlosky, or CSP, condition) is that the potential vorticity change sign in the interior of the fluid, or that the potential vorticity gradient in the interior has a particular sign with

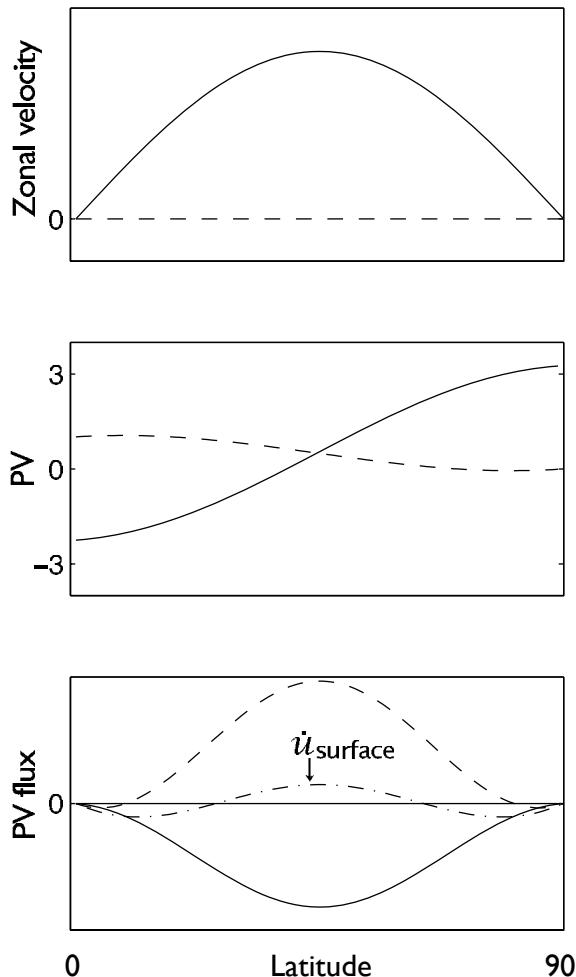


Fig. 12.15 Results from a diffusive closure in a two-layer zonally averaged model. Upper-layer quantities are solid lines and lower-layer quantities are dashed. The initial zonally averaged zonal velocity (top panel) is large in the upper layer and zero in the lower layer. This produces the potential vorticity structure illustrated in the middle panel, in units in which $\beta = 1$. A diffusive closure of the form (12.99) then produces the potential vorticity fluxes illustrated at the bottom, where the sum of the fluxes in the two layers is the dot-dashed line, and this is proportional to the tendency of the zonal wind. The residual circulation is proportional to the negative of the PV flux, and so is polewards in the upper layer and equatorwards in the lower layer. The vertical structure of the diffusivity is chosen such that (12.82) is satisfied.

respect to the buoyancy gradient at horizontal bounding surfaces, as discussed in chapters 6 and 7. These conditions do not apply in the statistical steady state of the forced-dissipative problem, but we may derive related conditions that do, although they are not completely general. We will focus on the interior condition and not the boundary conditions, as is appropriate in a layered model, but the argument could be extended to cover boundary issues explicitly.

The linear perturbation potential vorticity equation is

$$\frac{\partial q'}{\partial t} = -\bar{u} \frac{\partial q'}{\partial x} - v' \frac{\partial \bar{q}}{\partial y} - D', \quad (12.100)$$

where D' represents dissipative processes. From this we form the enstrophy equation

$$\frac{1}{2} \frac{\partial \bar{q}^2}{\partial t} + \bar{D}' q' = -v' q' \frac{\partial \bar{q}}{\partial y}. \quad (12.101)$$

In the standard linear problem we take $D = 0$ and then for growing waves the right-hand side

is positive. But the integral of $\overline{v'q'}$ over latitude and height is zero, and thus $\overline{v'q'}$ takes both positive and negative signs. Hence, $\partial\overline{q'}/\partial y$ must also take both positive and negative signs, and we recover the CSP condition that $\partial\overline{q'}/\partial y$ must change sign for an instability. (Note that we need not assume that the instabilities have normal form. A very similar argument was given in section 7.6.)

In a statistically steady state the production of variance by the terms on the right-hand side is balanced by a cascade of variance and dissipation at small scales. Just as in the linear instability problem the left-hand side is positive, now because $\overline{D'q'} > 0$, and therefore, once more, we see that $\partial\overline{q'}/\partial y$ must change sign somewhere. Furthermore, the eddy flux must be downgradient everywhere, i.e., $\overline{v'q'}\partial\overline{q'}/\partial y < 0$, because the left-hand side is positive everywhere.

If we now include nonlinear terms the zonally averaged perturbation enstrophy equation becomes

$$\frac{1}{2} \frac{\partial \overline{q'^2}}{\partial t} = -\overline{v'q'} \frac{\partial \overline{q'}}{\partial y} - \frac{1}{2} \frac{\partial}{\partial y} \overline{v'q'^2} - \overline{D'q'}. \quad (12.102)$$

On integrating in y the third-order term vanishes and we obtain

$$\int \left(\frac{1}{2} \frac{\partial}{\partial t} \overline{q'^2} + \overline{D'q'} \right) dy = - \int \overline{v'q'} \frac{\partial \overline{q'}}{\partial y} dy, \quad (12.103)$$

and so, if the left-hand side is positive, the flux must still be downgradient in the integrated sense that

$$\int \overline{v'q'} \frac{\partial \overline{q'}}{\partial y} dy < 0. \quad (12.104)$$

If the flux is *locally* downgradient, and in the nonlinear case this is an additional physical assumption, then because $\overline{v'q'}$ has both positive and negative values (because its integral is zero), then so must the potential vorticity gradient, $\partial\overline{q'}/\partial y$. That is, *when dissipation is present and if the potential vorticity fluxes are downgradient, a statistically steady state can only be maintained if the potential vorticity gradient changes sign somewhere*. In the continuously stratified case, this condition is replaced by ones involving a combination of the interior potential vorticity gradient and the buoyancy gradient at the boundary, the conditions being the same as necessary conditions for instability.

12.4 A STRATIFIED MODEL AND THE REAL ATMOSPHERE

In the previous section we introduced the effects of stratification by way of a two-layer model. Let us now discuss, albeit rather qualitatively, the dynamics of a continuously stratified model more relevant to the real atmosphere. These dynamics are generally similar to that of the two-layer model, although a number of differences in interpretation do arise. In particular, rather than the potential vorticity flux in the two layers, it is the potential vorticity flux in the interior and the buoyancy flux near the boundary that are the key aspects in producing the mean circulation.

12.4.1 Potential vorticity and its fluxes

The observed zonally averaged potential vorticity field is shown in Fig. 12.16. Of interest to us here is the fact that over most of the atmosphere, over most of the year, the potential

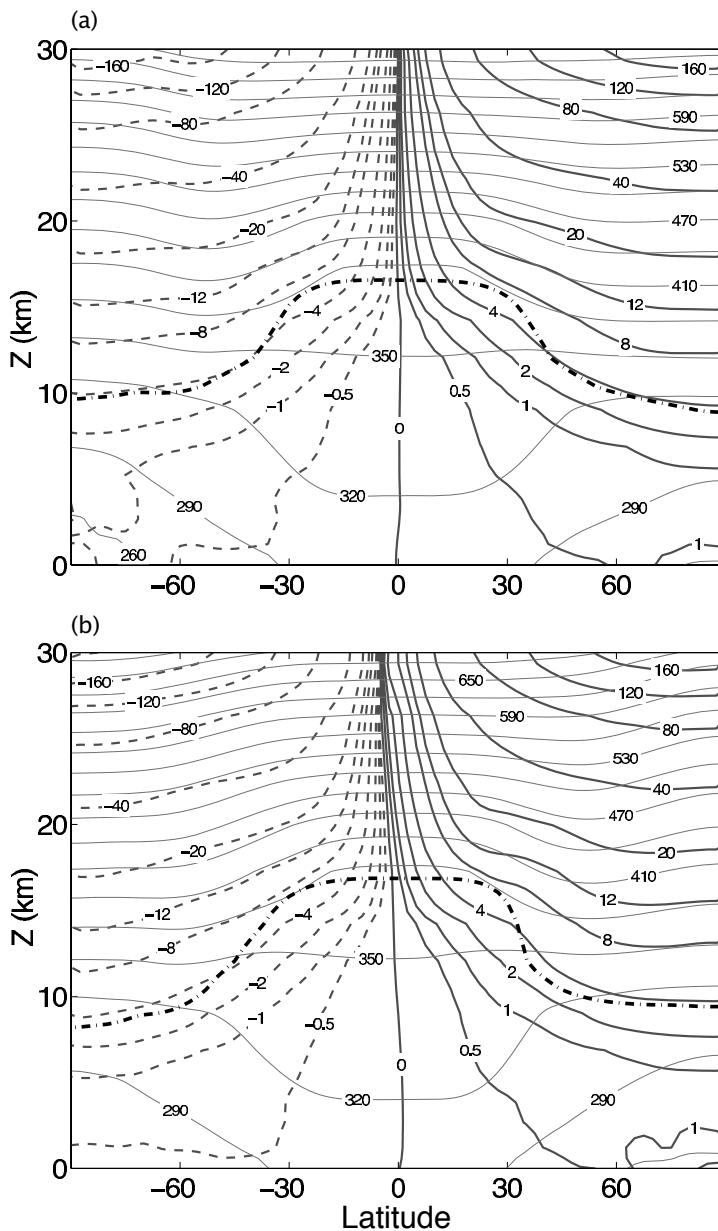


Fig. 12.16 The observed zonally averaged Ertel potential vorticity distribution (thicker, solid and dashed lines, peaking up at the equator) and the potential temperature (lighter lines) for (a) annual mean, (b) December–January–February. Also shown is the position of the WMO thermal tropopause (dot-dashed line). The potential vorticity is in ‘PV units’: $1 \text{ PVU} \equiv 1.0 \times 10^{-6} \text{ m}^2 \text{ K s}^{-1} \text{ kg}^{-1}$. Note the uneven contour interval for potential vorticity. The vertical coordinate is log pressure, with $Z = -7.5 \log(p/p_R) \text{ km}$, where $p_R = 10^5 \text{ Pa}$.

vorticity gradient is monotonic, with the potential vorticity increasing northwards. (The potential vorticity in the troposphere also increases moving northwards along isentropes.) How, then, can the atmosphere be baroclinically unstable? It is because the surface buoyancy (or temperature) decreases polewards, and thus the atmosphere becomes unstable via the interaction of a surface edge wave with an interior Rossby wave (see the conditions on page 265 or page 326). This baroclinic instability may then excite Rossby waves which propagate meridionally, producing a momentum convergence and westward surface flow, and an associated meridional circulation or Ferrel Cell, much as described in section 12.1. Let us explore these phenomena in a little more detail.

Surface winds

Consider the zonally averaged, continuously stratified momentum equations with quasi-geostrophic scaling,

$$\frac{\partial \bar{u}}{\partial t} = \bar{v}' \bar{\zeta}' + f_0 \bar{v} + F = \bar{v}' \bar{q}' + f_0 \bar{v}^* + F, \quad (12.105)$$

where F represents frictional effects and the residual velocity \bar{v}^* is given by

$$\bar{v}^* = -\frac{\partial \psi^*}{\partial z} = \bar{v} - \frac{\partial}{\partial z} \left(\frac{1}{N^2} \bar{v}' \bar{b}' \right). \quad (12.106)$$

The friction is given by the vertical gradient of a stress, $F = \partial \tau / \partial z$, and at the surface we may parameterize the stress, following (2.309) on page 112, by $\tau = r \bar{u}$ where r is a constant. Then, vertically integrating (12.105) from the surface to the top of the atmosphere (where frictional stresses and the buoyancy flux both vanish) we find, in steady state,

$$r \bar{u}(0) = \langle \bar{v}' \bar{\zeta}' \rangle = \langle \bar{v}' \bar{q}' \rangle + \frac{f_0}{N^2} \bar{v}' \bar{b}'(0), \quad (12.107)$$

where the angle brackets denote a vertical integral and (0) denotes surface values. Thus, the surface winds are determined, analogously to (12.84), by the vertically integrated relative vorticity fluxes, or equivalently by the integral of the interior potential vorticity fluxes and the buoyancy fluxes at the surface. The advantage of the latter representation is that both potential vorticity and buoyancy are materially conserved variables and it may be easier to deduce some properties of their fluxes than of the fluxes of relative vorticity. Compared to the two-layer formulation, the interior fluxes are analogous to those of the upper layer whereas the surface fluxes are analogous to those of the lower layer, especially as the lower layer becomes thin.

Potential vorticity and Eliassen–Palm fluxes

As in section 7.2, the quasi-geostrophic potential vorticity flux may be written as the divergence of the Eliassen–Palm (EP) vector,

$$\bar{v}' \bar{q}' = \nabla_x \cdot \mathcal{F} \quad (12.108)$$

where $\nabla_x \cdot \equiv \mathbf{j} \partial / \partial y + \mathbf{k} \partial / \partial z$ and

$$\mathcal{F} \equiv -\bar{u}' \bar{v}' \mathbf{j} + \frac{f_0}{N^2} \bar{v}' \bar{b}' \mathbf{k}. \quad (12.109)$$

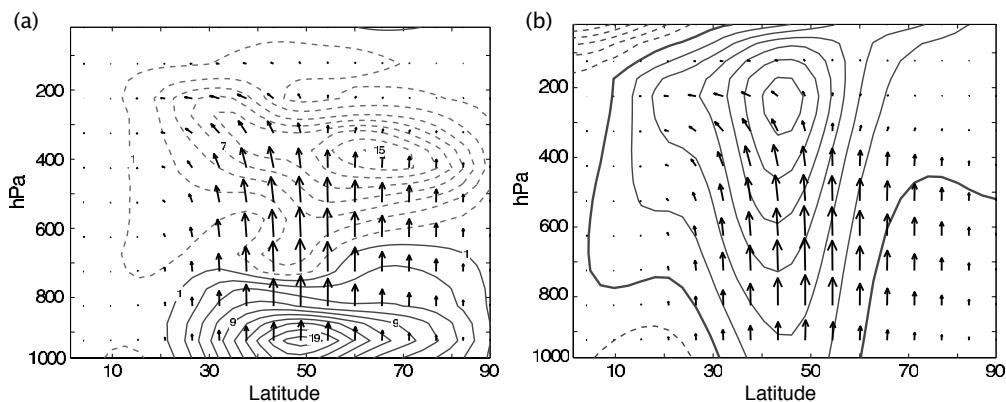


Fig. 12.17 The Eliassen-Palm flux in an idealized primitive equation of the atmosphere. (a) The EP flux (arrows) and its divergence (contours, with intervals of $2 \text{ m s}^{-1}/\text{day}$). The solid contours denote flux divergence, a positive PV flux, and eastward flow acceleration; the dashed contours denote flux convergence and deceleration. (b) The EP flux (arrows) and the time and zonally averaged zonal wind (contours). See the appendix for details of plotting EP fluxes.

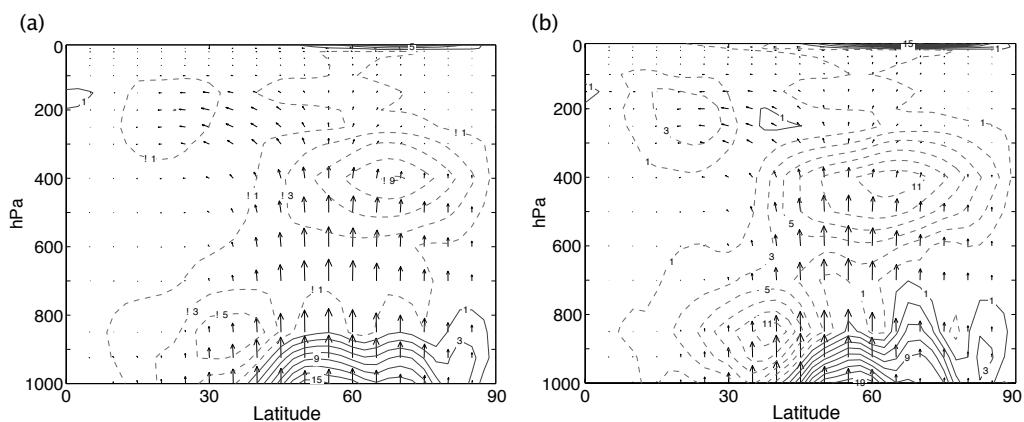


Fig. 12.18 The observed Eliassen-Palm flux (arrows) and its divergence (contours, with intervals of $2 \text{ m s}^{-1}/\text{day}$, zero contour omitted) in the Northern Hemisphere. Solid contours denote divergence, a positive (eastward) torque on the flow, and dashed contours denote convergence, a westward torque. (a) Annual mean, (b) DJF (December-January-February).

(See the appendix to this chapter for primitive equation and spherical coordinate versions; in practice the quasi-geostrophic expression qualitatively captures the dominant terms in the primitive equation expressions.) The EP vector as obtained from an idealized primitive equation general circulation model integration is shown in Fig. 12.17, and the EP vector from observations is shown in Fig. 12.18, and both show qualitatively similar properties — a generally upwards-pointing vector in mid-latitude, veering equatorwards aloft.

The upward component represents the meridional transfer of heat, and this occurs during the growth phase of the baroclinic lifecycle and is qualitatively captured by linear models — for example, in the Eady problem the EP flux is directed purely vertically (Fig. 7.2), and this resembles the vertical components of Figs. 12.17 and 12.18. But why should the average over a complete baroclinic lifecycle (which Figs. 12.17 and 12.18 represent) even approximately resemble that of the growing phase of the baroclinic lifecycle? After all, the eddies must subsequently decay, and one might imagine that the fluxes would then reverse themselves. In fact, this is not the case: eddies do not decay baroclinically, and their lifecycle is not reversible. Rather, there is an *irreversible* transfer to barotropic modes (as described in chapter 9) followed by a barotropic decay. Thus, there is no downward heat transfer in the cycle, and on average the heat transfer balances the net atmospheric heating. The lateral component of the EP flux is a consequence of Rossby wave propagation, much as was described in section 12.1. Baroclinic instability now plays the role of the mid-latitude wavemaker, and the EP flux emanates laterally from the baroclinic zone. The propagation is an irreversible process, with the Rossby waves breaking some distance from their source, and it is this that breaks the non-acceleration conditions and provides the mean flow acceleration and, consequentially, the observed zonal wind.

The divergence of the EP flux — that is, the potential vorticity flux — accelerates (or decelerates) the mean flow, as can be seen from (12.105) and Fig. 12.19. Broadly speaking, the EP flux decelerates the flow aloft (where it is balanced by the Coriolis force on the poleward residual flow) but provides an eastward acceleration at the surface (where it is largely balanced by friction). However, the two components of the flux (Fig. 12.20) have rather different effects on the mean flow. The horizontal component acts to extract momentum from the subtropics and deposit it in mid-latitudes, and so accelerate the flow producing a fairly barotropic eastward jet. (It is this component that gives rise to so-called negative viscosity, in which the eddies transfer momentum upgradient.) The vertical component of the EP flux arises from the meridional buoyancy flux and acts to reduce the intensity of the mid-latitude westerlies aloft, transferring momentum to the surface where it may be balanced by friction, and producing the surface westerlies.

Two questions spring to mind.

- (i) Why is the meridional wave-activity propagation predominantly in the upper atmosphere? That is to say, why do the EP vectors only veer laterally above about 400 hPa, and not in the lower atmosphere?
- (ii) Why is the wave-activity propagation (the direction of the EP flux vectors) predominantly equatorward?

As regards item (i), the propagation is mainly in the upper atmosphere because it is here that the potential vorticity gradient is strongest, as can be seen from Fig. 12.16. This can be understood on the basis of the two-layer model (for example, Fig. 12.15). Thus, wave propagation is more efficient in the upper troposphere, whereas the lower troposphere is more nonlinear and so here the enstrophy cascade, and wavebreaking, occur locally and closer to the region of baroclinic instability itself. Regarding item (ii), the proximate reason is that waves predominantly break on the equatorial side of the instability, and this in turn is for two possible reasons. One is that β increases towards the equator, so that linear propagation is more efficient. The other is that there is often a critical layer in the subtropics,

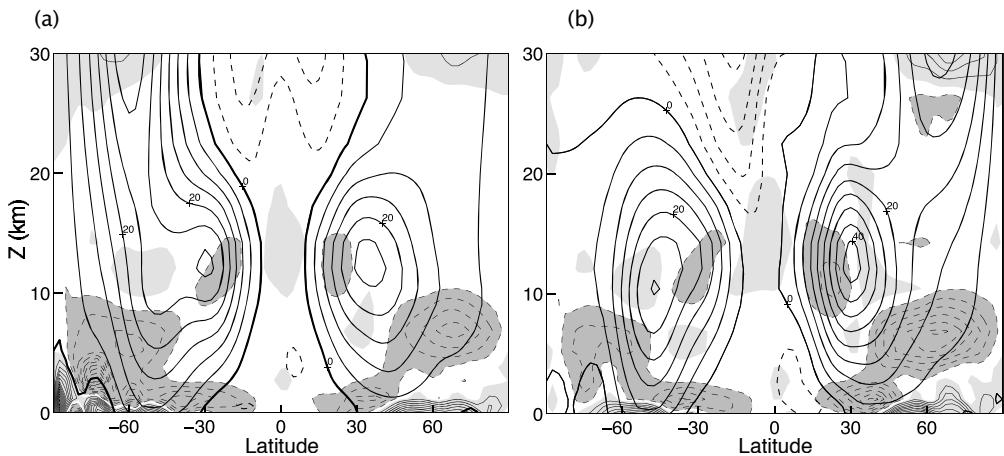


Fig. 12.19 The observed zonally averaged zonal wind (thicker contours, interval 5 m s^{-1}), and the Eliassen-Palm flux divergence (contour interval $2 \text{ m s}^{-1}/\text{day}$, zero contour omitted). Regions of positive EP flux divergence (eastward acceleration) are lightly shaded; regions less than $-2 \text{ m s}^{-1}/\text{day}$ are more darkly shaded. (a) Annual mean, (b) DJF (December–January–February).

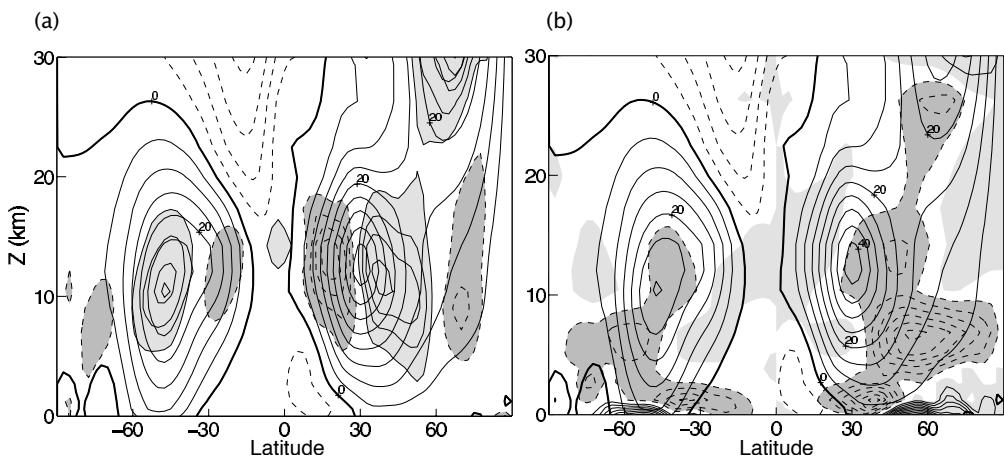


Fig. 12.20 The divergence of the two components of the EP flux (shaded), and the zonally averaged zonal wind (thicker contours) for DJF. (a) The momentum fluxes, $-\partial_y \bar{u}' \bar{v}'$, contour interval is $1 \text{ m s}^{-1}/\text{day}^{-1}$, light shaded for positive values > 1 , dark shaded for negative values < -1 . (b) The buoyancy flux, $f \partial_z (\bar{v}' \bar{b}' / N^2)$, with contour interval and shading convention as in Fig. 12.19.

where the speed of the waves equals that of the flow itself ($\bar{u} = c$), and here breaking can efficiently occur.

12.4.2 Overturning circulation

The Eulerian overturning circulation (meaning the circulation from a conventional zonal average at constant height) in mid-latitudes is a single indirect cell, the Ferrel Cell, with rising motion at high latitudes and sinking in the subtropics (top panel of Fig. 12.21). The residual circulation is direct and, consistent with the theory of section 7.3.3, resembles closely the thickness-weighted circulation (middle and lower panels Fig. 12.21). Qualitatively these features are captured by the two-layer dynamics of section 12.2.2, but the continuously stratified case differs in one or two respects.

The main difference between the continuous and two-layer cases is that in the former the return flows — both the lower branch of the Ferrel Cell and the equatorial branch of the residual circulation — are not distributed over the lower troposphere, but are confined to a relatively thin layer. In the lower branch of the Ferrel Cell the dynamical balance is between friction and the Coriolis force on the meridional flow, so that its thickness is that of a turbulent Ekman layer and about a kilometre. To understand this better, let us take a quasi-geostrophic perspective. The mean potential vorticity gradient in the free atmosphere is nearly everywhere polewards and the potential vorticity flux is largely downgradient and equatorwards. This means that here the residual circulation is largely polewards, satisfying the balance

$$f\bar{v}^* \approx -\bar{v}'q'. \quad (12.110)$$

In a multi-layer quasi-geostrophic model, with friction acting only in the lowest layer, the circulation is closed by return flow in the lowest layer; thus, as the number of layers increases the return flow is carried in an ever-thinner layer, this becoming a delta-function in the continuous limit, just as in the example of residual flow in the Eady problem (section 7.5). In the real atmosphere, the return flow cannot be confined to a delta-function, but this argument suggests that it will occur close to the surface and this expectation is borne out in the lower panels of Fig. 12.21 and in Fig. 11.14. In fact, much of the equatorial return flow occurs in isentropic layers that have a potential temperature below the mean value at the surface — that is, in cold air outbreaks.

12.5 † THE TROPOPAUSE AND THE STRATIFICATION OF THE ATMOSPHERE

As You Like It.

William Shakespeare, c. 1599.

In the previous sections we have, by and large, taken the stratification of the atmosphere as given. Let us now explore the physical processes that determine this stratification, beginning with a brief observational overview.

The atmosphere may be divided by stratification into certain distinct regions, illustrated in Fig. 12.22. The figure shows the so-called ‘U.S. standard atmosphere’, a rough average temperature profile and a sometimes-useful standard, as well as actual observed values in the lower atmosphere. In the lower 10 km or so of the atmosphere we have the *troposphere*, a dynamically active region wherein most of the weather and the vast predominance of heat

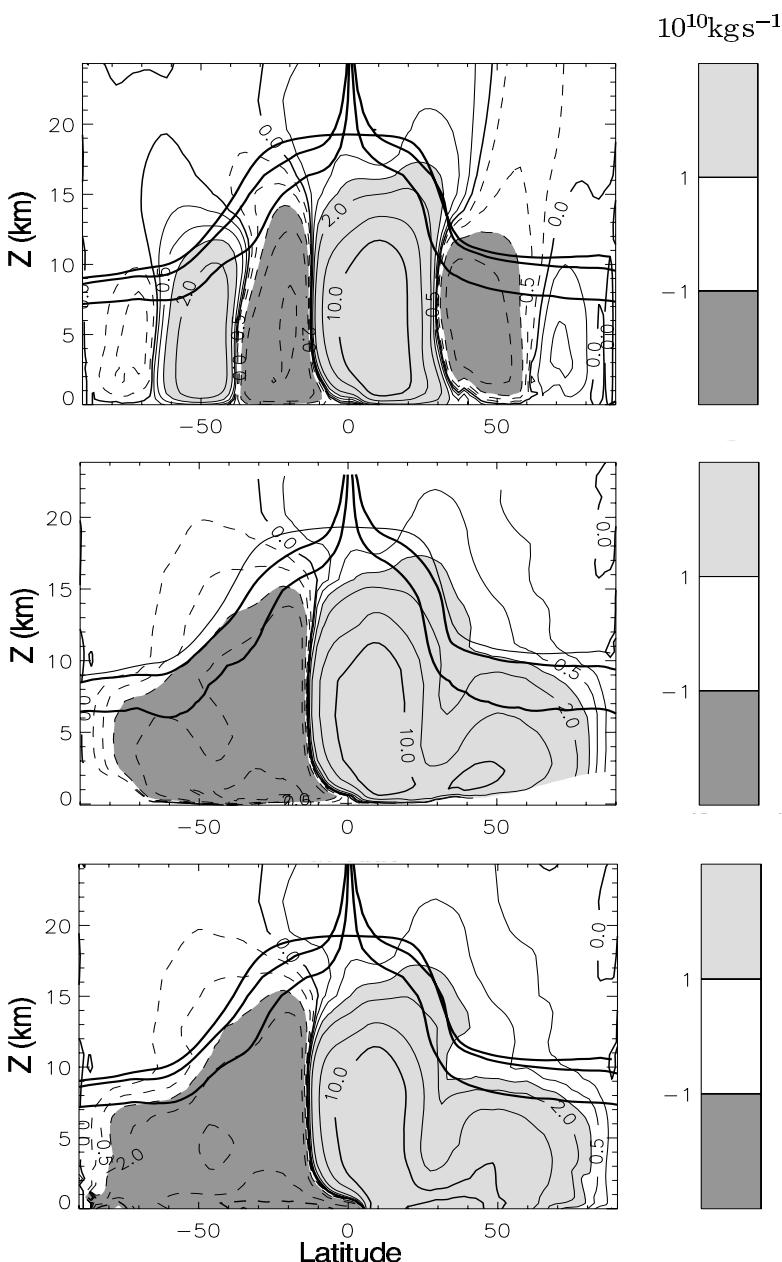
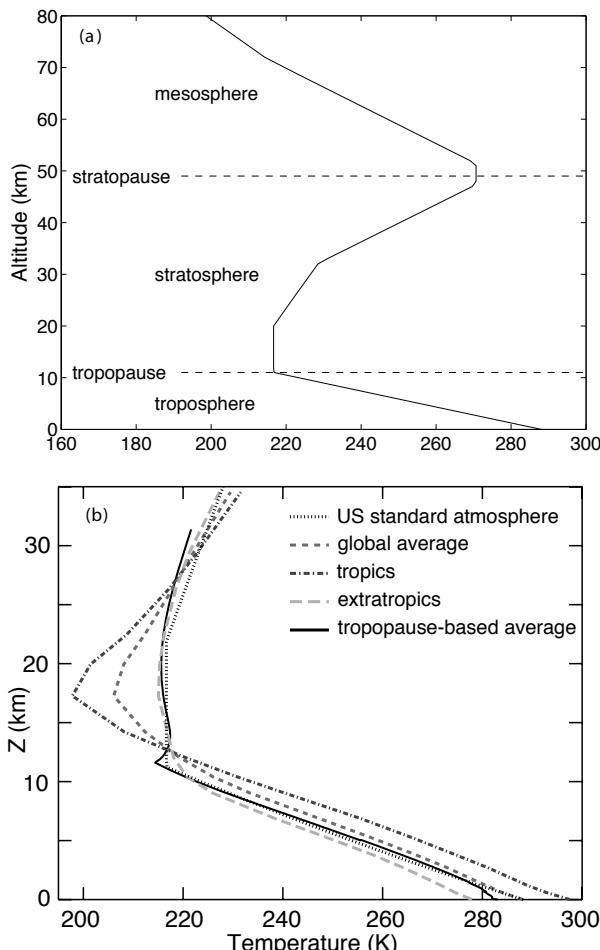


Fig. 12.21 Top: The observed zonally averaged, Eulerian-mean, streamfunction in northern hemisphere winter (DJF, 1994–1997). Negative contours are dashed, and values greater or less than $10^{10} \text{ kg s}^{-1}$ (10 Sv) are shaded, darker for negative values. The circulation is clockwise around the lighter shading. The three thick solid lines indicate various measures of the tropopause (section 12.5): the two that peak at the equator are isolines of potential vorticity, $Q = \pm 1.5, \pm 4$ PV units, and the flatter one is the WMO thermal tropopause. Middle: The thickness-weighted, or isentropic-mean, meridional mass streamfunction. After calculation in isentropic coordinates, the streamfunction is projected back on to log-pressure coordinates, Z (km), for display. Bottom: the residual streamfunction calculated from the Eulerian circulation and the eddy fluxes.⁶

Fig. 12.22 (a) The temperature profile of the ‘U.S. standard atmosphere’, marking the standard regions of the atmosphere below 80 km. In addition to the regions shown, the top of the mesosphere is marked by the mesopause, at about 80 km, above which lies the ‘thermosphere’, in which temperatures rise again into the ‘exosphere’, extending a few thousand kilometres and where the atmospheric temperature ceases to have a useful meaning. (b) Observed, annually averaged profiles of temperature in the atmosphere, where the ordinate is log-pressure. ‘Tropics’ is the average from 30°S to 30°N , and the extratropics is the average over the rest of the globe. The observations are from a reanalysis over 1958–2003 that extends upwards to about 35 km. See text for the meaning of ‘tropopause-based average’.



transport occurs. The troposphere is capped by the *tropopause*, above which lies the *stratosphere*, a region of stable stratification extending upwards to about 50 km. (Troposphere means ‘turning sphere’, appropriately so as within it dynamical overturning is prevalent. Stratosphere means ‘layered sphere’, and here there is much less vertical motion.) The stratosphere is capped by the *stratopause*, above which are the mesosphere, thermosphere and exosphere, regions of the upper atmosphere that do not concern us here. Our focus will be on the processes that determine the stratification of the lower atmosphere and the height of the tropopause.⁷

In the troposphere temperature generally falls with height, whereas in the stratosphere it increases with height, and this gives rise to a thermal definition of the tropopause:⁸ *The tropopause is the lowest level at which the lapse rate decreases to 2 K km^{-1} or less, provided also that the average lapse rate between this level and all higher levels within 2 km does not exceed 2 K km^{-1} .* At any particular time there might also be a second tropopause: if above the first tropopause the average lapse rate between any level and all higher levels within 1 km exceed 3 K km^{-1} , then a second (higher) tropopause is defined by that same criterion.

Potential Vorticity Fluxes and the Extratropical Atmosphere

The extratropical circulation of the atmosphere is driven by the differential heating between equator and pole, mediated by fluxes of potential vorticity. Thus, in a layered model we have the following.

- (i) *Zonal winds.* At each level the acceleration of the zonal winds is governed by the potential vorticity fluxes:

$$\frac{\partial \bar{u}_i}{\partial t} = \overline{v'_i q'_i} + f_0 \bar{v}_i^* + F_i. \quad (\text{PV.1})$$

where \bar{v}_i^* is the residual meridional flow and F_i represents friction.

- (ii) *Surface winds.* In steady state, the surface winds are produced by the vertically integrated potential vorticity fluxes:

$$r H_s \bar{u}_s = \sum_i H_i \overline{v'_i q'_i}, \quad (\text{PV.2})$$

where u_s is the surface wind, H_s the thickness of the lowest layer, and r is a frictional coefficient.

- (iii) *Meridional transport.* The total (or residual) meridional transport is, proximately, forced by the potential vorticity fluxes. For example, in a two-layer model

$$f_0^2 \frac{H}{H_2} \overline{v}_1^* - H_1 g' \frac{\partial^2 \overline{v}_1^*}{\partial y^2} = +g' \frac{\partial S}{\partial y} - f_0 \left(\overline{v'_1 q'_1} - \overline{v'_2 q'_2} \right) - f_0 (F_1 - F_2), \quad (\text{PV.3})$$

where S is proportional to the diabatic forcing, and this equation holds at all times. In a steady state the momentum equation gives simply

$$f_0 \bar{v}_i^* = -\overline{v'_i q'_i} - F_i. \quad (\text{PV.4})$$

Above the surface layer friction is negligible and the meridional transport responds almost solely to the potential vorticity fluxes.

Finally, such definitions are presumed not to apply if they are satisfied below 500 mb. As so defined, the thermal tropopause typically varies in height from about 16 km at low latitudes to about 8 km near the poles.

One may wonder whether such statements reflect a robust property of the atmosphere, or are merely a practical definition of the tropopause, and certainly we should not expect these quantitative definitions to hold in a changed climate, but regardless of that the tropopause is a distinct boundary separating two differently stratified regions, the troposphere and the stratosphere. The thermal tropopause is marked in Fig. 12.21 and, as we see there and in Fig. 12.16, in the extratropics it is almost parallel to isolines of potential vorticity, and sometimes an isoline of potential vorticity (say $Q = 3$ or 4 PV units) is used as a somewhat ad hoc definition of the extratropical tropopause.

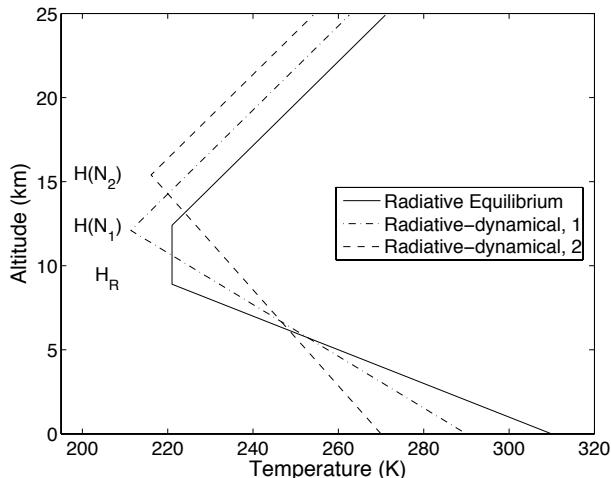


Fig. 12.23 Schema of the radiative equilibrium profile (solid), and two profiles of a radiative-dynamical equilibrium in which the lower atmosphere is adjusted to some specified lapse rate N_1 or N_2 . Here, $N_2 > N_1$ and $H(N_2) > H(N_1)$.

Finally, and interestingly, we note that the tropopause appears as a very sharp feature when viewed instantaneously, although this sharpness is often blurred when time or spatial averages are taken. The solid line in Fig. 12.22, denoted ‘tropopause-based average’, shows the profile obtained when the tropopause height itself is taken as a common reference level, using data from individual radiosonde ascents over the United States.⁹

12.5.1 A radiative-convective model

The radiative-equilibrium temperature is that temperature which arises from a pure radiative balance. As we noted in section 11.1.1, a black body subject to a net incoming radiation of S has a radiative-equilibrium temperature T_{rad} given by $\sigma T_{rad}^4 = S$ where σ is the Stefan-Boltzmann constant. For partially absorbing media like the Earth’s atmosphere the radiative equations are correspondingly more complicated, but nevertheless the vertical structure of the radiative equilibrium temperature may be calculated and its qualitative features are easy to understand. The atmosphere is largely transparent to solar radiation and thus it is largely heated from below by the ground, both through the latter’s emission of infrared radiation and by latent and sensible heat transfer. The atmosphere absorbs and re-emits infrared radiation, and the upshot is that the radiative-equilibrium temperature falls rapidly with height in the lowest several kilometres of the atmosphere before rising again in the stratosphere, this being in part due to a layer of ozone, concentrated between 25 and 30 kilometres altitude, that absorbs solar radiation. A radiative-equilibrium profile is sketched in Fig. 12.23 and, although schematic, it illustrates an important point — the radiative equilibrium temperature falls so rapidly in the lower atmosphere that it would be convectively unstable. Detailed calculations show that $-\partial T_{rad}/\partial z$ is often greater than 10 K km^{-1} , thus exceeding even the dry adiabatic lapse rate and far exceeding the moist adiabatic lapse rate of about 6 K km^{-1} (depending on temperature; see section 2.9.3 for a description of convective instabilities). Thus, if a radiative-equilibrium profile could be initially set up somehow, convection would quickly ensue, transferring heat vertically in an attempt to stabilize the profile.

The above observations suggest a simple radiative-convective model of the structure of the stratification, as follows. Starting with a radiative-equilibrium profile, the temperature in the lower atmosphere is modified until it becomes statically neutral; that is, the lapse rate is adjusted until it becomes equal to the dry or, if the air is saturated, the moist adiabatic lapse rate. We allow such an adjustment to occur up to such a height as is needed, above which the radiative-equilibrium temperature is maintained, and consequently there will typically be a sharp transition between the two regimes which may be identified with the tropopause. As we see in Fig. 12.23 the height of the tropopause will depend on the lapse rate in the troposphere, with a larger lapse rate (i.e., a more negative $\partial T/\partial z$) leading to a lower tropopause. This result does not depend on the adjustment mechanism being convective *per se*, that is on being due to small-scale essentially vertical motion. Any dynamical process that adjusts the lapse rate in the troposphere may produce a similar effect. The resulting lapse rate, determined by dynamical effects below some height (which we might associate with the tropopause) and radiative effects above, is called a radiative-dynamical profile, and a radiative-convective profile if the dynamical process is convection.

The height of the tropopause can be calculated if we know how to calculate the radiative equilibrium temperature and if we are given the tropospheric lapse rate. The simplest calculation assumes that the radiative equilibrium temperature of the stratosphere is unaltered by the adjustment process. Then, suppose that the initial temperature profile, $T_i(z)$ is given as a function of height by

$$T_i(z) = \begin{cases} T_{si} - Az & z \leq z_1 \\ (T_{si} - Az_1) + B(z - z_1) & z \geq z_1 \end{cases} \quad (12.111)$$

where the constants A , B and T_{si} are given, z_1 is the height at which the radiative equilibrium temperature starts to increase, and below z_1 the profile is convectively unstable. If convection adjusts this profile to become

$$T_f(z) = T_{sf} - Cz \quad z < H \quad (12.112)$$

until it intersects the stable profile at some height $H > z_1$, then we require

$$T_{sf} - CH = (T_{si} - Az_1) + B(H - z_1), \quad (12.113)$$

whence

$$H = \frac{T_{sf} - T_{si} + (A + B)z_1}{B + C}. \quad (12.114)$$

The final surface temperature is given by an energetic argument, that assuming there is no significant conversion to kinetic energy then the internal energy plus the potential energy is conserved, or $\int_0^H \rho(c_v T + gz) dz$ is fixed. (If condensation occurs there is an increase in energy equal to the latent heat released.)

The argument above is oversimplified in assuming that the radiative equilibrium temperature above the adjusted region is not altered by the adjustment process, but nevertheless there are a couple of robust conclusions.

- (i) The mechanism produces a relatively sharp tropopause, the existence of which does not depend on the stratospheric temperature actually increasing with height. The tropopause is the boundary between radiative equilibrium temperatures aloft, and dynamically influenced temperatures below.

- (ii) The smaller the lapse rate (i.e., the lower the value of $-\partial T/\partial z$, or equivalently the larger the stratification N^2) to which the lower atmosphere is adjusted, the higher and the warmer the tropopause. In a dry atmosphere, the lapse rate is adjusted to that of a dry adiabat, with $N^2 = 0$ and $\partial T/\partial z = -g/c_p$. In a saturated atmosphere adjustment is towards the moist adiabat (see section 2.9.3).

12.5.2 Radiative and dynamical constraints

We assumed in our simple calculation that the radiative equilibrium temperature profile of the stratosphere was given, independently of the troposphere. This is not quite right, but nevertheless if we can calculate the radiative equilibrium profile as a function of temperature then we can calculate the height of the tropopause if we are given the lapse rate in the troposphere and the surface (or average) temperature. The tropopause height is then the height at which the tropospheric temperature profile matches the radiative equilibrium temperature. The calculation can, at least in principle, be inverted: if the tropopause height is given (perhaps via purely dynamical reasoning), and the surface or average tropospheric temperature is also given, and one assumes a stratosphere in radiative equilibrium, then the tropospheric lapse rate follows. To illustrate this without the complication of a full radiative model, consider the heat balance at a particular latitude. The outgoing infrared radiation, I , is in balance with the incoming solar radiation, S , the convergence of horizontal energy flux C , and the energy flux from the surface, F . The temperature profile must adjust itself so that $I = S + F + C$, so that we may think of I as being effectively given — a radiative constraint. Suppose that the temperature profile has a constant lapse rate below some height H_T (the tropopause) and that above this radiative equilibrium holds. Given this, we may parameterize the outgoing infrared radiation by a specification of the tropopause height H_T , the surface temperature T_s , and the temperature at the tropopause, T_t . If dynamical processes serve to specify the lapse rate and the surface temperature, then the tropospheric height is determined from the radiative constraint. Alternatively, if the tropospheric height and surface temperature are specified, the lapse rate follows.

It is clear from the above arguments that we do not necessarily need convection to produce a tropopause — a boundary will in general occur separating a dynamically influenced troposphere and a stratosphere in near-radiative equilibrium. We may more generally think of the troposphere as that region of the atmosphere in which a redistribution of heat occurs, much of which may be *lateral*. This picture suggests itself naturally from the lower two panels of Fig. 12.21 where we see the height of the tropopause roughly coinciding with the height attained by the overturning circulation, and the troposphere is then a kind of boundary layer to the atmosphere above (Fig. 12.24). The top of the boundary layer is the tropopause but (unlike idealizations of the oceanic thermocline) it is not marked by a discontinuity in potential temperature; this is not because potential temperature is not necessarily being mixed in the troposphere, but because radiation keeps the profile continuous. However, the vertical temperature *gradient* is (approximately) discontinuous.

If the large-scale overturning circulation and associated lateral transport of heat is also able to transfer sufficient heat vertically so that a statically stable lapse rate can be maintained, then small-scale convective events and convective adjustment need play no role in determining the lapse rate of the troposphere and the height of the tropopause. Nevertheless, it is generally believed that within the tropics predominantly vertical, moist convection is the

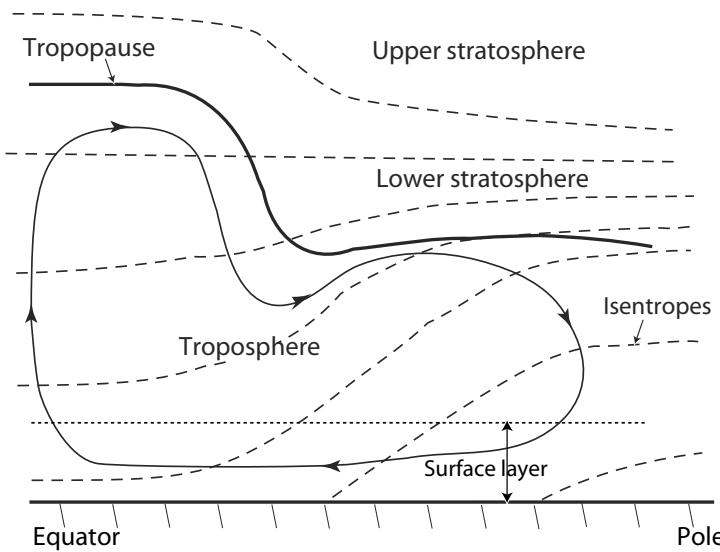


Fig. 12.24 A sketch of the stratification and overturning circulation in the lower atmosphere. The thick line marks the tropopause, the closed contour marks the residual overturning circulation and the dashed lines are isentropes. The overturning circulation has two distinct parts, a tropical Hadley Cell where most of the mass is carried by mean motions, and a shallower extratropical cell in which most of the mass transfer occurs via eddy motion. The equatorward return flow is mostly confined to a shallow surface layer. The ‘lower stratosphere’ is ventilated by the troposphere along isentropic surfaces, whereas in the ‘upper stratosphere’ isentropes do not intersect the tropopause. The tropopause is the boundary between the partially mixed troposphere and the near-radiative equilibrium stratosphere.

main process determining the stratification and the height of the tropopause.¹⁰ In contrast, the processes determining the extratropical stratification are still a matter of some debate, and in the next two sections we outline some of the dynamics relevant to this problem. The theories fall into two general camps, one related to the redistribution of potential vorticity by baroclinic eddies, and the second related to convection, and we deal with them in turn in the next two sections.

12.6 † BAROCLINIC EDDIES AND POTENTIAL VORTICITY TRANSPORT

The fundamental idea behind the arguments in this section is that the troposphere is that region in which a dynamical distribution of energy takes place, that redistribution is effected by baroclinic eddies and that potential vorticity dynamics then provides a natural height for these eddies and hence of the tropopause.

In mid-latitudes baroclinic eddies transfer heat upwards, and if they do this efficiently enough there is no need for convection (meaning predominantly vertical convection occurring on small scales) to maintain a statically stable lapse rate (see Fig. 12.24). Nonetheless, the baroclinic eddies do not extend infinitely upwards, and so we expect a boundary — a

tropopause — between a dynamical troposphere and a radiative stratosphere. How high is this tropopause, and what is its nature?

12.6.1 A linear argument

On the β -plane, linear baroclinic instability can produce a height scale that is different from the height of any pre-existing ‘lid’ or from the density scale height. This height scale is (section 6.9.1)

$$h = \frac{\Lambda f^2}{\beta N^2}, \quad (12.115)$$

where $\Lambda = \partial \bar{u} / \partial z$. That is to say, if $h < H$, where H is the density scale height (or the height of some lid in a Boussinesq model), then the baroclinic eddies will extend upwards to a height h , and this will be the vertical extent of significant heat fluxes. Thus (one might argue), below h the thermal structure is determined by the dynamical effects of baroclinic instability, whereas above h the atmosphere is more nearly in radiative equilibrium. Using $\Lambda = (15 \text{ m s}^{-1}) / (10 \text{ km})$, $\beta = 1.6 \times 10^{-11} \text{ s}^{-1} \text{ m}^{-1}$, $f = 1 \times 10^{-4} \text{ s}^{-1}$ and $N = 10^{-2} \text{ s}^{-1}$ gives $h \approx 10 \text{ km}$, which approximates the height of the tropopause in mid-latitudes, but this can only be a rough estimate, especially as our value of Λ is a little low.

Allowing the parameters f and β to vary with latitude, and using the thermal wind relation, (12.115) may be written in the form

$$h = s \frac{f}{\beta} = s \alpha \tan \vartheta, \quad (12.116)$$

where s is the isentropic slope, $-(\partial_y \theta) / (\partial_z \theta)$, ϑ is latitude and α is the Earth’s radius. This equation, when written as $s \sim h/a$, suggests that the isentropic slope is roughly such that isentropes extend from the surface at low latitudes to the tropopause at the poles, and this is more-or-less true in the present atmosphere (Fig. 12.16). This is still insufficient information to determine the lapse rate itself, for (12.115) or (12.116) provide only a single relationship between h and N^2 , and a radiative or diabatic constraint is also needed, just as with the simple convective adjustment argument. However, even given such a constraint, a linear argument is open to criticism on two grounds.

- (i) The amplitude of meridional heat transfer, which determines the meridional temperature gradient, is really determined by nonlinear effects.
- (ii) Two dimensional (latitude-height) radiative equilibrium calculations typically indicate a radiative tropopause (i.e., a level at which the temperature begins to increase) whose height is typically *less than* that calculated using (12.115). Such a tropopause would provide a lid on linear baroclinic waves, and above this their amplitude would fall rapidly (e.g., Fig. 6.21), and in that case h would play no significant role.

Let us turn, then, to nonlinear arguments.

12.6.2 Mixing potential vorticity and baroclinic adjustment

Let us suppose that baroclinic eddies mix potential vorticity (see section 10.5 for a general discussion of potential vorticity mixing). Such mixing will try to homogenize potential vorticity, or equivalently to expel potential vorticity gradients to a boundary, and the

(extratropical) tropopause would then occur at an isoline of potential vorticity and be marked by a near-discontinuity in the potential vorticity distribution. Because $Q \approx (f/\rho)\partial\theta/\partial z$, the tropopause would also correspond to a discontinuity in stratification. Supporting this notion is the fact that the potential vorticity distribution in the troposphere is indeed significantly more homogeneous than in the stratosphere (see Fig. 12.16, noting the unequal contour intervals of PV). The correspondence is perhaps not so marked as to convince the sceptic, and the lower stratosphere is often locally characterized by a more homogeneous potential vorticity than such Eulerian averages imply. Putting such objections aside, let us consider to what height such potential vorticity mixing might occur.

Potential vorticity mixing will occur only so far as needed in order to stabilize the mean flow. (We know that the meridional surface temperature gradient remains negative, so that if the flow is stabilized it must involve changes in the interior potential vorticity distribution.) Neglecting the contribution of relative vorticity, the quasi-geostrophic potential vorticity is given by

$$q = \beta y + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial \psi}{\partial z} \right), \quad (12.117)$$

so that

$$\frac{\partial \bar{q}}{\partial y} = \beta - \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \Lambda \right). \quad (12.118)$$

We might hypothesize that the vertical extent to which mixing occurs is just sufficient to make the two terms on the right-hand side a similar size in order that the potential vorticity can become homogeneous, or that it can change sign and be just unstable. This gives $\beta \sim f_0^2 \Lambda / (N^2 H_T)$, where H_T is the vertical extent of the instability and, we assume, the height of the tropopause; that is

$$H_T \sim \frac{f_0^2 \Lambda}{N^2 \beta}. \quad (12.119)$$

Put another way, in this theory the troposphere extends vertically as far as baroclinic waves can alter the potential vorticity from its planetary value. (A similar depth scale occurs when evaluating the depth of the wind's influence in an ocean circulation model, section 14.8.1.) This height determines the tropopause, and the radiative constraint then determines the lapse rate within the troposphere. The shear itself is proportional to the horizontal temperature gradient, and the height of the tropopause and the meridional temperature gradient may adjust together to ensure the satisfaction of (12.119). Note the similarity of (12.119) to (12.115) — a similarity that is unsurprising given that we are constructing a height scale from a shear, f , β and N^2 using potential vorticity dynamics in both cases.

Equilibration by potential vorticity mixing is closely related to a process known as *baroclinic adjustment*, by analogy with convective adjustment.¹¹ The essential idea is that baroclinic eddies are sufficiently efficient that they can stabilize the mean flow by transferring heat polewards and upwards until the necessary condition for instability (the Charney-Stern-Pedlosky condition, in so far as the flow is quasi-geostrophic and inviscid) is barely satisfied, and the atmosphere is marginally supercritical to baroclinic instability. The adjustment might proceed predominantly by changes in the static stability, N^2 , or predominantly by changes in the horizontal temperature gradient, but the arguments surrounding (12.116) suggest that, if adjustment-like arguments do hold, the isentropic slope will remain roughly constant over a broad range of parameters. The numerical and observational evidence

for such an adjustment is mixed, although it is plausible that a weaker version may hold in which potential vorticity is imperfectly homogenized and (12.119) provides a plausible scaling, but not a precise prediction, for the height of the tropopause.¹²

12.6.3 Diffusive transport of potential vorticity

A related approach to the problem is to assume that potential vorticity and potential temperature at the surface are transported diffusively but not completely homogenized. (The observational support for potential vorticity homogenization is in any case somewhat weak, and certainly there is no useful sense in which the surface temperature is homogenized. Relatedly, the observed mean atmosphere is baroclinically unstable. A complete solution of the problem would involve determining the diffusivities, and this is itself a problem of turbulence, as discussed in chapter 10. If the diffusivities increase rapidly with the shear of the zonal wind, as in the model of section 9.3, then it may be difficult for the supercriticality to become very large because the meridional heat fluxes will also increase rapidly and reduce the meridional temperature gradient, much as in adjustment ideas. In the absence of a wholly satisfactory theory for the diffusivities, let us see what might be understood without solving the problem in its entirety. We will take the circulation of the atmosphere to be that illustrated in Fig. 12.24 (see also Fig. 11.14), in which the polewards residual circulation in the free atmosphere is balanced by a shallow return flow near the surface, and make the following assumptions.

- (i) In the free atmosphere potential vorticity is diffusively transported downgradient, whereas near the surface potential temperature is diffused downgradient.
- (ii) The eddy diffusivities are vertically uniform, and in particular the diffusivities of interior potential vorticity and surface potential temperature are equal to each other.
- (iii) The height to which the eddy fluxes reach adjusts so that mass balance and the above conditions can be satisfied, and this height is to be associated with the tropopause.

The second assumption is perhaps the simplest one can make about eddy diffusivities, and the third assumption essentially defines our troposphere; nevertheless, as with all theories based on turbulent diffusion, the assumptions are to some degree ad hoc. Let us now make an idealized model based upon them; the model is informal and illustrative rather than quantitative, and the assumptions above are more important than the details of the model below.¹³

An idealized model

For illustrative purposes consider a multi-layer model in which the lowest model layer (the surface layer) is quite thin, and let us suppose that it is in this layer that all the equatorward return flow of the residual circulation occurs. The vertically integrated mass balance condition is that the poleward interior flow is equal to the equatorward surface flow, and if the interior flow is approximately vertically uniform we write this as $H_i \bar{v}_i^* + H_s \bar{v}_s^* \approx 0$, where the subscripts i and s denote the interior and the surface, respectively, with H_i being the total depth of the interior flow and H_s being the depth of the surface layer. In the interior the zonal momentum equation with quasi-geostrophic scaling (see, for example, the shaded box on page 510) implies that, in a steady state, the residual meridional velocity satisfies $-f_0 \bar{v}_i^* = \bar{v}_i' \bar{q}_i'$. If we assume that the potential vorticity is diffused downgradient with an

eddy diffusivity κ_i , and that the potential vorticity gradient is approximately equal to β , the residual flow is given by

$$\bar{v}_i^* \approx \frac{\kappa_i \beta}{f_0}. \quad (12.120)$$

If the surface layer is thin, its potential vorticity is, as in the delta-function construction of section 5.4.3, dominated by the contribution from the surface temperature, and from (5.88b), $q_s \approx (f_0^2/H_s g')(\psi_{s-1} - \psi_s)$, where the subscript 's - 1' denotes the layer above the surface layer, and $(f_0/g')(\psi_{s-1} - \psi_s)$ is the perturbation layer thickness h'_s , and is, in our model, proportional to the surface temperature. The residual flow in the surface layer is given by

$$\bar{v}_s^* = \frac{1}{H_s} \bar{v}_s' h'_s \approx -\frac{\kappa_s}{H_s} \frac{\partial \bar{h}_s}{\partial y}, \quad (12.121)$$

using assumption (i) above, where \bar{h}_s is the zonally averaged surface thickness gradient and κ_s is the surface diffusivity. Written in terms of surface temperature (using, for example, the 'level' rather than the 'layer' versions of the quasi-geostrophic equations) the corresponding result is

$$\bar{v}_s^* = \frac{\kappa_s}{H_s \partial \bar{\theta} / \partial z} \frac{\partial \bar{\theta}_s}{\partial y}. \quad (12.122)$$

If we now invoke assumption (ii), that $\kappa_i = \kappa_s$, and use the mass balance condition $H_i \bar{v}_i^* = -H_s \bar{v}_s^*$, then (12.120) and (12.122) combine to give

$$H_i \frac{\partial \bar{\theta}}{\partial z} \approx -\frac{f_0}{\beta} \frac{\partial \bar{\theta}_s}{\partial y}. \quad (12.123)$$

The left-hand side is approximately the potential temperature difference between surface and tropopause, and $\bar{\theta}_s$ is the zonally-averaged surface potential temperature. Equation (12.123) thus provides a relation between the stratification of the troposphere and the surface temperature gradient. With the thermal wind relation, the result is in turn very similar to (12.119), except that it is the shear at the surface that now appears. The differences between this argument and that of the previous subsection lie not so much in their respective results, which are rather similar in some respects. Rather, the differences lie in the assumptions needed to derive the results and in the interpretation of the results. Transport of potential vorticity is a common aspect, but in the arguments leading to (12.123) we do not make the a priori assumption that potential vorticity will be completely mixed, or that the flow will be marginally unstable. Rather, we make an observation about the nature of the circulation (that the return flow is close to the surface) and certain assumptions about the nature of the diffusion, and a related result emerges. (Note also that in a quasi-geostrophic model with fixed layer depths the depth of the eddying region cannot adjust in the way we have assumed; rather, the eddy diffusivities must adjust to satisfy the constraints.)

Details aside, the heart of all of the above arguments is that the troposphere is that part of the atmosphere within which the energy and entropy are redistributed by dynamical mechanisms, and that the depth of the tropopause is determined by the constraints arising from the potential vorticity dynamics that effect this redistribution.

12.7 † EXTRATROPICAL CONVECTION AND THE VENTILATED TROPOSPHERE

A point of view that differs *qualitatively* from the potential vorticity one is to suppose that the mid-latitude tropospheric lapse rate is maintained by convection, the convection occurring predominantly in the warm sector of mature baroclinic waves.¹⁴

In reality such convection will involve moisture, but for simplicity let us first consider a dry atmosphere. In a given baroclinic zone the minimum potential temperature difference between the tropopause and the surface, $\Delta_z \theta = \theta_T - \theta_S$ is approximately zero: if the tropopause were colder than this the column would be convectively unstable, and the difference would become zero. The essential assumption that we make is that within a baroclinic zone there generally *does* exist a region that is convectively unstable, and that convection then ensues with sufficient efficiency to partially fill the troposphere with air with that surface value of potential temperature. The process differs from convection in the tropics because it is organized by baroclinic waves and, if we imagine a succession of baroclinic waves around a latitude band the mean value of $\Delta_z \theta$ will be approximately, we assume, its minimum (zero) plus a fraction of its standard deviation. The standard deviation in turn is a consequence of the pre-existing meridional temperature gradient and meridional advection across that gradient, and therefore

$$\text{standard deviation}(\Delta_z \theta) \propto \Delta_y \theta, \quad (12.124)$$

where the term on the right-hand side is the meridional temperature difference at the surface across the baroclinic zone. The mean potential temperature difference between the surface and tropopause is then simply proportional to the meridional temperature gradient at that latitude, with an undetermined constant of proportionality and so

$$\Delta_z \theta \propto \Delta_y \theta. \quad (12.125)$$

Finally, if moisture is present (as it is!) the vertical potential temperature difference should be replaced by the equivalent potential temperature difference — the equivalent potential temperature being essentially the potential temperature achieved when all the water vapour in a parcel of air condenses and the latent heat of condensation is used to heat the parcel (section 2.9.3).

The physical hypothesis is essentially that within a baroclinic wave the advection of warm air into a cold region *necessarily* leads to convection, and that this convection then efficiently fills the available volume, to the extent possible, with the warmest possible fluid. Oceanographers will find this a comfortable concept, for they are used to the notion of convection filling the domain with the densest available fluid (densest in the oceanic case because oceanic convection usually occurs from the top, with cold, dense water sinking). However, unlike the ocean in which the bottom of the container limits the volume of dense water that can be made, here it is the tropopause that provides the upper lid; the height of this is determined by the dynamics itself, in conjunction with the radiative constraint discussed previously. Thus, the baroclinic zone becomes, in oceanographic parlance, *ventilated* by the warmest air at the surface. However, the entire baroclinic zone does not completely fill with this warm air because the convection is maintained by a meridional temperature gradient and it is necessarily intermittent: baroclinic instability would shut off if the entire baroclinic zone were filled with homogeneous warm fluid, and the zone would then meridionally restratify. It is this continual maintenance of variance that leads to (12.125).

The Stratification of the Troposphere and Stratosphere

We may summarize our ideas about stratification as follows.

- ★ The troposphere is that region of the atmosphere where dynamics plays an important role in stratification, whereas the stratosphere is more nearly in radiative equilibrium (although in winter in the lower stratosphere, radiative equilibrium is not a very good approximation). The tropopause is the change in stratification between the two regions.
- ★ The tropospheric lapse rate and the height of the tropopause are determined by a combination of dynamics and radiation. If a dynamical process predicts the height of the tropopause then, if the surface temperature is also given, the stratification follows via radiative considerations. Similarly, if the lapse rate is given, the tropopause height follows.
- ★ There are two general classes of theory for tropospheric stratification.
 - (i) Potential vorticity transport and its variants. The common theme of this class of theory is that the lateral and upward heat transport by baroclinic instability determines the mid-latitude meridional temperature gradient and stratification, and potential vorticity dynamics plays a dominant role. In one incarnation ('baroclinic adjustment') baroclinic activity is so efficient that the atmosphere becomes only marginally supercritical to baroclinic instability, a process related to homogenization of potential vorticity. In a distinct variation on this theme, potential vorticity and surface potential temperature are diffused downgradient, but potential vorticity is not necessarily homogenized. A complete theory would require estimates of the structure and magnitude of the eddy diffusivities. In all these ideas, the depth to which the transport and mixing occurs determines the position of the tropopause.
 - (ii) Moist convection. It is generally thought that moist convection does play the dominant role in determining tropical stratification, leading to a lapse rate that is moist-neutral. In the extratropics such convection acts in concert with baroclinic eddies to produce a stratification that is related to the meridional temperature difference across a baroclinic zone.

In both of the above, the troposphere is the region in which a dynamical redistribution of heat occurs. In the real atmosphere, both mechanisms may operate, and/or be limiting cases, each applicable in different parameter regimes.

- ★ In (i), the constraints arising from potential vorticity dynamics suggest the importance of the height scale $H_T \sim (f^2 \Lambda) / (N^2 \beta)$, or equivalently that $s \sim \beta H_T / f$, where s is the isentropic slope, at least to the extent that relative vorticity gradients are much smaller than planetary vorticity gradients.
- ★ In (ii), the lapse rate is bounded from below by the moist adiabatic lapse rate and, on average, the (equivalent) potential temperature difference between the surface and the troposphere is proportional to the meridional temperature difference across the baroclinic zone.

The above picture does make a couple of assumptions that are by no means self-evident, or necessarily generally applicable. First, only certain distributions have the property that the mean is the minimum plus a fraction of the variance, although qualitatively similar properties may hold more generally and the qualitative thrust of argument does not depend on it holding exactly. Second, and more importantly, the argument depends on the heat flux due to convection being comparable to or controlled by the baroclinic eddies themselves. If convective heat fluxes dominate, as in the tropics, then the lapse rate will be approximately moist neutral. If the baroclinic heat fluxes dominate, we would (without a third alternative) return to the potential vorticity arguments of the previous section.

The ultimate consequence of these convection arguments is that the moist isentropic slope is proportional to that slope which would take an isentrope at the surface to the tropopause across a baroclinic zone. This contrasts with the potential vorticity mixing ideas, which suggest (at least in so far as relative vorticity fluxes are neglected) that the dry isentropic slope is proportional to the slope that goes from the ground to the tropopause over a horizontal scale f/β , the equator-to-pole scale. However, differentiating between these two predictions may be difficult: in today's climate the two predictions are similar, and using a numerical model to go to a parameter regime in which they differ may lead to results that are not relevant for today's climate. We stated on page 529 that the arguments for the maintenance of extratropical convection fell into two general camps; it may be that these camps represent limiting cases, each of which may hold in different parameter regimes. The real atmosphere may lie in between, or it may have different dynamics in different seasons, or it may be disdainful of all our ideas.

APPENDIX: TEM AND ELIASSEN-PALM FLUX FOR THE PRIMITIVE EQUATIONS IN SPHERICAL COORDINATES

In spherical and log-pressure coordinates let us define the residual streamfunction for the ideal-gas primitive equations by¹⁵

$$\psi^* \equiv \psi + \overline{\frac{v' \theta'}{\partial_z \theta}}. \quad (12.126)$$

Here, an overbar denotes a conventional (Eulerian) zonal average, ψ is the streamfunction of the zonally averaged flow, $Z = -H \ln(p/p_R)$ where p_R is a reference pressure and H is a scale height, and $\rho_R = \rho_0 \exp(-Z/H)$ where ρ_0 is a constant. The associated transformed, or residual, velocities are:

$$\bar{v}^* = -\frac{1}{\rho_R} \frac{\partial}{\partial Z} (\psi^* \rho_R), \quad \bar{w}^* = \frac{1}{a \cos \vartheta} \frac{\partial}{\partial \vartheta} (\psi^* \cos \vartheta), \quad (12.127)$$

with an equivalent expression for \bar{v} and \bar{w} in terms of ψ . (The notation for log-pressure coordinates follows section 2.6.3 on page 80, except here we use a lowercase w for the vertical velocity.) If we write the equations of motion in terms of the residual velocities instead of the Eulerian velocities we obtain the 'transformed Eulerian mean', or TEM, equations (section 7.3). The TEM forms of the zonally averaged thermodynamic and zonal momentum equations are:

$$\frac{\partial \bar{\theta}}{\partial t} + \frac{\bar{v}^*}{a} \frac{\partial \bar{\theta}}{\partial \vartheta} + \bar{w}^* \frac{\partial \bar{\theta}}{\partial Z} = \frac{1}{\rho_R} \frac{\partial G}{\partial z}, \quad (12.128a)$$

$$\frac{\partial \bar{u}}{\partial t} + \bar{v}^* \left(\frac{1}{a \cos \vartheta} \frac{\partial}{\partial \vartheta} (\bar{u} \cos \vartheta) - f \right) + \bar{w}^* \frac{\partial \bar{u}}{\partial Z} = \frac{1}{\rho_R \cos \vartheta} \nabla \cdot \mathbf{F}. \quad (12.128b)$$

The transformed equations of motion are completed by the meridional momentum, mass continuity and hydrostatic equations:

$$\bar{u} \left(f + \frac{\bar{u}}{a} \tan \vartheta \right) = - \frac{1}{a} \frac{\partial \Phi}{\partial \vartheta} + \hat{S}, \quad (12.129a)$$

$$\frac{1}{a \cos \vartheta} \frac{\partial}{\partial \vartheta} (\bar{v}^* \cos \vartheta) + \frac{1}{\rho_R} \frac{\partial}{\partial Z} (\rho_R \bar{w}^*) = 0, \quad (12.129b)$$

$$\frac{\partial \Phi}{\partial Z} = \frac{R \bar{T}}{H} T = \frac{R}{H \bar{\theta}} e^{-\kappa Z/H}. \quad (12.129c)$$

In (12.128a), $\mathbf{F} = (\mathcal{F}^{\vartheta}, \mathcal{F}^Z)$ is the Eliassen-Palm flux, given by

$$\mathcal{F}^{\vartheta} = \rho_R \cos \vartheta \left[\bar{u}_Z \frac{\bar{v}' \bar{\theta}'}{\partial_Z \bar{\theta}} - \bar{u}' \bar{v}' \right], \quad (12.130a)$$

$$\mathcal{F}^Z = \rho_R \cos \vartheta \left[\left(f - \frac{\partial_{\vartheta} (\bar{u} \cos \vartheta)}{a \cos \vartheta} \right) \frac{\bar{v}' \bar{\theta}'}{\partial_Z \bar{\theta}} - \bar{u}' \bar{w}' \right], \quad (12.130b)$$

with

$$\nabla \cdot \mathbf{F} = \frac{1}{a \cos \vartheta} \frac{\partial}{\partial \vartheta} (\mathcal{F}^{\vartheta} \cos \vartheta) + \frac{\partial}{\partial Z} \mathcal{F}^Z. \quad (12.130c)$$

In (12.128a)

$$G = \frac{\rho_R}{\partial_Z \bar{\theta}} \left(\bar{v}' \bar{\theta}' \frac{1}{a} \frac{\partial \bar{\theta}}{\partial \vartheta} + \bar{w}' \bar{\theta}' \frac{\partial \bar{\theta}}{\partial Z} \right), \quad (12.130d)$$

and \hat{S} in (12.129a) contains various, generally small, terms that lead to departures from gradient-wind balance between \bar{u} and the geopotential Φ . Expressions very similar to the ones above also arise in pressure coordinates.

In many circumstances, the EP flux is well approximated by

$$\mathbf{F} = \left(-\rho_R \cos \vartheta \bar{u}' \bar{v}', f \rho_R \cos \vartheta \frac{\bar{v}' \bar{\theta}'}{\partial_Z \bar{\theta}} \right), \quad (12.131)$$

in which case the zonal flow is accelerated by the EP flux according to

$$\frac{\partial \bar{u}}{\partial t} + \dots = \frac{1}{a \cos^2 \vartheta} \frac{\partial}{\partial \vartheta} \left(-\bar{u}' \bar{v}' \cos^2 \vartheta \right) + \frac{1}{\rho_R} \frac{\partial}{\partial Z} \left(\rho_R f \frac{\bar{v}' \bar{\theta}'}{\partial_Z \bar{\theta}} \right). \quad (12.132)$$

With $f = f_0$ (12.131) becomes the quasi-geostrophic EP flux, and in this limit G is also neglected.

In the figures that show the EP vectors, the horizontal and vertical components of the EP flux are scaled by a (the Earth's radius) and by $H = 1000$ hPa (the pressure depth of the atmosphere), respectively. The scaling determines the direction of the arrows and makes it possible to see the divergence by eye, and which component dominates in producing that divergence. In the figures that show the EP flux divergence, we plot the right-hand side of (12.128b), namely the EP flux divergence divided by $\rho_R \cos \vartheta$, this being the quantity that directly contributes to the acceleration of the zonal flow.¹⁶

Notes

- 1 The modern view of the mid-latitude general circulation — the largely zonally asymmetric motion that provides the bulk of the meridional transport of heat and momentum in the extratropics — began to take form in the 1920s in papers by Defant (1921) and Jeffreys (1926). Defant regarded the mid-latitude circulation as turbulence on a large scale (albeit without realizing the important organizing effects of waves), and calculated the horizontal eddy-diffusivities using Prandtl-like mixing length arguments. Soon after, Jeffreys presciently wrote of ‘the dynamical necessity for a continual exchange of air between high and low latitude’ and that ‘no general circulation of the atmosphere without cyclones is dynamically possible when friction is taken into account.’ This point of view slowly gained ground, with, for example, Starr (1948) advocating the point of view that large-scale eddies were responsible for the bulk of the meridional transport of momentum in mid-latitudes, and Rossby (1949) eventually noting in a review article that ‘One is forced to conclude that there no longer exists a compelling reason to build the theory of the maintenance of the general circulation exclusively on *meridional* solenoidal circulations’. Soon after this came a pair of discussion papers by Eady (1950, 1954), that, setting the stage for the modern viewpoint, struggle with the turbulent transport of mid-latitude eddies and the maintenance of the surface currents — the importance of the enstrophy budget is discussed, for example, and Eady comes close to deriving wave activity conservation. Around that time Kuo (1951) discussed the maintenance of zonal flows by the mechanism of vorticity transfer in a state with a meridional background gradient, similar to mechanism I of section 12.1.2. Another landmark is the influential monograph by Lorenz (1967) that summarized, clarified and added to progress to that date, noting (in his last paragraph) that the cause of the poleward eddy momentum transport across mid-latitudes (and hence the cause of the surface eastward winds) had not at that time been rigorously explained. If perhaps not rigorous, we do now have a qualitative explanation of these dynamics by way of potential vorticity dynamics and the momentum transport in Rossby waves, as described in this chapter and chapter 7.
- 2 The mechanism producing a westerly jet in the atmosphere, and the associated surface westerlies, used to be referred to as ‘negative viscosity’ (Starr 1968), because it is associated with an upgradient transfer of momentum. The generation of a zonal flow by rearrangements of vorticity on a background state with a meridional gradient was noted by Kuo (1951). Thompson (1971) (see also Thompson 1980) calculated the momentum transport by Rossby waves; in particular he noted that the zonal momentum flux was in the opposite direction to the group velocity, and noted the potential for upgradient transfer and mean flow generation, as experimentally verified by Whitehead (1975). These ideas were developed further by Green (1970), Rhines (1979), Rhines & Holland (1979), among others, and we also draw from the review of Held (2000).
- 3 This technique is noted by Lighthill (1965), who remarks that the idea goes back to Rayleigh.
- 4 Following Held (2000).
- 5 Models of the general circulation of this ilk were introduced by Green (1970). Dickinson (1969) also considered the potential vorticity transport in planetary waves.
- 6 Adapted from Juckes (2001).
- 7 Early evidence that the temperature increases above about 11 km came from the balloon measurements of Tesserenc De Bort (1902), who also suggested the names tropopause and stratosphere, and Assmann (1902). See Hoinka (1997) for a historical account. More recently, radiative and dynamical issues relevant to this topic are discussed by, among others, Stone

- (1972), Held (1982), Juckes (2000), Shepherd (2002) and some of the articles in Schneider & Sobel (2007).
- 8 Paraphrasing World Meteorological Organization (1957); see also Lewis (1991).
- 9 Thomas Birner calculated the tropopause-based averages. See also Birner *et al.* (2002) and Birner (2006).
- 10 However, it is sometimes said that the height of the tropical tropopause is rather higher than the depth to which deep convection penetrates, and if true this is in need of explanation.
- 11 Stone (1978). See Zurita-Gotor & Lindzen (2007) for a review.
- 12 Evidence is mixed that the atmosphere, or models thereof, seek to become marginally supercritical in any sense. For example, and in ascending order of complexity: (i) In quasi-geostrophic flows (in which static stability is fixed) Salmon (1980) and Vallis (1988) found that the flow could be strongly supercritical. (ii) Using an idealized dry GCM, Schneider & Walker (2006) found that their model atmosphere was only marginally supercritical over a very broad parameter regime (although their criterion for supercriticality differed somewhat from the quasi-geostrophic one, and the paper specifically does not imply that atmosphere is baroclinically neutral or that potential vorticity is homogenized; see section 12.6.3). (iii) Using a comprehensive, moist, GCM Thuburn & Craig (1997) found results that they interpreted as not being supportive of baroclinic adjustment, using quasi-geostrophic criteria to determine the predicted supercriticality and tropopause height. Thuburn and Craig's results do support the notion of radiative and dynamical constraints discussed earlier. (iv) Stone & Nemet (1996) found that the isentropic slope of the real atmosphere does not vary strongly with season, even though the heat flux does, a result supportive of baroclinic adjustment ideas. However, linear calculations show that the mean atmospheric state, certainly in winter, is baroclinically unstable, with growth rates of about 0.2 day^{-1} or more (e.g., Valdes & Hoskins 1989). None of the above-mentioned studies is inconsistent with the notion that downgradient fluxes of potential vorticity are important in determining the thermal stratification of the atmosphere.
- 13 See Schneider & Walker (2006) and references therein for a continuously stratified and, notably, non-quasi-geostrophic derivation, and associated numerical simulations.
- 14 Following Juckes (2000).
- 15 For more detail see Edmon *et al.* (1980) or Andrews (1987).
- 16 E. Gerber kindly constructed these figures.

Further reading

Green, J. S. A., 1999. *Atmospheric Dynamics*.

A somewhat personal view of dynamics, with a number of interesting perspectives and insights on how the atmosphere works.

James, I., 1994. *An Introduction to Circulating Atmospheres*.

An introductory book on the global circulation of the atmosphere, with discussions of theory, observations and models.

Schneider, T. & Sobel, A., 2007. *The Global Circulation of the Atmosphere: Phenomena, Theory, Challenges*.

This book contains several interesting review articles on the large-scale atmospheric circulation that will bring the reader to the frontiers of knowledge.

Problems

- 12.1 ♦ Construct a numerical model (e.g., in Fortran, C, or Matlab) that steps (12.97) forward in time, using a diffusive closure for the potential vorticity fluxes, a frictional term equivalent to a linear drag on velocity in the lower layer, and thermodynamic source term equivalent to a relaxation back to a temperature that monotonically decreases with latitude. Explore the effects of different lateral boundary conditions on the potential vorticity flux, the effects of satisfying or not the kinematic constraint (12.82), and the effects of various horizontal structures and amplitudes of the transfer coefficients. Discuss whether this is a useful model of the zonally averaged circulation.
- 12.2 Suppose the Earth's atmosphere was heated at the pole and cooled at the equator, but that is was otherwise similar to the present atmosphere. Suppose also that this pattern of heating leads to the formation of a single baroclinic zone in mid-latitudes. Discuss the general circulation that might result, including the pattern of surface winds, the shear, and the meridional overturning circulation.

CHAPTER THIRTEEN

Planetary Waves and the Stratosphere

IN TRYING TO BUILD A THEORY of the general circulation of the extratropical atmosphere it is useful to divide that task into two. The first task (chapters 11 and 12) is to understand the zonally averaged circulation and the transient zonal asymmetries by supposing that, to a first approximation, this circulation is qualitatively the same as it would be if the boundary conditions were zonally symmetric, with no mountains or land-sea contrasts. Given the statistically zonally symmetric circulation, the second task is to understand the zonally asymmetric circulation, and we may do this by supposing that the latter is a perturbation on the former, and using a theory linearized about the zonally symmetric state. It is by no means obvious that such a procedure will be successful, for it depends on the nonlinear interactions among the zonal asymmetries being weak. We might make some *a priori* estimates that suggest that this might be the case, but the ultimate justification for the approach lies in its *a posteriori* success.

In practice, the procedure involves calculating the forced, stationary Rossby waves that are generated by the interaction of surface topography and zonally asymmetric thermal forcing with the zonally averaged flow; that is, we compute the forced *stationary waves* of the system. Such stationary waves are examples of *planetary waves*, which are just waves at the planetary scale, with the definition usually being restricted to Rossby waves or similar. Properly including the effects of transient eddies — equilibrated, finite-amplitude baroclinic systems — is the most difficult aspect of calculating the stationary wave response, although their effects may be included diagnostically by evaluating their associated heat and momentum fluxes from observations and adding them to the right-hand sides of the appropriate equations. However, we will find that the calculations are quite revealing even if the effects of transient eddies are omitted entirely. We focus first on the response to orography at the lower boundary, and then consider thermodynamic forcing — arising,

for example, from an inhomogeneous surface temperature field.¹ Because these waves propagate upwards as well as laterally, this will lead us into the second topic of this chapter, a discussion of stratospheric dynamics.

Even though stationary wave theory is quite successful in describing the zonal asymmetries of the mid-latitude flow, there are a number of zonal asymmetries, particularly at low latitudes, about which the theory is silent. These include monsoon circulations caused by land-sea contrasts, and the Walker circulation of the equatorial Pacific. Their dynamics is beyond our scope, and the reader should consult the original literature to learn more.

13.1 FORCED AND STATIONARY ROSSBY WAVES

13.1.1 A simple one-layer case

Many of the essential ideas can be illustrated by a one-layer quasi-geostrophic model, with potential vorticity equation

$$\frac{Dq}{Dt} = 0, \quad q = \zeta + \beta y - \frac{f_0}{H}(\eta - h_b), \quad (13.1)$$

where H is the mean thickness of the layer, η is the height of the free surface, h_b is the bottom topography, and the velocity and vorticity are given by $\mathbf{u} = (g/f_0)\nabla^\perp\eta \equiv (g/f_0)\mathbf{k} \times \nabla\eta$ and $\zeta = (\partial v/\partial x - \partial u/\partial y) = (g/f_0)\nabla^2\eta$. Linearizing (13.1) about a flat-bottomed state with zonal flow $\bar{u}(y) = -(g/f_0)\partial\bar{\eta}/\partial y$ gives

$$\frac{\partial q'}{\partial t} + \bar{u} \frac{\partial q'}{\partial x} + v' \frac{\partial \bar{q}}{\partial y} = 0, \quad (13.2)$$

where $q' = \zeta' - (f_0/H)(\eta' - h_b)$ and $\partial\bar{q}/\partial y = \beta + \bar{u}/L_d^2$ with $L_d = \sqrt{gH}/f_0$, the radius of deformation. Equation (13.2) may be written, after the cancellation of a term proportional to $\bar{u}\partial\eta'/\partial x$, as

$$\frac{\partial}{\partial t} \left(\zeta' - \frac{\psi'}{L_d^2} \right) + \bar{u} \frac{\partial \zeta'}{\partial x} + \beta v' = -\bar{u} \frac{\partial \hat{h}}{\partial x}, \quad (13.3)$$

where $\psi' = (g/f_0)\eta'$ and $\hat{h} = h_b f_0 / H = h_b g / (L_d^2 f_0)$.

The solution of this equation consists of the solution to the homogeneous problem (with the right-hand side equal to zero, as considered in section 5.7 on Rossby waves) and the particular solution. We proceed by decomposing the variables into their Fourier components

$$(\zeta', \psi', \hat{h}) = \text{Re}(\tilde{\zeta}, \tilde{\psi}, \tilde{h}_b) \sin ly e^{ikx}. \quad (13.4)$$

where such decomposition is appropriate for a channel, periodic in the x -direction and with no variation at the meridional boundaries, $y = (0, L)$. The full solution will be a superposition of such Fourier modes and, because the problem is linear, these modes do not interact. The free Rossby waves, the solution to the homogeneous problem, evolve according to

$$\psi = \text{Re} \tilde{\psi} \sin ly e^{i(kx - \omega t)}, \quad (13.5)$$

where ω is given by the dispersion relation [cf. (5.185)]

$$\omega = k\bar{u} - \frac{k\partial_y \bar{q}}{K^2 + k_d^2} = \frac{k(\bar{u}K^2 - \beta)}{K^2 + k_d^2}, \quad (13.6a,b)$$

where $K^2 = k^2 + l^2$ and $k_d = 1/L_d$. Stationary waves occur at the wavenumbers for which $K = K_s \equiv \sqrt{\beta/\bar{u}}$. To the free waves we add the solution to the steady problem,

$$\bar{u} \frac{\partial \zeta'}{\partial x} + \beta v' = -\bar{u} \frac{\partial \hat{h}}{\partial x}, \quad (13.7)$$

which is, using the notation of (13.4)

$$\tilde{\psi} = \frac{\tilde{h}_b}{(K^2 - K_s^2)}. \quad (13.8)$$

Now, \tilde{h}_b is a complex amplitude; thus, for $K > K_s$ the streamfunction response is *in phase* with the topography. For $K^2 \gg K_s^2$ the steady equation of motion is

$$\bar{u} \frac{\partial \zeta'}{\partial x} \approx -\bar{u} \frac{\partial \hat{h}}{\partial x}, \quad (13.9)$$

and the topographic vorticity source is balanced by zonal advection of relative vorticity. For $K^2 < K_s^2$ the streamfunction response is *out of phase* with the topography, and the dominant balance for very large scales is between the meridional advection of planetary vorticity, $v \partial f / \partial y$ or βv , and the topographic source. For $K = K_s$ the response is infinite, with the stationary wave resonating with the topography. Now, any realistic topography can be expected to have contributions from *all* Fourier components. Thus, for *any* given zonal wind there will be a resonant wavenumber and an infinite response. This, of course, is not observed, and one reason is that the real system contains friction. The simplest way to include this is by adding a linear damping to the right-hand side of (13.3), giving

$$\frac{\partial}{\partial t} \left(\zeta' - \frac{\psi'}{L_d^2} \right) + \bar{u} \frac{\partial \zeta'}{\partial x} + \beta v' = -r \zeta' - \bar{u} \frac{\partial \hat{h}}{\partial x}. \quad (13.10)$$

The free Rossby waves all decay monotonically to zero (problem 13.11). However, the steady problem, obtained by omitting the first term on the left-hand side, has solutions

$$\tilde{\psi} = \frac{\tilde{h}_b}{(K^2 - K_s^2 - iR)}, \quad (13.11)$$

where $R = (rK^2/\bar{u}k)$, and the singularity has been removed. The amplitude of the response is still a maximum for the stationary wave, and for this wave the phase of the response is shifted by $\pi/2$ with respect to the topography. The solution is shown in Fig. 13.1.² It is typical that for a mountain range whose Fourier composition contains all wavenumbers, there is a minimum in the streamfunction a little downstream of the mountain ridge.

13.1.2 Application to Earth's atmosphere

With three parameters, I can fit an elephant.

William Thomson, Lord Kelvin (1824–1907).

Rather surprisingly, given the complexity of the real system and the simplicity of the model, when used with realistic topography a one-layer model gives reasonably realistic answers for

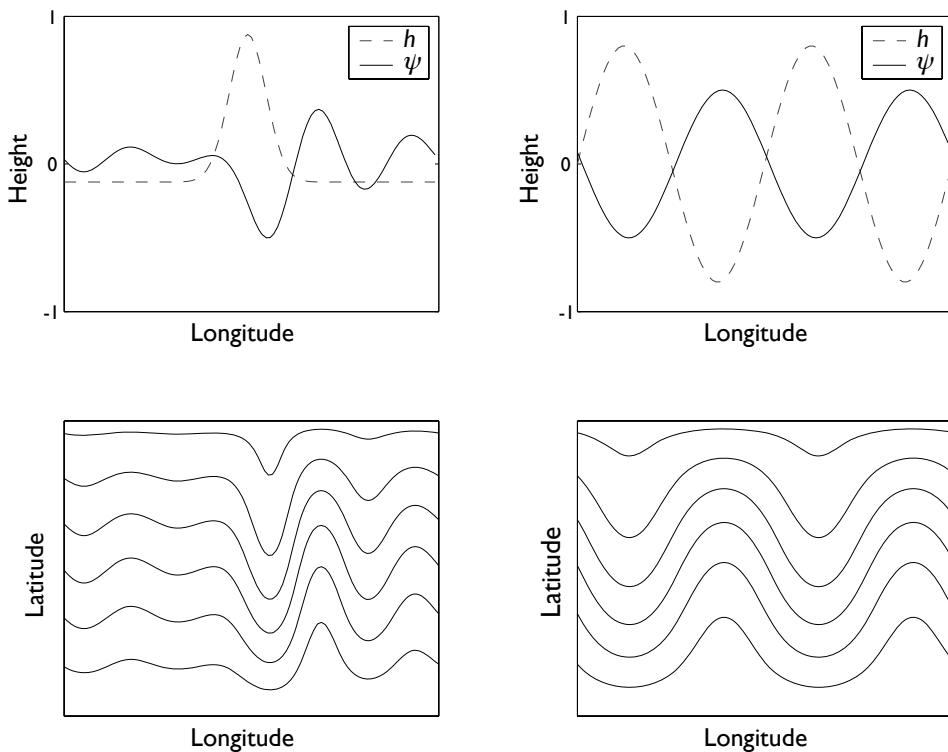


Fig. 13.1 The response to topographic forcing, i.e., the solution to the steady version of (13.10), for topography consisting of an isolated Gaussian ridge (left panels) and a pure sinusoid (right panels). The wavenumber of the stationary wave is about 4 and $r/(\bar{u}k) = 1$. The upper panels show the amplitude of the topography (dashed curve) and the perturbation streamfunction response (solid curve). The lower panels are contour plots of the streamfunction, including the mean flow. With the ridge, the response is dominated by the resonant wave and there is a streamfunction minimum, a ‘trough’, just downstream of the ridge. In the case on the right, the flow cannot resonate with the topography, which consists only of wavenumber 2, and the response is exactly out of phase with the topography.

the Earth's atmosphere. Thus, we calculate the stationary response to the Earth's topography using (13.10), using a reasonably realistic representation of the Earth's topography and, with qualification, the zonal wind. The zonal wind on the left-hand side of (13.10) is interpreted as the wind in the mid-troposphere, whereas the wind on the right-hand side is better interpreted as the surface wind, and so perhaps is about 0.4 times the mid-troposphere wind. Since the problem is linear, this amounts to tuning the amplitude of the response. The results, obtained using a rather crude representation of the Earth's topography, are plotted in Fig. 13.2. Also plotted is the observed time averaged response of the real atmosphere (the 500 mb height field at 45°N). The agreement between model and observation is quite good, but this must be regarded as somewhat fortuitous if only because the other main source of the stationary wave field — thermal forcing — has been completely omitted from

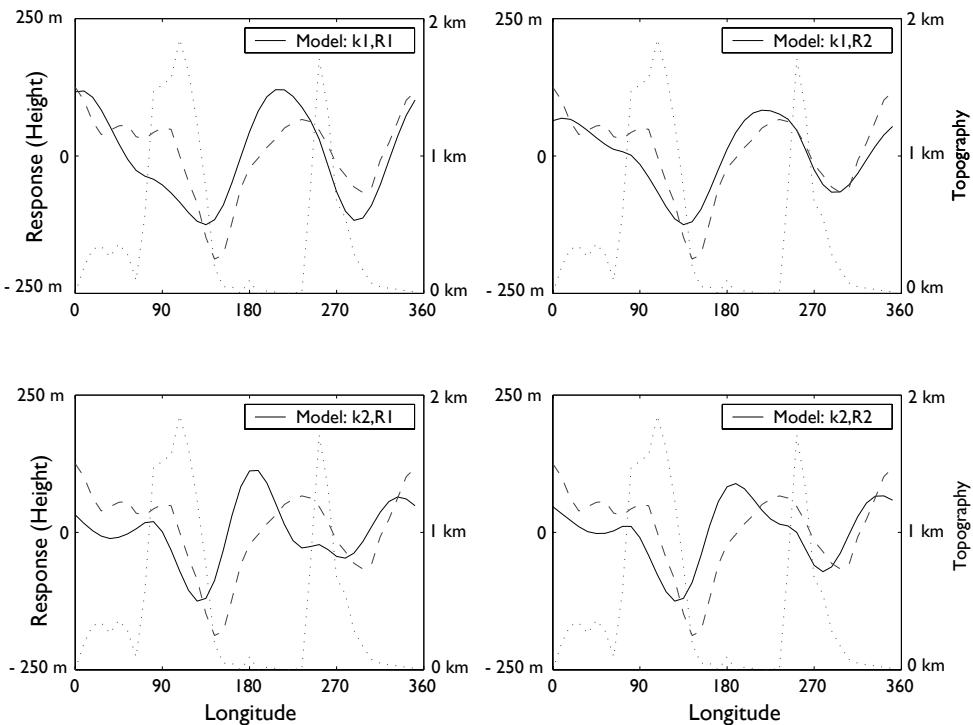


Fig. 13.2 Solutions of the Charney-Eliassen model. The solid lines are the steady solution of (13.10) using the Earth's topography at 45°N with two values of friction ($R1 \approx 6$ days, $R2 \approx 3$ days) and two values of resonant zonal wavenumber (2.5 for $k1$, 3.5 for $k2$), corresponding to zonal winds of approximately 17 and 13 $m s^{-1}$. The solutions are given in terms of height, η' , where $\eta' = f_0 \psi' / g$, with the scale on the left of each panel. The dashed line in each panel is the observed average height field at 500 mb at 45°N in January. The dotted line is the topography used in the calculations, with the scale on the right of each panel.

the calculation; the agreement is also a consequence of the aforesaid tuning. Nevertheless, the calculation does suggest that the stationary, zonally asymmetric, features of the Earth's atmosphere arise via the interaction of the zonally symmetric wind field and the zonally asymmetric lower boundary, and that these may be calculated to a reasonable approximation with a linear model.

13.1.3 * One-dimensional Rossby wave trains

Although the Fourier analysis above gives exact (linear) results, it is not particularly revealing of the underlying dynamics. We see from Fig. 13.1 that the response to the Gaussian ridge is largely downstream of the ridge, and this suggests that it will be useful to consider the response as being due to Rossby *wavetrains* being excited by local features. This is also suggested by Fig. 13.3, which shows that the response to realistic topography is relatively

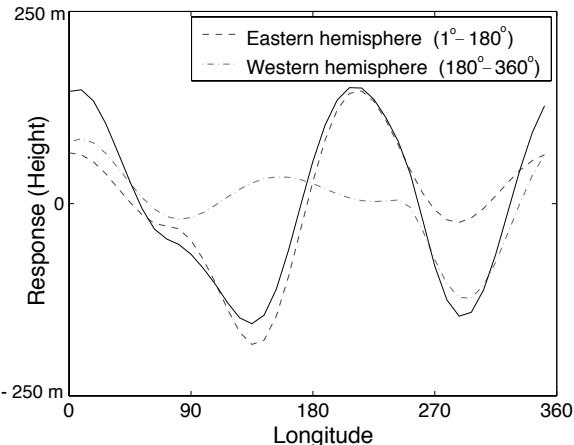


Fig. 13.3 The solution of the upper left-hand panel Fig. 13.2 (solid line), and the solution divided into two contributions (dashed lines), one due to the topography only of the western hemisphere (i.e., with the topography in the east set to zero) and the other due to the topography only of the eastern hemisphere.

local, and may be considered to arise from two relatively well-defined wavetrains each of finite extent one coming from the Rockies and the other from the Himalayas.

One way to analyse these wavetrains, and one which also brings up the concept of group velocity in a natural way, is to exploit (as in section 12.1.2) a connection between changes in wavenumber and changes in frequency. Consider the linear barotropic vorticity equation in the form

$$\frac{\partial}{\partial t}(\zeta - k_d^2 \psi) + \bar{u} \frac{\partial \zeta}{\partial x} + \beta \frac{\partial \psi}{\partial x} = -r\zeta, \quad (13.12)$$

where r is a frictional coefficient, which we presume to be small. Setting $k_d = 0$ for simplicity, the linear dispersion relation is

$$\omega = \bar{u}k - \frac{\beta k}{K^2} - ir \equiv \omega_R(k, l) - ir, \quad (13.13)$$

where $K^2 = k^2 + l^2$ and $\omega_R(k, l)$ is the inviscid dispersion relation for Rossby waves. Now, if there is a local source of the waves, for example an isolated mountain, we may expect to see a *spatial* attenuation of the wave as it moves away from the source. We may then regard the system as having a fixed, real frequency, but a changing, possibly complex, wavenumber. To determine this wavenumber for stationary waves (and so with $\omega = 0$), for small friction we expand the dispersion relation in a Taylor series about the inviscid value of ω_R at the real stationary wavenumber k_s , where $k_s = (K_s^2 - l^2)^{1/2}$ and $K_s = \sqrt{\beta/\bar{u}}$. This gives

$$\omega + ir = \omega_R(k, l) \approx \omega_R(k_s, l) + \left. \frac{\partial \omega_R}{\partial k} \right|_{k=k_s} k' + \dots. \quad (13.14)$$

Thus, $k' \approx ir/c_g^x$, where c_g^x is the zonal component of the group velocity evaluated at a fixed position and at the stationary wavenumber; using (5.183b) this is given by

$$c_g^x = \left. \frac{\partial \omega_R}{\partial k} \right|_{k=k_s} = \frac{2\bar{u}k_s^2}{k_s^2 + l^2}. \quad (13.15)$$

The solution therefore decays away from a source at $x = 0$ according to

$$\psi \sim \exp(ikx) = \exp[i(k_s + k')x] \approx \exp(ik_s x - rx/c_g^x) \quad (13.16)$$

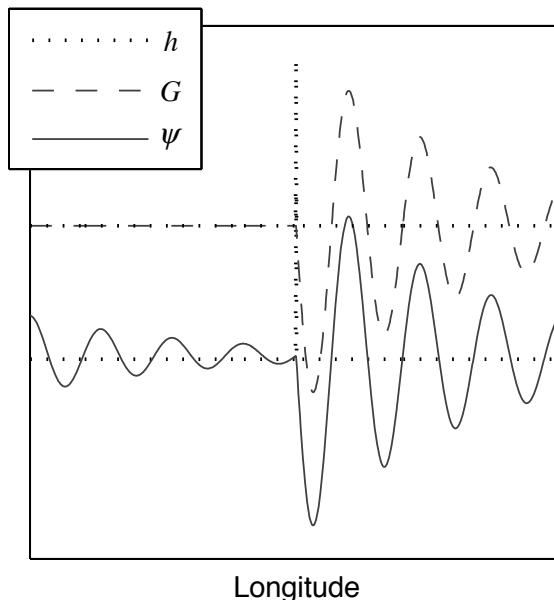


Fig. 13.4 A one-dimensional Rossby wave train excited by uniform eastward flow over a δ -function mountain ridge (h) in the centre of the domain. The upper curve, G , shows the Green's function (13.17), whereas the lower curve shows the exact (linear) response, ψ , in a re-entrant channel calculated numerically using the Fourier method. The two solutions are both centred around zero and offset for clarity; the only noticeable difference is upstream of the ridge, where there is a finite response in the Fourier case because of the progression of the wavetrain around the channel. The stationary wavenumber is 7.5.

and, because $c_g^x > 0$, the response is east of the source. The approximate solution for the streamfunction (denoted ψ_δ) of (13.10) in an infinite channel, with the topography being a δ -function mountain ridge at $x = x'$, and with all fields varying meridionally like $\sin ly$, is thus

$$\psi_\delta(x - x', y) \sim \begin{cases} 0 & x \leq x' \\ -\frac{1}{k_s} \sin ly \sin[k_s(x - x')] \exp[-r(x - x')/c_g^x] & x \geq x'. \end{cases} \quad (13.17)$$

In the more general problem in which the topography is a general function of space, every location constitutes a separate source of wavetrains, and the complete (approximate) solution is given by the integral

$$\psi'(x, y) = \int_{-\infty}^{\infty} \hat{h}(x') \psi_\delta(x - x', y) dx'. \quad (13.18)$$

The field $\psi_\delta(x - x', y)$, is known as ‘Green’s function’ for the problem, sometimes denoted $G(x - x', y)$.

Example solutions calculated using both the Fourier and Green’s function methods are

illustrated in Fig. 13.4. As in Fig. 13.1 there is a trough immediately downstream of the mountain, a result that holds for a broad range of parameters. In these solutions, the streamfunction decays almost completely in one circumnavigation of the channel, and thus, downstream of the mountain, both methods give virtually identical results. Such a correspondence will not hold if the wave can circumnavigate the globe with little attenuation, for then resonance will occur and the Greens-function method will be inaccurate; thus, whether the resonant picture or the wavetrain picture is more appropriate depends largely on the frictional parameter. A frictional time scale of about 10 days is often considered to approximately represent the Earth's atmosphere, in which case waves are only slightly damped on a global circumnavigation, and the Fourier picture is natural with the possibility of resonance. However, the smaller (more frictional) value of 5 days seems to give quantitatively better results in the barotropic problem, and the solution is more evocative of wavetrains. The larger friction may perform better because it is crudely parameterizing the meridional propagation and dispersion of Rossby waves that is neglected in the one-dimensional model.³

13.1.4 The adequacy of linear theory

Having calculated some solutions, we are in a position to estimate, *post facto*, the adequacy of the linear theory by calculating the magnitude of the omitted nonlinear terms. The linear problem here differs in kind from that which arises when using linear theory to evaluate the stability of a flow, as in chapter 6. In that case, we assume a small initial perturbation and the initial evolution of that perturbation is then accurately described, by construction, using linear equations. However, the amplitude of the perturbation is arbitrary, for it may grow exponentially and its size at any given time is proportional to the magnitude of the initial perturbation, which is assumed small but which is otherwise unconstrained. In contrast, when we are calculating the stationary linear response to flow over topography or to a thermal source, the amplitude of the solution is *not* arbitrary; rather, it is determined by the parameters of the problem, including the size of the topography, and represents a real quantity that might be compared to observations. Of course, because the problem is linear, the amplitude of the solution is directly proportional to the magnitude of the topography or thermal perturbation.

From (13.11), and recalling that the amplitude of \tilde{h} is scaled relative to the real topography by the factor $(g/L_d^2 f_0)$, we crudely estimate the amplitude of the response to topography to be

$$|\psi'| \sim \frac{\alpha g h_b}{f_0} \approx \alpha \times 10^8 \text{ m}^2 \text{ s}^{-1} = 2 \times 10^7 \text{ m}^2 \text{ s}^{-1}, \quad |\eta'| = \frac{f_0 |\psi'|}{g} \sim \alpha h_b \approx 0.2 \text{ km.} \quad (13.19)$$

where the non-dimensional parameter α accounts for the distance of the response from resonance and the ratio of the length scale to the deformation scale. Choosing $\alpha = 0.2$ and $h_b = 1 \text{ km}$ gives the numerical values above, which are similar to those calculated more carefully, or observed (Fig. 13.2).

If linear theory is to be accurate, we must demand that the self advection of the response is much smaller than the advection by the basic state, and so that

$$|J(\psi', \nabla^2 \psi')| \ll |\bar{u} \frac{\partial}{\partial x} \nabla^2 \psi'|, \quad (13.20)$$

or, again rather crudely, that $|\psi'/L| \ll \bar{u}$. For $L = 5000$ km we have $\psi'/L = 4 \text{ m s}^{-1}$, which is a few times smaller than a typical mid-troposphere zonal flow of 20 m s^{-1} , suggesting that the linear approximation may hold water. However, the inequality is by no means so well satisfied that we can state without equivocation that the linear approximation is a good one, especially as a different choice of numerical factors would give a different answer, and the use of a simple barotropic model also implies inaccuracies. Rather, we conclude that we must carefully calculate the linear response, and both compare it with the observations and calculate the implied nonlinear terms, before concluding that linear theory is appropriate. In fact linear theory seems to do quite well, and certainly gives qualitative insight into the nature of the mean mid-latitude zonal asymmetries.

13.2 * MERIDIONAL PROPAGATION AND DISPERSION

Rossby waves, of course, propagate meridionally as well as zonally. Furthermore, one of the major mountain ranges on the Earth — the Himalayas — is fairly localized in the meridional direction, and even though the Rockies and Andes do form a convenient meridional ridge, the Rossby waves they generate will still propagate both zonally and meridionally. Furthermore, the coefficients of the linear equations of motion vary with space: on the sphere β is a function of latitude and in general topography is a function of both latitude and longitude. Given this complexity, we cannot solve the full problem except numerically, but a few ideas from wave tracing illustrate many of the features of the response, and indeed of the stationary wave pattern in the Earth's atmosphere.⁴

13.2.1 Ray tracing

We first, informally and without proof, review some results about rays and ray tracing. The appendix on page 236 and section 7.2.2 both contain material on group velocity, and the reader may wish to review them before proceeding.⁵

Most of the important properties of a wave, such as the energy (if conserved) and the wave activity, propagate at the group velocity, and rays are lines that are parallel to the group velocity, generally emanating from some wave source. A ray is perpendicular to the local wave front, and in a homogeneous medium a wave propagates in a straight line. In non-homogeneous media the group velocity varies with position; however, if the medium varies only slowly, on a scale much larger than that of the wavelength of the waves, the wave activity still propagates along rays at the group velocity.

Let us represent a wave by

$$\psi(x_i, t) = \Psi(x_i, t) e^{i\theta(x_i, t)} \quad (13.21)$$

where the amplitude, Ψ , varies more slowly than the phase, θ . [We use subscripts $(i, j, \text{etc.})$ to denote Cartesian axes, repeated subscripts are to be summed over, and (x_i, t) means (x, y, z, t) .] Locally, the frequency, ω , and wavenumber, k_i satisfy

$$k_i = \frac{\partial \theta}{\partial x_i}, \quad \omega = -\frac{\partial \theta}{\partial t}, \quad (13.22a,b)$$

and these imply

$$\frac{\partial k_i}{\partial t} = -\frac{\partial \omega}{\partial x_i}. \quad (13.23)$$

The frequency is, in general, a function of wavenumber, position and time, with the time dependence arising if the medium varies in time. A relation of the form $\omega = \omega(k_i, x_i, t)$ constitutes the dispersion relation, and the local group velocity is given by $c_{gi} = \partial \omega / \partial k_i$. Assuming no explicit time dependence in the dispersion relation, and using (13.22), we can write the frequency as

$$\omega = \omega(k_i, x_i) = \omega(\partial \theta / \partial x_i, x_i). \quad (13.24)$$

Then, using (13.24), (13.23) becomes

$$\frac{\partial k_i}{\partial t} = - \left(\frac{\partial \omega}{\partial x_i} \right)_{k_i} - \left(\frac{\partial \omega}{\partial k_j} \right)_{x_i} \frac{\partial k_j}{\partial x_i} = - \left(\frac{\partial \omega}{\partial x_i} \right)_{k_i} - \left(\frac{\partial \omega}{\partial k_j} \right)_{x_i} \frac{\partial}{\partial x_i} \frac{\partial \theta}{\partial x_j}, \quad (13.25)$$

or, using the definition of the group velocity,

$$\frac{\partial k_i}{\partial t} + c_{gj} \frac{\partial k_i}{\partial x_j} = - \left(\frac{\partial \omega}{\partial x_i} \right)_{k_i}. \quad (13.26)$$

The left-hand side is the change in wavenumber along a ray. If the frequency is constant in space the right-hand side vanishes and the wavenumber is simply propagated at the group velocity. If the frequency is independent of a particular coordinate then the corresponding wavenumber is constant along the ray. If the frequency changes with position (as in general it will), then the wavenumber will change along a ray, and thus so will the direction of propagation — the wave is *refracted*. Note that we can write (13.26), and the definition of group velocity, in the compact forms

$$\frac{D_{c_g} k_i}{Dt} = - \frac{\partial \omega}{\partial x_i}, \quad \frac{D_{c_g} x_i}{Dt} = \frac{\partial \omega}{\partial k_i}, \quad (13.27a,b)$$

where $D_{c_g} / Dt \equiv \partial / \partial t + (\mathbf{c}_g \cdot \nabla)$.

One other consequence of these manipulations is that the frequency is constant along a ray. This follows by noting that frequency is a function both of space and wavenumber so that

$$\frac{D_{c_g} \omega}{Dt} = \left(\frac{\partial \omega}{\partial k_i} \right) \frac{D_{c_g} k_i}{Dt} + \left(\frac{\partial \omega}{\partial x_i} \right) \frac{D_{c_g} x_i}{Dt} = 0, \quad (13.28)$$

where the second equality follows by use of (13.27). One practical result of all this that in problems of the form

$$\frac{\partial}{\partial t} \nabla^2 \psi + \beta(y) \frac{\partial \psi}{\partial x} = 0, \quad (13.29)$$

both the frequency and the x -wavenumber are constant along a ray.

13.2.2 Rossby waves and Rossby rays

If the topography is localized, then ray theory provides a useful way of calculating and interpreting the response to a flow over that topography. On the β -plane and away from the orographic source the steady linear response to a zonally uniform but meridionally varying zonal wind will obey

$$\bar{u}(y) \frac{\partial}{\partial x} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \psi' + \beta \frac{\partial \psi'}{\partial x} = 0. \quad (13.30)$$

In fact, an equation of this form applies on the sphere. To see this, we transform the spherical coordinates (λ, ϑ) into Mercator coordinates with the mapping⁶

$$x = a\lambda, \quad \frac{1}{a} \frac{\partial}{\partial \lambda} = \frac{\partial}{\partial x}, \quad y = \frac{a}{2} \ln \left(\frac{1 + \sin \vartheta}{1 - \sin \vartheta} \right), \quad \frac{1}{a} \frac{\partial}{\partial \vartheta} = \frac{1}{\cos \vartheta} \frac{\partial}{\partial y}. \quad (13.31)$$

The spherical coordinate vorticity equation then becomes

$$\bar{u}_M \frac{\partial}{\partial x} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \psi' + \beta_M \frac{\partial \psi'}{\partial x} = 0, \quad (13.32)$$

where $\bar{u}_M = \bar{u} / \cos \vartheta$ and

$$\beta_M = \frac{2\Omega}{a} \cos^2 \vartheta - \frac{d}{dy} \left[\frac{1}{\cos^2 \vartheta} \frac{d}{dy} (\bar{u}_M \cos^2 \vartheta) \right] = \cos \vartheta \left(\beta_s + \frac{1}{a} \frac{\partial \bar{\zeta}}{\partial \vartheta} \right), \quad (13.33)$$

where $\beta_s = 2a^{-1}\Omega \cos \vartheta$. Thus, β_M is the meridional gradient of the absolute vorticity, multiplied by the cosine of latitude. An advantage of Mercator coordinates over their spherical counterparts is that (13.32) has a Cartesian flavour to it, with the metric coefficients being absorbed into the parameters \bar{u}_M and β_M . Of course, unlike the case on the true β -plane, the parameter β_M is not a constant, but this is not a particular disadvantage if \bar{u}_{yy} is also varying with y .

Having noted the spherical relevance we revert to the β -plane and seek solutions of (13.30) with the form $\psi' = \tilde{\psi}(y) \exp(ikx)$, whence

$$\frac{d^2 \tilde{\psi}}{dy^2} = \left(k^2 - \frac{\beta}{\bar{u}} \right) \tilde{\psi} = \left(k^2 - K_s^2 \right) \tilde{\psi}, \quad (13.34)$$

where $K_s = (\beta/\bar{u})^{1/2}$. From this equation it is apparent that if $k < K_s$ the solution is harmonic in y and Rossby waves may propagate away from their source. On the other hand, wavenumbers $k > K_s$ are trapped near their source; that is, short waves are meridionally trapped by eastward flow.

Without solving (13.34), we can expect an isolated mountain to produce two wavetrains, one for each meridional wavenumber $l = \pm(K_s^2 - k^2)^{1/2}$. These wavetrains will then propagate along a ray, and given the dispersion relation this trajectory can be calculated (usually numerically) using the expressions of the previous section. The local dispersion relation of Rossby waves is

$$\omega = \bar{u}k - \frac{\beta k}{k^2 + l^2}, \quad (13.35)$$

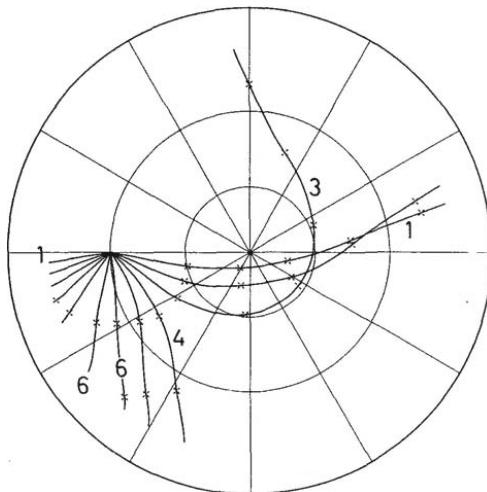
so that their group velocity is

$$c_g^x = \frac{\partial \omega}{\partial k} = \bar{u} - \frac{\beta(l^2 - k^2)}{(k^2 + l^2)^2} = \frac{\omega}{k} + \frac{2\beta k^2}{(k^2 + l^2)^2}, \quad (13.36a)$$

$$c_g^y = \frac{\partial \omega}{\partial l} = \frac{2\beta k l}{(k^2 + l^2)^2}. \quad (13.36b)$$

The sign of the meridional wavenumber thus determines whether the waves propagate polewards (positive l) or equatorwards (negative l). Also, because the dispersion relation (13.36) is independent of x and t , the zonal wavenumber and frequency in the wave group

Fig. 13.5 The rays emanating from a point source at 30°N and 180° (nine o'clock), calculated using the observed value of the wind at 300 mb.⁷ The crosses mark every 180° of phase, and mark the positions of successive positive and negative extrema. The numbers indicate the zonal wavenumber of the ray. The ray paths may be compared with the full linear calculation shown in Fig. 13.6.



are constant along the ray, and the meridional wavenumber then adjusts to satisfy the local dispersion relation (13.35). Thus, from (13.34), the meridional scale becomes larger as K_s approaches k from above and an incident wavetrain is reflected, its meridional wavenumber changes sign, and it continues to propagate eastwards.

Stationary waves have $\omega = 0$, and the trajectory of a ray is parameterized by

$$\frac{dy}{dx} = \frac{c_g^y}{c_g^x} = \frac{l}{k}. \quad (13.37)$$

For a given zonal wavenumber the trajectory is then fully determined by this condition and that for the local meridional wavenumber which from (13.35) is

$$l^2 = K_s^2 - k^2. \quad (13.38)$$

Finally, from (13.36) the magnitude of the group velocity is

$$|c_g| = [(c_g^x)^2 + (c_g^y)^2]^{1/2} = 2 \frac{k}{K_s} \bar{u}, \quad (13.39)$$

which is double the speed of the projection of the basic flow, \bar{u} , onto the wave direction. Given the above relations, and the zonal wind field, we can compute rays emanating from a given source, although the calculation must still be done numerically. One example is given in Fig. 13.5.

* *A JWKB solution*

Some information about the wave amplitudes along a ray can be obtained using a JWKB approach.⁸ Let us write (13.34) as

$$\frac{d^2 \tilde{\psi}}{dy^2} + l^2 \tilde{\psi} = 0 \quad (13.40)$$

where

$$l^2(y) = K_s^2 - k^2. \quad (13.41)$$

If $l(y)$ is sufficiently slowly varying in y (i.e., if $|dl^{-1}/dy| \ll 1$) then we may seek a solution of the form

$$\tilde{\psi} = A e^{ig(y)}. \quad (13.42)$$

This leads to an approximate solution for $g(y)$, namely

$$g(y) = \int^y l(y) dy + \frac{i}{2} \ln l(y), \quad (13.43)$$

and the approximate solution for the stationary streamfunction is then

$$\psi(x, y) = Al^{-1/2} \exp \left[i \left(kx + \int^y l(y) dy \right) \right], \quad (13.44)$$

where A is a constant. Consider, for example, the disturbance excited by an isolated low-latitude peak, with \bar{u} increasing, and so K_s decreasing, polewards of the source. Assuming that initially there exists a zonal wavenumber k less than K_s then two eastward propagating wavetrains are excited. The meridional wavenumber of the poleward wavetrain diminishes according to (13.41), so that, using (13.37), the ray becomes more zonal. At the latitude where $k = K_s$, the ‘turning latitude’ the wave is reflected but continues propagating eastwards. The southward propagating wavetrain is propagating into a medium with smaller \bar{u} and larger K_s . At the critical latitude, where $\bar{u} = 0$, $l \rightarrow \infty$ but c_g^x and c_g^y both tend to zero, but [using (13.36)] in such a way that $c_g^x/c_g^y \rightarrow 0$. That is, the rays become meridionally oriented and their speed tends to zero. At this latitude the waves may be absorbed, but the analysis is specialized and beyond our scope.⁹ Finally, we mention without proof that for zonal flows with constant angular velocity the trajectories are great circles.

13.2.3 Application to an idealized atmosphere

Given the complexity of the real atmosphere, and the availability of computers, it is probably best to think of the remarks above as helping us interpret more complete numerical, but still linear, calculations of stationary Rossby waves — for example, solutions of the stationary barotropic vorticity equation in spherical coordinates,

$$\frac{\bar{u}}{a \cos \vartheta} \frac{\partial \zeta'}{\partial \lambda} + v' \left(\frac{1}{a} \frac{\partial \bar{\zeta}}{\partial \vartheta} + \beta \right) = - \frac{\bar{u} f_0}{a H \cos \vartheta} \frac{\partial h_b}{\partial \lambda} - r \zeta', \quad (13.45)$$

where $[u, v] = a^{-1}[-\partial \psi / \partial \vartheta, (\partial \psi / \partial \lambda) / \cos \vartheta]$, $\beta = 2\Omega a^{-1} \cos \vartheta$ and $\zeta = \nabla^2 \psi$. The last term in (13.45) crudely represents the effects of friction and generally reduces the sensitivity of the solutions to resonances. Solutions to (13.45) may be obtained first by discretizing and then numerically inverting a matrix, and although the actual procedure is quite involved it is analogous to the Fourier methods used earlier for the simpler one-dimensional problem. Such linear calculations, in turn, help us interpret the stationary wave pattern from still more complete models and in the Earth’s atmosphere.

Figure 13.6 shows the stationary solution to the problem with a realistic northern hemisphere zonal flow and an isolated, circular mountain at 30° N. The topography excites two wavetrains, both of which slowly decay downstream because of frictional effects, rather like the one-dimensional wavetrain in Fig. 13.4. The polewards propagating wavetrain develops a more meridional orientation, corresponding to a smaller meridional wavenumber

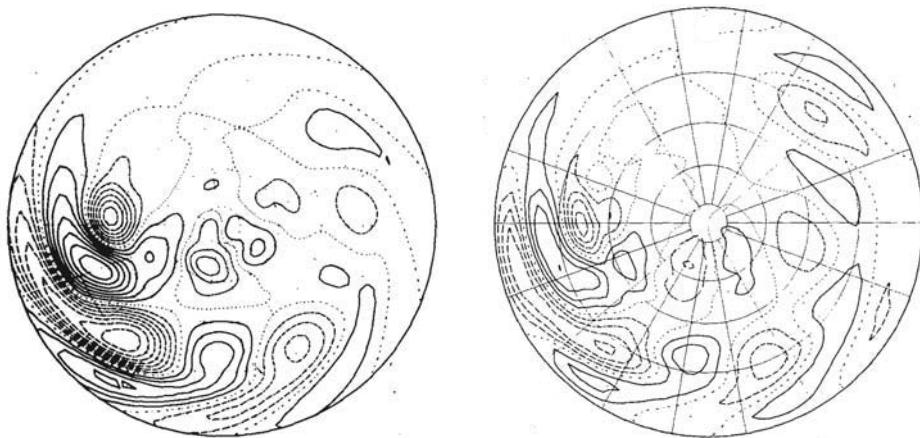


Fig. 13.6 The linear stationary response induced by circular mountain at 30°N and at 180° longitude (nine o'clock). The figure on the left uses a barotropic model, whereas the figure on the right uses a multi-layer baroclinic model.¹⁰ In both cases the mountain excites a low-wavenumber polar wavetrain and a higher-wavenumber subtropical train.

l , before moving southwards again, developing a much more zonal orientation eventually to decay completely as it meets the equatorial westward flow. The equatorially propagating train decays a little more rapidly than its polewards moving counterpart because of its proximity to the critical latitude. More complicated patterns naturally result if a realistic distribution of topography is used, as we see later in Fig. 13.10. We can see wavetrains emanating from both the Rockies and the Himalayas, but distinct poleward and equatorward wavetrains are hard to discern.

13.3 * VERTICAL PROPAGATION OF ROSSBY WAVES IN A STRATIFIED MEDIUM

We now consider the propagation and stationary wave structure of Rossby waves in a stratified atmosphere, focusing on their vertical propagation.¹¹ The vertical propagation is important not just because it must be taken into account to obtain an accurate picture of the tropospheric response to topographic and thermal forcing, but because it can excite motion in the stratosphere, as considered in section 13.5. We will continue to use the stratified quasi-geostrophic equations, but we now allow the model to be compressible and semi-infinite, extending from $z = 0$ to $z = \infty$.

13.3.1 Model formulation

The potential vorticity equation describes motion in the fluid interior, with a surface boundary condition of vertical velocity being determined by the thermodynamic equation, and the upper boundary condition being determined by a radiation condition. Guided by the barotropic problem, we will allow for the possibility of Ekman friction and topography at the surface, but otherwise the flow is presumed inviscid and adiabatic. Our vertical coordinate,

z , may be regarded as being either height or log-pressure, with a slight difference in the lower boundary condition.¹² The potential vorticity equation is

$$\frac{\partial q}{\partial t} + J(\psi, q) = 0, \quad q = \nabla^2 \psi + f + \frac{f_0^2}{\rho_R} \frac{\partial}{\partial z} \left(\frac{\rho_R}{N^2} \frac{\partial \psi}{\partial z} \right), \quad (13.46)$$

where we take $\rho_R = \rho_0 e^{-z/H}$ with H being a specified density scale height, typically $RT(0)/g$. We linearize this equation about a zonal wind that depends only on z ; that is, we let

$$\psi = -\bar{u}(z)\gamma + \psi', \quad (13.47)$$

and obtain

$$\frac{\partial q'}{\partial t} + \bar{u} \frac{\partial q'}{\partial x} + \nu' \frac{\partial \bar{q}}{\partial y} = 0, \quad \frac{\partial \bar{q}}{\partial y} = \beta - \frac{f_0^2}{\rho_R} \frac{\partial}{\partial z} \left(\frac{\rho_R}{N^2} \frac{\partial \bar{u}}{\partial z} \right), \quad (13.48)$$

or equivalently, in terms of streamfunction,

$$\begin{aligned} \left(\frac{\partial}{\partial t} + \bar{u} \frac{\partial}{\partial x} \right) \left[\nabla^2 \psi' + \frac{f_0^2}{\rho_R} \frac{\partial}{\partial z} \left(\frac{\rho_R}{N^2} \frac{\partial \psi'}{\partial z} \right) \right] \\ + \frac{\partial \psi'}{\partial x} \left[\beta - \frac{f_0^2}{\rho_R} \frac{\partial}{\partial z} \left(\frac{\rho_R}{N^2} \frac{\partial \bar{u}}{\partial z} \right) \right] = 0. \end{aligned} \quad (13.49)$$

The lower boundary is obtained using the thermodynamic equation,

$$\frac{\partial}{\partial t} \left(\frac{\partial \psi}{\partial z} \right) + J \left(\psi, \frac{\partial \psi}{\partial z} \right) + \frac{N^2}{f_0} w = 0, \quad (13.50)$$

along with an equation for the vertical velocity, w , at the lower boundary. This is

$$w = \mathbf{u} \cdot \nabla h_b + r\zeta, \quad (13.51)$$

where the two terms respectively represent the kinematic contribution to vertical velocity due to flow over topography and the contribution from Ekman pumping, with r a constant, and the effects are taken to be additive. Linearizing the thermodynamic equation about the zonal flow and using (13.51) gives

$$\frac{\partial}{\partial t} \left(\frac{\partial \psi'}{\partial z} \right) + \bar{u} \frac{\partial}{\partial x} \frac{\partial \psi'}{\partial z} - \nu' \frac{\partial \bar{u}}{\partial z} = -\frac{N^2}{f_0} \left(\bar{u} \frac{\partial h_b}{\partial x} + r \nabla^2 \psi' \right), \quad \text{at } z = 0. \quad (13.52)$$

13.3.2 Model solution

We look for solutions of (13.48) and (13.52) in the form

$$\psi' = \text{Re } \tilde{\psi}(z) \sin ly e^{ik(x-ct)} \quad \text{with} \quad h_b = \text{Re } \tilde{h}_b \sin ly e^{ikx} \quad (13.53)$$

Solutions must then satisfy

$$\left[\frac{f_0^2}{\rho_R} \frac{\partial}{\partial z} \left(\frac{\rho_R}{N^2} \frac{\partial \tilde{\psi}}{\partial z} \right) \right] = \tilde{\psi} \left(K^2 - \frac{\partial \bar{q}/\partial y}{\bar{u} - c} \right) \quad (13.54)$$

in the interior, and the boundary condition

$$(\bar{u} - c) \frac{\partial \tilde{\psi}}{\partial z} - \tilde{\psi} \frac{\partial \bar{u}}{\partial z} + \frac{irN^2K^2}{kf_0} \tilde{\psi} = -\frac{N^2 \bar{u} \tilde{h}_b}{f_0}, \quad \text{at } z = 0, \quad (13.55)$$

as well as a radiation condition at plus infinity (and we must have that $\rho_0 \tilde{\psi}^2$ be finite). Let us simplify by considering the case of constant \bar{u} and N^2 and with $r = 0$. We also let

$$\Phi(z) = \tilde{\psi}(z) \left(\frac{\rho_R}{\rho_R(0)} \right)^{1/2} = \tilde{\psi}(z) e^{-z/2H} \quad (13.56)$$

and obtain the interior equation

$$\frac{d^2\Phi}{dz^2} + m^2\Phi = 0, \quad \text{where } m^2 = \frac{N^2}{f_0^2} \left(\frac{\beta}{\bar{u} - c} - K^2 - \gamma^2 \right), \quad (13.57a,b)$$

and $\gamma^2 = f_0^2/(4N^2H^2) = 1/(2L_d)^2$, where L_d is the deformation radius. The surface boundary condition is

$$(\bar{u} - c) \left(\frac{d\Phi}{dz} + \frac{\Phi}{2H} \right) = -\frac{N^2 \bar{u} \tilde{h}_b}{f_0}, \quad \text{at } z = 0. \quad (13.58)$$

Oscillating waves

From (13.57b) we obtain the dispersion relation for Rossby waves, namely

$$\omega = \bar{u}k - \frac{\beta k}{K^2 + \gamma^2 + m^2 f_0^2 / N^2}. \quad (13.59)$$

The three components of the group velocity for these waves are:

$$c_g^x = \bar{u} + \frac{\beta [k^2 - (l^2 + m^2 f_0^2 / N^2 + \gamma^2)]}{(K^2 + m^2 f_0^2 / N^2 + \gamma^2)^2}, \quad (13.60a)$$

$$c_g^y = \frac{2\beta kl}{(K^2 + m^2 f_0^2 / N^2 + \gamma^2)^2}, \quad (13.60b)$$

$$c_g^z = \frac{2\beta km f_0^2 / N^2}{(K^2 + m^2 f_0^2 / N^2 + \gamma^2)^2}. \quad (13.60c)$$

The propagation in the horizontal is analogous to the propagation in a shallow water model [c.f. (5.186b)]; note also that higher baroclinic modes (bigger m) will have a more westward group velocity. The vertical group velocity is proportional to m , and therefore for waves that are excited at the surface we must choose m to be positive, for positive k .

Stationary waves

Stationary waves have $\omega = ck = 0$. In this case (13.57) has a solution $\Phi = \Phi_0 \exp(imz)$ provided m^2 is positive where

$$m = \pm \frac{N}{f_0} \left(\frac{\beta}{\bar{u}} - K^2 - \gamma^2 \right)^{1/2}. \quad (13.61)$$

Furthermore, m itself must be positive. For non-steady waves we must choose the sign of m to ensure that the group velocity, and hence the wave activity, is directed away from the

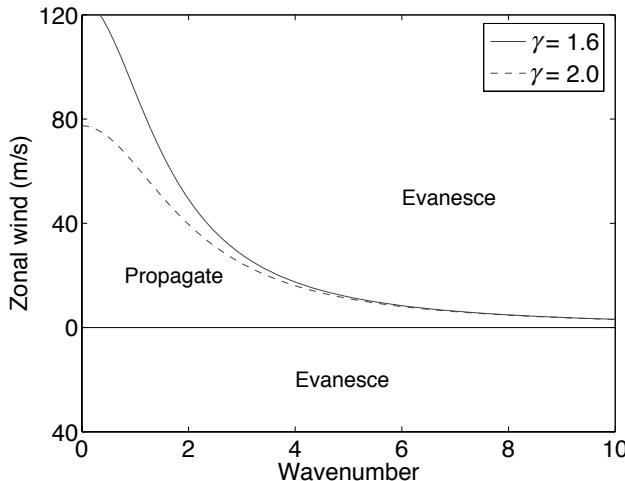


Fig. 13.7 The boundary between propagating waves and evanescent waves as a function of zonal wind and wavenumber, using (13.62), for a couple of values of γ . With $N = 2 \times 10^{-2} \text{ s}^{-1}$, $\gamma = 1.6$ ($\gamma = 2$) corresponds to a scale height of 7.0 km (5.5 km) and a deformation radius NH/f of 1400 km (1100 km).

energy source. This must still hold as $m \rightarrow 0$, and therefore the positive sign in (13.61) corresponds to the physically realizable solution.¹³

The condition $m^2 > 0$ holds if

$$0 < \bar{u} < \frac{\beta}{K^2 + \gamma^2}, \quad (13.62)$$

and this is illustrated in Fig. 13.7. Stationary, vertically oscillatory modes can exist only for zonal flows that are eastwards and that are less than the critical velocity $U_c = \beta/(K^2 + \gamma^2)$. To interpret this condition, note that in a resting medium the Rossby wave frequency has a minimum value (and maximum absolute value), when $m = 0$, of

$$\omega = -\frac{\beta k}{K^2 + \gamma^2}. \quad (13.63)$$

Note too that in a frame moving with speed \bar{u} our Rossby waves (stationary in the Earth's frame) have frequency $-\bar{u}k$, and this is the forcing frequency arising from the now-moving bottom topography. Thus, (13.62) is equivalent to saying that for oscillatory waves to exist *the forcing frequency must lie within the frequency range of vertically propagating Rossby waves*.

For westward flow, or for sufficiently strong eastward flow, the waves decay exponentially as $\Phi = \Phi_0 \exp(-\alpha z)$ where

$$\alpha = \frac{N}{f_0} \left(K^2 + \gamma^2 - \frac{\beta}{\bar{u}} \right)^{1/2}. \quad (13.64)$$

Note that the critical velocity U_c is a function of wavenumber, and that it increases with horizontal wavelength. Thus, for a given eastward flow long waves may penetrate vertically when short waves are trapped, an effect sometimes referred to as 'Charney-Drazin filtering'. One important consequence of this is that the stratospheric motion is typically of larger scales than that of the troposphere, because waves tend to be excited first in the troposphere (by baroclinic instability and by flow over topography, among other things), but the shorter

Stationary, Topographically Forced Solutions

Collecting the results in section 13.3.2, the stationary solutions of (13.48) and (13.52) are:

$$\psi'(x, y, z) = \operatorname{Re} e^{imz} e^{z/2H} e^{ikx} \sin ly \frac{f_0 \tilde{h}_b [im - (2H)^{-1}]}{K_s^2 - K^2}, \quad m^2 > 0 \quad (\text{T.1a})$$

$$\psi'(x, y, z) = \operatorname{Re} e^{[(2H)^{-1} - \alpha]z} e^{ikx} \sin ly \frac{N^2 \tilde{h}_b}{f_0 [\alpha - (2H)^{-1}]}, \quad m^2 < 0 \quad (\text{T.1b})$$

where

$$m = + \frac{N}{f_0} \left(\frac{\beta}{\bar{u}} - K^2 - y^2 \right)^{1/2} \quad (\text{T.2})$$

and

$$\alpha = + \frac{N}{f_0} \left(K^2 + y^2 - \frac{\beta}{\bar{u}} \right)^{1/2}, \quad (\text{T.3})$$

and $y = f_0/(2NH)$. If $m^2 > 0$ the solutions are propagating, or radiating, waves in the vertical direction. If $m^2 < 0$ the energy of the solution, $|\rho_R \psi'^2|$, is vertically evanescent. The condition $m^2 > 0$ is equivalent to

$$0 < \bar{u} < \frac{\beta}{K^2 + (f_0/2NH)^2}, \quad (\text{T.4})$$

so that vertical penetration is favoured when the winds are weakly eastwards, and the range of \bar{u} -values that allows this is larger for longer waves.

waves are trapped and only the longer ones reach the stratosphere. In the summer, the stratospheric winds are often westwards and all waves are trapped in the troposphere; the eastward stratospheric winds that favour vertical penetration occur in the other three seasons, although very strong eastward winds can suppress penetration in mid-winter.

Finally, the surface boundary condition, (13.58) gives

$$\Phi_0 = \frac{N^2 \tilde{h}_b / f_0}{(\alpha, -im) - (2H)^{-1}}, \quad (\text{T.65})$$

where $(\alpha, -im)$ refers to the (trapped, oscillatory) case. Equation (13.65) indicates that resonance is possible when $\alpha = 1/(2H)$, and from (13.64) this occurs when $K^2 = \beta/\bar{u}$, that is when barotropic Rossby waves are stationary. This wave resonates because the wave is a solution of the unforced (and inviscid) equations and, because $\tilde{\psi} = \Phi \exp[z/(2H)]$, $\tilde{\psi} = 1$ and has uniform vertical structure. If $K > K_s$ then $\alpha > 1/(2H)$ and the forced wave (i.e., the amplitude of ψ) decays with height with no phase variation. If $\alpha < 1/(2H)$ then $\tilde{\psi}$ increases with height (although $\rho_R |\tilde{\psi}|^2$ decreases with height), and this occurs when $(K_s^2 - y^2)^{1/2} < K < K_s$. If $(K_s^2 - y^2)^{1/2} > K$ then the amplitude of ϕ , (i.e., $\rho_R |\tilde{\psi}|^2$) is independent of height; their vertical structure is oscillatory, like $\exp(imz)$. The complete solutions are collected for convenience in the box above.

13.3.3 Properties of the solution

The various dynamical fields associated with the solution can all be easily constructed from (T.1), and a few simple properties of the solution are worth noting explicitly. In some cases the explicit calculation is left as a problem to the reader — see problems 13.6 and 13.7.

Amplitudes and phases. The decaying solutions have no vertical phase variations (a property known as ‘equivalent barotropic’) and the streamfunction is exactly in phase or out of phase with the topography according as $K > K_s$ and $\alpha > (2H)^{-1}$, or $K < K_s$ and $\alpha < (2H)^{-1}$. In the latter case the amplitude of the streamfunction actually increases with height, but the energy, proportional to $\rho_R |\psi'|^2$ falls. The oscillatory solutions have constant energy with height but a shifting phase. The phase of the streamfunction at the surface may be in or out of phase with the topography, depending on m , but the potential temperature, $\partial\psi/\partial z$ is always out of phase with the topography. That is, positive values of h_b are associated with cool fluid parcels.

Vertical energy propagation. As noted, the energy propagates upwards for the oscillatory waves. This may be verified by calculating $\overline{p'w'}$ (the vertical component of the energy flux), where p' is the pressure perturbation, proportional to ψ' , and w' is the vertical velocity perturbation. To this end, linearize the thermodynamic equation (13.50) to give

$$\frac{\partial}{\partial t} \left(\frac{\partial\psi'}{\partial z} \right) + \overline{u} \frac{\partial}{\partial x} \frac{\partial\psi'}{\partial z} - \frac{\partial \overline{u}}{\partial z} \frac{\partial\psi'}{\partial x} + \frac{N^2}{f_0} w' = 0. \quad (13.66)$$

Then, multiplying by ψ' and integrating by parts gives a balance between the second and fourth terms,

$$N^2 \overline{\psi' w'} = \overline{u b' v'}, \quad (13.67)$$

where $b' = f_0 \partial\psi'/\partial z$ and $v' = \partial\psi'/\partial x$. Thus, the upward transfer of energy is proportional to the poleward heat flux.

Meridional heat transport. The meridional heat transport associated with a wave is

$$\rho_R \overline{v' b'} = \rho_R f_0 \frac{\partial\psi'}{\partial x} \frac{\partial\psi'}{\partial z}. \quad (13.68)$$

For an oscillatory wave this can readily be shown to be positive. In particular, it is proportional to $km/(K_s^2 - K^2)$, and this is positive because $km > 0$ is the condition that energy is directed upwards, and $K_s^2 > K^2$ for oscillatory solutions. The meridional transport associated with a trapped solution is identically zero.

Form drag. If the waves propagate energy upwards, there must be a surface interaction to supply that energy. There is a force due to *form drag* associated with this interaction, given by

$$\text{form drag} = p' \frac{\partial h_b}{\partial x} \quad (13.69)$$

(see chapter 3). In the trapped case, the streamfunction is either exactly in or out of phase with the topography, so this interaction is zero. In the oscillatory case

$$\overline{\psi' \frac{\partial h_b}{\partial x}} = \frac{f_0 \tilde{h}_b^2 km}{4(K_s^2 - K^2)}, \quad (13.70)$$

where the factor of 4 arises from the x and y averages of the squares of sines and cosines. The rate of doing work is \overline{u} times (13.70).

13.4 * EFFECTS OF THERMAL FORCING

How does thermal forcing influence the stationary waves? To give an accurate answer for the real atmosphere is a little more difficult than for the orographic case where the forcing can be included reasonably accurately in a quasi-geostrophic model with a term $\bar{u} \cdot \nabla h_b$ at the lower boundary. Anomalous (i.e., variations from a zonal or temporal mean) thermodynamic forcing typically also arises initially at the lower boundary through, for example, variations in the surface temperature. However, such anomalies may be felt throughout the lower troposphere on a relatively short time scale by way of such non-geostrophic phenomena as convection, so that the effective thermodynamic source that should be applied in a quasi-geostrophic calculation has a finite vertical extent. However, an accurate parameterization of this may depend on the structure of the atmospheric boundary layer and this cannot always be represented in a simple way.¹⁴ Because of such uncertainties our treatment concentrates on the fundamental and qualitative aspects of thermal forcing.

The quasi-geostrophic potential vorticity equation, linearized around a uniform zonal flow, is [cf. (13.49)]

$$\left(\frac{\partial}{\partial t} + \bar{u} \frac{\partial}{\partial x} \right) \left[\nabla^2 \psi' + \frac{f_0^2}{\rho_R} \frac{\partial}{\partial z} \left(\frac{\rho_R}{N^2} \frac{\partial \psi'}{\partial z} \right) \right] + \frac{\partial \psi'}{\partial x} \left[\beta - \frac{f_0^2}{\rho_R} \frac{\partial}{\partial z} \left(\frac{\rho_R}{N^2} \frac{\partial \bar{u}}{\partial z} \right) \right] = \frac{f_0}{N^2} \frac{\partial Q}{\partial z} \equiv T, \quad (13.71)$$

where Q is the source term in the (linear) thermodynamic equation,

$$\frac{\partial}{\partial t} \left(\frac{\partial \psi'}{\partial z} \right) + \bar{u} \frac{\partial}{\partial x} \frac{\partial \psi'}{\partial z} - \nu' \frac{\partial \bar{u}}{\partial z} + \frac{N^2}{f_0} w' = \frac{Q}{f_0}. \quad (13.72)$$

A particular solution to (13.71) may be constructed if \bar{u} and N^2 are constant, and if Q has a simple vertical structure. If we again write $\psi' = \text{Re } \tilde{\psi}(z) \sin ly \exp(ikx)$ and let $\Phi(z) = \tilde{\psi}(z) \exp(-z/2H)$ we obtain

$$\frac{d^2 \Phi}{dz^2} + m^2 \Phi = \frac{T}{ik\bar{u}} e^{-z/2H}, \quad \text{where } m^2 = \frac{N^2}{f_0^2} \left(\frac{\beta}{\bar{u}} - K^2 - y^2 \right). \quad (13.73)$$

If we let $T = T_0 \exp(-z/H_Q)$, so that the heating decays exponentially away from the Earth's surface, then the particular solution to the stationary problem is found to be

$$\tilde{\psi} = \text{Re} \frac{i \hat{T} e^{-z/H_Q}}{k\bar{u} \left[(N/f_0)^2 (K_s^2 - K^2) + H_Q^{-2} (1 + H_Q/H) \right]}, \quad (13.74)$$

where \hat{T} is proportional to T . This solution does not satisfy the boundary condition at $z = 0$, which in the absence of topography and friction is

$$\bar{u} \frac{\partial}{\partial x} \frac{\partial \psi'}{\partial z} - \nu' \frac{\partial \bar{u}}{\partial z} = \frac{Q(0)}{f_0}. \quad (13.75)$$

A homogeneous solution must therefore be added, and just as in the topographic case this leads to a vertically radiating or a surface trapped response, depending on the sign of m^2 . One way to calculate the homogeneous solution is to first use the linearized thermodynamic

equation (13.72), or the linearized vorticity equation (13.77), to calculate the vertical velocity at the surface implied by (13.74), $w_p(0)$ say. We then notice that the homogeneous solution is effectively forced by an equivalent topography given by $h_e = -w_p(0)/(ik\bar{u}(0))$, and so proceed as in the topographic case. The complete solution is rather hard to interpret, and is in any case available only in special cases, so it is useful to take a more qualitative approach.

13.4.1 Thermodynamic balances

It is the properties of the particular solution that distinguish the response to thermodynamic forcing from that due to topography, because the homogeneous solutions of the two cases are similar. And far from the source region, the homogeneous solution will dominate, giving rise to wavetrains as discussed previously.

We can determine many of the properties of the response to thermodynamic forcing by considering the balance of terms in the steady linear thermodynamic equation, which we write as

$$\bar{u} \frac{\partial}{\partial x} \frac{\partial \psi'}{\partial z} - \frac{\partial \psi'}{\partial x} \frac{\partial \bar{u}}{\partial z} + \frac{N^2}{f_0} w' = \frac{Q}{f_0} \equiv R \quad (13.76a)$$

or

$$f_0 \bar{u} \frac{\partial v'}{\partial z} - f_0 v' \frac{\partial \bar{u}}{\partial z} + N^2 w' = Q. \quad (13.76b)$$

The vorticity equation is

$$\bar{u} \frac{\partial \zeta'}{\partial x} + \beta v' = \frac{f_0}{\rho_R} \frac{\partial \rho_R w'}{\partial z}. \quad (13.77)$$

Assuming that the diabatic forcing is significant, we may imagine three possible simple balances in the thermodynamic equation:

- (i) zonal advection dominates, and $v' = \partial \psi' / \partial x \sim QH_Q / (f_0 \bar{u})$;
- (ii) meridional advection dominates, and $v' \sim QH_u / (f_0 \bar{u})$;
- (iii) vertical advection dominates, and $w' \sim Q/N^2$. Then, for large enough horizontal scales the balance in the vorticity equation is $\beta v' \sim f_0 w'_z$ and $v' \sim f_0 Q / (\beta N^2 H_Q)$. For smaller horizontal scales advection of relative vorticity may dominate that of planetary vorticity, and β is replaced by $\bar{u} K^2$.

Here, H_Q is the vertical scale of the source (so that $\partial Q / \partial z \sim Q/H_Q$) and H_u is the vertical scale of the zonal flow (so that $\partial \bar{u} / \partial z \sim \bar{u}/H_u$). We also assume that the vertical scale of the solution is H_Q , so that $\partial v' / \partial z \sim v'/H_Q$. Which of the above three balances is likely to hold? Heuristically, we might suppose that the balance with the smallest v' will dominate, if only because meridional motion is suppressed on the β -plane. Then, zonal advection dominates meridional advection if $H_u > H_Q$, and vice versa. Defining $\hat{H} = \min(H_u, H_Q)$ then horizontal advection will dominate vertical advection if

$$\mu_1 = \frac{\beta N^2 H_Q \hat{H}}{\bar{u} f_0^2} \ll 1. \quad (13.78)$$

More systematically, we can proceed in *reductio ad absurdum* fashion by first neglecting the vertical advection term in (13.76), and seeing if we can construct a self-consistent solution. If $\psi' = \operatorname{Re} \tilde{\psi}_p(z) e^{ikx}$, and noting that $\bar{u} \partial \tilde{\psi}_p / \partial z - \tilde{\psi}_p \partial \bar{u} / \partial z = \bar{u}^2 (\partial / \partial z)(\tilde{\psi}_p / \bar{u})$ we obtain

$$\tilde{\psi}_p = \frac{i\bar{u}}{k f_0} \int_z^\infty \frac{\tilde{Q}}{\bar{u}^2} dz, \quad (13.79)$$

where \tilde{Q} denotes the Fourier amplitude of Q . Then, from the vorticity equation (13.77), we obtain the (Fourier amplitude of the) vertical velocity

$$\tilde{w}_p = \frac{-ik}{f_0 \rho_R} \int_z^\infty \rho_R \bar{u} (K_s^2 - K^2) \tilde{\psi}_p \, dz. \quad (13.80)$$

Using this one may, at least in principle, check whether the vertical advection in (13.76) is indeed negligible. If \bar{u} is uniform (and so $H_u \gg H_Q$) then we find

$$\tilde{\psi}_p \propto \frac{iQH_Q}{kf_0 \bar{u}} \quad \text{and} \quad \tilde{w}_p \propto \frac{QH_Q^2(K_s^2 - K^2)}{f_0^2}. \quad (13.81a,b)$$

Using this, vertical advection indeed makes a small contribution to the thermodynamic equation provided that

$$\mu_2 = \frac{N^2 H_Q^2 |K_s^2 - K^2|}{f_0^2} \ll 1 \quad (13.82)$$

If $K_s^2 \gg K^2$ and $\hat{H} = H_Q$ then (13.82) is equivalent to (13.78). If \bar{u} is not constant and if $H_u \ll H_Q$ then H_u replaces H_Q and the criterion for the dominance of horizontal advection becomes

$$\mu = \frac{N^2 \hat{H} H_Q |K_s^2 - K^2|}{f_0^2} \ll 1. \quad (13.83)$$

This is the condition that the first term in the denominator of (13.74) is negligible compared with the second. For a typical tropospheric value of $N^2 = 10^{-4} \text{ s}^{-1}$ and for $K > K_s$ we find that $\mu \approx (H_Q/7 \text{ km})^2$, and so we can expect $\mu < 1$ in extra-equatorial regions where the heating is shallow. At low latitudes f_0 is smaller, β is bigger and $\mu \approx (H_Q/1 \text{ km})^2$, and we can expect $\mu > 1$. However, there is both uncertainty and variation in these values.

Equivalent topography

In the case in which zonal advection dominates, the equivalent topography is given by

$$h_e = \frac{-\tilde{w}_p(0)}{ik\bar{u}(0)} = \frac{1}{\bar{u}(0) f_0 \rho_R(0)} \int_0^\infty \rho_R \bar{u} (K_s^2 - K^2) \tilde{\psi}_p \, dz, \quad (13.84)$$

where $\tilde{\psi}_p$ is given by (13.79). The point to notice here is that if $K < K_s$ the equivalent topography is in phase with ψ_p .

13.4.2 Properties of the solution

In the tropics μ may be large for H_Q greater than a kilometre or so. Heating close to the surface cannot produce a large vertical velocity and will therefore produce a meridional velocity. However, away from the surface the heat source will be balanced by vertical advection. For scales such that $K < K_s$, a criterion that might apply at low latitudes for wavelengths longer than a few thousand kilometres, the associated vortex stretching $f \partial w / \partial z > 0$ is balanced by βv and a poleward meridional motion occurs. This implies a trough west of the heating and/or a ridge east of the heating, although the use of quasi-geostrophic theory to draw tropical inferences may be a little suspect.

In mid-latitudes μ is typically small and horizontal advection locally balances diabatic

heating. In this case there is a trough a quarter-wavelength downstream from the heating, and equatorward motion at the longitude of the source. [To see this, note that if the heating has a structure like $\cos kx$ then from either (13.74) or (13.79) the solution goes like $\psi_p \propto -\sin kx$.] The trough may be warm or cold, but is often warm. If $H_Q \ll H_u$, as is assumed in obtaining (13.74), then θ is positive and warm. This is because zonal advection dominates and so the effect of the heating is advected downstream. If $H_Q \gg H_u$ and meridional advection is dominant, then the trough is still warm provided Q decreases with height. The vertical velocity can be inferred from the vorticity balance. If $f_0 \partial w / \partial z \approx \beta v$ and if $w = 0$ at the surface (in the absence of Ekman pumping and any topographic effects) there is *descent* in the neighbourhood of a heat source. This counter-intuitive result arises because it is the horizontal advection that is balancing the diabatic heating. (This result cannot be inferred from the particular solution alone.) If the advection of relative vorticity balances vortex stretching, the opposite may hold.

The homogeneous solution can be inferred from (13.84) and (T.1). Consider, for example, waves that are trapped ($m^2 < 0$) but still have $K < K_s$; that is $K^2 < K_s^2 < K^2 + y^2$. The homogeneous solution forced by the equivalent topography is out of phase with that topography, and so out of phase with ψ_p , using (13.84). For still shorter waves, $K > K_s$, the homogeneous solution is in phase with the equivalent topography, and so again out of phase with ψ_p . Thermal sources produced by large-scale continental land masses may have $K^2 < K_s^2$ and, if $K^2 + y^2 < K_s^2$ they will produce waves that penetrate up into the stratosphere and typically these solutions will dominate far from the source. Evidently though, the precise relationship between the particular and homogeneous solution is best dealt with on a case-by-case basis. A few more general points are summarized in the box on page 566.

13.4.3 Numerical solutions

The numerically calculated response to an isolated heat source is illustrated in Figs. 13.8 and 13.9. The first figure shows the response to a 'deep' heating at 15°N. As the reasoning above would suggest, the vertical velocity field (not shown) is upwards in the vicinity of the source. Away from the source, the solution is dominated by the homogeneous solutions in the form of wavetrains, as described in section 13.1.3, with a simple vertical structure. (In fact, the pattern is quite similar to that obtained with a suitably forced barotropic model, as was found earlier in the topographically forced case.)

Figure 13.9 shows the response to a perturbation at 45°N, and again the solutions are qualitatively in agreement with the reasoning above. The local heating is balanced by an equatorward wind, and there is a surface trough about 20° east of the source, and an upper-level pressure maximum, or ridge, about 60° east. The scale height of the wind field, H_u is about 8 km, greater than that of the source, and the balance in the thermodynamic equation is between the zonal advection of the temperature anomaly $\bar{u}\theta'_x$ and the heat source, so producing a temperature maximum downstream. Again, the far field is dominated by the wavetrain of the homogeneous solution.

Finally, we show a calculation (Fig. 13.10) that, although linear, includes realistic forcing from topography, heat sources and observed transient eddy flux convergences, and uses a realistic zonally averaged zonal flow, although some physical parameters (e.g., representing frictional and diffusion) in the calculation must be changed in order that a steady solution

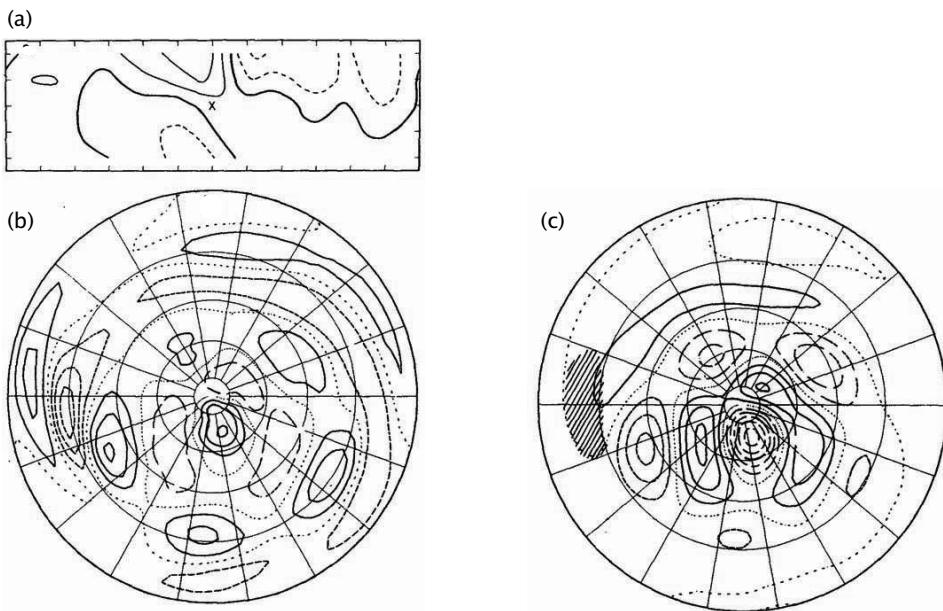


Fig. 13.8 Numerical solution of a baroclinic primitive equation model with a deep heat source at 15°N and a zonal flow similar to that of northern hemisphere winter. (a) Height field in a longitude height at 18°N (vertical tick marks at 100, 300, 500, 700 and 900 mb); (b) 300 mb vorticity field; (c) 300 mb height field. The cross in (a) and the hatched region in (c) indicate the location of the heating.¹⁵

can be achieved. Such a calculation is likely to be the most accurate achievable by a linear model, and discrepancies from observations indicate the presence of nonlinearities that are neglected in the calculation. In fact, a generally good agreement with the observed fields is found, and provides some *post facto* justification for the use of linear, stationary wave models.¹⁸

In such realistic calculations it is virtually impossible to see the wavetrains emerging from isolated features like the Rockies or Himalayas, because they are combined with the responses from all the other sources included in the calculation. Indeed, breaking up the forcing into separate contributions from orographic forcing, heating, and the time averaged momentum and heat fluxes from transient eddies reveals that all of these separate contributions have a non-negligible influence. We should also remember that the effects of the fluxes from the transient eddies are not explained by such a calculation, merely included in a diagnostic sense. Nevertheless, the agreement does reveal the extent to which we might understand the steady zonally asymmetric circulation of the real atmosphere as the response due to the interaction of a zonally uniform zonal wind with the asymmetric features of the Earth's geography and transient eddy field. The quasi-stationary response of the planetary waves to surface anomalies, and the interaction of transient eddies with the large-scale planetary wave field, are important factors in the natural variability of climate, and their understanding remains a difficult challenge and a topic of research for dynamical meteorologists.

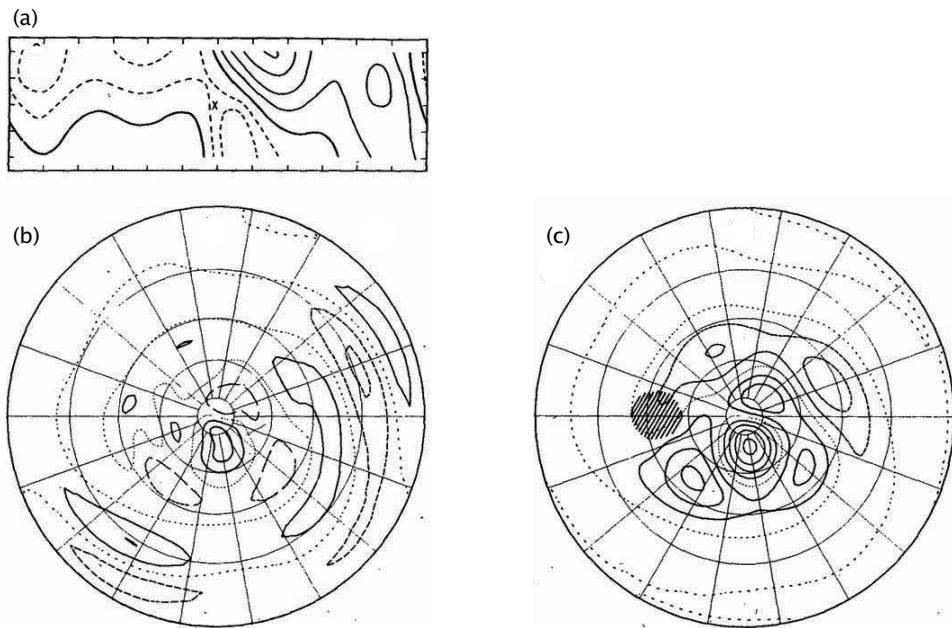


Fig. 13.9 As for Fig. 13.8, but now the solution of a baroclinic primitive equation model with a deep heat source at 45°N. (a) Height field in a longitude height at 18°N; (b) 300 mb vorticity field; (c) 300 mb height field. The cross in (a) and the hatched region in (c) indicate the location of the heating.¹⁶

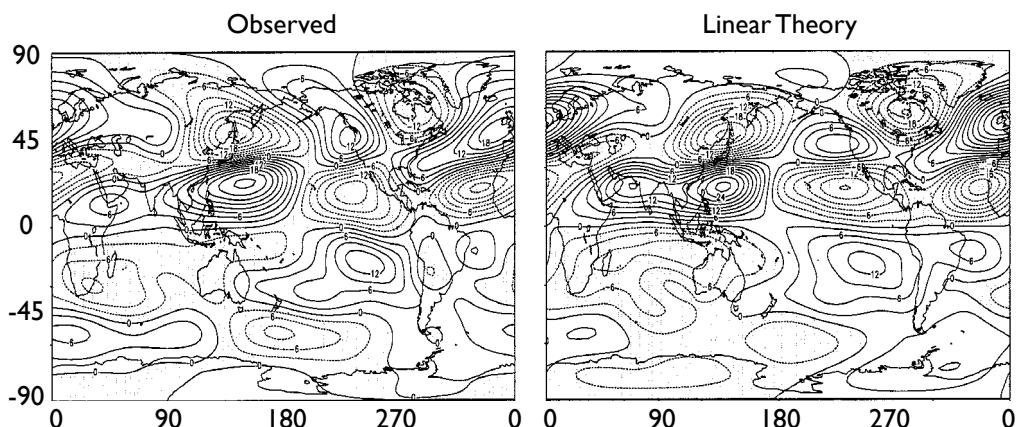


Fig. 13.10 Left: the observed stationary (i.e., time-averaged) streamfunction at 300 mb (about 7 km altitude) in northern hemisphere winter. Right: the steady, linear response to forcing by orography, heat sources and transient eddy flux convergences, calculated using a linear model with the observed height-varying zonally averaged zonal wind. Contour interval is $3 \times 10^6 \text{ m}^2 \text{ s}^{-1}$, and negative values are shaded. Note the generally good agreement, and also the much weaker zonal asymmetries in the southern hemisphere.¹⁷

Thermal Forcing of Stationary Waves: Salient Points

- (i) The solution is composed of a particular solution and a homogeneous solution.
- (ii) The homogeneous solution may be thought of as being forced by an ‘equivalent topography’, chosen so that the complete solution satisfies the boundary condition on vertical velocity at the surface.
- (iii) For a localized source, the far field is dominated by the homogeneous solution. This solution has the same properties as a solution forced by real topography. Thus, it may include waves that penetrate vertically into the stratosphere as well as wavetrains propagating around the globe with an equivalent barotropic structure.
- (iv) In the extratropics, a diabatic heating is typically balanced by horizontal advection, producing a trough a quarter wavelength east (downstream) of a localized heat source. The heat source is balanced by advection of cooler air from higher latitudes, and there may be sinking air over the heat source. This can occur when $\mu \ll 1$ [see (13.83)].
- (v) In the tropics, a heat source may be locally balanced by vertical advection, that is adiabatic cooling as air ascends. This can occur when $\mu \gg 1$.
- (vi) In the real atmosphere, the stationary solutions must coexist with the chaos of time-dependent, nonlinear flows. Thus, they are likely to manifest themselves only in time averaged fields and in a modified form.

13.5 STRATOSPHERIC DYNAMICS

In our final topic of this chapter we look, all too briefly, at the circulation in the stratosphere. (We will draw on results from earlier starred sections but the less technically inclined reader who may have skipped them can simply refer back as needed.) It is convenient to divide this circulation into two components: (i) the zonally averaged meridional overturning circulation and (ii) the quasi-horizontal circulation. There is also a region of the lower stratosphere that interacts directly with the troposphere and where fluid properties are exchanged; however, the dynamics of this region are complex and a topic of research we shall not explore them here.¹⁹

13.5.1 A descriptive overview

The radiative forcing of the stratosphere is effectively illustrated in Fig. 13.11, and the observed zonally averaged temperature and zonal-wind structure are plotted in Fig. 13.12. From these we infer the following.

- ★ The stratosphere is very stably stratified, with a lapse rate corresponding to $N \approx$

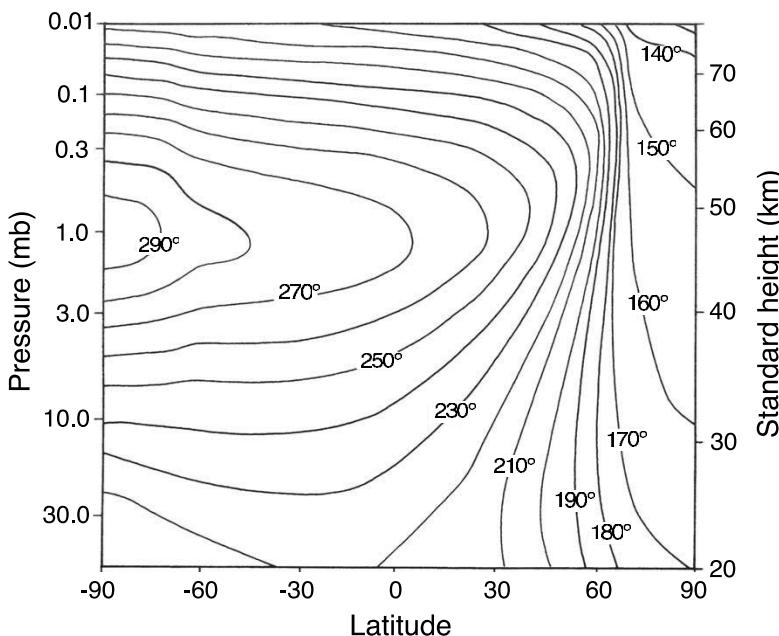


Fig. 13.11 The zonally averaged radiative-equilibrium temperature in January. The downward solar radiation at the top of the atmosphere is given, and the upward radiative flux into the stratosphere is based on observed properties, including temperature, of the troposphere.²⁰

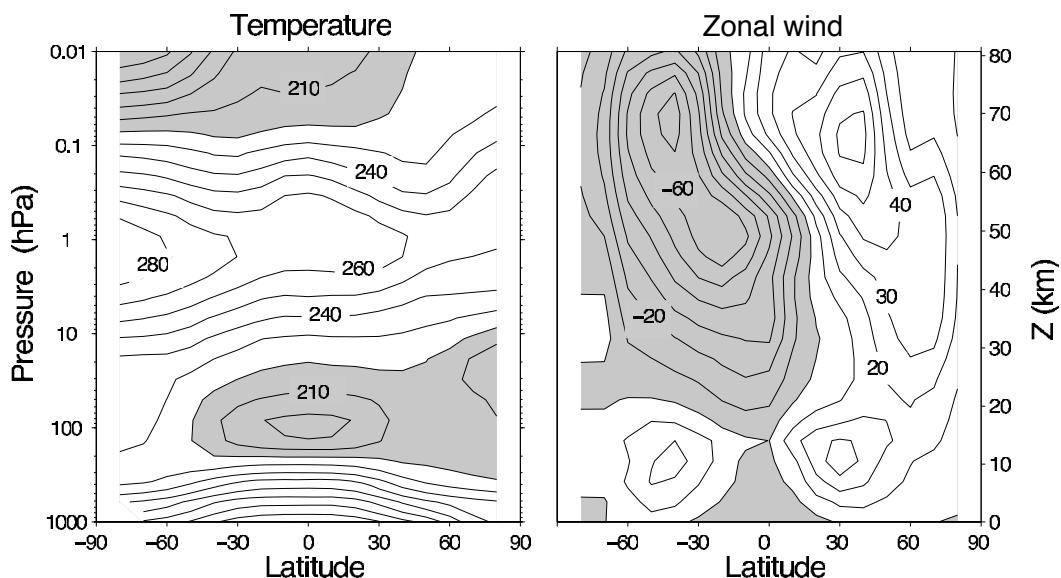


Fig. 13.12 The zonally averaged temperature and zonal wind in January. The temperature contour interval is 10 K, and values less than 220 K are shaded. Zonal wind contours are 10 m s^{-1} and negative (westward) values are shaded.²¹

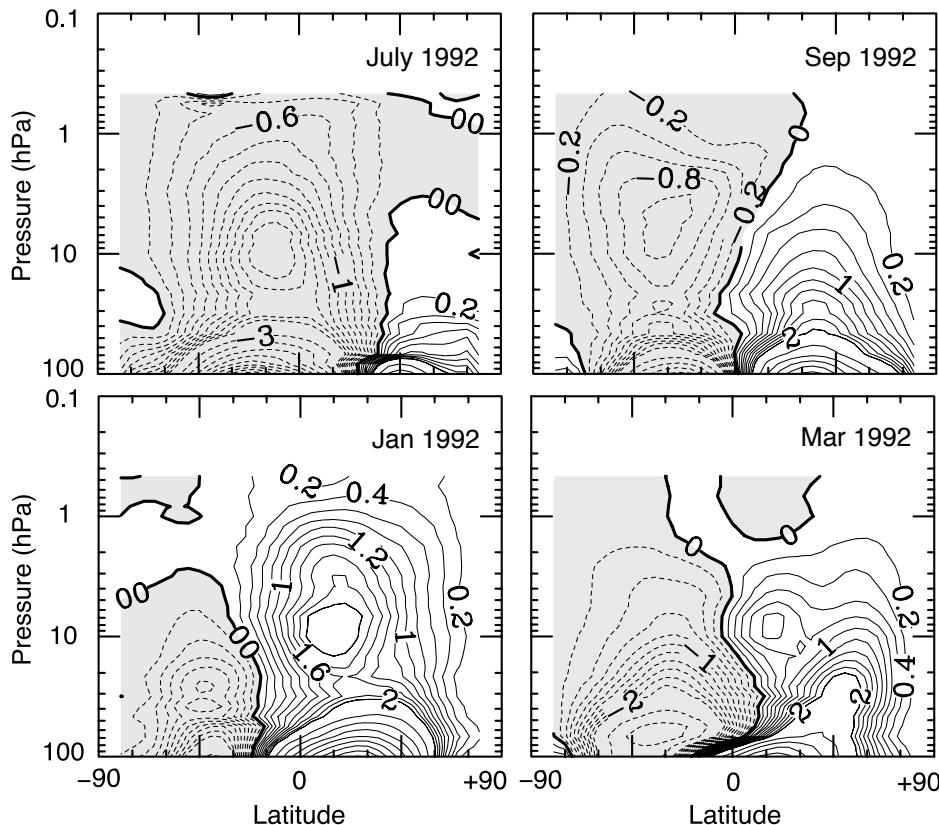


Fig. 13.13 The observed thickness-weighted (residual) streamfunction in the stratosphere, in Sverdrups (10^9 kg s^{-1}). The circulation is clockwise where the contours are solid. Note the stronger circulation in the winter hemispheres, whereas the equinoctial circulations (September, March) are more inter-hemispherically symmetric.²²

$2 \times 10^{-2} \text{ s}$, about twice that of the troposphere on average. This is in part due to the absorption of solar radiation by ozone between 20 and 50 km.

- ★ In the summer the solar absorption at high latitudes leads to a reversed temperature gradient (warmer pole than equator) and, by thermal wind balance, a negative vertical shear of the zonal wind. The temperature distribution is, in fact, not far from the radiative equilibrium distribution, and over much of the summer stratosphere the mean winds are indeed negative (westwards).
- ★ In winter high latitudes receive very little solar radiation and there is a strong meridional temperature gradient and consequently a strong vertical shear in the zonal wind. Nevertheless, this temperature gradient is significantly weaker than the radiative equilibrium temperature gradient, implying a poleward heat transfer by the fluid motions.

There must be, then, a circulation that keeps the stratosphere from radiative equilibrium, and one that is weakest in summer. In fact, a stratospheric meridional overturning circulation

was inferred by A. Brewer and G. Dobson based on observations of water vapour and chemical transport, and is often called the *Brewer-Dobson circulation*.²³ It is depicted in Fig. 13.13; this shows the observed thickness-weighted circulation, almost equivalent to the residual circulation, and so represents both the Eulerian mean and eddy-contributed components. It comprises a single, equator-to-pole cell in each hemisphere, stronger in the winter hemisphere where it goes high into the stratosphere. There is also a distinct lower branch to the circulation, present in all seasons although strongest in winter, that is confined to the lower stratosphere and is in some ways a vertical extension of (the residual circulation of) the tropospheric Ferrel Cell. Not all the upper circulation is ventilated by the troposphere — some of it recirculates within the stratosphere. This circulation and some of the associated dynamics is illustrated schematically in Fig. 13.14,²⁴ and three regions may usefully be delineated: (i) A tropical region; (ii) a mid-latitude region; (iii) the polar vortex. The tropical region is relatively quiescent, an area of upward motion where air is drawn up from the troposphere. In mid-latitudes the residual flow is generally polewards before sinking at high latitudes. In winter the extreme cold leads to the formation of a *polar vortex*, a strong cyclonic vortex that appears quite isolated from mid-latitudes although, especially in the Northern Hemisphere, it is not always centred over the pole.

13.5.2 † Dynamics of the overturning circulation

Wave breaking and residual flow

Based on our discussion in the previous chapter of the wave-mean flow interaction and the (residual) Ferrel Cell, we might expect the Brewer-Dobson circulation to be a consequence of wave-breaking and enstrophy dissipation in the stratosphere. We therefore ask:

- (i) What is the source of such waves?
- (ii) Does such wavebreaking give rise to a circulation of the right sense?
- (iii) Why is the circulation weakest in summer?

The equations of motion governing the mean fields are the zonally averaged momentum and thermodynamic equations, which with quasi-geostrophic scaling and in residual form (see section 7.3.1) may be written as

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v}^* = \nabla \cdot \mathbf{F} + \bar{F}, \quad (13.85a)$$

$$\frac{\partial \bar{\theta}}{\partial t} + \frac{\partial \bar{\theta}}{\partial z} \bar{w}^* = \bar{J}, \quad (13.85b)$$

where \bar{F} represents frictional effects (for example, due to small-scale turbulence) and \bar{J} represents heating, and on the β -plane the residual velocities are related to the Eulerian means by

$$\bar{v}^* = \bar{v} - \frac{1}{\rho_R} \frac{\partial}{\partial z} \left(\rho_R \frac{\bar{v}' \bar{\theta}'}{\partial_z \bar{\theta}} \right), \quad \bar{w}^* = \bar{w} + \frac{\partial}{\partial y} \left(\frac{\bar{v}' \bar{\theta}'}{\partial_z \bar{\theta}} \right). \quad (13.86)$$

The vector \mathbf{F} is the Eliassen-Palm flux, and this is related to the meridional flux of potential vorticity by $\nabla \cdot \mathbf{F} = \bar{v}' \bar{q}'$. The wave activity itself obeys the Eliassen-Palm relation

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla \cdot \mathbf{F} = \mathcal{D}, \quad (13.87)$$

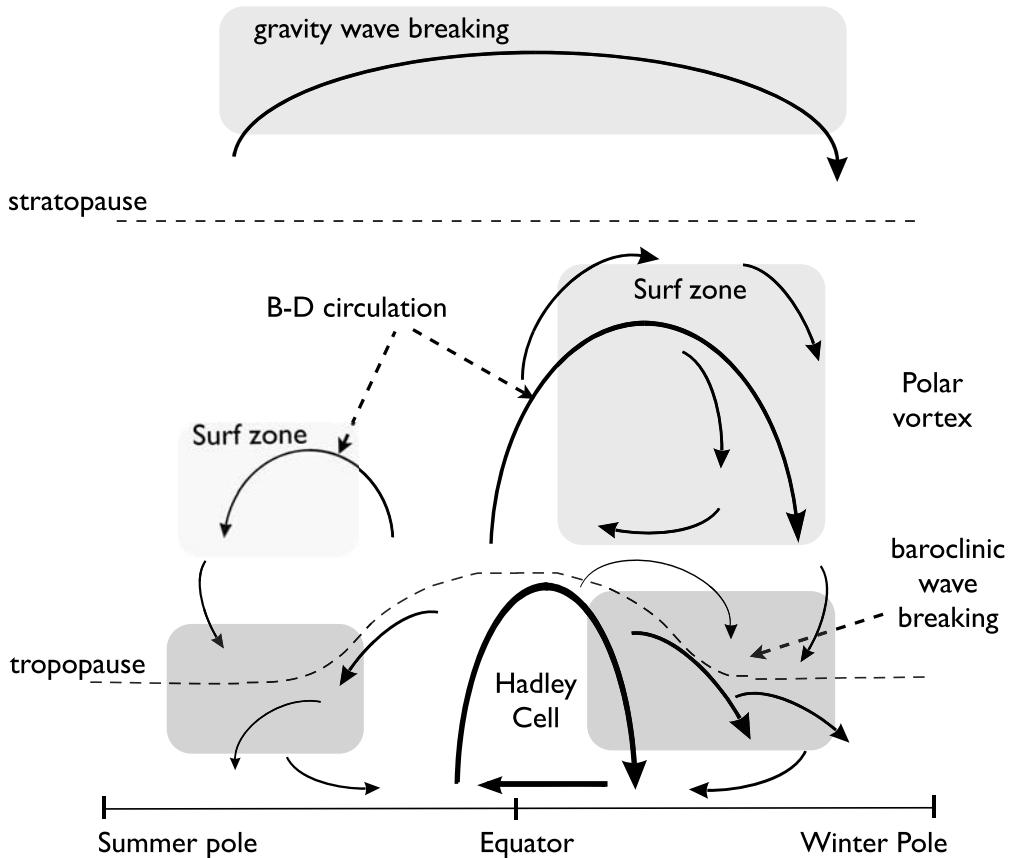


Fig. 13.14 A schema of the residual mean meridional circulation of the atmosphere. The solid arrows indicate the residual circulation (B-D for Brewer-Dobson) and the shaded areas the main regions of wave breaking (i.e., enstrophy dissipation) associated with the circulation. In the surf zone the breaking is mainly that of planetary Rossby waves, and in the troposphere and lower stratosphere the breaking is that of baroclinic eddies. The surf zone and residual flow are much weaker in the summer hemisphere. Only in the Hadley Cell is the residual circulation comprised mainly of the Eulerian mean; elsewhere the eddy component dominates.

where \mathcal{A} is the wave activity, \mathcal{F} is its flux and \mathcal{D} is its dissipation.

Now, in the stratosphere baroclinic instability is relatively weak, certainly compared to the troposphere (e.g., Fig. 6.21), and the main source of wave activity is upwards propagation from the turbulent troposphere, where both baroclinic instability and flow over zonal asymmetries can generate wave activity. From the autumn to the spring, the zonal wind in the stratosphere is generally receptive to planetary-scale Rossby waves propagating up from the troposphere (Fig. 13.7), although at high latitudes in winter there may be a period when the eastward zonal winds are too strong for waves to propagate. If these waves break in the stratosphere then there will be an enstrophy flux to small scales and dissipation. In a

quasi-statistically-steady state and with small frictional effects the dominant balance in the zonal momentum equation (13.85a) is

$$-f_0 \bar{v}^* \approx \bar{v}' \bar{q}', \quad (13.88)$$

where \bar{v}^* is the residual velocity and the potential vorticity flux on the right-hand side is induced by the Rossby wave breaking. In dissipative regions the zonally averaged potential vorticity flux will tend to be down its mean gradient and, if the potential vorticity gradient is polewards (largely because of the β -effect), the residual velocity will be positive if f_0 is positive. That is, the residual flow will be *polewards*, in both hemispheres, and the mechanism giving rise to this is called the ‘Rossby wave pump’. Put another way, Rossby waves propagating up from the troposphere break and deposit westward momentum in the stratosphere, and this wave drag is largely balanced by the Coriolis force on the residual meridional circulation.

This meridional circulation is weakest in summer mainly because linear Rossby waves cannot propagate upwards through the westward mean winds, as illustrated in Fig. 13.15. It is quite striking how the EP vectors avoid the region of westward winds in the summer hemisphere, even though the level of wave activity at low elevations is relatively similar in the summer and winter hemispheres (look between 10 and 15 km in the figure). We can interpret this by noting that for nearly plane waves the EP flux obeys the group velocity property, meaning that $\mathcal{F} = c_g \mathcal{A}$; however, as discussed in section 13.3, if the mean winds are westward the waves evanesce instead of propagate, and thus almost the entire summer hemisphere is shielded from upwardly propagating waves, leaving it in a near-radiative equilibrium state. In the other seasons, the EP flux is able to propagate into the stratosphere and a circulation is generated. This acts to weaken the pole-equator temperature gradient, as we see by inspection of the thermodynamic equation: if the heating is represented by a simple relaxation to a radiative equilibrium state, θ_E , then in a steady state we have

$$N^2 \bar{w}^* = \frac{\theta_E - \theta}{\tau}. \quad (13.89)$$

Poleward flow in mid-latitudes must be supplied by rising air at low latitudes, and sinking air at high latitudes. Thus, from autumn to spring, at low latitudes we have $\theta < \theta_E$ and at high latitudes $\theta > \theta_E$.

Although cause and effect can be very difficult to disentangle in fluid dynamical problems, and the ultimate cause of nearly all fluid motions in the atmosphere is the differential heating from the Sun, it is important to realize that the meridional overturning in the stratosphere is not a direct response to differential solar heating: note that the most intense solar heating is over the summer pole, yet here there is little or no ascent. Rather, the circulation is more usefully thought of as a response to potential vorticity fluxes which in turn are determined by the upward propagation of Rossby waves from the troposphere combined with the poleward gradient of potential vorticity in the stratosphere.

Meridional overturning circulation and downward control

In our discussions of the meridional overturning circulation of the troposphere in section 12.4, we found that a meridional overturning circulation was induced by a momentum source in the zonally averaged zonal momentum equation; this circulation effectively brings the mid-troposphere momentum source to the surface where momentum may be dissipated,

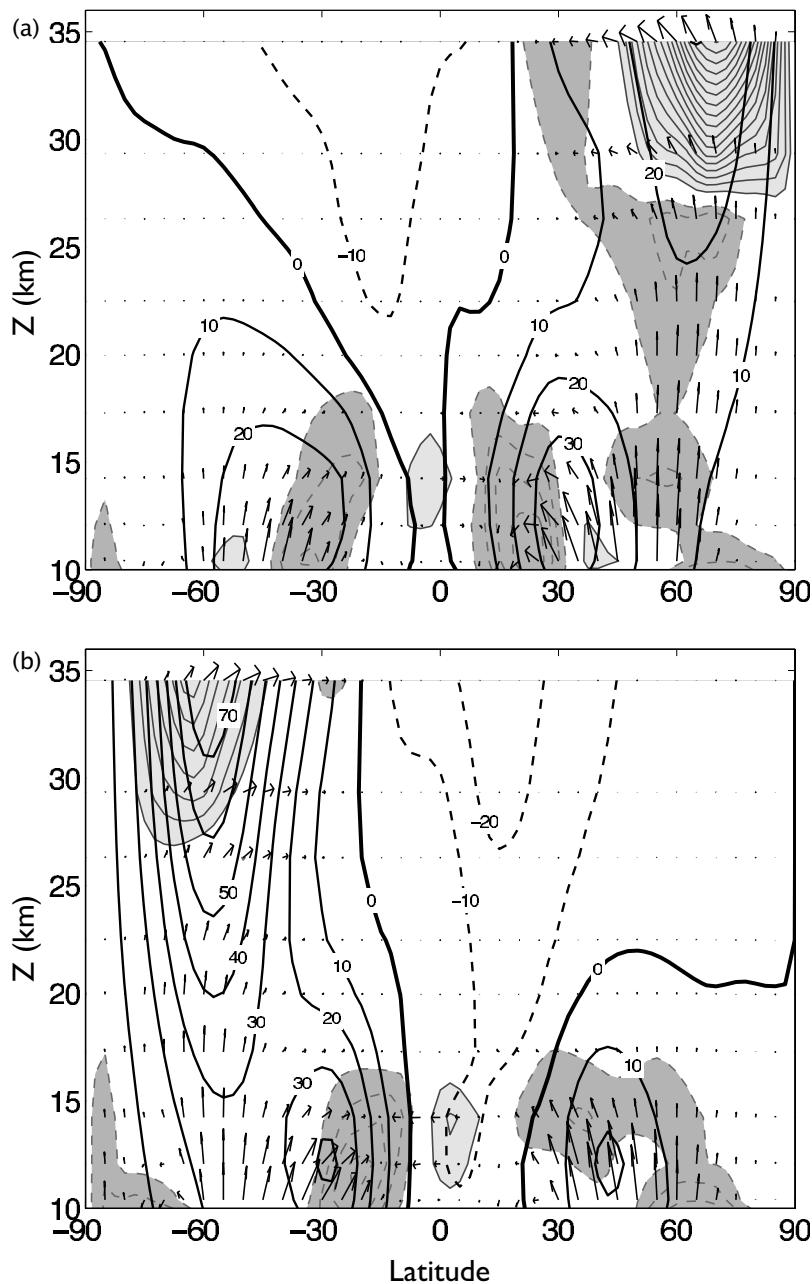


Fig. 13.15 The EP flux vectors (arrows), the EP flux divergence (shaded and light contours) and the zonally averaged zonal wind (heavy contours) for (a) northern hemisphere winter; (b) southern hemisphere winter. Note the almost zero EP values in the summer hemispheres, and strong convergence at high latitudes in the winter hemispheres, leading to poleward residual flow and/or zonal flow acceleration. The EP divergence is shaded for values greater than $+1 \text{ m s}^{-1}/\text{day}$, light solid contours) and for values less than $-1 \text{ m s}^{-1}/\text{day}$ (light dashed contours). The vertical coordinate is log pressure, extending between about 260 and 10 mb.

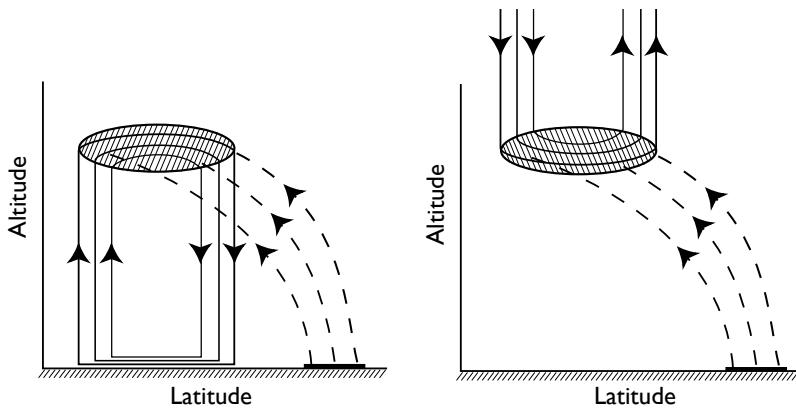


Fig. 13.16 Idealized example of downward control. Left panel: wave activity propagates upwards (dashed lines) from a tropospheric source, breaking and depositing zonal momentum in the shaded region. This induces an overturning circulation (solid lines) below the region of momentum deposition, connecting the region of wave breaking with a frictional boundary layer. Right panel: putative 'upward control', which would require a frictional sink above the wave breaking region.

allowing an overall momentum balance to be achieved. A similar effect, known as 'downward control', may occur in the stratosphere, potentially allowing the wave breaking that occurs in the stratosphere to influence the troposphere.²⁵

The essential idea is well illustrated using the quasi-geostrophic equations on the β -plane. In TEM (residual) form, the zonally averaged zonal momentum equation may be written as [cf. (13.85a)]

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v}^* = \mathcal{F} + \mathcal{D}, \quad (13.90)$$

where \mathcal{F} is the zonal force per unit mass associated with wave breaking (it is the divergence of the Eliassen-Palm flux, \mathcal{F}), and \mathcal{D} is a frictional drag that is important, if at all, only in a bottom boundary layer. (The drag might be approximately represented by $-r\bar{u}(z=0)$, and is only non-zero if the surface wind is non-zero.) The mass continuity equation is

$$\frac{\partial \bar{v}^*}{\partial y} + \frac{1}{\rho_R} \frac{\partial}{\partial z} (\rho_R \bar{w}^*) = 0. \quad (13.91)$$

From (13.90) and (13.91) we obtain, for a steady state,

$$\frac{1}{\rho_R} \frac{\partial}{\partial z} (\rho_R \bar{w}^*) = \frac{1}{f_0} \frac{\partial}{\partial y} (\mathcal{F} + \mathcal{D}). \quad (13.92)$$

If we now assume that $\rho_R \bar{w}^* \rightarrow 0$ as $z \rightarrow \infty$ then it follows that, above the frictional layer,

$$\bar{w}^*(z) = -\frac{1}{\rho_R} \frac{\partial}{\partial y} \int_z^\infty \rho_R \frac{\mathcal{F}}{f_0} dz. \quad (13.93)$$

That is, the vertical velocity at a level z is determined by the momentum sources *above* that level. The physical situation is illustrated in Fig. 13.16. Here, a wave source in the troposphere

propagates upwards and breaks in the middle atmosphere depositing momentum. This induces a meridional circulation as illustrated in the left panel, with a response *below* the momentum source.

The above derivation is disingenuous in its simplicity, for it seems that we might just as well have assumed $w = 0$ at $z = 0$, leading to

$$\bar{w}^*(z) = \frac{1}{\rho_R} \frac{\partial}{\partial y} \int_0^z \rho_R \frac{\mathcal{F} + \mathcal{D}}{f_0} dz. \quad (13.94)$$

This might appear to give ‘upward control’. However, note the appearance of the drag term on the right-hand side of (13.94). Furthermore, mass conservation demands that

$$\int_0^\infty \rho_R \bar{v}^* dz = 0, \quad \text{implying} \quad \int_0^\infty \rho_R (\mathcal{F} + \mathcal{D}) dz = 0, \quad (13.95a,b)$$

using (13.90), for steady state conditions. Thus, above the level of the momentum source \mathcal{F} , (13.94) also implies that the vertical velocity is zero, because \mathcal{F} and \mathcal{D} have cancelling effects. Thus, it is the location of the frictional boundary layer where the momentum is removed that distinguishes up from down. Equation (13.95b) tells us that the frictional boundary layer at the bottom must adjust itself to remove the same amount of momentum that is deposited by wave breaking higher up, if there is to be a steady state. If there were a momentum sink above the momentum deposition region there would be no justification for downward control. However, it is hard to envision how such a sink could exist without violating angular momentum conservation (see also problem 13.10).

It is instructive to compare downward control with the Stommel problem in oceanography (section 14.1.1). The steady version of (13.90) and the equation for the streamfunction for the *horizontal* flow, ψ , in the ocean, (14.6), may respectively be written

$$f_0 \frac{\partial \psi^*}{\partial z} = \mathcal{F} + \mathcal{D}, \quad \beta \frac{\partial \psi}{\partial x} = \mathcal{F}_w + r \nabla^2 \psi, \quad (13.96a,b)$$

where \mathcal{F}_w represents the wind forcing at the ocean surface, and the second term on the right-hand side of (13.96b) represents friction. In (13.96a), $\bar{v}^* = -\partial \psi^*/\partial z$ and in (13.96b), $v = \partial \psi / \partial x$. In the ocean interior, the frictional term is negligible, and in solving the resulting first-order equation ($\beta \partial \psi / \partial x = \mathcal{F}_w$) we may apply the boundary condition of $\psi = 0$ only at one meridional boundary. The natural choice is to choose the eastern boundary for this, and then invoke frictional processes to bring ψ to zero on the west. It is a natural choice because Rossby waves propagate westwards (‘westward control’, as in Fig. 14.12); thus, the boundary current (e.g., the Gulf Stream) is on the west *because* the wind’s influence is carried westwards by Rossby waves, not vice versa. Westward control is an enormously robust effect that pervades almost every aspect of large-scale physical oceanography. In the atmospheric case there is no similar mechanism that demands that the influence of the momentum source be propagated downwards. Rather, downward control results *because* the frictional boundary layer is at the bottom, not vice versa.

As we mentioned, the mechanism of downward control is related to that which gives rise to the Ferrel Cell in the troposphere, and that is certainly a strong and robust effect. Whether the downward control effect following wavebreaking *in the stratosphere* is strong enough to influence circulation in the troposphere, or the structure of the tropopause, remains an open question.

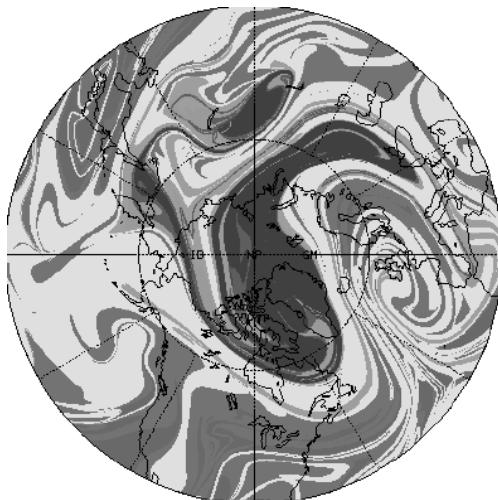


Fig. 13.17 The tracer distribution in the northern hemisphere lower stratosphere on 28 January 1992. The tracer was initialized on 16 January by setting it equal to the potential vorticity field calculated from an observational analysis, and then advected for 12 days by the observed wind fields.²⁶

13.5.3 † The polar vortex and the quasi-horizontal circulation

Let us now shift our perspective and consider the quasi-horizontal circulation in the stratosphere. Stratospheric dynamics are rather more two-dimensional than those in the troposphere because the high stratification inhibits vertical motion, and the vortex stretching term in the quasi-geostrophic potential vorticity equation is relatively small. In any case, because diabatic effects occur on a somewhat longer time scale than advective processes, the flow may be characterized by the advection of potential vorticity on more slowly evolving isentropic surfaces. In mid-latitudes this flow is forced by wave propagation from below, and the upshot is that the mid-latitude stratospheric circulation is a good example of geostrophic turbulence, as illustrated in Figs. 13.17 and 13.18. Both the potential vorticity and the tracer are evocative of the flows in chapters 8 and 9. We see Rossby waves breaking and vortices stretched into filaments and tendrils, the features of an enstrophy cascade. We also perceive some idea of the spectral non-locality of the enstrophy transfer — a single large vortex overturns and breaks and there is little sense of an orderly cascade of enstrophy to dissipative scales. For this reason, the mid-latitude region is known as the *surf zone*. It is precisely this wave breaking that gives rise to the enstrophy flux to small scales and its dissipation, and which in turn gives rise to the residual flow that is the Brewer-Dobson circulation.

This surf-zone does not usually extend to the pole, and in winter dense cold air over the pole forms itself into a cyclonic vortex, the *polar vortex* (see Figs. 13.18 and Fig. 13.19). Although the vortex is diabatically forced, and has a preferred location, the tendency of quasi-two-dimensional flow to organize itself into vortices (as we see in Figs. 6.6 and 8.9) undoubtedly contributes to its coherence and isolation from the rest of the hemisphere. The boundary of the vortex, as measured by the value of the potential vorticity or of the tracer, is quite sharp with the value of PV often jumping by a factor of 2 or so, and the vortex is quite persistent — it is a near-permanent feature of the winter hemisphere. Within the vortex potential vorticity tends to homogenize, and once formed the main communication that the vortex has with the surf zone is via occasional wave breaking at its boundary. It is

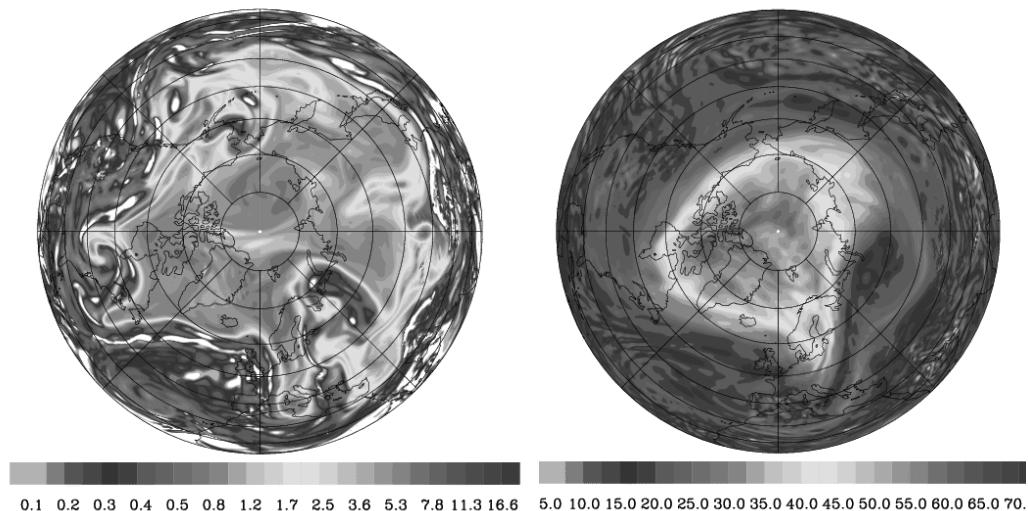
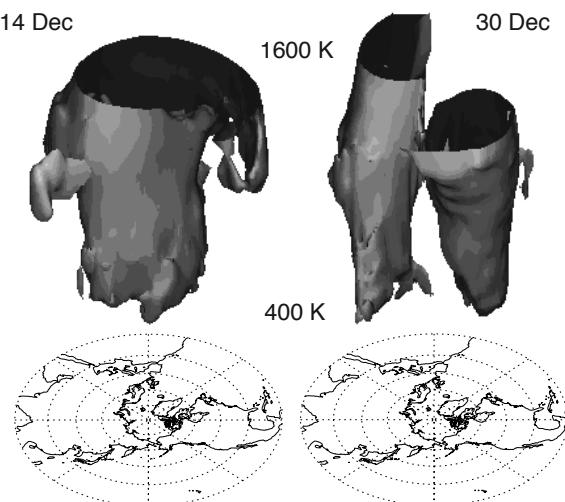


Fig. 13.18 The potential vorticity on two isentropic surfaces, the 310 K surface (left) and the 475 K surface (right), on January 19, 2005. The shaded bar is in PV units. The 310 K surface is mainly in the troposphere (see Fig. 12.16) where baroclinic instability is abundant. The 475 K surface is at about 20 km altitude, and on it we see a polar stratospheric vortex with a fairly sharp boundary where the PV gradient is high, and a mid-latitude region of smaller-scale features and wave breaking.²⁷

Fig. 13.19 The edge of the stratospheric polar vortex in 1984. Plotted is the 35 PVU isosurface of $Q^* = Q(\theta/\theta_0)^{4.1}$, where Q is Ertel PV and $\theta_0 = 475$ K. The vertical coordinate is potential temperature. Like Q , Q^* is materially conserved in adiabatic flow. It is approximately constant at the vortex edge, roughly compensating for the change in density with height that affects the Ertel PV. The left panel shows the vortex in a fairly usual state, and the right panel shows a split vortex following a stratospheric sudden warming.²⁸



interesting that, although the potential vorticity gradient is strong at the edge of the vortex, the exchange of properties is weak, implying a failure of notions of diffusion.

Stable as it is, the polar vortex is nevertheless sometimes disrupted by wave activity from below; this tends to occur when the wave activity itself is quite strong, and when the mean conditions are such as to steer that wave activity polewards. Occasionally, this activity is sufficiently strong so as to cause the vortex to break down, or to split into two smaller vortices, and so allow warm mid-latitude air to reach polar latitudes — an event known as a *stratospheric sudden warming*.

Notes

- 1 Much of our basic understanding in this area stems from conceptual and numerical work on forced Rossby waves by Charney & Eliassen (1949), who looked at the response to orography using a barotropic model. This was followed by a study by Smagorinsky (1953) on the response to thermodynamic forcing using a baroclinic, quasi-geostrophic model. Seeking more realism later studies have employed the primitive equations and spherical coordinates in studies that are at least partly numerical (e.g., Egger 1976, and a host of others), although most theoretical studies still use the quasi-geostrophic equations. We also draw from various review articles, among them Smith (1979), Dickinson (1980), Held (1983) (particularly for sections 13.1.3 and 13.4) and Wallace (1983). See also the collection in the *Journal of Climate*, vol. 15, no. 16, 2002.
- 2 To obtain the solutions shown in Fig. 13.1 and Fig. 13.2, the topography is first specified in physical space. Its Fourier transform is taken and the streamfunction in wavenumber space is calculated using (13.11). The inverse Fourier transform of this gives the streamfunction in physical space.
- 3 The difference between wavetrains emanating from an isolated topographic feature and a global resonant response is relevant for intra-seasonal variability, which might be considered a quasi-stationary response to slowly changing boundary conditions like the sea-surface temperature. If resonance is important, we might expect to see global-scale anomalies, whereas the viewpoint of damped wave-trains is more local. This whole area is one of continuing, active, research with deep roots going back to Namias (1959) and Bjerknes (1959) and beyond.

A different point of view, one that we do not explore in this book, is that the zonally asymmetric features of the Earth's atmosphere are predominantly due to *nonlinear* effects. One possibility is that eddies might significantly modify (and perhaps amplify and sustain) stationary patterns through their large-scale turbulent transfers; see, for example, Green (1977) and Shutts (1983). We could incorporate such effects into a linear model by including the eddy effects as forcing term on the right-hand side of a linear equation such as (13.3), or its two- or three-dimensional analogue, although the forcing term would have to be calculated using a nonlinear theory or taken from observations. Different again is the notion, inspired by models of low-order dynamical systems, that the atmosphere might have *regimes* of behaviour, and that the zonally asymmetric patterns are manifestations of the time spent in a particular regime before transitioning to another. See for example Kimoto & Ghil (1993) and Palmer (1997).

- 4 The description of the stationary waves in terms of wavetrains comes from Hoskins & Karoly (1981), with some earlier theoretical results having been derived by Longuet-Higgins (1964).
- 5 See also Lighthill (1978).
- 6 Steers (1962) and Phillips (1973).

- 7 From Hoskins & Karoly (1981).
- 8 A description of JWKB methods, more commonly called WKB methods, can be found in Bender & Orszag (1978) and a number of other textbooks on approximate mathematical methods. 'WKB' stands for Wentzel, Kramers and Brillouin. They presented the technique in 1926 as a way to find approximate solutions of the Schrödinger equation. Harold Jeffrey's, a mathematical geophysicist, had proposed a similar technique in 1924, and Rayleigh in 1912 had already addressed some aspects of the theory. A general mathematical treatment of the topic was in fact given by Joseph Liouville and George Green in the first half of the nineteenth century. The story thus affirms the hypothesis that methods are often named after the last people to discover them.
- 9 At the critical latitude the JWKB analysis fails and both dissipative and nonlinear effects are likely to play a role. See Dickinson (1968) and Tung (1979).
- 10 From Grose & Hoskins (1979) and Hoskins & Karoly (1981).
- 11 Much of this development stems from Charney & Drazin (1961).
- 12 See also White (1977) and problem 13.5.
- 13 This argument may seem to some readers like prestidigitation. If so, problem 13.6 provides another approach that some may find more transparent.
- 14 Because of such difficulties, understanding the effects of sea-surface temperature anomalies on the atmosphere has become largely the subject of GCM experiments, and one plagued with ambiguous results that depend in part on the particular configuration of the GCM. Some of the modelling issues are reviewed by Kushnir *et al.* (2002).
- 15 From Hoskins & Karoly (1981).
- 16 From Hoskins & Karoly (1981).
- 17 Adapted from Held *et al.* (2002).
- 18 Such solutions are nearly always most easily obtained numerically. One way is to use a Fourier method described earlier. A related method is to write the equations in finite difference form, schematically as $AX = F$, where X is the vector of all the model fields, F represents the known forcing and A is a matrix obtained from the equations of motion and boundary conditions, and solve for X . A quite different method is to use a nonlinear time-dependent model, such as a GCM: prescribe or hold steady the zonally averaged zonal flow as well as all the zonally asymmetric forcing terms, but multiply the asymmetric terms by a small number (e.g., 0.01) to ensure the response is linear; then calculate the steady response by forward time integration, and then divide that solution by the small number to obtain the final solution.
- 19 To read more about the middle atmosphere see, for example, the review by Hamilton (1998), the collection of articles in *Journal of the Meteorological Society of Japan*, vol. 80, no. 4B, 2002, and Andrews *et al.* (1987).
- 20 Adapted from Fels (1985) with the help of K. Hamilton.
- 21 Figure courtesy of J. Wilson of GFDL, using data from Fleming *et al.* (1988).
- 22 Adapted from Eluszkiewicz *et al.* (1997).
- 23 Brewer (1949) and Dobson (1956). Brewer deduced upward motion into the stratosphere at low latitudes based on the water vapour distribution, while Dobson deduced a poleward transport within the stratosphere based on the ozone distribution — the circulation takes ozone from the low latitudes toward the poles. Although originally the Brewer–Dobson

circulation was taken to mean the chemical transport circulation, it is now usually taken to mean the residual (thickness or mass) circulation. The two may differ if there is mixing of chemical without mixing of mass, and the chemical transport may differ among chemicals.

- 24 See also Plumb (2002), which motivated this figure.
- 25 Haynes *et al.* (1991), who also consider the non-quasi-geostrophic case.
- 26 Courtesy of D. Waugh. Colour plots of figures may be downloaded from the web site of this book.
- 27 Courtesy of A. Dörnbrack.
- 28 Courtesy of G. Esler. See also Lait (1994) for discussion of the alternative PV.

Problems

- 13.1 Consider the barotropic vorticity equation on the β -plane, with an uneven lower surface, satisfying the equation of motion

$$\frac{Dq}{Dt} = 0, \quad q = \nabla^2 \psi + \beta y + h_b(x, y), \quad (\text{P13.1})$$

where h_b is proportional to the bottom topography, assumed to be small. Linearize this about a constant zonal flow, U , and seek *steady-state* solutions of the form $\psi = \text{Re } \tilde{\psi} e^{i(kx+ly)}$, with the topography similarly represented. Show that the response is infinite (a resonance) if its wavenumber is equal to that of stationary, free, barotropic Rossby waves.

Suppose that friction is introduced, so that the equation of motion becomes $Dq/Dt = -r\zeta$. Show that the response is now always finite. If the mountain is a single sinusoid, $h_b = H \sin kx$, sketch the response (i.e., the streamfunction field).

- 13.2 (a) Explore the response of the single-layer quasi-geostrophic system to flow over topography. Using Matlab, or otherwise, first obtain a response similar to that in Fig. 13.1. Then vary the frictional time scale, the wavenumber of the stationary Rossby wave and the structure of the topography. Show that when the topography contains a resonant wavenumber a trough in the streamfunction often occurs just downstream of the mountain peak, and that this is to be expected from the analytic solution.
- (b) Suppose that the basic flow is uniform and *westwards*. Obtain and discuss the form of any stationary solutions on the f -plane and β -plane.
- 13.3 ♦ Using an atlas, or obtaining the information from the literature or on-line, obtain a rough representation of the Earth's topography at 45°S and express it as a Fourier series. Then obtain (e.g., using Matlab) the barotropic stationary response to this topography — that is, the solution to (13.10). Explore the sensitivity of the solution to variations in \bar{u} , to using a different \bar{u} on the left- and right-hand sides of (13.10), to the frictional parameter r , and to the deformation radius L_d . Artificially flatten the topography over South America and comment on how the results vary. Finally, discuss whether your calculations are qualitatively and quantitatively in accord with observations. [This problem develops a calculation similar to that of the well-known paper of Charney & Eliassen (1949).]
- 13.4 Obtain an expression analogous to (13.18) for the case with a finite deformation radius ($k_d \neq 0$). Compare the two results and explain the differences, if any.
- 13.5 Using log-pressure coordinates, show that the surface boundary condition analogous to (13.52) is

$$\frac{\partial}{\partial t} \left(\frac{\partial \psi'}{\partial Z} - \frac{N^2}{g} \psi' \right) + \bar{u} \frac{\partial}{\partial x} \frac{\partial \psi'}{\partial Z} - \nu' \frac{\partial \bar{u}}{\partial Z} = -\frac{N^2}{f_0} \left(\bar{u} \frac{\partial h_b}{\partial x} + \alpha \nabla^2 \psi' \right) \quad \text{at } Z = 0, \quad (\text{P13.2})$$

where here Z is proportional to log-pressure.

Hint: Note that the relation between $W = DZ/Dt$ and the real vertical velocity is $w = (f_0/g)\partial\psi/\partial t + RT/(gH)W$, and choose $H = RT(0)/g$.

- 13.6 The vertical energy flux in a radiating wave is proportional to $\rho_R \overline{p'w'}$ where the overline denotes a horizontal average and p' and w' are the pressure and the vertical velocity, respectively.
- For the oscillatory solution, with $m^2 > 0$, show that if the energy flux is to be directed upwards then the product km must be positive, where k and m are the zonal and vertical wavenumbers.
 - For the trapped solution with $m^2 < 0$, show that the vertical propagation of energy is zero.

N.B. In this problem and the next it is important to take the real part of each field properly before evaluating the averages. Relatedly, if $h = \operatorname{Re} h_b e^{ikx}$ where $h_b = h_{br} + i h_{bi}$ then $h_b = h_{br} \cos kx - h_{bi} \sin kx$, but with little loss of generality one may choose either h_{br} or h_{bi} equal to zero.

- 13.7 Obtain an expression for the meridional heat flux associated with the solutions (T.1). In particular, show that for $m^2 > 0$ it is proportional to $|h_b|^2 km/(K_s^2 - K^2)$ and therefore deduce that it is positive for an upwardly propagating wave. Show that for the trapped solutions the meridional heat flux is zero.
- 13.8 Evaluate the wave activity density (pseudomomentum) associated with the solutions (T.1), and the associated EP flux. Show that the group velocity property is satisfied, and that the transport of wave activity is directed upwards for oscillatory solutions.

- 13.9 ♦ Obtain the homogeneous solution to (13.71) that, when added to the particular solution, properly satisfies the boundary condition (13.75). Discuss the solution, and in particular show that the total response remains bounded even as the denominator in (13.74) goes to zero.

13.10 *OVERTURNING CIRCULATION AND DOWNWARD CONTROL*

- (a) Beginning with momentum and thermodynamic equations in the residual quasi-geostrophic forms

$$\frac{\partial \bar{u}}{\partial t} + \frac{f_0}{\rho} \frac{\partial \psi^*}{\partial z} = \mathcal{F} + \mathcal{D}, \quad \frac{\partial \bar{b}}{\partial t} + \frac{N^2}{\rho} \frac{\partial \psi^*}{\partial y} = H, \quad (13.3a,b)$$

and using thermal wind relation to eliminate time derivatives, obtain an elliptic equation for ψ^* . (You may use Cartesian geometry throughout this problem.)

- Derive a diagnostic relation that must be satisfied between H (the heating) and $\mathcal{F} + \mathcal{D}$ that must be satisfied in a steady state.
- ♦ In a steady state, suppose that \mathcal{F} is one-signed and non-zero only in the middle of the domain, and that \mathcal{D} appropriately balances it in an integral sense. Obtain approximate or numerical solutions of your elliptic equation with \mathcal{D} non-zero at the top or with \mathcal{D} non-zero at the bottom of the domain, and $\psi^* = 0$ at the domain boundaries. You may assume the domain has finite vertical extent and that $\rho = 1$. Make and state appropriate assumptions as to the nature of H , or anything else, as needed.
- ♦ Suppose that wave activity is generated near the Earth's surface and propagates to the stratosphere where the waves break. Assume that \mathcal{D} is non-zero only in an isolated region above the region of wave breaking. Discuss whether angular momentum conservation can be satisfied. Compare this case with the one in which \mathcal{D} is non-zero in a surface boundary layer. As examples, suppose that \mathcal{D} is a drag or a viscous force on the zonal wind.

Part IV

LARGE-SCALE OCEANIC

CIRCULATION

To increase our knowledge the subject [oceanography] must be made attractive to men who do not mind facing up to the difficulties of fluid mechanics.

From *Nature*, 25 December 1954.

CHAPTER FOURTEEN

Wind-Driven Gyres

UNDERSTANDING THE CIRCULATION OF THE OCEAN involves a combination of observations, comprehensive numerical modelling, and more conceptual modelling, or ‘theory’. All of these have become essential, but in this chapter, and the ones following, our emphasis is on the last of the tripos. Its (continuing) role is not to explain every feature of the observed ocean circulation, nor to necessarily describe details best left to numerical simulations. It is to provide a conceptual and theoretical framework for understanding the circulation of the ocean, for interpreting observations and suggesting how new observations may best be made, and to aid the development and interpretation of experiments with numerical models.

The main currents of the world's oceans are sketched in Fig. 14.1.¹ (Over most of the ocean, the vertically averaged currents have a similar sense to the surface currents, one exception being at the equator where the surface currents are mainly westwards but the vertical integral is dominated by the eastward undercurrent.) Two seemingly dichotomous aspects stand out: (i) the complexity of the currents as they interact with topography and the geography of the continents; (ii) the simplicity and commonality of the large-scale structures in the major ocean basins, and in particular the ubiquity of subtropical and subpolar gyres. Indeed these gyres, sweeping across the great oceans carrying vast quantities of water and heat, are perhaps the single most conspicuous feature of the circulation. The subtropical gyres are anticyclonic, extending polewards to about 45°, and the subpolar gyres are cyclonic and polewards of this, primarily in the Northern Hemisphere. The existence of gyres, and that they are strongest in the west, has been known for centuries; this *western intensification* leads to such well-known currents as the Gulf Stream in the Atlantic (charted by Benjamin Franklin), the Kuroshio in the Pacific, and the Brazil Current in the South Atlantic. Although even today we barely have sufficient observations to produce a synoptic map of the ocean

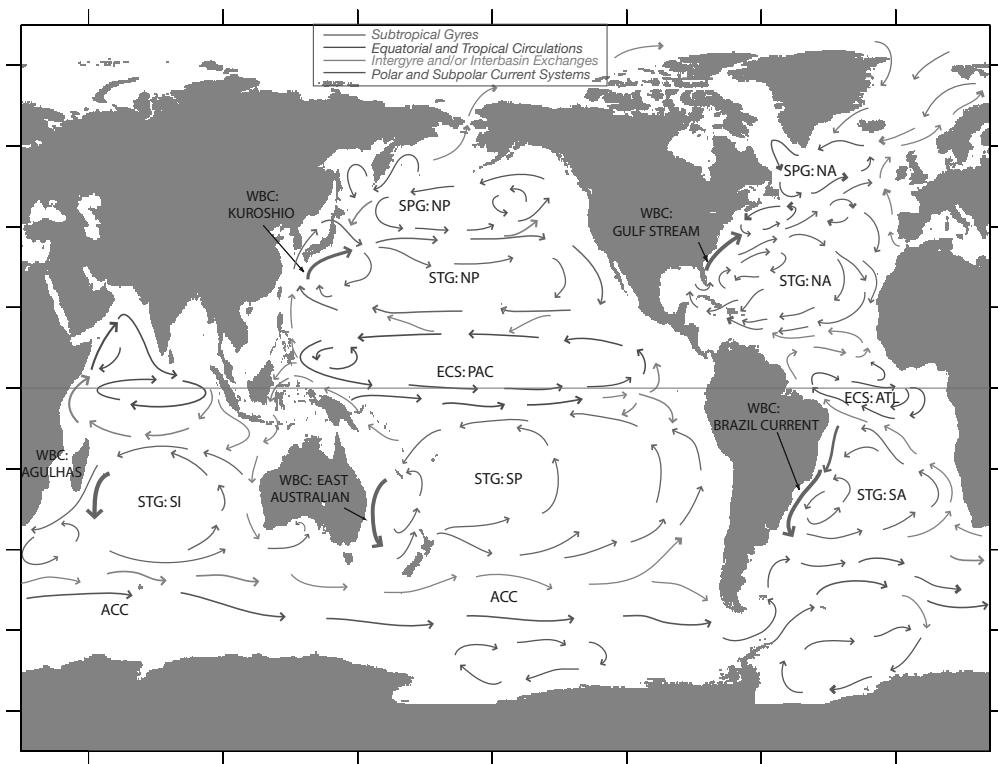


Fig. 14.1 A schema of the main currents of the global ocean. Key: STG – Subtropical Gyre; SPG – Subpolar Gyre; WBC – Western Boundary Current; ECS – Equatorial Current System; NA – North Atlantic; SA – South Atlantic; NP – North Pacific; SP – South Pacific; SI – South Indian; ACC – Antarctic Circumpolar Current; ATL – Atlantic; PAC – Pacific. The figure is a qualitative, and not quantitative, representation of the actual flow.

currents, except at the surface, the large-scale mean currents are fairly well mapped and Fig. 14.2 illustrates the average current pattern of the North Atlantic using a combination of observations and a numerical model, and the Gulf Stream is clearly visible.²

For much of this chapter we consider a model, and variations about it, that explains the large-scale features of ocean gyres and that lies at the core of ocean circulation theory — the steady, forced-dissipative, homogeneous model of the ocean circulation first formulated by Stommel.⁴ Such models explain many of the zeroth-order features of the ocean, in particular the existence of gyres and the appearance of intense western boundary currents. In the later part of the chapter we examine the vertical structure of the wind-driven circulation, taking the stratification as given. In the two chapters following we consider the maintenance of that stratification, first considering buoyancy forcing (i.e., heat and salinity fluxes at the surface) alone, and then the combined effects of wind and buoyancy forcing.

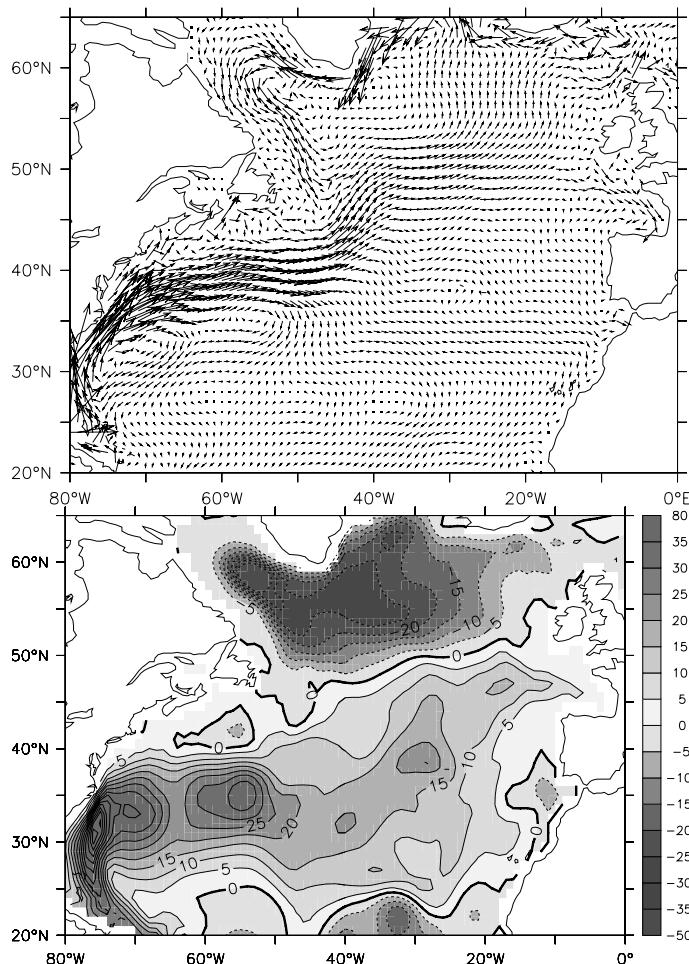


Fig. 14.2 Top: the time averaged velocity field at a depth of 75 m in the North Atlantic, obtained by constraining a numerical model to observations. Bottom: the streamfunction of the vertically integrated flow, in Sverdrups ($1 \text{ Sv} = 10^9 \text{ kg s}^{-1}$). Note the presence of an anticyclonic subtropical gyre, a cyclonic subpolar gyre, and intense western boundary currents.³

14.1 THE DEPTH INTEGRATED WIND-DRIVEN CIRCULATION

The equations that govern the large-scale flow in the oceans are the planetary-geostrophic equations. Greatly simplified as these are compared to the Navier-Stokes equations, or even the hydrostatic Boussinesq equations, they are still quite daunting: a prognostic equation for buoyancy is coupled to the advecting velocity via hydrostatic and geostrophic balance, and the resulting problem is formidably nonlinear. However, it turns out that thermodynamic effects can effectively be eliminated by the simple device of vertical integration; the resulting equations are linear, and the only external forcing is that due to the wind stress. The

resulting model then, at the price of some comprehensiveness, gives a useful picture of the *wind-driven* circulation of the ocean.

14.1.1 The Stommel model

The planetary-geostrophic equations for a Boussinesq fluid are:

$$\frac{Db}{Dt} = \dot{b}, \quad \nabla_3 \cdot \mathbf{v} = 0, \quad (14.1a,b)$$

$$\mathbf{f} \times \mathbf{u} = -\nabla \phi + \frac{1}{\rho_0} \frac{\partial \boldsymbol{\tau}}{\partial z}, \quad \frac{\partial \phi}{\partial z} = b. \quad (14.2a,b)$$

These equations are, respectively, the thermodynamic equation (14.1a), the mass continuity equation (14.1b), the horizontal momentum equation (14.2a), (i.e., geostrophic balance, plus a stress term), and the vertical momentum equation (14.2b), (i.e., hydrostatic balance). These equations are derived more fully in chapter 5, but they are essentially the Boussinesq primitive equations with the advection terms omitted in the horizontal momentum equation, on the basis of small Rossby number. In this chapter we will henceforth absorb the factor of ρ_0 into the $\boldsymbol{\tau}$, so that $\boldsymbol{\tau}$ denotes the kinematic stress, and the gradient operator will be two dimensional, in the x - y plane, unless noted.

Take the curl of (14.2a) (that is, cross differentiate its x and y components) and integrate over the depth of the ocean to give

$$\int \mathbf{f} \nabla \cdot \mathbf{u} dz + \frac{\partial f}{\partial y} \int v dz = \text{curl}_z(\boldsymbol{\tau}_T - \boldsymbol{\tau}_B), \quad (14.3)$$

where the operator curl_z is defined by $\text{curl}_z \mathbf{A} \equiv \partial A^y / \partial x - \partial A^x / \partial y = \mathbf{k} \cdot \nabla \times \mathbf{A}$, and the subscripts T and B are for top and bottom. The divergence term vanishes if the vertical velocity is zero at the top and bottom of the ocean. Strictly, at the top of the ocean the vertical velocity is given by the material derivative of height of the ocean's surface, Dh/Dt , but on the large-scales this has a negligible effect and we may make the rigid-lid approximation and set it to zero. At the bottom of the ocean the vertical velocity is only zero if the ocean is flat-bottomed; otherwise it is $\mathbf{u} \cdot \nabla \eta_B$, where η_B is the orographic height at the ocean floor. The neglect of this topographic term is probably the most restrictive single approximation in the model. Given this neglect, (14.3) becomes

$$\beta \bar{v} = \text{curl}_z(\boldsymbol{\tau}_T - \boldsymbol{\tau}_B) \quad (14.4)$$

where henceforth, in this section, quantities with an overbar are understood to be the vertical integral over the depth of the ocean. If the stresses depend only on the velocity fields then thermodynamic fields do not affect the vertically integrated flow.

At the top of the ocean, the stress is given by the wind. At the bottom, in the absence of topography we assume that the stress may be parameterized by a linear drag, or Rayleigh friction, as might be generated by an Ekman layer; it is this assumption that particularly characterizes this model as being due to Stommel. (Note that we parameterize the friction by a drag acting on the vertically integrated velocity. Using the velocity at the bottom of the ocean would be more realistic, but this wrinkle is beyond the scope of vertically integrated models.) Equation (14.4) then becomes

$$\beta \bar{v} = -r \bar{\zeta} + F_\tau(x, y), \quad (14.5)$$

Formulation of the Stommel/Munk model

- ★ Vertically integrated planetary-geostrophic equations, or a homogeneous fluid with nonlinearity neglected.
- ★ Friction parameterized by a linear drag (Stommel model) or a harmonic Newtonian viscosity (Munk model) or both (Stommel-Munk model).
- ★ Flat bottomed ocean.

Variations on the theme include allowing nonlinearity in the vorticity equation, posing the problem in domains of various shapes, and allowing bottom topography (in particular sloping sidewalls). The solutions are normally calculated in the boundary-layer approximation, but some exact solutions exist.

where F_τ is the wind-stress curl at the top of the ocean and is a known function. Because the velocity is divergence-free, we can define a streamfunction ψ such that $\bar{u} = -\partial\psi/\partial y$ and $\bar{v} = \partial\psi/\partial x$. Equation (14.5) then becomes

$$r\nabla^2\psi + \beta\frac{\partial\psi}{\partial x} = F_\tau(x, y). \quad (14.6)$$

This equation is often referred to as the *Stommel problem* or the *Stommel model*, and may be posed in a variety of two dimensional domains.

14.1.2 Alternative formulations

A number of alternative formulations leading to (14.6) are possible. None are perhaps as well justified as the derivation via the planetary-geostrophic equations but the differences in the specific assumptions made give some indication of the robustness of the derivation, and show how the model might be extended to include topographic or nonlinear effects.

I. A Homogeneous model

Rather than vertically integrating, we may suppose that the ocean is a homogeneous fluid obeying the shallow water equations (chapter 3). The potential vorticity equation [cf. (3.78) on page 138] is:

$$\frac{D}{Dt}\left(\frac{\zeta + f}{h}\right) = \frac{F}{h}, \quad (14.7)$$

where F represents friction and forcing. In an ocean with a rigid-lid and flat bottom (14.7) gives the barotropic vorticity equation,

$$\frac{D\zeta}{Dt} + \beta v = F, \quad (14.8)$$

where the term F again represents the wind-stress curl and a linear drag. Further, since the horizontal velocity is divergence-free (because of the flat-bottom and rigid-lid) we may represent it as a streamfunction, whence we obtain the closed equation

$$\frac{D}{Dt}\nabla^2\psi + \beta\frac{\partial\psi}{\partial x} = F_\tau(x, y) - r\nabla^2\psi, \quad (14.9)$$

where F_τ again represents the wind forcing. This equation is the ‘time-dependent nonlinear Stommel problem’. The steady nonlinear problem is sometimes of interest, too, and this is just

$$J(\psi, \nabla^2 \psi) + \beta \frac{\partial \psi}{\partial x} = F_\tau(x, y) - r \nabla^2 \psi. \quad (14.10)$$

To obtain the original Stommel model we just ignore the advective derivative, which will be valid if $|\zeta| \sim Z \ll \beta L$ where $Z = U/L$ is a representative value of vorticity. This condition is equivalent to

$$R_\beta \equiv \frac{U}{\beta L^2} \ll 1. \quad (14.11)$$

R_β is called the *beta-Rossby number*. On sufficiently large scales, $\beta \sim f/L$ and (14.11) is similar to a small Rossby number assumption.

II. Quasi-geostrophic formulation

In the planetary-geostrophic formulation, the horizontal velocity is divergence-free only because we have vertically integrated. If the scales of motion are not too large the horizontal flow at every level is divergence-free for another reason: because it is in geostrophic balance. In reality, over a single oceanic gyre (say from 15° to 40° latitude), variations in Coriolis parameter are not large, and this prompts us to formulate the model in terms of the quasi-geostrophic equations. Formally, such a model would then be restricted to length scales, L , of no more than $\mathcal{O}(Ro^{-1})$ larger than the deformation radius, and for gyre scales this criterion is marginally satisfied if $L_d = 100$ km. An advantage of the quasi-geostrophic equations is that they readily allow for the inclusion of both nonlinearity and stratification. [For an informal summary of quasi-geostrophy, see the box on page 222.] The quasi-geostrophic vorticity equation for a Boussinesq system is:

$$\frac{D\zeta}{Dt} + \beta v = f_0 \frac{\partial w}{\partial z} + \text{curl}_z \frac{\partial \tau}{\partial z}. \quad (14.12)$$

If we neglect the advective derivative, and vertically integrate, we obtain

$$\beta \int_B^T v \, dz = f_0 [w]_B^T + \text{curl}_z [\tau]_B^T. \quad (14.13)$$

where T denotes the ocean top and B the bottom. We now make one of two virtually equivalent choices.

- (i) We suppose that the integration is over the entire depth of the ocean, in which case the term $[w]_B^T$ vanishes given a rigid lid and a flat bottom. If the stress at the top of the ocean is given by the wind stress, and at the bottom of the ocean it is parameterized by a linear drag, then we obtain

$$\beta \bar{v} = F_\tau(x, y) - r \bar{\zeta} \quad (14.14)$$

just as in (14.5), and where an overbar denotes a vertical integral. Writing $\bar{v} = \partial \psi / \partial x$ and $\bar{\zeta} = \nabla^2 \psi$ then gives (14.6).

- (ii) We suppose that the integration is between two thin Ekman layers at the top and bottom of the ocean. The stress is zero at the interior edge of these layers, but the vertical

velocity is not. At the base of the upper Ekman layer, at $z = -\delta_T$, the vertical velocity is given by:

$$w(x, y, -\delta_T) = \text{curl}_z(\mathbf{\tau}_T/f_0), \quad (14.15)$$

where the top of the ocean is at $z = 0$ and δ_T is the thickness of the top Ekman layer. Similarly, at the top of the lower Ekman layer, the vertical velocity is:

$$w(x, y, -H + \delta_B) = \delta_B \zeta, \quad (14.16)$$

where $z = -H$ at the ocean bottom and δ_B is the thickness of the bottom Ekman layer. Neglecting the advective derivative, and integrating over the ocean between the two Ekman layers, (14.12) becomes

$$\beta \bar{v} = \text{curl}_z \mathbf{\tau}_T - f_0 \delta_B \zeta = \text{curl}_z \mathbf{\tau}_T - f_0 \delta_B \bar{\zeta}/H, \quad (14.17)$$

where, to obtain the second equality, we assume that the bottom drag may be parameterized using the vertically integrated vorticity, $\bar{\zeta}$. Defining the drag coefficient r by $r = f_0 \delta_B / H$, and introducing a streamfunction gives

$$r \nabla^2 \psi + \beta \frac{\partial \psi}{\partial x} = \text{curl}_z \mathbf{\tau}_T, \quad (14.18)$$

as before.

14.1.3 Approximate solution of Stommel model

Sverdrup balance

Equation (14.6) is linear and it is possible to obtain an exact, analytic solution. However, it is more insightful to approach the problem perturbatively, by supposing that the frictional term is small, meaning there is an approximate balance between wind stress and the β -effect.⁵ Friction is small if $|r\zeta| \ll |\beta v|$ or

$$\frac{r}{L} = \frac{f \delta_B}{HL} \ll \beta \quad (14.19)$$

using $r = f \delta_B / H$, and where L is the horizontal scale of the motion, and generally speaking this inequality is well satisfied for large-scale flow. The vorticity equation becomes

$$\beta \bar{v} \approx \text{curl}_z \mathbf{\tau}_T, \quad (14.20)$$

which is known as *Sverdrup balance*.⁶ The observational support for Sverdrup balance is rather mixed, discrepancies arising not so much from the failure of (14.19), but from the presence of small-scale eddying motion with concomitantly large nonlinear terms, and the presence of non-negligible vertical velocities induced by the interaction with bottom topography.⁷ Nevertheless, Sverdrup balance provides a useful, if not impregnable, foundation on which to build. (Sometimes ‘Sverdrup balance’ is taken to mean the linear geostrophic vorticity balance $\beta v = f \partial w / \partial z$, but we will restrict its use to mean a balance between the beta effect and wind stress curl.)

Boundary-layer solution

For simplicity, consider a square domain of side a and rescale the variables by setting

$$x = a\hat{x}, \quad y = a\hat{y}, \quad \tau = \tau_0\hat{\tau}, \quad \psi = \hat{\psi}\frac{\tau_0}{\beta}, \quad (14.21)$$

where τ_0 is the amplitude of the wind stress. The hatted variables are non-dimensional and, assuming our scaling to be sensible, these are $\mathcal{O}(1)$ quantities in the interior. Equation (14.18) becomes

$$\frac{\partial \hat{\psi}}{\partial \hat{x}} + \epsilon_S \nabla^2 \hat{\psi} = \text{curl}_z \hat{\boldsymbol{\tau}}_T, \quad (14.22)$$

where $\epsilon_S = (r/a\beta) \ll 1$, in accord with (14.19). For the rest of this section we will drop the hats over non-dimensional quantities. Over the interior of the domain, away from boundaries, the frictional term in (14.22) is small. We can take advantage of this by writing

$$\psi(x, y) = \psi_I(x, y) + \phi(x, y), \quad (14.23)$$

where ψ_I is the interior streamfunction and ϕ is a boundary layer correction. Away from boundaries ψ_I is presumed to dominate the flow, and this satisfies

$$\frac{\partial \psi_I}{\partial x} = \text{curl}_z \boldsymbol{\tau}_T. \quad (14.24)$$

The solution of this equation (called the ‘Sverdrup interior’) is

$$\psi_I(x, y) = \int_0^x \text{curl}_z \boldsymbol{\tau}(x', y) dx' + g(y), \quad (14.25)$$

where $g(y)$ is an arbitrary function of integration that gives rise to an arbitrary zonal flow. The corresponding velocities are

$$v_I = \text{curl}_z \boldsymbol{\tau}, \quad u_I = -\frac{\partial}{\partial y} \int_0^x \text{curl}_z \boldsymbol{\tau}(x', y) dx' - \frac{dg(y)}{dy}. \quad (14.26)$$

The dynamics is most clearly illustrated if we now restrict our attention to a wind-stress curl that is zonally uniform, and that vanishes at two latitudes, $y = 0$ and $y = 1$. An example is

$$\boldsymbol{\tau}_T^y = 0, \quad \boldsymbol{\tau}_T^x = -\cos(\pi y), \quad (14.27)$$

for which $\text{curl}_z \boldsymbol{\tau}_T = -\pi \sin(\pi y)$. The Sverdrup (interior) flow may then be written as

$$\psi_I(x, y) = [x - C(y)] \text{curl}_z \boldsymbol{\tau}_T = \pi[C(y) - x] \sin \pi y, \quad (14.28)$$

where $C(y)$ is the arbitrary function of integration [$C(y) = -g(y)/\text{curl}_z \boldsymbol{\tau}_T$]. If we choose C to be a constant, the zonal flow associated with it is $C \text{curl}_z \boldsymbol{\tau}_T$. We can then satisfy $\psi = 0$ at either $x = 0$ (if $C = 0$) or $x = 1$ (if $C = 1$). These solutions are illustrated in Fig. 14.3 for the particular stress (14.27).

Regardless of our choice of C we cannot satisfy $\psi = 0$ at both zonal boundaries. We must choose one, and then construct a *boundary layer* solution (i.e., determine ϕ) to satisfy the other condition. Which choice do we make? On intuitive grounds it seems that we

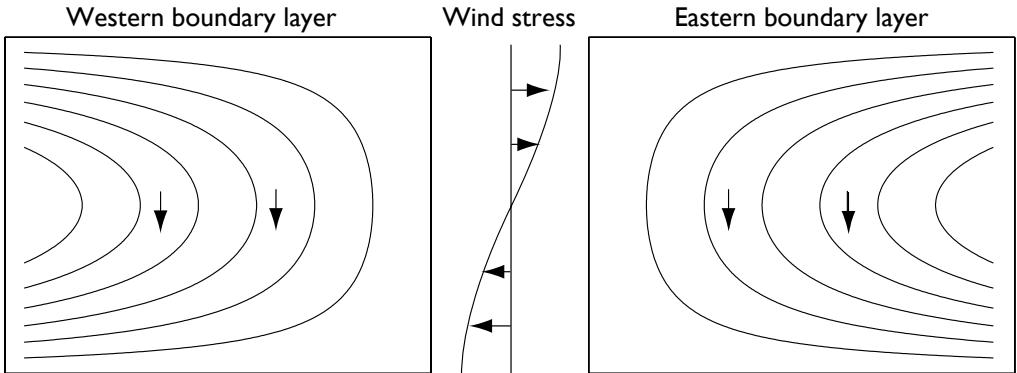


Fig. 14.3 Two possible Sverdrup flows, ψ_I , for the wind stress shown in the centre. Each solution satisfies the no-flow condition at either the eastern or western boundary, and a boundary layer is therefore required at the other boundary. Both flows have the same, equatorward, meridional flow in the interior. Only the flow with the western boundary current is physically realizable, however, because only then can friction produce a curl that opposes that of the wind stress, so allowing the flow to equilibrate.

should choose the solution that satisfies $\psi = 0$ at $x = 1$ (the solution on the left in Fig. 14.3), for then the interior flow then goes round in the same direction as the wind: the wind is supplying a clockwise torque, and to achieve an angular momentum balance anticlockwise angular momentum must be removed by friction. We can imagine that this would be provided by the frictional forces at the western boundary layer if the interior flow is clockwise, but not by friction at an eastern boundary layer when the interior flow is anticlockwise. Note that this argument is not dependent on the sign of the wind-stress curl: if the wind blew the other way a similar argument still implies that a western boundary layer is needed. We will now see if and how the mathematics reflects this intuitive but non-rigorous argument.

Asymptotic matching

Near the walls of the domain the boundary layer correction $\phi(x, y)$ must become important in order that the boundary conditions may be satisfied, and the flow, and in particular $\phi(x, y)$, will vary rapidly with x . To reflect this, let us *stretch* the x -coordinate near this point of failure (i.e., at either $x = 0$ or $x = 1$, but we do not know at which yet) and let

$$x = \epsilon\alpha \quad \text{or} \quad x - 1 = \epsilon\alpha. \quad (14.29a,b)$$

Here, α is the stretched coordinate, which has values $\mathcal{O}(1)$ in the boundary layer, and ϵ is a small parameter, as yet undetermined. We then suppose that $\phi = \phi(\alpha, y)$, and using (14.23) in (14.22), we obtain

$$\epsilon_S(\nabla^2\psi_I + \nabla^2\phi) + \frac{\partial\psi_I}{\partial x} + \frac{1}{\epsilon}\frac{\partial\phi}{\partial\alpha} = \text{curl}_z\boldsymbol{\tau}_T, \quad (14.30)$$

where $\phi = \phi(\alpha, y)$ and $\nabla^2 \phi = \epsilon^{-2} \partial^2 \phi / \partial \alpha^2 + \partial^2 \phi / \partial y^2$. Now, by choice, ψ_I exactly satisfies Sverdrup balance, and so (14.30) becomes

$$\epsilon_S \left(\nabla^2 \psi_I + \frac{1}{\epsilon^2} \frac{\partial^2 \phi}{\partial \alpha^2} + \frac{\partial^2 \phi}{\partial y^2} \right) + \frac{1}{\epsilon} \frac{\partial \phi}{\partial \alpha} = 0. \quad (14.31)$$

We now choose ϵ to obtain a physically meaningful solution. An obvious choice is $\epsilon = \epsilon_S$ (see problem 14.1), for then the leading-order balance in (14.31) is

$$\frac{\partial^2 \phi}{\partial \alpha^2} + \frac{\partial \phi}{\partial \alpha} = 0, \quad (14.32)$$

the solution of which is

$$\phi = A(y) + B(y) e^{-\alpha}. \quad (14.33)$$

Evidently, ϕ grows exponentially in the negative α direction. If this were allowed, it would violate our assumption that solutions are small in the interior, and we must eliminate this possibility by allowing α to take only positive values in the interior of the domain, and by setting $A(y) = 0$. We therefore choose $x = \epsilon \alpha$ so that $\alpha > 0$ for $x > 0$; the boundary layer is then at $x = 0$, that is, it is a *western boundary*, and it decays eastwards in the direction of increasing α — that is, into the ocean interior. We now choose $C = 1$ in (14.28) to make $\psi_I = 0$ at $x = 1$ in (14.28) and then, for the wind stress (14.27), the interior solution is given by

$$\psi_I = \pi(1 - x) \sin \pi y. \quad (14.34)$$

This alone satisfies the boundary condition at the eastern boundary. The function $B(y)$ is chosen to satisfy the additional condition that

$$\psi = \psi_I + \phi = 0 \quad \text{at } x = 0, \quad (14.35)$$

and using (14.34) this gives

$$\pi \sin \pi y + B(y) = 0. \quad (14.36)$$

Using this in (14.33), with $A(y) = 0$, then gives the boundary layer solution

$$\phi = -\pi \sin \pi y e^{-x/\epsilon_S}. \quad (14.37)$$

The composite (boundary layer plus interior) solution is the sum of (14.34) and (14.37), namely

$$\psi = (1 - x - e^{-x/\epsilon_S}) \pi \sin \pi y.$$

(14.38)

With dimensional variables this is

$$\psi = \frac{\tau_0 \pi}{\beta} \left(1 - \frac{x}{a} - e^{-x/(a\epsilon_S)} \right) \sin \frac{\pi y}{a}. \quad (14.39)$$

This is a ‘single gyre’ solution. Two or more gyres can be obtained with a different wind forcing, such as $\tau^x = -\tau_0 \cos(2\pi y)$, as in Fig. 14.4.

It is a relatively straightforward matter to generalize to other wind stresses, provided

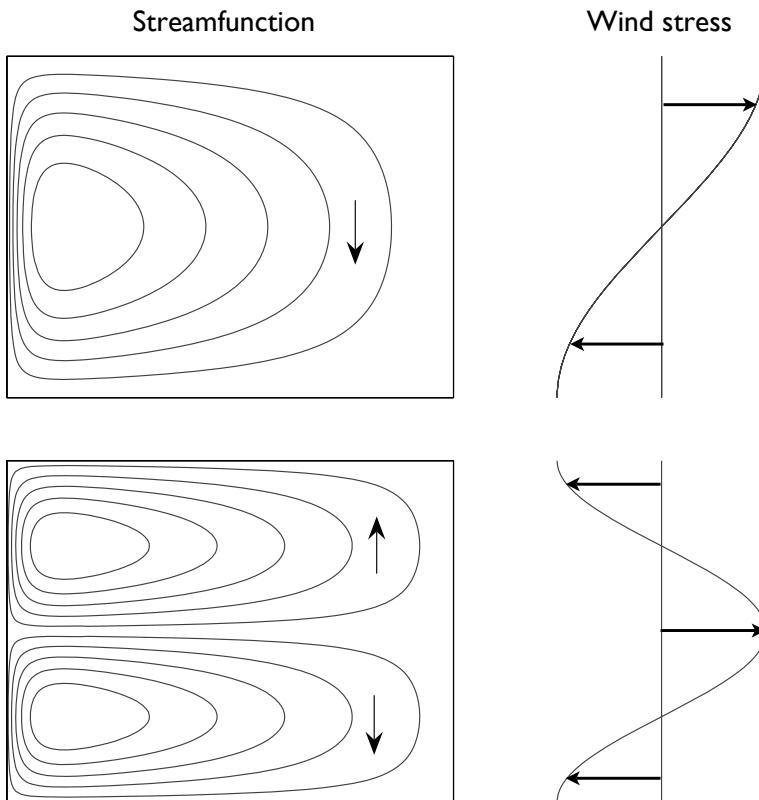


Fig. 14.4 Two solutions of the Stommel model. Upper panel shows the streamfunction of a single-gyre solution, with a wind stress proportional to $-\cos(\pi y/a)$ (in a domain of side a), and the lower panel shows a two-gyre solution, with wind stress proportional to $\cos(2\pi y/a)$. In both cases $\epsilon_S = 0.04$.

these also vanish at the two latitudes between which the solution is desired. It is left as a problem to show that in general

$$\psi_I = \int_{x_E}^x \operatorname{curl}_z \boldsymbol{\tau}(x', y) dx', \quad (14.40)$$

and that the composite solution is

$$\psi = \psi_I - \psi_I(0, y) e^{-x/(x_E \epsilon_S)}. \quad (14.41)$$

14.2 USING VISCOSITY INSTEAD OF DRAG

A natural variation on the Stommel problem is to use a harmonic viscosity, $\nu \nabla^2 \zeta$, in place of the drag term $-r\zeta$ in the vorticity equation, the argument being that the wind-driven circulation does not reach all the way to the ocean bottom so that an Ekman drag is not

Some Properties of the Stommel Model

- (i) The transport in the Sverdrup interior is equatorwards for an anti-cyclonic wind-stress curl. This transport is exactly balanced by the poleward transport in the western boundary layer.
- (ii) There must be a boundary layer to satisfy mass conservation, and this must be a western boundary layer if the friction acts to provide a force of opposite sign to the motion itself. As there is a balance between friction and the β -effect, it is a ‘frictional boundary layer’. The western location does not depend on the sign of the wind stress, nor on the sign of the Coriolis parameter, but it does depend on the sign of β , and so on the direction of rotation of the Earth.
- (iii) The boundary layer width arises by noting that the terms $r\nabla^2\psi$ and $\beta\partial\psi/\partial x$ are in approximate balance in the western boundary layer, implying a length scale of $L_w = (r/\beta)$. If r , the inverse frictional time, is 20 days⁻¹ then $L_w \approx 60$ km, similar to the width of the Gulf Stream.

appropriate. This variation is called the ‘Munk problem’ or ‘Munk model’,⁸ and if both drag and viscosity are present we have the ‘Stommel–Munk’ model. The particular form of the lateral friction used in the Munk problem is still somewhat hard to justify because it relies on an ill-founded eddy diffusion of relative vorticity (chapter 10). Our treatment of this problem is relatively brief, concentrating only on those areas where the problem differs from the Stommel problem. The problem is to find and understand the solution to the (dimensional) equation

$$\beta \frac{\partial \psi}{\partial x} = \text{curl}_z \mathbf{T}_T + \nu \nabla^2 \zeta = \text{curl}_z \mathbf{T}_T + \nu \nabla^4 \psi \quad (14.42)$$

in a given domain, for example a square of side a . We need two boundary conditions at each wall to solve the problem uniquely, and as before for one of them we choose $\psi = 0$ to satisfy the no-normal-flow condition. For the other condition, two possibilities present themselves.

- (i) Zero vorticity, or $\zeta = 0$. Since $\psi = 0$ along the boundary, this possibility is equivalent to $\partial^2\psi/\partial n^2 = 0$ where $\partial/\partial n$ denotes a derivative normal to the boundary. This is known as the ‘free-slip’ condition. At $x = 0$, for example, the condition becomes $\partial v/\partial x = 0$; that is, there is no horizontal shear at the boundary.
- (ii) No flow along the boundary (the ‘no slip’ condition). That is $\psi_n = 0$ where the subscript denotes the normal derivative of the streamfunction. At $x = 0$ we have $v = 0$.

There is little a priori justification for choosing either of these. The second choice would be demanded if ν were a molecular viscosity, but then we would have to resolve a molecular boundary layer which perhaps would be a few millimetres thick. Instead, ν must be interpreted as some form of eddy viscosity, as we discuss in chapter 10. In that case, one might argue that the free-slip condition should be preferred, but in the absence of a proper

theoretical basis for such eddy viscosities there is no truly rational way to make the choice. We will solve the no-slip problem; see exercise 14.4 for the free-slip problem.

Let the wind stress be the canonical $\tau^x = -\cos(\pi y/a)$. Then the interior (Sverdrup) flow is given by (14.34), as for the Stommel problem. This satisfies the free-slip boundary conditions at $y = 0, 1$, namely $\partial_x^2 \psi = 0$, automatically. However, we will need boundary layers at both the western and eastern boundaries, because the interior solution cannot satisfy all four boundary conditions required by (14.42). The eastern boundary layer will be relatively weak, and needed only to satisfy the no-slip or free-slip condition but, as in the Stommel problem, there will be a strong western boundary layer, needed to satisfy the no-normal flow condition. How thick will this be? Inspection of (14.42) suggests that the frictional term and the β term will balance in a boundary layer of thickness of order L_M where

$$L_M = \left(\frac{\nu}{\beta} \right)^{1/3}. \quad (14.43)$$

Non-dimensionalizing (14.42) in a similar way to the Stommel problem yields

$$-\epsilon_M \nabla^4 \hat{\psi} + \frac{\partial \hat{\psi}}{\partial \hat{x}} = \text{curl}_z \hat{\mathbf{T}}_T, \quad (14.44)$$

where $\epsilon_M = (\nu/\beta a^3)$. Considering only the western boundary layer correction, we let the solution be the sum of an interior (Sverdrup) streamfunction plus a boundary layer correction:

$$\hat{\psi} = \psi_I + \phi_W(\alpha, \hat{y}) \quad (14.45)$$

where α is a stretched coordinate such that $\hat{x} = \epsilon \alpha$ where ϵ is some small parameter. Substituting (14.45) into (14.44) and subtracting the Sverdrup balance gives

$$-\epsilon_M \left(\nabla^4 \psi_I + \frac{1}{\epsilon^4} \frac{\partial^4 \phi_W}{\partial \alpha^4} \right) + \frac{1}{\epsilon} \frac{\partial \phi_W}{\partial \alpha} = 0. \quad (14.46)$$

A non-trivial balance is obtained when $\epsilon = \epsilon_M^{1/3}$ (also see problem 14.1), implying a dimensional western boundary thickness of $\epsilon a = (\nu/\beta)^{1/3}$, consistent with (14.43). Equation (14.46) then becomes, at leading order,

$$-\frac{\partial^4 \phi_W}{\partial \alpha^4} + \frac{\partial \phi_W}{\partial \alpha} = 0. \quad (14.47)$$

The boundary conditions on (14.47) are that:

- (i) $\phi_W \rightarrow 0$ as $\alpha \rightarrow \infty$: this states that the perturbation decays as it extends into the interior;
- (ii) $\phi_W = -\psi_I$ at $\hat{x} = \alpha = 0$: this is the no-normal-flow condition on the meridional boundary.
- (iii) $\partial \phi_W / \partial \hat{x} = -\partial \psi_I / \partial \hat{x}$ at $\hat{x} = \alpha = 0$: this is the no-slip condition.

In addition, zonal boundary layers must exist at $\hat{y} = 0, 1$ and another meridional boundary layer must exist at $\hat{x} = 1$, in order to satisfy the no-slip condition. Solving the full problem is a straightforward albeit non-trivial algebraic exercise and, omitting the weak zonal boundary

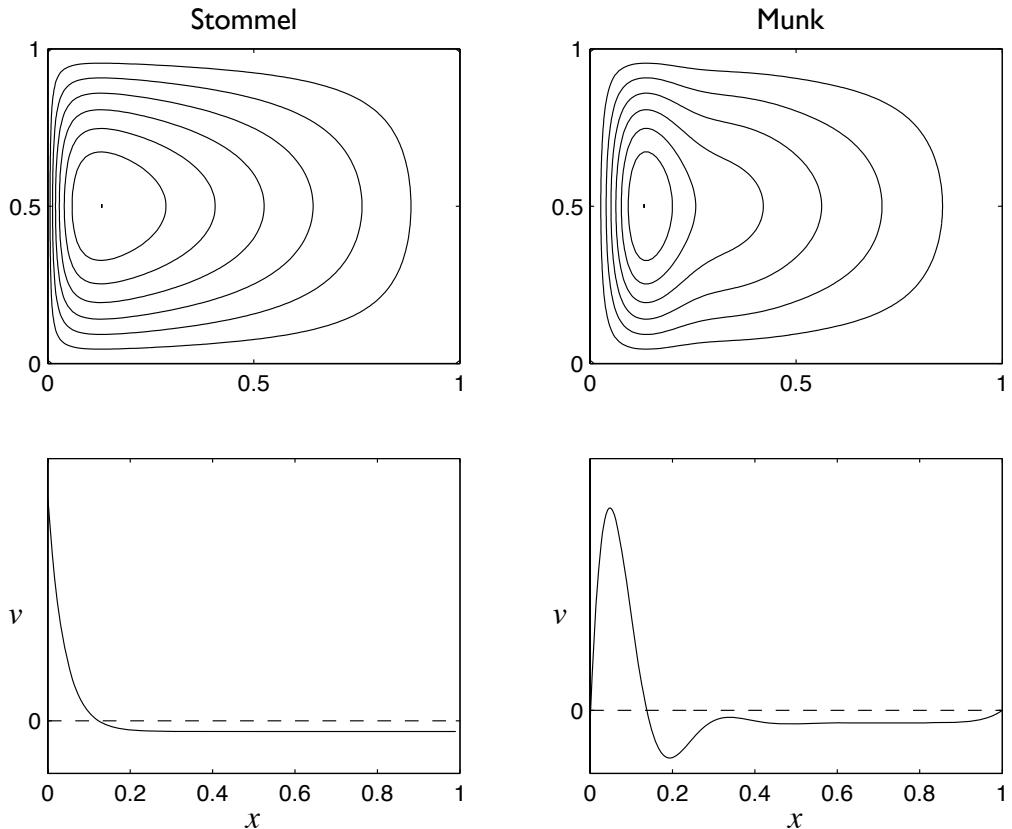


Fig. 14.5 The Stommel and Munk solutions, (14.49) with $\epsilon_S = \epsilon_M^{1/3} = 0.04$, with the wind stress $\tau = -\cos \pi y$, for $x, y \in (0, 1)$. Upper panels are contours of streamfunction in the x - y plane, and the flow is clockwise. The lower panels are plots of meridional velocity, v , as a function of x , in the centre of the domain ($y = 0.5$). The Munk solution can satisfy both no-normal flow and one other boundary condition at each wall, here chosen to be no-slip.

layers at $\hat{y} = 0, 1$ but including the eastern boundary layer correction, we eventually find the solution

$$\hat{\psi} = \psi_I - \psi_I(0, \hat{y}) e^{-\hat{x}/(2\epsilon)} \left[\cos\left(\frac{\sqrt{3}\hat{x}}{2\epsilon}\right) + \frac{1}{\sqrt{3}} \sin\left(\frac{\sqrt{3}\hat{x}}{2\epsilon}\right) \right] - \epsilon e^{(\hat{x}-1)/\epsilon} \frac{\partial}{\partial \hat{x}} \psi_I(1, \hat{y}). \quad (14.48)$$

where $\epsilon = (\nu/\beta a^3)^{1/3}$. With the canonical wind stress (14.27) this solution becomes

$$\hat{\psi} = \pi \sin(\pi \hat{y}) \left\{ 1 - \hat{x} - e^{-\hat{x}/(2\epsilon)} \left[\cos\left(\frac{\sqrt{3}\hat{x}}{2\epsilon}\right) + \frac{1}{\sqrt{3}} \sin\left(\frac{\sqrt{3}\hat{x}}{2\epsilon}\right) \right] + \epsilon e^{(\hat{x}-1)/\epsilon} \right\}. \quad (14.49)$$

The solutions of this are plotted in Fig. 14.5. Note how the Munk layers bring the tangential as well as the normal velocity to zero. The eastern boundary layer has a similar thickness to the western boundary layer, but is not as dynamically important since its *raison d'être*

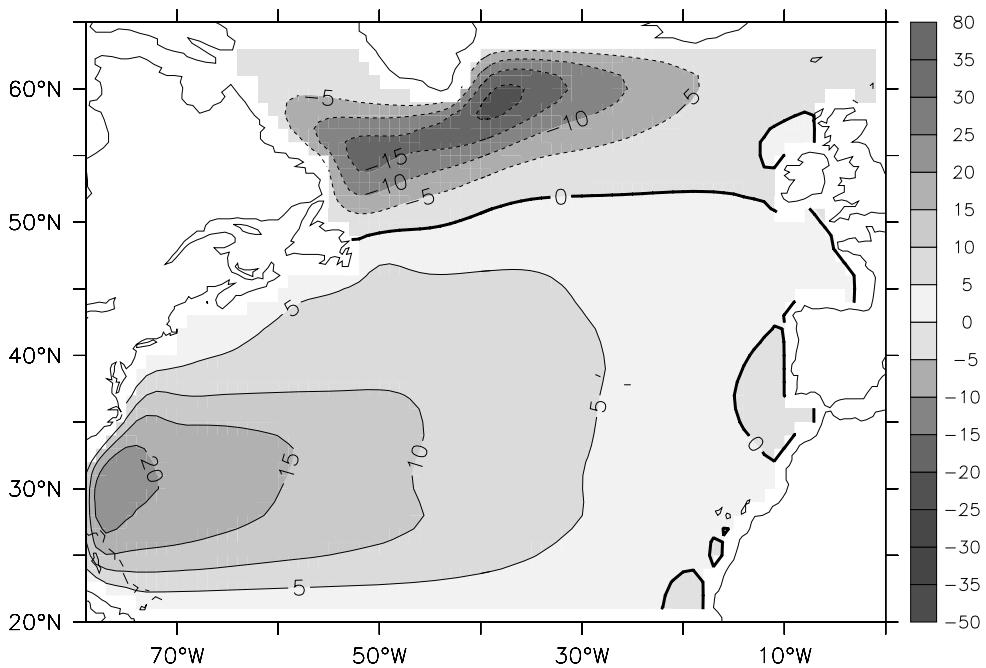


Fig. 14.6 The solution (streamfunction, in Sverdrups) to the Stommel-Munk problem numerically calculated for the North Atlantic, using the observed wind field. The model ocean has realistic geometry, but is flat-bottomed. The calculation qualitatively reproduces the large-scale patterns, including the subtropical and subpolar gyre and the western intensification of both, although the Gulf Stream and its eastward extension are rather diffuse, and its separation from the coast is a little too far north. Compare with Fig. 14.2.

is to enable the no-slip condition to be satisfied, a relatively weak frictional constraint that manifests itself by a boundary layer in which the flow parallel to the boundary is slowed down. On the other hand the western boundary layer exists in order that the no-normal flow condition can be satisfied, which causes a qualitative change in the flow pattern. It should be emphasized that neither the Stommel nor the Munk models are quantitative descriptors of the real ocean, but taken together the similarities of their solutions are a powerful argument for the relative insensitivity of the qualitative form of the solution to the detailed form of the friction. The models do in fact produce qualitatively realistic patterns of large-scale flow in the major basins of the world, as illustrated in Fig. 14.6.

14.3 ZONAL BOUNDARY LAYERS

The canonical wind stress [$\tau^x = -\tau_0 \cos(\pi y/a)$] is special because its curl vanishes at $y = 0$ and a , and so the interior solution satisfies $\psi_I = 0$ at $y = 0$ and a . We cannot expect the real wind to be so accommodating. Consider, then, the problem forced by the linear

wind profile,

$$\tau^x = \frac{\tau_0}{a} \left(y - \frac{a}{2} \right). \quad (14.50)$$

Scaling the variables in the usual way leads to the non-dimensional problem

$$\epsilon_S \nabla^2 \hat{\psi} + \frac{\partial \hat{\psi}}{\partial \hat{x}} = -1. \quad (14.51)$$

The interior flow obeys $\partial \hat{\psi} / \partial \hat{x} = -1$ everywhere, and evidently does not satisfy $\hat{\psi} = 0$ at either $\hat{y} = 0$ or $\hat{y} = 1$; thus, boundary layers are needed at the meridional and zonal boundaries and we let the solution be the sum of five parts,

$$\hat{\psi} = \psi_I + \phi_W + \phi_E + \phi_N + \phi_S, \quad (14.52)$$

with self-explanatory notation. The interior solution, ψ_I , is the solution of $\partial \psi_I / \partial \hat{x} = -1$, so that a solution satisfying $\psi_I = 0$ at $\hat{x} = 1$ is $\psi_I = (1 - \hat{x})$ and there is no need for an eastern boundary layer. By the same methods we used in section 14.1.3 the western boundary layer correction is easily found to be

$$\phi_W = -e^{-\hat{x}/\epsilon_S}. \quad (14.53)$$

It remains to find ϕ_N and ϕ_S , the boundary layer corrections at the northern and southern boundaries.

Consider the boundary layer correction at $y = 1$, and introduce the stretched coordinate α where $\epsilon' \alpha = y - 1$ and where ϵ' is a small parameter. Thus, we let $\phi_N = \phi_N(x, \alpha)$, and on substituting (14.52) into (14.51) we find

$$\epsilon_S \left(\nabla^2 \psi_I + \frac{\partial^2 \phi_N}{\partial \hat{x}^2} + \frac{1}{\epsilon'^2} \frac{\partial^2 \phi_N}{\partial \alpha^2} \right) + \frac{\partial \psi_I}{\partial \hat{x}} + \frac{\partial \phi_N}{\partial \hat{x}} = -1, \quad (14.54)$$

having neglected the small contributions from the other boundary layer streamfunctions (such as ϕ_S). To obtain a non-trivial balance we choose $\epsilon'^2 = \epsilon_S$ and obtain the dominant balance

$$\frac{\partial^2 \phi_N}{\partial \alpha^2} + \frac{\partial \phi_N}{\partial \hat{x}} = 0. \quad (14.55)$$

The boundary conditions necessary to complete the solution are:

- (i) the total streamfunction (interior plus boundary correction) must vanish at the northern boundary; that is, $\phi_N(\hat{x}, \alpha = 0) = -\psi_I(\hat{x}, \hat{y} = 1) = -(1 - \hat{x})$;
- (ii) the boundary solution should match to the interior streamfunction far from the boundary; that is $\phi_N(\hat{x}, \alpha \rightarrow \infty) = 0$;
- (iii) at the eastern boundary ϕ_N should also vanish, else it would provide a velocity into the eastern wall; that is $\phi_N(1, \alpha) = 0$.

(Actually obtaining the solution is quite algebraically tedious.) The solution at the southern boundary is obtained in an analogous way, and the complete solution is illustrated in Fig. 14.7. The non-dimensional thickness of the zonal (northern and southern) boundary layers is $\epsilon_S^{1/2}$, which, because it scales like the half power of a small number, is much thicker than the western boundary current. The thickness arises from the dimensional equations that dominate at the boundary, namely

$$r \frac{\partial^2 \psi}{\partial y^2} + \beta \frac{\partial \psi}{\partial x} = 0. \quad (14.56)$$

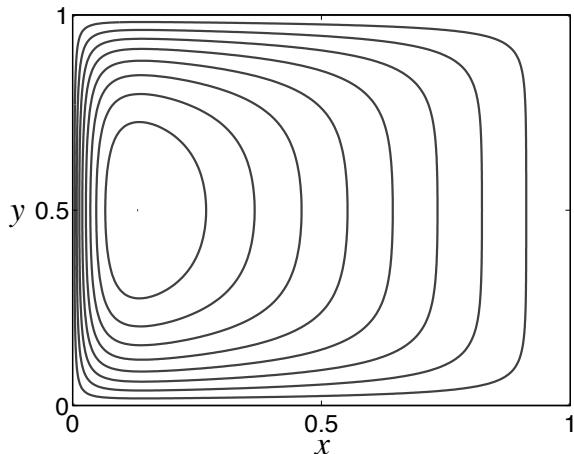


Fig. 14.7 Solutions to the Stommel problem with a wind stress that increases linearly from $y = 0$ to $y = 1$, as in (14.50). The interior solution is $\psi_I = (1 - x)$, or $v_I = -1$, necessitating zonal boundary layers at $y = 0$ and $y = 1$, as well as a western boundary layer at $x = 0$.

This follows from (14.18) by noting that the Laplacian operator must be dominated by the derivatives in the y -direction, and the β term has an (interior) component that annihilates the wind-stress curl, plus a boundary layer correction to balance the Laplacian. Inspection of (14.56) yields a dimensional thickness $L_Z \sim \sqrt{ra/\beta} = \epsilon_S^{1/2} a$, where a is the length scale in the x -direction.

14.4 * THE NONLINEAR PROBLEM

In the nonlinear problem we seek solutions to

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) + \beta \frac{\partial \psi}{\partial x} = -r \nabla^2 \psi + \text{curl}_z \mathbf{\tau}_T + \nu \nabla^2 \zeta, \quad (14.57)$$

which we have written in dimensional form. In the Stommel problem we set $\nu = 0$ and in the Munk problem we set $r = 0$. In general solutions will be time-dependent and turbulent and this will create motion on small scales, so that ν cannot be neglected. The 'steady nonlinear Stommel–Munk problem' is

$$J(\psi, \zeta) + \beta \frac{\partial \psi}{\partial x} = -r \nabla^2 \psi + \text{curl}_z \mathbf{\tau}_T + \nu \nabla^2 \zeta. \quad (14.58)$$

We can scale this by first supposing that the leading order balance is Sverdrupian (i.e., $\beta \partial \psi / \partial x \sim \text{curl}_z \mathbf{\tau}_T$), from which we obtain the scales $\Psi = |\tau| / \beta$ and $U = |\tau| / (\beta L)$. Equation (14.58) may then be non-dimensionalized to yield

$$R_\beta J(\hat{\psi}, \hat{\zeta}) + \frac{\partial \hat{\psi}}{\partial \hat{x}} = -\epsilon_S \nabla^2 \hat{\psi} + \text{curl}_z \hat{\mathbf{\tau}}_T + \epsilon_M \nabla^2 \hat{\zeta}, \quad (14.59)$$

where $R_\beta = U / \beta L^2 = |\tau| / (\beta^2 L^3)$, the β -Rossby number for this problem, is a measure of the nonlinearity. Evidently, the nonlinear term increases in importance with increasing wind stress and for a smaller domain (see also problem 14.9).

14.4.1 A perturbative approach

A direct attack on the full nonlinear problem (14.58) is possible only through numerical methods, so first we shall explore the problem perturbatively, assuming the nonlinear term to be small; the analysis is straightforward, albeit messy. We begin with the Stommel problem, (14.59) with $\epsilon_M = 0$, and expand the streamfunction in terms of R_β ,

$$\hat{\psi} = \psi_0 + R_\beta \psi_1 + \dots \quad (14.60)$$

Now substitute this into (14.59) and equate powers of R_β . The lowest-order problem is simply

$$\epsilon_S \nabla^2 \psi_0 + \frac{\partial \psi_0}{\partial \hat{x}} = \text{curl}_z, \hat{\tau} \quad (14.61)$$

which is the Stommel problem we have already solved. At the next order,

$$\epsilon_S \nabla^2 \psi_1 + \frac{\partial \psi_1}{\partial \hat{x}} = J(\psi_0, \zeta_0). \quad (14.62)$$

This equation has precisely the same form as the Stommel problem, with the known nonlinear term on the right-hand side playing the part of the wind stress. The algebra to obtain the solution is rather tedious, because the right-hand side varies with both \hat{x} and \hat{y} , but this is much ameliorated by the use of computer algebraic manipulation languages. For the canonical wind stress $\tau^x = -\tau_0 \cos(\pi \hat{y})$ the corrected solution, in the boundary layer approximation and ignoring any corrections at the zonal boundaries, is found to be⁹

$$\hat{\psi} \approx \sin(\pi \hat{y})(1 - \hat{x} - e^{-\hat{x}/\epsilon_S}) - \frac{R_\beta \pi^3}{2\epsilon_S^3} \sin(2\pi \hat{y}) \hat{x} e^{-\hat{x}/\epsilon_S}. \quad (14.63)$$

The solution is illustrated in Fig. 14.8. The perturbation is antisymmetric about $\hat{y} = 1/2$, being positive for $\hat{y} > 1/2$ and negative for $\hat{y} < 1/2$. This tends to move the centre of the gyre polewards, narrowing and intensifying the flow in the poleward half of the western boundary current, whereas the western boundary current equatorwards of $\hat{y} = 1/2$ is broadened and weakened. The net effect is that the centre of the gyre is pushed polewards — essentially because the western boundary current is advecting the vorticity of the gyre polewards. In the perturbation solution the advection is both by and of the linear Stommel solution; thus, negative vorticity is advected polewards, intensifying the gyre in its poleward half, weakening it in its equatorward half. The solution illustrated in Fig. 14.8 has $\epsilon = 0.04$ and $R_\beta = 10^{-4}$; for larger values of R_β the perturbation itself starts to dominate.

The problem with this perturbative approach is that a boundary layer solution to the Stommel problem does not calculate derivatives accurately, so that the nonlinear term $J(\psi, \nabla^2 \psi)$ is poorly approximated in the western boundary layer; however, in the interior where the errors are small the perturbative correction is negligible. A more accurate perturbative approach begins with the *exact* solution to the Stommel problem, and then proceeds in the same way. However, the analytic effort is considerable, and the intuitive sense of the way nonlinearities affect the solution is not apparent. (See also problem 14.7.)

14.4.2 A numerical approach

Fully nonlinear solutions show qualitatively similar effects to those seen in the perturbative solutions, as we see in Fig. 14.9, where the solution to (14.59) for the Stommel and Munk problems are obtained numerically by Newton's method.¹⁰

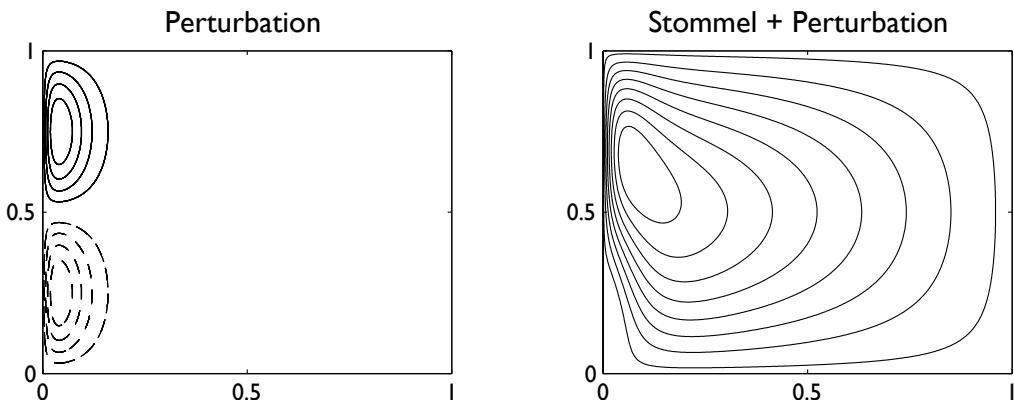


Fig. 14.8 The nonlinear perturbation solution of the Stommel problem, calculated according to (14.63). On the left is the perturbation, $-R_\beta \pi^3 / (2\epsilon_s^2) \sin(2\pi y) x e^{-x/\epsilon_s}$, and on the right is the reconstituted solution, using $R_\beta = 10^{-4}$ and $\epsilon = 0.04$. Dashed contours are negative.

Just as with the perturbative procedure, for both the Stommel and Munk problems small values of nonlinearity lead to the poleward advection of the gyre's anticyclonic vorticity in the western boundary current, strengthening and intensifying the boundary current in the northwest corner. A higher level of nonlinearity results in a strong recirculating regime in the upper westward quadrant, and ultimately much of the gyre's transport is confined to this regime. The western boundary current itself becomes less noticeable as nonlinearity increases, the more nonlinear solutions have a much greater degree of east-west symmetry than the linear ones, just as the fully nonlinear Fofonoff solutions (section 14.5.3).

The qualitative effects above do not depend on the precise formulation of the model, but the boundary conditions do play an important role in the detailed solution. For example, for a given value of R_β , nonlinearity has a stronger effect in the Munk problem with slip boundary conditions than with no-slip, because in the latter the velocity is reduced to zero at the boundary with a corresponding reduction in the advection term. However, these solutions themselves are unlikely to be relevant for larger values of nonlinearity, because then the flow becomes hydrodynamically unstable

14.5 * INERTIAL SOLUTIONS

In this section we further explore inertial effects and ask: might purely inertial effects be sufficient to satisfy boundary conditions at the western boundary? Can we envision a purely inertial gyre circulation? Our question is motivated by the steady wind-driven, Rayleigh-damped quasi-geostrophic equation, namely

$$J(\psi, \nabla^2 \psi) + \beta \frac{\partial \psi}{\partial x} = F_\tau - r \nabla^2 \psi, \quad (14.64)$$

where F_τ is the wind forcing. Since the inertial (advective) terms are of a higher order than the linear terms, indeed they are of a higher order than the Rayleigh drag, it is natural

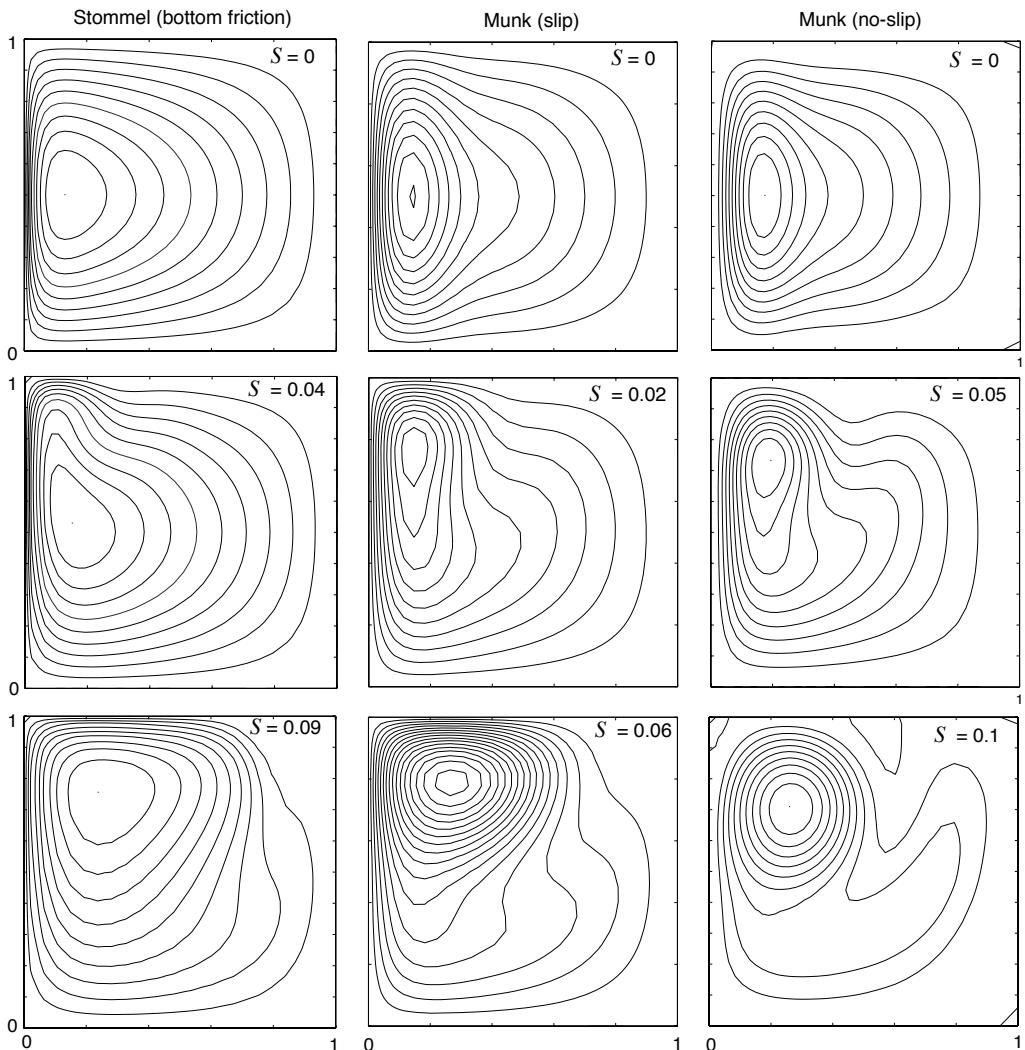


Fig. 14.9 Streamfunctions in solutions of the nonlinear Stommel and Munk problems, obtained numerically with a Newton's method, for various values of the nonlinearity parameter $S = R_\beta^{1/2}$. As in the perturbation solution, for small values of nonlinearity the centre of the gyre moves polewards, strengthening the boundary current in the north-western quadrant (for a northern-hemisphere solution). As nonlinearity increases, the recirculation of the gyre dominates, and the solutions become increasingly inertial.¹¹

to wonder if they themselves might serve to satisfy the no-normal flow condition on the western boundary, without recourse to rather ill-defined frictional terms. The answer is no, as we see below, but nevertheless nonlinear effects may be important in the western boundary layer even if they are small in the interior.

14.5.1 Roles of friction and inertia

The need for friction

Consider the steady barotropic flow satisfying

$$\mathbf{u} \cdot \nabla q = \text{curl}_z \boldsymbol{\tau}_T + Fr, \quad (14.65)$$

where $q = \nabla^2 \psi + \beta y$ and Fr represents frictional effects. Noting that \mathbf{u} is divergence-free and integrating the left-hand side of (14.65) over the area, A , between two closed streamlines (ψ_1 and ψ_2 , say) and using the divergence theorem we find

$$\int_A \nabla \cdot (\mathbf{u}q) \, dA = \oint_{\psi_1} \mathbf{u}q \cdot \mathbf{n} \, dl - \oint_{\psi_2} \mathbf{u}q \cdot \mathbf{n} \, dl = 0. \quad (14.66)$$

Here, \mathbf{n} is the unit vector normal to the streamline so that $\mathbf{u} \cdot \mathbf{n} = 0$. The integral of the wind-stress curl over the same area will not, in general, be zero. Now, we can take these two streamlines as close together as we wish; thus, a balance in (14.65) can only be achieved if *every* closed contour passes through a region where frictional effects are non-zero. This does not mean that nonlinear terms may not locally dominate the friction, just that friction must be somewhere important. In the Stommel and Munk problems, it means that every streamline must pass through the frictional western boundary layer.

Frictional and inertial scales

The ratio of the size of the nonlinear terms to the linear terms is given by the β -Rossby number, $R_\beta = U/(\beta L^2)$. In the western boundary layer the length scales can be expected to be much smaller than the basin scale, and if the balance in the western boundary layer were between the nonlinear and beta term, as in

$$u \frac{\partial}{\partial x} \nabla^2 \psi \sim \beta \frac{\partial \psi}{\partial x}, \quad (14.67)$$

then the *inertial boundary layer thickness*, δ_I , is given by

$$\delta_I = \left(\frac{U}{\beta} \right)^{1/2}. \quad (14.68)$$

This, of course, gives $R_\beta = 1$ if $L = \delta_I$. The more energetic the flow, the wider the region where nonlinearity is important, and the corresponding scale is sometimes called the Charney thickness.¹²

The linearized Stommel equation has a boundary layer of dimensional thickness of order $\delta_S = (r/\beta)$, obtained by equating $\beta \partial \psi / \partial x$ and $r \nabla^2 \psi$. This thickness is equal to the inertial boundary layer thickness when $U = (r^2/\beta)$. If $\delta_I > \delta_S$, i.e., if $U > (r^2/\beta)$, then nonlinearity *must* be important in the western boundary layer, because the nonlinear terms are at least

as important as the beta term in (14.64). On the other hand, if the Stommel boundary layer is wider than the inertial boundary layer, there is no obvious need for nonlinearity to be important, since the Stommel boundary layer generates no length scales smaller than δ_I and R_β remains small.

14.5.2 Attempting an inertial western boundary solution

Although friction must be important, it is nevertheless instructive to try to find a purely inertial solution for the western boundary layer, and to see if and how the attempt fails. Let us suppose that the interior solution has purely zonal flow, with flow either towards or away from the western boundary current, and consider the dimensional equation of motion

$$J(\psi, \nabla^2 \psi + \beta y) = 0. \quad (14.69)$$

This has the general solution

$$\nabla^2 \psi + \beta y = G(\psi), \quad (14.70)$$

where G is an arbitrary function of its argument. Consider first the case with flow entering the boundary layer with a local velocity $-U$ (i.e., westwards); that is, $\psi = Uy$. The potential vorticity in this region (just outside our putative boundary layer) is $Q_I = \beta y$. Thus, the relation between potential vorticity and streamfunction is $Q_I = \beta \psi_I / U$ and so

$$G(\psi) = \frac{\beta \psi}{U}. \quad (14.71)$$

Because we are assuming the flow is inviscid, the fluid will preserve this relationship between potential vorticity and streamfunction *even as it moves through the western boundary layer*. Thus, using (14.70) and (14.71), the flow in the interior *and* in the boundary layer are given by solutions of

$$\nabla^2 \psi - \frac{\beta \psi}{U} = -\beta y, \quad (14.72)$$

with $\psi = 0$ on the boundary. To obtain a solution, we let $\psi = \psi_I + \phi$, where $\psi_I = Uy$ is the particular solution to (14.72) (and, of course, the interior flow). The boundary layer correction then obeys

$$\nabla^2 \phi - \frac{\beta \phi}{U} = 0, \quad (14.73)$$

with $\phi = -\psi_I$ at $x = 0$. Now, in the boundary layer, length scales in the x -direction are much smaller than length scales in the y -direction, and so (14.73) becomes approximately

$$\frac{\partial^2 \phi}{\partial x^2} - \frac{\beta \phi}{U} = 0. \quad (14.74)$$

(We could formalize this procedure by non-dimensionalizing and introducing a stretched coordinate, as we did for the Stommel problem.) Solutions of (14.74) are $\phi = -\psi_I e^{-x/\delta_I}$, where $\delta_I = (U/\beta)^{1/2}$, and so the full solution is

$$\psi = \psi_I (1 - e^{-x/\delta_I}). \quad (14.75)$$

Clearly, this solution smoothly transitions into the interior solution for large x .

What about solutions exiting the boundary layer? We might attempt a similar procedure, but now the interior boundary condition that we must match is that of eastward flow, namely $\psi_I = -Uy$. Thus, analogously (14.72), to we seek solutions to the problem,

$$\nabla^2 \psi + \frac{\beta \psi}{U} = -\beta y \quad (14.76)$$

for which the homogeneous (boundary layer) equation is

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\beta \phi}{U} = 0. \quad (14.77)$$

Solutions of (14.77) are [cf. (14.75)]

$$\psi = \psi_I(1 - e^{-x/\delta_I^*}), \quad (14.78)$$

where $\delta_I^* = i\delta_I = i(U/\beta)^{1/2}$. The solution is therefore wave-like, and does not transition smoothly to the interior flow. This suggests that friction must be important in allowing the boundary layer to smoothly connect to the interior for reasons connected to the conservation of potential vorticity, as explained heuristically in the next subsection.¹³

In addition to this transition problem, we note that (14.75) and (14.78) together do not constitute a globally acceptable inviscid solution, for the simple reason that the flow in the westwards flowing interior region has a different potential vorticity-streamfunction (q - ψ) relationship than does the eastwards flowing interior. If these two regions are connected by a boundary layer, the flow in that boundary layer must be viscous or unsteady since the value of potential vorticity on the streamlines has changed, whereas inviscid flow conserves potential vorticity.

The connection between the boundary layer and the interior

We have seen that solutions with a boundary layer character that blend smoothly to a flow interior exist only for westward interior flow ($u_I = -U < 0$): an inertial western boundary layer on a beta-plane evidently cannot easily release fluid into the interior. The underlying reason for this stems from the conservation of potential vorticity, as we now explain (and see Fig. 14.10).

Conservation of potential vorticity demands that, in barotropic flow, $v_x - u_y + \beta y$ is a constant on streamlines. In the interior, relative vorticity, ζ , is zero, and in the meridional boundary layers it is effectively v_x . Consider first the case with westward flow in the interior (left panel of Fig. 14.10). Fluid from the irrotational interior approaches the western boundary where it is deflected either polewards or equatorwards. In the former case its relative vorticity falls, so allowing potential vorticity to be conserved because the reduction of ζ can be balanced by an increase in the planetary vorticity, f . Similarly, flow deflected southwards produces positive relative vorticity, which is compensated for by the reduced value of f . In the eastern boundary layer, southwards (northwards) moving flow has negative (positive) relative vorticity. As it emerges into the interior its relative vorticity increases (decreases), this being balanced by a fall (rise) in the value of f . Thus, we see that the solution with a westwards flowing interior can indeed conserve potential vorticity, at both eastern and western boundaries.

On the other hand suppose the interior flow were eastwards (right panel of Fig. 14.10). Flow moving polewards in the western boundary layer has negative relative vorticity. It

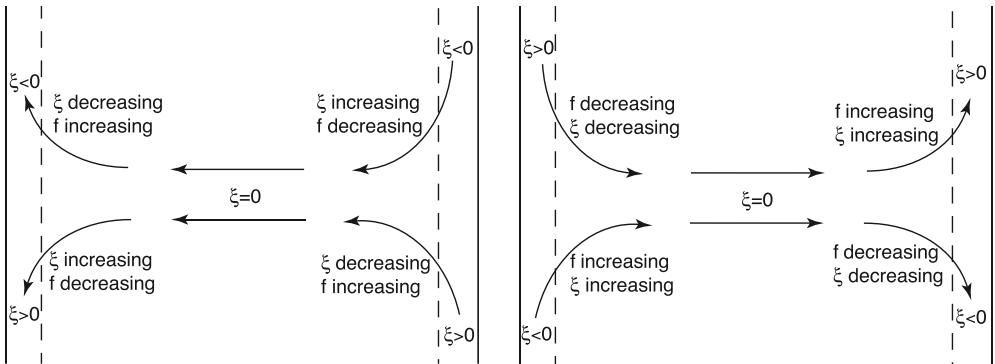


Fig. 14.10 Putative inertial boundary layers connected to a westwards flowing interior flow (left panel) or eastwards flowing interior flow (right panel), in the Northern Hemisphere. Westward flow into the western boundary layer, or flow emerging from an eastern boundary layer, is able to conserve its potential vorticity through a balance between changes in relative vorticity and Coriolis parameter. But flow cannot emerge smoothly from a western boundary layer into an eastwards flowing interior and still conserve its potential vorticity. The right panel thus has inconsistent dynamics.

cannot be freely released into an irrotational interior because f and ζ would then both need to increase, violating potential vorticity conservation. Flow moving southwards with positive relative vorticity similarly is trapped within the western boundary current, unless it meets a zonal boundary which allows a eastwards moving boundary current with positive relative vorticity. Similar arguments show that an eastern boundary current cannot entrain fluid from an eastwards flowing irrotational interior.

One might ask, why cannot we simply reverse the trajectory of all the fluid parcels in an inviscid flow and thereby obtain a solution with an eastward-flowing interior? The answer is that such a flow will only be a solution if we also reverse the direction of the Earth's rotation, and so reverse the sign of β , and hence the location of the frictional boundary current.

14.5.3 A fully inertial approach: the Fofonoff model

Rather than attempt to match an inertial boundary layer with an interior Sverdrup flow, we may look for a purely inertial solution that holds basinwide, and such a construction is known as the *Fofonoff model*.¹⁴ That is, we seek global solutions to the inviscid, unforced problem,

$$J(\psi, \nabla^2 \psi + \beta y) = 0. \quad (14.79)$$

We should not regard this problem as representing even a very idealized wind-driven ocean; rather, we may hope to learn about the properties of purely inertial solutions and this might, in turn, tell us something about the ocean circulation.

The general solution to (14.79) is

$$\nabla^2 \psi + \beta y = Q(\psi), \quad (14.80)$$

where $Q(\psi)$ is an arbitrary function of its argument. For simplicity we choose the linear

form,

$$Q(\psi) = A\psi + B, \quad (14.81)$$

where $A = \beta/U$ and $B = \beta y_0$, where U and y_0 are arbitrary constants. Thus, (14.80) becomes

$$\left(\nabla^2 - \frac{\beta}{U}\right)\psi = \beta(y_0 - y). \quad (14.82)$$

We will further choose $\beta/U > 0$, which we anticipate will provide a westward-flowing interior flow, and which (from our experience in the previous section) is more likely to provide a meaningful solution than an eastward interior, and we will use boundary-layer methods to find a solution. A natural scaling for ψ is UL , where L is the domain size, and with this the non-dimensional problem is

$$(\epsilon_F \nabla^2 - 1)\hat{\psi} = \hat{y}_0 - \hat{y}, \quad (14.83)$$

where $\epsilon_F = U/(\beta L^2)$ and $\hat{y} = y/L$. If we take ϵ_F to be small (note that $\epsilon_F = R_\beta$) we can find a solution by boundary-layer methods similar to those used in section 14.1.3 for the Stommel problem. Thus, we write $\hat{\psi} = \hat{\psi}_I + \hat{\phi}$ where $\hat{\psi}_I = \hat{y} - \hat{y}_0$ [dimensionally, $\psi_I = U(y - y_0)$] and $\hat{\phi}$ is the boundary layer correction, to be calculated separately for each boundary using

$$\epsilon_F \nabla^2 \hat{\phi} - \hat{\phi} = 0 \quad (14.84)$$

and the boundary condition that $\hat{\phi} + \hat{\psi}_I = 0$. For example, at the northern boundary, $\hat{y} = \hat{y}_N$, the y -derivatives will dominate and (14.84) may be approximated by

$$\epsilon_F \frac{\partial^2 \hat{\phi}}{\partial \hat{y}^2} - \hat{\phi} = 0, \quad (14.85)$$

with solution

$$\hat{\phi} = B \exp[-(\hat{y}_N - \hat{y})/\epsilon_F^{1/2}], \quad (14.86)$$

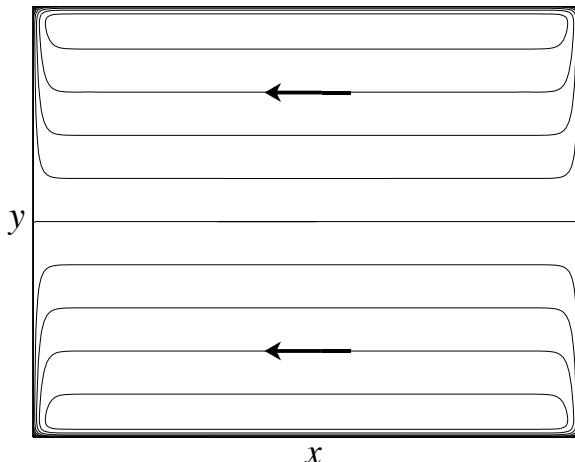
where $B = \hat{y}_0 - \hat{y}_N$, hence satisfying the boundary condition that $\hat{\phi}(\hat{x}, \hat{y}_N) = -\hat{\psi}_I(\hat{x}, \hat{y}_N)$. We follow a similar procedure at the other boundaries to obtain the full solution, and in dimensional form this is

$$\psi = U(y - y_0) \left[1 - e^{-x/\delta_I} - e^{-(x_E - x)/\delta_I} \right] + U(y_0 - y_N) e^{-(y_N - y)/\delta_I} + U y_0 e^{-y/\delta_I}. \quad (14.87)$$

where $\delta_I = (U/\beta)^{1/2}$ is the boundary layer thickness. Evidently, only positive values of U corresponding to a westward interior flow give boundary-layer solutions that decay into the interior.

A typical solution is illustrated in Fig. 14.11. On approaching the western boundary layer, the interior flow bifurcates at $y = y_0$. The western boundary layer, of width δ_I , accelerates away from this point, being constantly fed by the interior flow. The westward return flow occurs in zonal boundary layers at the northern and southern edges, also of width δ_I . Flow along the eastern boundary layers is constantly being decelerated, because it is feeding the interior. If one of the zonal boundaries corresponds to y_0 (e.g., if $y_N = y_0$) there would be no boundary layer along it, since ψ is already zero at $y = y_0$. Rather, there would be westward flow along it, just as in the interior. Indeed, a slippery wall placed at $y = 0.5$ would have no effect on the solution illustrated in Fig. 14.11.

Fig. 14.11 The Fofonoff solution. Plotted are contours (streamlines) of (14.87) in the plane $0 < x < x_E$, $0 < y < y_N$ with $U = 1$, $y_N = 1$, $x_E = y_N = 1$, $y_0 = 0.5$ and $\delta_I = 0.05$. The interior flow is westwards everywhere, and $\psi = 0$ at $y = y_0$. In addition, boundary layers of thickness $\delta_I = \sqrt{U/\beta}$ bring the solution to zero at $x = (0, x_E)$ and $y = (0, y_N)$, excepting small regions at the corners.



14.6 TOPOGRAPHIC EFFECTS ON WESTERN BOUNDARY CURRENTS

The above sections have emphasized the role of friction in satisfying the boundary conditions in the west. However, we should certainly not think of friction as being the *cause* of the western boundary layer and in this section we shall show that if there are sloping sidewalls the role of friction is significantly different, and the western boundary current may even be largely inviscid!¹⁵ The key point is that the flow may be inviscid if it is able to follow potential vorticity contours. In a flat-bottomed western boundary layer the flow is moving to larger values of f (and not as a direct response to the wind) and so the flow *must* be frictional. However, if the sidewalls are sloping, then the flow may preserve its potential vorticity (approximately, its value of f/h) if it moves offshore as it moves polewards. In the treatment below, we will focus on homogeneous fluids, noting that the interaction of topography and stratification is a subtle and rather complex problem.

14.6.1 Homogeneous model

The potential vorticity evolution equation for a homogeneous model with topography and a rigid lid [cf. (14.7)] may be written as:

$$\frac{Dq}{Dt} = \frac{F}{h}, \quad (14.88)$$

where $q = (\zeta + f)/h$ where $h = h(x, y)$ is the time-independent depth of the fluid, and F represents vorticity forcing and frictional terms. The advecting velocity is determined by noting that the mass conservation equation is just $\nabla \cdot [\mathbf{u}h(x, y)] = 0$, which allows us define the mass-transport streamfunction ψ such that

$$u = -\frac{1}{h} \frac{\partial \psi}{\partial y}, \quad v = \frac{1}{h} \frac{\partial \psi}{\partial x}. \quad (14.89)$$

The streamfunction itself is obtained from the vorticity by solving the elliptic equation,

$$\nabla \cdot \left(\frac{1}{h} \nabla \psi \right) = \zeta = qh - f. \quad (14.90)$$

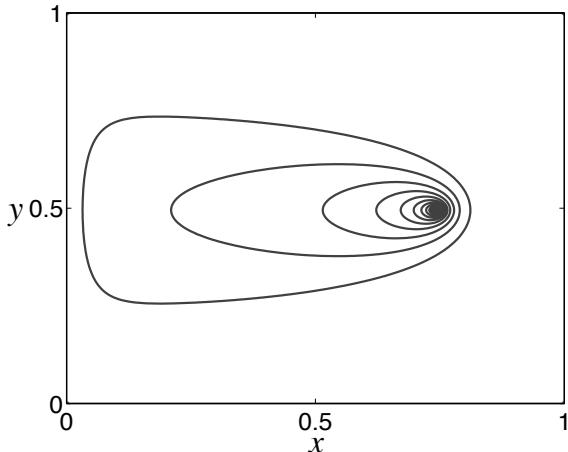


Fig. 14.12 The β -plume, namely Green's function for the Stommel problem. Specifically we plot the solution of (14.93) with $\psi = 0$ at the walls, and a delta-function source at $x = 0.75$, $y = 0.5$. The streamfunction trails westwards from the source, as if it were a tracer being diffused while being advected westwards along lines of constant f .

Equations (14.88), (14.90) and (14.89) form a closed system. Unlike the quasi-geostrophic case, neither the Rossby number nor the topography need be small for the model to be valid. Including finite size topography but not stratification is not especially realistic *vis-à-vis* the real ocean, but nevertheless this model is physically realizable and a useful tool.

The ‘topographic Stommel problem’ is obtained by neglecting relative vorticity in the potential vorticity in (14.88) (i.e., let $q = f/h$) and using a linear drag acting on the vertically integrated fields for friction, and a wind-stress curl forcing. Multiplying (14.88) by h , expanding the advective term and omitting time dependence gives

$$J\left(\psi, \frac{f}{h}\right) = -r\nabla^2\psi + \text{curl}_z(\boldsymbol{\tau}_T/h), \quad (14.91)$$

where the boundary conditions are $\psi = 0$ at the domain edges and $\boldsymbol{\tau}_T$ represents the wind stress at the top.

14.6.2 Advection dynamics

An illuminating way to begin to study the problem is to write (14.91) in the form¹⁶

$$J(\Psi, \psi) = +r\nabla^2\psi - \text{curl}_z(\boldsymbol{\tau}_T/h), \quad (14.92)$$

where $\Psi \equiv f/h$. Thus, noting that $J(\Psi, \psi) = \mathbf{U} \cdot \nabla \psi$ where $\mathbf{U} = \mathbf{k} \times \nabla(f/h)$, we regard ψ as being advected by the pseudovelocity \mathbf{U} . This advection is along f/h contours and is quasi-westwards, meaning that high values of potential vorticity lie to the right. Equation (14.92) is then an advection-diffusion equation for the tracer ψ , with the ‘source’ of ψ being the wind stress curl with ψ being diffused by the first term on the right-hand side of (14.92), and advected by \mathbf{U} . This same interpretation applies to the original Stommel problem, of course, where the pseudovelocity, $-\beta\mathbf{i}$, is purely westwards, and it is useful to first revisit this problem.

Consider, then, a flat-bottomed ocean, where the wind-stress curl is just a point source at \mathbf{x}_0 . With $\Psi = f$, (14.92) becomes

$$r\nabla^2\psi + \beta \frac{\partial\psi}{\partial x} = \delta(\mathbf{x} - \mathbf{x}_0). \quad (14.93)$$

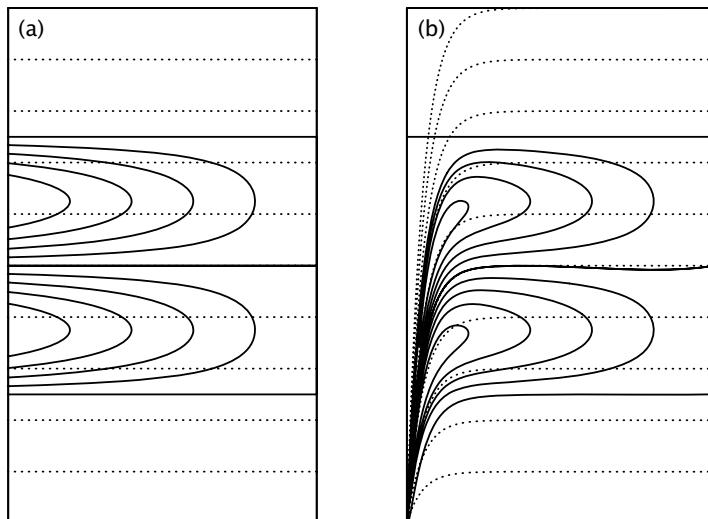


Fig. 14.13 The two-gyre Sverdrup flow (solid contours) for (a) a flat-bottomed domain and (b) a domain with sloping sidewalls. The f/h contours are dotted.¹⁷

This can be transformed to a Helmholtz equation by writing $\phi = \psi \exp(\beta x/2r)$, giving $\nabla^2 \phi - [\beta/(2r)]^2 \phi = \delta(\mathbf{x} - \mathbf{x}_0)$. This may then be solved exactly, and the solution (for ψ) is illustrated in Fig. 14.12 — this is Green's function for the Stommel problem. The tracer ψ is ‘adverted’ westwards along f/h contours — lines of latitude in this case — spreading diffusively as it goes; the resulting structure is called a β -plume. The western boundary layer results as a consequence of the f/h contours colliding with the western boundary, along with the need to satisfy the boundary condition $\psi = 0$. If there were no diffusion at all, ψ would just propagate westwards from the source, and if in addition the source were spatially distributed, for example as $\sin \pi y$, the solution streamfunction would represent Sverdrup interior flow. This case is illustrated in Fig. 14.13a; this is the solution to (14.93) with $r = 0$ and with the right-hand side replaced by a conventional ‘two-gyre’ wind stress.

Now consider the case with a sloping sidewall on the western boundary. The f/h contours (the dotted lines in Fig. 14.13b) tend to converge at the southwest corner of the domain, and only where f/h contours intersect the boundary is a diffusive boundary layer required. In terms of the interpretation above, wind stress provides a source for the streamfunction ψ and the latter is advected *pseudowestwards* — i.e., along potential vorticity contours, with higher values to the right. The source in this case is distributed over the entire domain, but the contours all converge in the southwest corner. The (numerically obtained) solution to the associated Stommel problem is illustrated in Fig. 14.14, and the western boundary current in this case is no longer a frictional boundary layer. Friction is necessarily important where the flow crosses f/h contours; linear theory suggests that this will occur at the southwest corner. It also occurs on the western boundary where the f/h contours are densely packed and the vorticity in the topographic Sverdrup flow is large, and the friction enables the flow to move across the f/h contours. (In the flat bottomed case the f/h contours are zonal and friction allows the flow to move meridionally.)

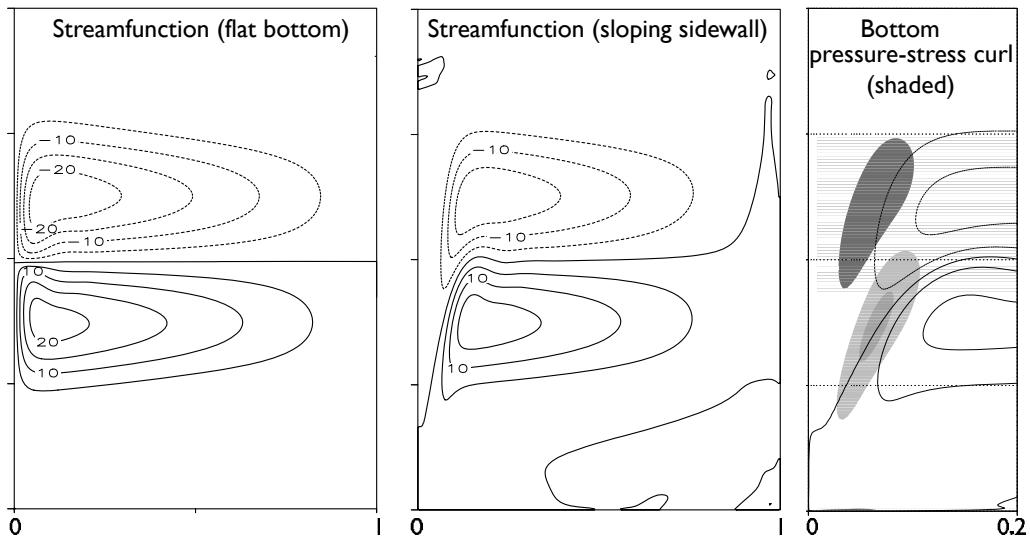


Fig. 14.14 The numerically obtained steady solution to the homogeneous problem with a two-gyre forcing and friction, for a flat-bottomed domain and a domain with sloping western sidewall. The shaded regions in the right panel show the regions where bottom pressure-stress curl is important in the meridional flow of the western boundary currents.¹⁸

14.6.3 Bottom pressure stress and form drag

In the homogeneous problem f and h appear only in the combination f/h , and we may solve the problem entirely without considering pressure effects. It is nevertheless informative to think about how the pressure interacts with the topography to produce meridional flow, and as way toward addressing the effects of stratification. The geostrophic momentum equation is

$$\mathbf{f} \times \mathbf{u} = -\nabla \phi + \mathbf{F}, \quad (14.94)$$

where \mathbf{F} represents both wind forcing and frictional terms. Integrating this over the depth of the ocean (with $z = 0$ at the top), and using the Leibnitz rule ($\nabla \int_{\eta_B}^0 \phi \, dz = \int_{\eta_B}^0 \nabla \phi \, dz - \phi_B \nabla \eta_B$) to evaluate the pressure term, gives

$$\mathbf{f} \times \bar{\mathbf{u}} = -\nabla \bar{\phi} - \phi_B \nabla \eta_B + \bar{\mathbf{F}}. \quad (14.95)$$

Here, the overbar denotes a vertical integral (e.g., $\bar{\mathbf{u}} = \int_{\eta_B}^0 \mathbf{u} \, dz$), ϕ_B is the pressure at $z = \eta_B$ and η_B is the z -coordinate of the bottom topography. We take the top of the ocean at $z = 0$, and note that $\nabla \eta_B = -\nabla h$ where h is the fluid thickness. The second term on the right-hand side of (14.95) is the stress on the fluid due to a correlation between the pressure gradient and the topography at the ocean bottom — it is the bottom *form drag*, first encountered in section 3.5.

Taking the curl of (14.95) gives

$$\beta \bar{v} = -\text{curl}_z(\phi_B \nabla \eta_B) + \text{curl}_z \bar{\mathbf{F}} = -J(\phi_B, \eta_B) + \text{curl}_z \bar{\mathbf{F}}. \quad (14.96)$$

This equation holds for both a stratified and homogeneous fluid. The first term on the

right-hand side is the bottom pressure-stress curl, or the form-drag curl. (It is also sometimes informally referred to as the bottom pressure torque.)

Equation (14.96) is similar to (14.91), in that both arise from (14.95). To derive an equation with the same form as (14.91) but valid for a stratified fluid, we write the vertical integral of the pressure as

$$\int_{-h}^0 \phi \, dz = \int_{-h}^0 [d(\phi z) - z(\partial \phi / \partial z) \, dz] = \int_{-h}^0 [d(\phi z) - z b \, dz] = h \phi_B + \Gamma, \quad (14.97)$$

using hydrostatic balance, and where $\Gamma \equiv - \int_{-h}^0 z b \, dz$. Using (14.97) in (14.95) we obtain:

$$\mathbf{f} \times \bar{\mathbf{u}} = -h \nabla \phi_B + \nabla \Gamma + \bar{\mathbf{F}}. \quad (14.98)$$

The curl of this equation just gives back (14.96), but if we divide by h before taking the curl we obtain

$$J(\psi, f/h) + J(h^{-1}, \Gamma) = \text{curl}_z(\bar{\mathbf{F}}/h). \quad (14.99)$$

The second term in (14.99) is known as the JEBAR term — joint effect of baroclinicity and relief — and it couples the depth integrated flow with the baroclinic flow.¹⁹ Thus, if the bottom is not flat, the Stommel-Munk models are not solutions for the vertically integrated flow. If the stratification vanishes (i.e., b is constant) then Γ is function of h alone and $J(h^{-1}, \Gamma) = 0$, and (14.99) reprises (14.91), given an appropriate choice of $\bar{\mathbf{F}}$.

Bottom pressure stress in a homogeneous gyre

For the remainder of this section we restrict attention to a homogeneous (i.e., unstratified) gyre. If there is no forcing or friction, then from (14.96) and (14.98) we see that the flow simultaneously satisfies

$$\beta v = -\nabla \phi_B \times \nabla \eta_B, \quad \text{and} \quad J(\psi, f/h) = 0. \quad (14.100a,b)$$

The right-hand side of (14.100a), the form-drag curl, is non-zero whenever the pressure gradient has a component parallel to the topographic contour. From (14.100) we may conclude that if there is meridional flow in an unforced, inviscid fluid it must be along f/h contours, and this this meridional flow may be thought of as being driven the curl of the form drag. If the domain is flat-bottomed then the form drag is zero, and in that case all meridional flow is forced or viscous.

Real flows are, of course, both forced and viscous. Bottom stresses may — and likely do — locally dominate viscous stresses. However, the bottom pressure-stress curl cannot balance the wind-stress curl when integrated over the whole domain, or indeed when integrated over an area bounded by a line of constant ϕ_B or constant η_B , because its integral over such an area vanishes and (14.91) cannot be balanced if $r = 0$. In the numerical simulations shown in Fig. 14.14, it is the bottom pressure-stress curl term that largely balances the poleward flow term (βv) in the vorticity equation in parts of western boundary current, with friction acting in the opposite sense. That is, over some regions where the flow is crossing f/h contours we have the balance

$$[\beta v] \approx [\text{bottom pressure-stress curl}] - [\text{friction}], \quad (14.101)$$

where the terms in square brackets are positive, and friction is small. In contrast, in the

flat-bottomed case in the western boundary layer we have the classical balance $[\beta v] \approx +[\text{friction}]$, with both terms positive.

Now consider the balance of momentum, integrated zonally across the domain. We write the vorticity equation (14.96) in the form

$$\nabla \cdot (f \bar{\mathbf{v}}) = -\text{curl}_z(\phi_B \nabla \eta_B) + \text{curl}_z \boldsymbol{\tau}_T - \text{curl}_z \boldsymbol{\tau}_B, \quad (14.102)$$

where $\boldsymbol{\tau}_T$ is the wind stress at the top and $\boldsymbol{\tau}_B$ the frictional stress at the bottom. Integrate (14.102) over the area of a zonal strip bounded by two nearby lines of latitude, y_1 and y_2 , and the coastlines at either end. The term on the left-hand side vanishes by mass conservation and using Stokes' theorem we obtain:

$$\int_{y_1} \phi_B \frac{\partial \eta_B}{\partial x} dx - \int_{y_2} \phi_B \frac{\partial \eta_B}{\partial x} dx = \int_{y_1} (\tau_T^x - \tau_F^x) dx - \int_{y_2} (\tau_T^x - \tau_F^x) dx. \quad (14.103)$$

If the topography is non-zero, there is nothing in this equation to prevent the wind stress being balanced by the form stress terms, with the friction being a negligible contribution. If, for example, friction were to be confined to the southwest corner, then bottom pressure stress is the proximate driver of fluid polewards in the western boundary current. This may hold only if the scale of the sloping sidewall is greater than the thickness of the Stommel layer; if the converse holds then the sidewalls appear to be essentially vertical to the flow. If the sidewalls are truly vertical, then the form stress is confined to delta-functions at the walls. Friction must then be important even in the zonal balance, because if we restrict the integral in (14.103) to a strip that does not quite reach the sidewalls, the left-hand side vanishes identically and the wind stress can only be balanced by friction.

To conclude this discussion, we note that the effects of topography are likely greater in homogeneous fluids than in stratified fluids, because the stratification may partially shield the wind-driven upper ocean from feeling the topography, but we leave the exploration of that topic for another day.

14.7 * VERTICAL STRUCTURE OF THE WIND-DRIVEN CIRCULATION

We now examine the vertical structure of the wind-driven flow.²⁰ We pose the problem using the quasi-geostrophic equations, taking the overall stratification of the ocean as a given. Of course, the production of the stratification is itself one of the most important problems of physical oceanography, but we leave that to the next two chapters.

14.7.1 A two-layer quasi-geostrophic Model

Scales of motion

We will concern ourselves with scales that are sufficiently larger than the deformation radius that we can ignore the relative vorticity relative to vortex stretching and the β -effect. We must be careful in so doing, because quasi-geostrophic scaling applies only to scales that are not significantly larger than the deformation scale; thus, our analysis will be formally valid under the following set of inequalities.

$$\beta L \ll f_0 \quad (\text{small variations in Coriolis parameter})$$

$$\begin{aligned}\beta L > U/L & \quad (\text{to ignore relative vorticity compared to planetary vorticity}) \\ L^2 > L_d^2 & \quad (\text{to ignore relative vorticity compared to vortex stretching}) \\ Ro L^2 \ll L_d^2 & \quad (\text{to keep the variations in stratification small}),\end{aligned}$$

where L_d is the deformation radius and L the scale of the motion. Since in the mid-latitude ocean $L_d \approx 10^5$ m, then the above inequalities are reasonably well satisfied for $L \approx 10^6$ m and $U \approx 0.1$ m s $^{-1}$ with $\beta = 10^{-11}$ m $^{-1}$ s $^{-1}$ and $f_0 = 10^{-4}$ s $^{-1}$. The first and last of the inequalities are standard quasi-geostrophic requirements, with the '> \gg ' symbol denoting the asymptotic ordering. The middle two inequalities are taken within the quasi-geostrophic dynamics, and are needed in order to ignore relative vorticity and give a balance between the β -effect and vortex stretching. The simultaneous satisfaction of all these conditions may seem restrictive, but the plangent dynamics contained within the quasi-geostrophic equations and the generality of the method employed below will suggest that the principle results obtained may transcend the particular limitations of the equations used.

Constructing the model

We now make the following simplifications for our model ocean.

- (i) We use the two-layer quasi-geostrophic equations, with layers of equal thickness.
- (ii) We seek only statistically-steady solutions.
- (iii) We include a frictional term coming from a downgradient flux of potential vorticity. Given the neglect of relative vorticity, this is equivalent to an interfacial drag.
- (iv) We will neglect the western boundary layer.

Because of the equal-layer-thickness assumption, which we make primarily for algebraic simplicity, it is best considered as a model for the upper ocean above a level where the vertical velocity is approximately zero. The equations of motion are then

$$J(\psi_1, q_1) = \frac{1}{H_0} \text{curl}_z \boldsymbol{\tau}_T - \nabla \cdot \mathbf{T}_1, \quad J(\psi_2, q_2) = -\nabla \cdot \mathbf{T}_2 \quad (14.104a,b)$$

where

$$q_1 = \beta y + F(\psi_2 - \psi_1), \quad q_2 = \beta y + F(\psi_1 - \psi_2). \quad (14.105a,b)$$

Here, $F = f_0^2 / (g' H_0) = 1/L_d^2$ is a measure of the stratification, where H_0 is the thickness of either layer, and the $\nabla \cdot \mathbf{T}$ terms are represent interfacial eddy stresses, which we parameterize by a downgradient flux of potential vorticity as in

$$\mathbf{T}_1 = -\kappa \nabla q_1 = -R \nabla(\psi_2 - \psi_1), \quad \mathbf{T}_2 = -\kappa \nabla q_2 = -R \nabla(\psi_1 - \psi_2), \quad (14.106)$$

where R and κ are constants, and $R = \kappa F$. We will be interested in the limit of small R , or more specifically $UFL/R \gg 1$, which is a large Péclet number condition. (The Péclet number is similar to a Reynolds number, but with the diffusivity replacing the kinematic viscosity.) So first consider the case when R is identically zero. An *exact* solution to (14.104) has $\psi_2 = 0$, so that (14.104a) becomes $\beta \partial \psi_1 / \partial x = H_0^{-1} \text{curl}_z \boldsymbol{\tau}_T$, with solution

$$\psi_1 = -\frac{1}{H_0 \beta} \int_x^{x_E} \text{curl}_z \boldsymbol{\tau}_T \, dx. \quad (14.107)$$

That is, there is no flow in the lower layer, and the upper layer solution is given by Sverdrup balance. The solution satisfies $\psi_1 = 0$ at $x = x_E$ and, because $\psi_2 = 0$, the nonlinear term on the left-hand side of (14.104a) vanishes identically.

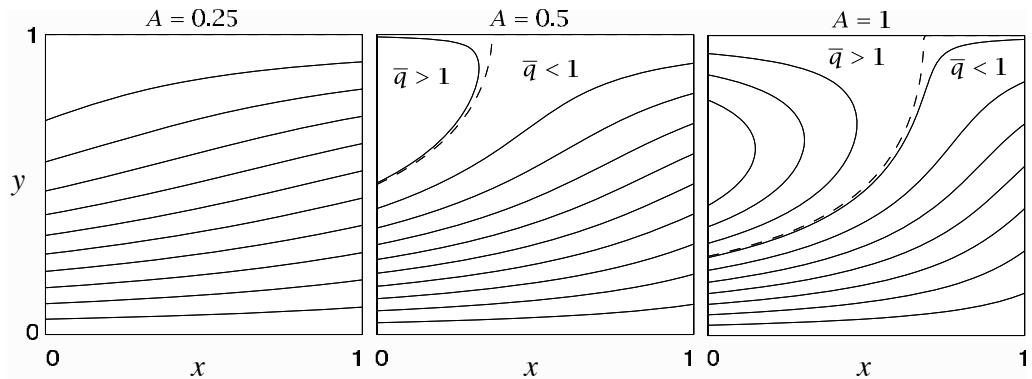


Fig. 14.15 Contours of $\bar{q} = \beta y + A \sin \pi y(1 - x)$, with $\beta = 1$. The dashed line is $\bar{q} = 1$, which separates the blocked region to the east ($\bar{q} < 1$) from the closed region to the west ($\bar{q} > 1$). See Fig. 14.16 for plots of the other fields.

General solution

We now construct the solution without assuming $\psi_2 = 0$. Although the equations are nonlinear, we obtain a linear equation for the lower-layer streamfunction by adding (14.104a) and (14.104b), giving

$$J(\psi_1, \beta y + F(\psi_2 - \psi_1)) + J(\psi_2, \beta y + F(\psi_1 - \psi_2)) = \frac{1}{H_0} \operatorname{curl}_z \boldsymbol{\tau}_T. \quad (14.108)$$

The nonlinear terms cancel leaving

$$J(\bar{\psi}, \beta y) = \frac{1}{H_0} \operatorname{curl}_z \boldsymbol{\tau}_T, \quad \text{where } \bar{\psi} = \psi_1 + \psi_2, \quad (14.109a,b)$$

with solution [cf. (14.107)]

$$\bar{\psi} = -\frac{1}{H_0 \beta} \int_x^{x_E} \operatorname{curl}_z \boldsymbol{\tau}_T \, dx'. \quad (14.110)$$

This simply says that the vertically integrated flow obeys Sverdrup balance. For the canonical wind stress

$$\boldsymbol{\tau}_T = -\boldsymbol{\tau}_0 \cos \pi y \mathbf{i} \quad (14.111)$$

and we obtain $\bar{\psi} = (\pi \tau_0 / \beta H_0)(x_E - x) \sin \pi y$. It is useful to define

$$\bar{q} \equiv (\beta y + F \bar{\psi}), \quad (14.112)$$

and then $\bar{q} = \beta[y + A(1 - x) \sin \pi y]$, where $A = \pi \tau_0 / (\beta H_0)$ parameterizes the wind strength, and this is plotted in Fig. 14.15. For $\bar{q} < 1$ (below and to the right of the dashed line) all the geostrophic contours intersect the eastern boundary and the flow is ‘blocked’. For $\bar{q} > 1$ the flow is ‘closed’.

Lower layer

Although the full equations are nonlinear, using (14.112) we can obtain a linear equation for the lower layer. Because the Jacobian of a field with itself vanishes, (14.104b) and (14.105b) imply that

$$J(\psi_2, \bar{q}) = -\nabla \cdot \mathbf{T}_2, \quad (14.113)$$

and this is useful because \bar{q} is a function of the wind, using (14.112). If $\nabla \cdot \mathbf{T}_2 = 0$ then

$$J(\psi_2, \bar{q}) = 0. \quad (14.114)$$

As well as the possibility that $\psi_2 = 0$ we now have the more general solution

$$\psi_2 = G(\bar{q}), \quad (14.115)$$

where G is an *arbitrary* function of its argument. Isolines of ψ_2 and \bar{q} are then coincident. (Contours that are isolines of both streamfunction and potential vorticity are generally known as geostrophic contours.)

Consider a blocked isoline of \bar{q} ; that is, one that intersects the eastern boundary (see Fig. 14.15). The ψ_2 contour coincident with this has a value of zero at the eastern boundary (by the no-normal flow condition). Thus $\psi_2 = 0$ *everywhere* in the blocked region, and $q_2 = \bar{q}$. In this region the Sverdrup transport is carried everywhere by the upper layer, and the lower layer is at rest. This region is called a ‘shadow zone’, for the fluid is in the shadow of the eastern boundary, and it will re-appear in a model of the thermocline in chapter 16. In the region of closed contours, ψ_2 cannot be given by this argument. But if R is small, we can expect (14.114) to approximately hold, and that the presence of a small amount of dissipation will determine the functional relationship between ψ_2 and \bar{q} . Thus, in summary, there are two regions of flow:

- (i) the blocked region in which $\psi_1 \approx \bar{\psi} \gg \psi_2$ and ψ_1 is approximately given by (14.107);
- (ii) a closed region in which $\psi_2 = G(\bar{q}) + \mathcal{O}(R)$.

14.7.2 The functional relationship between ψ and q

A general argument

In chapter 10 we showed that, within a region of closed contours, the values of a tracer that is materially conserved except for the effects of a small diffusion would become *homogeneous*. In the case at hand, potential vorticity is that tracer, so that within potential vorticity contours or closed streamlines potential vorticity will become homogenized. If we can determine the value of q_2 within the region of closed contours, then from (14.105) ψ_2 is given by

$$\psi_2 = (1/2F)(\bar{q} - q_2) \quad (14.116)$$

and the solution would be complete. Now, outside the closed region $\psi_2 \ll \psi_1$, so that the outermost contour of the closed region must be characterized by $q_2 \approx \bar{q}$, for this makes ψ_2 continuous between closed and blocked regions. Thus, the value of q_2 within the closed homogeneous region is that of \bar{q} (i.e., $\beta y + F\bar{\psi}$) on its boundary. Since this contour intersects the poleward edge of the domain, where $\bar{\psi}$ is zero, the value of this contour is just βy at $y = L$; that is, βL . Thus, within the closed region,

$$q_2 = \beta L. \quad (14.117)$$

Summary of Wind-Driven, Two-layer Solution

The vertically integrated flow in a wind-driven two-layer quasi-geostrophic model is determined by Sverdrup balance. The effects of eddies may be crudely parameterized by a downgradient diffusion of potential vorticity. If this is identically zero, then the lower-layer flow is identically zero and the upper-layer flow carries all the transport. If the diffusion is small but non-zero, the lower-layer streamfunction approximately satisfies $J(\psi_2, \bar{q})$, where \bar{q} is given by (14.112), and therefore ψ_2 is a function of \bar{q} — that is, $\psi_2 \approx G(\bar{q})$. For a typical subtropical wind, contours of \bar{q} , and therefore contours of ψ_2 , are naturally divided into two regions (Fig. 14.15).

- (i) A blocked region (the shadow zone), in which contours of \bar{q} intersect the eastern boundary, the lower layer flow is zero and the upper layer carries all the Sverdrup transport.
- (ii) A closed region in which (if we envision a nearly inviscid western boundary current) the flow recirculates. In this region we posit that the lower layer potential vorticity becomes homogeneous, with a value determined by the value at the region's boundary, and this in turn is determined by tracing \bar{q} back to the domain boundary.

To satisfy a circulation constraint the function $G(\bar{q})$ must be a linear function, and given this, the entire solution may be determined. If, for example, the wind is zonal and a function of y only, and $\text{curl}_z \mathbf{T}_T = g(y)$ then, in both regions:

$$\bar{\psi} \equiv \psi_1 + \psi_2 = -\frac{1}{\beta H_0} g(y)(x_E - x), \quad \bar{q} \equiv \beta y + F\bar{\psi}. \quad (\text{OC.1a,b})$$

In the blocked region:

$$\psi_2 = 0, \quad \psi_1 = -\frac{1}{\beta H_0} (x_E - x)g(y), \quad (\text{OC.2a})$$

$$q_1 = \beta y + F(\psi_1 - \psi_2), \quad q_2 = \beta y + F(\psi_2 - \psi_1). \quad (\text{OC.2b})$$

In the closed region:

$$q_2 = \beta L \quad (\text{by homogenization}), \quad (\text{OC.3a})$$

$$\psi_2 = \frac{1}{2F}(\bar{q} - \beta L), \quad \psi_1 = \bar{\psi} - \psi_2, \quad (\text{OC.3b})$$

$$q_1 = \beta y + F(\psi_2 - \psi_1) = 2\beta y - \beta L. \quad (\text{OC.3c})$$

For $g(y) = -\sin \pi y$ these solutions are illustrated in Figs. 14.15 and 14.16.

This approach provides a solution to the conundrum of what drives the lower ocean, for if there are no eddy effects at all [i.e., $T_1 = T_2 = 0$ in (14.106)], then the lower layer flow is stationary. This seems unrealistic, for the upper layer flow could be made quite shallow. Another solution to this issue is provided in section 16.4, wherein it is assumed that the lower layers may outcrop and so feel the wind directly.

A specific calculation

Now consider the steady, lower-layer potential vorticity equation (14.104b); noting that $J(\psi_2, F(\psi_1 - \psi_2)) = J(\psi_2, F(\psi_1 + \psi_2))$, (14.104b) may be written as

$$J(\psi_2, \bar{q}) = -\nabla \cdot \mathbf{T}_2. \quad (14.118)$$

Integrating around a closed contour of \bar{q} the left-hand side vanishes and

$$R \int (\nabla \psi_1 - \nabla \psi_2) \cdot \mathbf{n} \, dl = 0 \quad (14.119)$$

or

$$\oint \mathbf{u}_1 \cdot dl = \oint \mathbf{u}_2 \cdot dl \quad (14.120)$$

thus, the deep circulation around a mean geostrophic contour (i.e., isoline of \bar{q}) is equal to the upper-level circulation.

Previously we argued that

$$\psi_2 = G(\bar{q}) = G(\beta y + F(\psi_1 + \psi_2)). \quad (14.121)$$

where G is an arbitrary function of its argument. In order to satisfy (14.120) (a linear relation between \mathbf{u}_1 and \mathbf{u}_2) G must be a linear function, and so we write

$$\psi_2 = C \left[\frac{\beta y}{F} + (\psi_1 + \psi_2) \right] + B, \quad (14.122)$$

where C and B are constants. This may be rearranged to give

$$\psi_1 = -C \frac{\beta y}{F} + (\psi_1 + \psi_2)(1 - C) - B. \quad (14.123)$$

This is consistent with (14.120) if $C = 1/2$. With this, (14.122) gives

$$\bar{q} = 2F(\psi_2 - B) \quad (14.124)$$

and the potential vorticity in the closed contour region of the lower layer is

$$q_2 = \beta y + F\bar{q} - 2F\psi_2 = -2FB. \quad (14.125)$$

That is, it is constant. Outside the closed contours $\psi_2 \ll \psi_1$ so that $q_2 \approx \bar{q} = \beta y + F\psi_1$. If we trace this contour to the edge of the domain where $\psi_1 = 0$ and $y = L$ then we see that the value of \bar{q} on the contour, and hence q_2 in the closed region, is βL , as in (14.117), and $B = -\beta L/(2F)$. Using (14.124) then gives

$$\psi_2 = (2F)^{-1}(\bar{q} - \beta L). \quad (14.126)$$

Given ψ_2 and q_2 [from (14.117)], we obtain q_1 and ψ_1 using (14.105) and (14.109b), giving

$$q_1 = 2\beta y - \beta L, \quad \psi_1 = \bar{q} - \psi_2. \quad (14.127)$$

All these fields are illustrated in Fig. 14.16, and see the shaded box on the previous page for a summary.

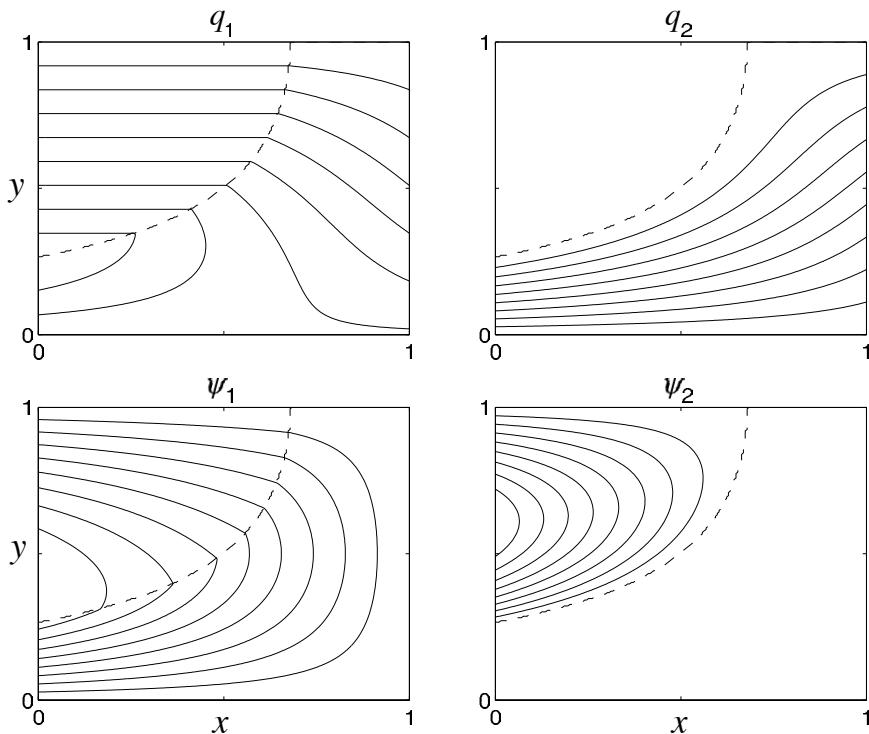


Fig. 14.16 Upper- and lower-level potential vorticity and streamfunction for the canonical wind stress (14.111). The field of \bar{q} is that of Fig. 14.15 with $A = 1$. The dashed line divides the blocked region from the closed region. The lower layer streamfunction ψ_2 is non-zero only in the closed region, and here $q_2 = \beta L$ and $q_1 = 2\beta y - \beta L$. In the blocked region the upper layer carries all of the Sverdrup transport. Both the streamfunction and potential vorticity are continuous at the divide: $\psi_2 = 0$ and $q_2 = \bar{q} = \beta L$.

14.8 * A MODEL WITH CONTINUOUS STRATIFICATION

We now look at the dynamics of the continuously stratified circulation, largely by way of an extension of our two-layer procedure. Let us first consider how deep the wind's influence is.

14.8.1 Depth of the wind's influence

The thermal wind-relationship in the form $f \partial u / \partial z = \partial b / \partial y$ implies a vertical scale H given by

$$H = \frac{fUL}{\Delta b}, \quad (14.128)$$

where Δb is a typical magnitude of the horizontal variation of the buoyancy. We can relate this to the Ekman pumping velocity W_E using the linear geostrophic vorticity equation, $\beta v = f \partial w / \partial z$, which, (assuming that the horizontal components of velocity are roughly

similar, i.e., $V = U$), implies that

$$U = \frac{f W_E}{\beta H}. \quad (14.129)$$

Equations (14.128) and (14.129) may be combined to give an estimate of the depth of the wind-driven circulation, namely

$$H = \left(\frac{f^2 W_E L}{\beta \Delta b} \right)^{1/2}, \quad (14.130)$$

where L may be interpreted as the gyre scale. We now use quasi-geostrophic scaling to relate the horizontal temperature gradient to the stratification using the thermodynamic equation,

$$\frac{D\theta}{Dt} + w N^2 = 0, \quad (14.131)$$

with implied scaling

$$\Delta b = \frac{W_E N^2 L}{U} = \frac{N^2 \beta H L}{f_0}, \quad (14.132)$$

where the second equality uses (14.129). Using (14.130) and (14.132) gives

$$H = \left(\frac{W_E f^3}{\beta^2 N^2} \right)^{1/3}. \quad (14.133)$$

Potential vorticity interpretation

The estimate (14.133) can be obtained and interpreted more directly: *the wind-driven circulation penetrates as far as it can alter the potential vorticity q from its planetary value $\beta\gamma$.* Recall that, ignoring relative vorticity,

$$q = \beta\gamma + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial \psi}{\partial z} \right). \quad (14.134)$$

The two terms are comparable if

$$\frac{f_0^2}{N^2 H^2} U L \approx \beta L \quad (14.135)$$

or

$$H^2 \approx \frac{f_0^2 U}{N^2 \beta}. \quad (14.136)$$

Using (14.129) to eliminate U in favour W_E recovers (14.133). Thus, for a given stratification, we have an estimate of the depth of the wind-driven circulation, or at least a scaling for depth of the vertical influence of the wind.

14.8.2 The complete solution

Armed with an estimate for the depth of the wind's influence, we can obtain a solution for the continuously stratified case analogous to that found in the two-layer case in section 14.7. Our assumptions are as follows.

- (i) In the limit of small dissipation, streamfunction and potential vorticity have a functional relationship with each other.
- (ii) Potential vorticity is homogenized within closed isolines of q or ψ . The value of q within the homogenized pool is that of the outermost contour, which here is the value of q at the poleward edge of the barotropic gyre.
- (iii) Outside of the pool region, (i.e., below the depth of the wind's influence) the streamfunction is zero, and the potential vorticity is given by the planetary value, i.e., βy .

Given these, finding a solution is not difficult. If N^2 is constant and neglecting relative vorticity, the expression for potential vorticity is

$$q = \frac{\partial^2}{\partial z^2} \left(\frac{f_0^2}{N^2} \psi \right) + \beta y. \quad (14.137)$$

We non-dimensionalize by writing

$$z = \left(\frac{f_0^2 U}{N^2 \beta} \right)^{1/2} \hat{z}, \quad q = \beta L \hat{q}, \quad \psi = \hat{\psi} U L, \quad y = L \hat{y}, \quad w = \frac{U^2 f_0}{N^2 H} \hat{w}, \quad (14.138)$$

where the hatted variables are non-dimensional, and the scaling for w arises from the thermodynamic equation $N^2 w \sim J(\psi, f_0 \psi_z)$. With this, (14.137) becomes

$$\hat{q} = \frac{\partial^2 \hat{\psi}}{\partial \hat{z}^2} + \hat{y}, \quad (14.139)$$

The flow is then given by solving the following equations:

$$\psi_{zz} + y = y_0, \quad -D(x, y) < z < 0, \quad (14.140a)$$

$$\psi = 0, \quad z \leq -D(x, y), \quad (14.140b)$$

where D is the (to be determined) depth of the bowl, y_0 is a constant, and we have dropped the hats over the non-dimensional variables. The solution in $D(x, y) < z < 0$ corresponds to the closed region of the two-layer model, and the solution $z \leq -D(x, y)$ corresponds to the blocked region of zero lower-layer flow. The constant y_0 is the non-dimensional value of potential vorticity within the pool region, and following our reasoning in the two-layer case this is the value of the potential vorticity at the northern boundary. Dimensionally this is βL , so that in non-dimensional units $y_0 = 1$.

The lower boundary condition on (14.140a) is that $\psi = \psi_z = 0$ at $z = -D$, because in the abyss $\psi = \partial \psi / \partial z = 0$ and we require that both ψ and $\partial \psi / \partial z$ be continuous (note that the buoyancy perturbation is proportional to $\partial \psi / \partial z$). The solution that satisfies this is

$$\psi = \frac{1}{2} (z + D)^2 (y_0 - y), \quad (14.141)$$

and $\psi = 0$ for $z < D$. To obtain an expression for D we first note that the non-dimensional vertical velocity at $z = 0$ is given by

$$w = -J(\psi, \psi_z), \quad (14.142)$$

which, using (14.141), gives

$$w = \frac{1}{2} (z + D)^2 (y_0 - y) \frac{\partial D}{\partial x}. \quad (14.143)$$

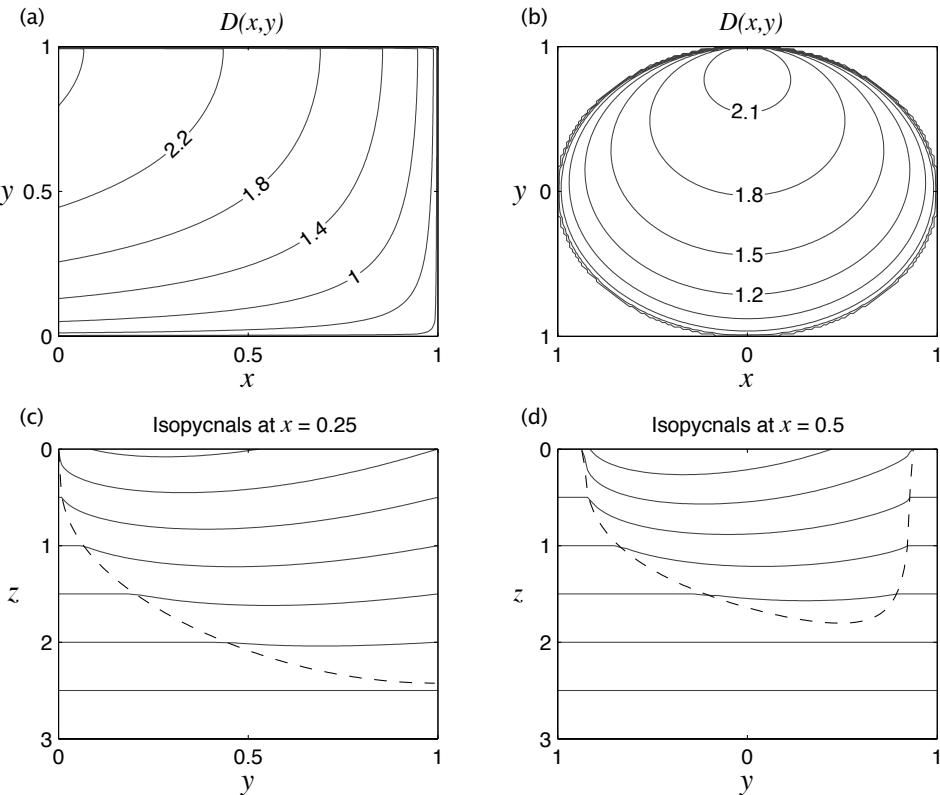


Fig. 14.17 Solutions of (14.112) for two different barotropic streamfunctions. On the left $\psi_B = (1 - x) \sin \pi y$ and on the right $\psi_B = 1 - (x^2 + y^2)$ for $x^2 + y^2 < 1$, zero elsewhere. The upper panels show contours of the depth of the wind-influenced region [solutions of (14.145)]. The depth increases to the northwest in the left panel, and to the north in the right panel, so that in both cases the area of the bowl shrinks with depth. The lower panels are contours of $z + (\beta L / f_0) \psi_z / 2$, with $\beta L / f_0 = 1/2$, obtained from (14.141) or (14.146), at $x = 0.25$ and $x = -0.5$ in the two cases. These are isopycnal surfaces, with a rather large value of $\beta L / f_0$ to exaggerate the displacement in the bowl region. The dashed lines indicate the boundary of the bowl region, outside of which the isopycnals are flat.

At $z = 0$ the vertical velocity is the Ekman pumping velocity and (14.143) becomes

$$D^2 \frac{\partial D}{\partial x} = \frac{2w_E}{(\gamma_0 - \gamma)} . \quad (14.144)$$

But the Ekman pumping velocity is related to the barotropic streamfunction, ψ_B , by the Sverdrup relationship, so that integrating (14.144) gives

$$D^3 = \frac{6\psi_B}{(\gamma_0 - \gamma)} = -\frac{6(x_E - x)w_E}{(\gamma_0 - \gamma)} , \quad (14.145)$$

where the second equality holds if w_E is not a function of x . This is a solution for the depth of moving region, the bowl in which potential vorticity is homogenized. An expression for

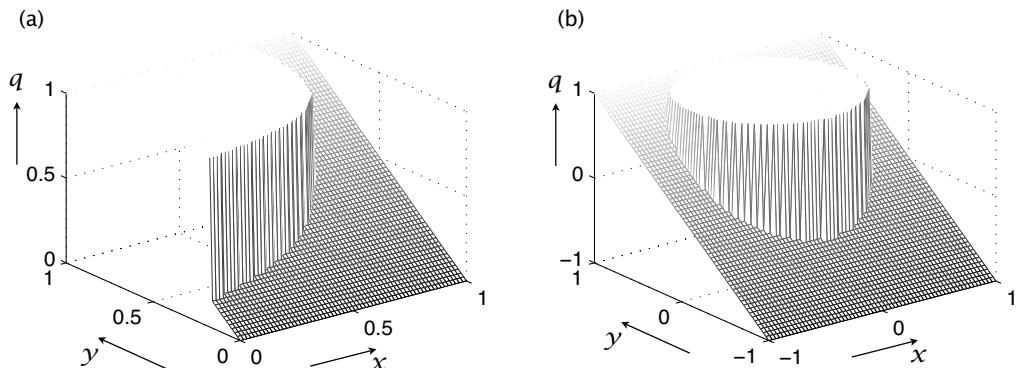


Fig. 14.18 As for Fig. 14.17, but now a perspective of potential vorticity [obtained from (14.147)]. Within the bowls the circulation is clockwise and the potential vorticity is uniform. Outside the bowls the fluid is stationary and the potential vorticity has the planetary value βy . The value of potential vorticity within the bowl is the planetary value at the poleward edge of the gyre, and is the same value at all depths.

the streamfunction is then obtained by using (14.145) in (14.141), and is found to be

$$\psi = \begin{cases} \frac{1}{2} [z(\gamma_0 - \gamma)^{1/2} + (6\psi_B)^{1/3}(\gamma_0 - \gamma)^{1/6}]^2 & -D < z < 0, \\ 0 & z < -D. \end{cases} \quad (14.146)$$

The potential vorticity corresponding to this solution is

$$q = \begin{cases} \gamma_0 & -D < z < 0, \\ \gamma & z < -D. \end{cases} \quad (14.147)$$

Solutions are illustrated in Figs. 14.17 and 14.18 for cases with two different barotropic streamfunctions.

It is possible to heuristically extend models such as the one described above by appending a western boundary layer, and indeed the homogenization of potential vorticity depends upon the presence of such a layer in allowing the flow to recirculate. However, as we saw in section 14.5.3, it is difficult for flow to leave a western boundary layer without the help of friction, and a neutrally stable, damped, stationary Rossby wave typically forms. The critical issue then is whether the presence of dissipation in the western boundary layer affects the homogenization of potential vorticity in the gyre itself. Numerical solutions with a quasi-geostrophic model do show that potential vorticity is able to homogenize under such circumstances. Homogenization in the real ocean has been observed in the Pacific and, somewhat less compellingly, in the Atlantic.²¹

Notes

- 1 This figure is based in part on earlier drawings by W. Schmitz and L. Talley that are largely based on observations, as well as on output from numerical models constrained to the observations. However, it is highly schematic and not quantitative.

- 2 The climatological data set of Conkright *et al.* (2001) was assimilated into a primitive equation numerical model (MOM) by strongly relaxing the model temperature and salinity fields to those of the data set, a method arcaneously known as ‘robust diagnostics’ or ‘nudging’. Larger scale features are broadly representative of reality but smaller scale features may be inaccurate.
- 3 Courtesy of R. Zhang. See also Zhang & Vallis (2007).
- 4 Henry Stommel (1920–1992) was one of the most creative and influential physical oceanographers of the twentieth century. Spending most of his career at Woods Hole Institute of Oceanography, his enduring contributions include the first essentially correct theory of western intensification (and so of the Gulf Stream), some of the first models of abyssal flow and the thermohaline circulation (chapter 15), and his foundational work on the thermocline. His *forté* was in constructing elegantly simple models of complex phenomena — often models that were physically realizable in the laboratory — and at the same time testing and encouraging the testing of the models against observations. This chapter might have been entitled ‘Variations on a theme of Stommel’.
- 5 The asymptotic solution to this boundary value problem was obtained by Wasow (1944), a few years prior to Stommel’s work, and further investigated by Levinson (1950). However, it seems unlikely these two investigators were motivated by the oceanographic problem.
- 6 Harald Sverdrup (1888–1957) was a Norwegian meteorologist/oceanographer who is most famous for the balance that now bears his name, but he also played a leadership role in scientific policy and was the director of Scripps Institution of Oceanography from 1936–1948. The Sverdrup unit is also named for him. Originally defined as a measure of volume transport, with $1 \text{ Sv} \equiv 10^6 \text{ m}^3 \text{ s}^{-1}$, it is more generally thought of as a mass transport with $1 \text{ Sv} \equiv 10^9 \text{ kg s}^{-1}$, in which case it can also be used as a measure of transport in the atmosphere. The Hadley Cell, for example, has an average transport of about 100 Sv (Figs. 11.3 and 12.21).
- 7 Leetmaa *et al.* (1977) and Wunsch & Roemmich (1985) offer complementary views on the matter.
- 8 After Munk (1950).
- 9 Hendershott (1987), Veronis (1966a).
- 10 Veronis (1966a,b) was one of the first to investigate nonlinear effects in wind-driven gyres. See also Fox-Kemper & Pedlosky (2004) and references therein.
- 11 Solutions kindly provided by B. Fox-Kemper.
- 12 After Charney (1955).
- 13 A more general discussion is given by Greenspan (1962). Il’in & Kamenkovich (1964) and Lerley & Ruehr (1986) show numerically that the friction must be sufficiently strong for steady boundary-layer solutions to exist.
- 14 After Fofonoff (1954).
- 15 Hughes & de Cuevas (2001).
- 16 Welander (1968).
- 17 Figure kindly provided by Laura Jackson.
- 18 Adapted from Jackson *et al.* (2006).
- 19 Sarkisyan & Ivanov (1971).

- 20 This section and the next draw from Rhines & Young (1982b). Young & Rhines (1982) also considered the problem of a western boundary layer.
- 21 See Rhines & Young (1982a) for a numerical example of PV homogenization (numerical simulation by W. Holland). See Keffer (1985), Talley (1988) and Lozier *et al.* (1996) for some observed PV fields. See Holland *et al.* (1984) for some of both.

Further reading

Abarbanel H. D. I. & Young, W. R., Eds., 1987. *General Circulation of the Ocean*.

Contains several useful review articles on the oceanic general circulation. Particularly relevant to this chapter is the article by Hendershott on homogeneous single-layer models, and the article by Young on vertical structure.

Pedlosky, J., 1996. *Ocean Circulation Theory*.

As well as covering the theory of wind-driven gyres, this book discusses cross-gyre flows and equatorial dynamics.

Problems

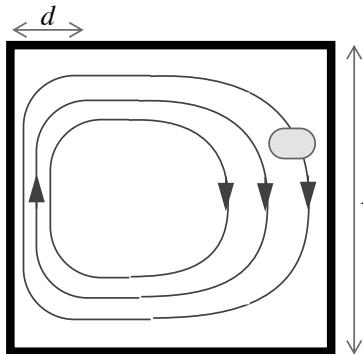
- 14.1 (a) Consider the Stommel equation (14.31), and suppose the boundary layer thickness ϵ is related to ϵ_S by $\epsilon = \epsilon_S^\gamma$. Show that we may obtain a solution that satisfies the boundary conditions if $\gamma = 1$. Is this the unique solution?
 (b) Repeat part (a), except for the Munk equation. Specifically, in (14.46) let $\epsilon = \epsilon_M^\gamma$, and show that $\gamma = 1/3$. Is this the unique solution?
- 14.2 ♦ Obtain the exact analytic solution to the Stommel problem (14.6) in a rectangular basin, and not by way of a boundary-layer approximation. You may use a simple, single-gyre, zonal wind stress.
- 14.3 ♦ Obtain the exact analytic solution to the Munk problem (14.42) in a rectangular basin, and not by way of a boundary-layer approximation. You may use a simple, single-gyre, zonal wind stress.
- 14.4 Obtain a boundary-layer solution of the Munk problem, (14.44), but with ‘free-slip’ boundary conditions instead of no-slip boundary conditions. Thus, instead of setting the normal derivative of ψ to be zero at the boundaries ($\partial\psi/\partial n = 0$), set the relative vorticity at the boundary to be zero, and thus $\partial^2\psi/\partial n^2 = 0$.
- 14.5 ♦ Obtain a solution of either the Stommel or Munk problems in a triangular domain, or in a circular domain.
- 14.6 What is the formal asymptotic accuracy of derivatives in the boundary-layer solution to the Stommel problem? Show that the nonlinear term $J(\psi_s, \nabla^2\psi_s)$, where ψ_s is the streamfunction, is estimated to $\mathcal{O}(1/\delta_W^3)$ accuracy, where δ_W is the width of the western boundary layer.
- 14.7 (a) ♦ Using regular perturbation theory explicitly obtain the nonlinear corrections to the Stommel solutions for a single gyre flow, and show that (14.63) is an appropriate approximation of this. Discuss the properties of the solution, and in particular, show where are the errors largest. Optional: Repeat this calculation using the exact solution to the Stommel problem, and compare the solutions to those obtained using the boundary-layer approximation. (This problem may be made easier by using algebraic manipulation software.)
 (b) ♦ Suggest a better approach to this problem that is still semi-analytic or that at least gives some insight into the nonlinear solution.
- 14.8 ♦ Do problem 14.7, but with lateral viscosity instead of bottom drag, as in the Munk problem.

- 14.9 Suppose that nonlinearity were expected to be of leading order importance in (14.58). Show that this suggests that the non-dimensional version of the equation should be:

$$J(\hat{\psi}, \hat{\zeta}) + \gamma \frac{\partial \hat{\psi}}{\partial \hat{x}} = \epsilon'_S \nabla^2 \hat{\psi} + \text{curl}_z \hat{\mathbf{r}}_T + \epsilon'_M \nabla^2 \hat{\zeta} \quad (\text{P14.1})$$

where $\gamma = \beta(L^3/|\tau|)^{1/2}$ and $\epsilon'_S = r(L/|\tau|)^{1/2}$. Friction becomes less important as τ increases. Is there a limit in which friction is nowhere important?

- 14.10 Consider the two-dimensional, non-divergent flow in a box of side L , as pictured. The equatorward flow, of magnitude U , is weak and spread over a lateral scale L , whereas the poleward flow is concentrated in a layer of width $d \ll L$. Suppose the initial condition is a spot of tracer, as shown.



Estimate how long it takes to homogenize the tracer in terms of U , L , κ and d . Hint: Consider the advection-diffusion equation,

$$\frac{\partial \phi}{\partial t} + \mathbf{u} \cdot \nabla \phi = \kappa \nabla^2 \phi. \quad (\text{P14.2})$$

and estimate advective and diffusive times in the western boundary layer and the interior.

- 14.11 For a two-layer wind-driven quasi-geostrophic model, obtain an expression for the minimum strength of the wind that leads to closed contours of $\bar{q} = \beta y + F\bar{\psi}$. In particular, for the solution illustrated in Fig. 14.15, at what values of A do closed contours first form?
(Answer: $A = \pi^{-1}$.)

- 14.12 Show by differentiating (14.146) that potential vorticity is uniform in the pool region. What is its value?

Paradise Lost.
John Milton, c. 1667.

CHAPTER FIFTEEN

The Buoyancy-Driven Ocean Circulation

OUR GOAL IN THIS CHAPTER AND THE NEXT is to gain a rudimentary understanding of the three-dimensional dynamics and structure of the ocean circulation. In this chapter we focus on the *meridional overturning* and the associated *abyssal* circulation of the ocean, treating them as if they were solely driven by buoyancy forces, and in chapter 16 we look at the combined effects of wind and buoyancy forcing.

The meridional overturning circulation (MOC) is so-called because it is associated with sinking at high latitudes and upwelling elsewhere, although as we shall see even such a seemingly simple matter as this is not wholly settled. The circulation is also sometimes known as the ‘thermohaline’ circulation, reflecting a belief that it is driven by gradients in temperature and salinity, but because other mechanisms are also important that name is not appropriate as a general descriptor. In fact, the theory explaining the MOC is not in as satisfactory a state as it is for the quasi-horizontal wind-driven circulation discussed in chapter 14. In the theory of the wind-driven circulation, a rational series of approximations from the governing Navier-Stokes equations leads to a sequence of simple models (e.g., the homogeneous models of the wind-driven circulation, layered quasi-geostrophic models, etc.), whose foundations are thus reasonably secure, whose shortcomings are understood, and whose behaviour can be fairly completely analysed. Attempts to proceed in a similar fashion with overturning circulation have been less successful; the reason is that the approximations required in order that a tractable conceptual model be constructed are unavoidably severe and, from a fluid-dynamicist’s perspective, unjustifiable. Thus, although the large-scale overturning flow is well described by the planetary-geostrophic equations whose complexity is similar to that of the quasi-geostrophic equations, there is in turn no rational simplification of these that leads to a model that is both as simple and as informative as the homogeneous Stommel model of the quasi-horizontal, wind-driven circulation. Nevertheless, progress *has*

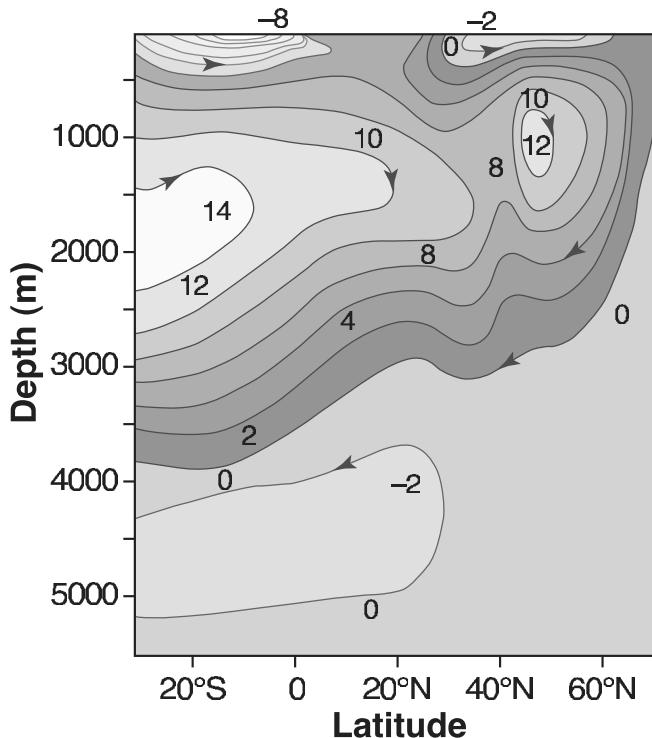


Fig. 15.1 The mean meridional overturning circulation in the North Atlantic, obtained with a combination of observations and a model. The contours are the streamfunction of the zonally averaged meridional flow. The units are Sverdrups, and the circulation is mostly clockwise, with sinking at high latitudes.¹

been made, through the use of both numerical models and simpler, conceptual, models, and in this chapter we concentrate on the foundations underlying these.

That there *is* a deep circulation has been known for a long time, largely from observations of tracers such as temperature, salinity, and constituents such as dissolved oxygen and silica.³ We can also take advantage of numerical models that are able to assimilate hydrographic and other observations and produce an estimate of the overturning circulation that is consistent with both the observations and the equations of motion, as illustrated in Fig. 15.1. Associated with the overturning circulation is a stratification that has a quite distinctive structure, as illustrated in Fig. 15.2. Most of the stratification is evidently concentrated in the upper kilometre or so of the ocean, with a relatively (although not completely) unstratified abyss full of dense water that has originated from high latitudes — the isopycnals shown in Fig. 15.2 all outcrop (i.e., intersect the surface) in the North Atlantic subpolar gyre and/or in the Antarctic Circumpolar Current (ACC). The region of high stratification near the surface is known as the *thermo*cline ('thermo' for temperature, 'cline' for changing); it is effectively synonymous with the *pycnocline* (for changing density), although the latter could exist without the former if there were a *halocline*, in which salinity changed rapidly.

What is the nature and underlying cause of this stratification and the associated circulation? We start by considering a simple but very revealing fluid model in which we ignore the effects of the wind and consider only buoyancy forcing at the surface.

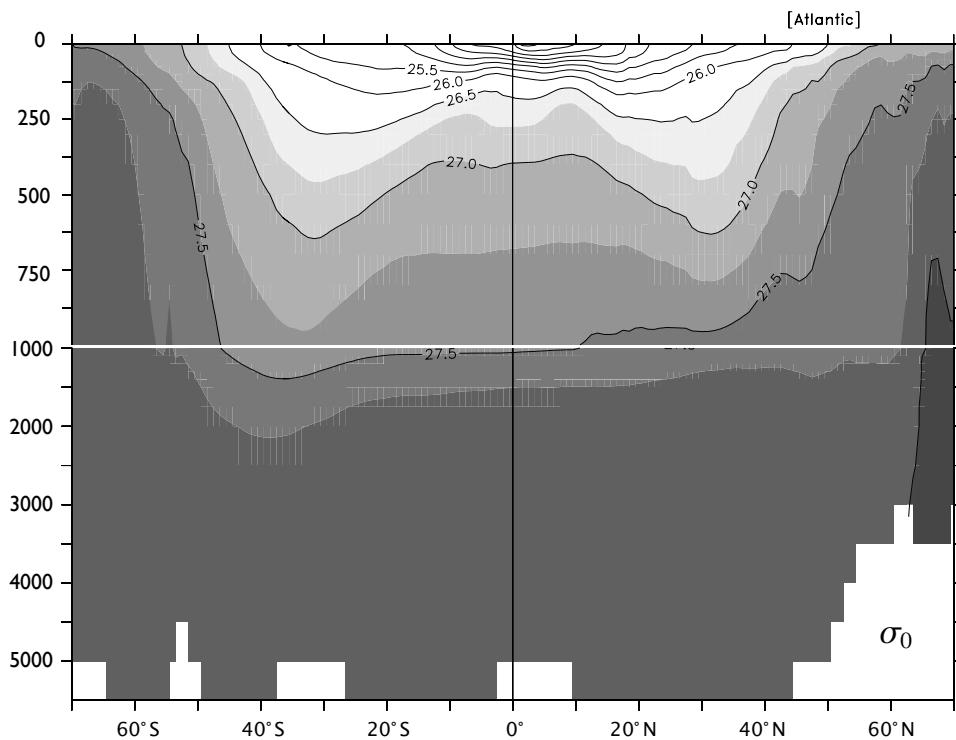


Fig. 15.2 The zonally averaged potential density (σ_0) in the Atlantic ocean, as a function of depth (m) and latitude. Note the break in the vertical scale at 1000 m. The region of rapid change of density (and temperature) is concentrated in the upper kilometre, in the *main thermocline*, below which the ocean has a much more uniform density.²

15.1 SIDEWAYS CONVECTION

Perhaps the simplest and most obvious fluid dynamical model of the overturning circulation is that of *sideways convection*. The physical situation is sketched in Fig. 15.3. A fluid (two- or three-dimensional) is held in a container that is insulated on all of its sides and its bottom, but its upper surface is non-uniformly heated and cooled. In the purest fluid dynamical problem the heat enters the fluid solely by conduction at the upper surface, and one may suppose that here the temperature is imposed. Thus, for a simple Boussinesq fluid the equations of motion are

$$\frac{D\mathbf{v}}{Dt} + \mathbf{f} \times \mathbf{v} = -\nabla\phi + b\mathbf{k} + \nu\nabla^2\mathbf{v}, \quad (15.1a)$$

$$\frac{Db}{Dt} = \kappa\nabla^2b, \quad (15.1b)$$

$$\nabla \cdot \mathbf{v} = 0, \quad (15.1c)$$

with boundary conditions

$$b(x, y, 0, t) = g(x, y), \quad (15.2)$$

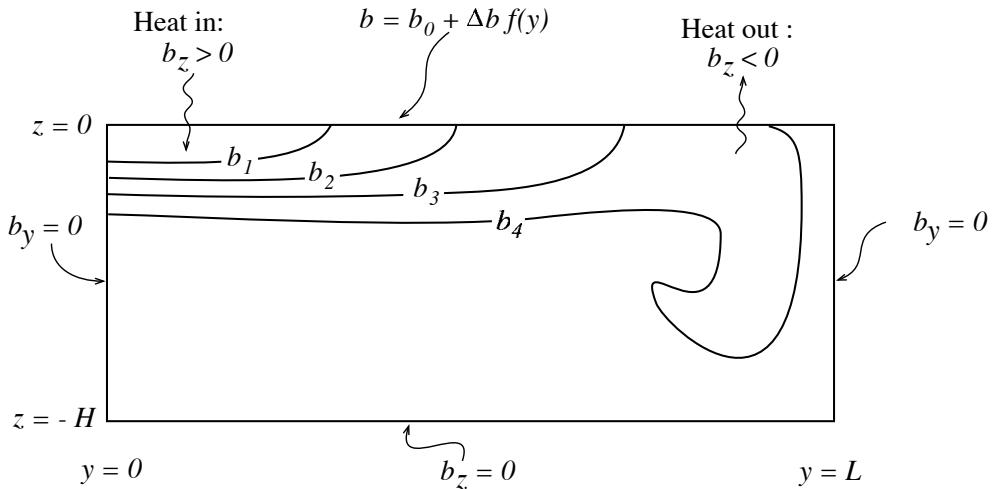


Fig. 15.3 A schema of sideways convection. The fluid is differentially heated and cooled along its top surface, whereas all the other walls are insulating. The result is, typically, a small region of convective instability and sinking near the coldest boundary, with generally upwards motion elsewhere.⁴

where $g(x, y)$ is a specified field, and $\partial_n b = 0$ on the other boundaries, meaning that the derivative normal to the boundary, and so the buoyancy flux, is zero. The oceanographic relevance of (15.1) and (15.2) should be clear: the ocean is heated *and* cooled from above, and although the thermal forcing in the real ocean may differ in detail (being in part a radiative flux, and in part a sensible and latent heat transfer from the atmosphere), (15.2) is a useful idealization. An alternative upper boundary condition would be to impose a flux condition whereby

$$\text{flux} = \kappa \frac{\partial}{\partial z} b(x, y, 0, t) = h(x, y), \quad (15.3a)$$

where $h(x, y)$ is given. In some numerical models of the ocean, the heat input at the top is parameterized by way of a relaxation to some specified temperature. This is a form of flux condition in which

$$\kappa \frac{\partial b}{\partial z} = C(b^*(x, y) - b), \quad (15.3b)$$

and C is an empirical constant and $b^*(x, y)$ is given.⁵ Although this may be a little more relevant than (15.2) for the real ocean, which of the three boundary conditions is chosen will not affect the arguments below, and we use (15.2).

15.1.1 Two-dimensional convection

We may usefully restrict attention to the two-dimensional problem, in latitude and height. The two-dimensional flow may then loosely be thought of as representing the large-scale, statistically steady, zonally averaged flow of the ocean. The zonally averaged zonal flow is then small, and concomitantly so is the Coriolis force. The incompressibility of the flow then

allows one to define a streamfunction such that

$$v = -\frac{\partial \psi}{\partial z}, \quad w = \frac{\partial \psi}{\partial y}, \quad \zeta = \nabla_x^2 \psi = \left(\frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right), \quad (15.4)$$

where ζ is the vorticity in the meridional plane. We will omit the subscript x on the Laplacian operator where there is no ambiguity. Taking the curl of Boussinesq equations of motion (15.1) then gives

$$\frac{\partial \nabla^2 \psi}{\partial t} + J(\psi, \nabla^2 \psi) = \frac{\partial b}{\partial y} + \nu \nabla^4 \psi \quad (15.5a)$$

$$\frac{\partial b}{\partial t} + J(\psi, b) = \kappa \nabla^2 b \quad (15.5b)$$

where $J(a, b) = (\partial_y a)(\partial_z b) - (\partial_z a)(\partial_y b)$.

Non-dimensionalization and scaling

We non-dimensionalize (15.5) by formally setting

$$b = \Delta b \hat{b}, \quad \psi = \Psi \hat{\psi}, \quad y = L \hat{y}, \quad z = H \hat{z}, \quad t = \frac{LH}{\Psi} \hat{t}, \quad (15.6)$$

where the hatted variables are non-dimensional, Δb is the temperature difference across the surface, L is the horizontal size of the domain, and Ψ , and ultimately the vertical scale H , are to be determined. Substituting (15.6) into (15.5) gives

$$\frac{\partial \hat{\nabla}^2 \hat{\psi}}{\partial \hat{t}} + \hat{J}(\hat{\psi}, \hat{\nabla}^2 \hat{\psi}) = \frac{H^3 \Delta b}{\Psi^2} \frac{\partial \hat{b}}{\partial \hat{y}} + \frac{\nu L}{\Psi H} \hat{\nabla}^4 \hat{\psi}, \quad (15.7a)$$

$$\frac{\partial \hat{b}}{\partial \hat{t}} + \hat{J}(\hat{\psi}, \hat{b}) = \frac{\kappa L}{\Psi H} \hat{\nabla}^2 \hat{b}, \quad (15.7b)$$

where $\hat{\nabla}^2 = (H/L)^2 \partial^2 / \partial \hat{y}^2 + \partial^2 / \partial \hat{z}^2$ and the Jacobian operator is similarly non-dimensional. If we now use (15.7b) to choose Ψ as

$$\Psi = \frac{\kappa L}{H}, \quad (15.8)$$

so that $t = H^2 \hat{t} / \kappa$, then (15.7) becomes

$$\frac{\partial \hat{\nabla}^2 \hat{\psi}}{\partial \hat{t}} + \hat{J}(\hat{\psi}, \hat{\nabla}^2 \hat{\psi}) = Ra \sigma \alpha^5 \frac{\partial \hat{b}}{\partial \hat{y}} + \sigma \hat{\nabla}^4 \hat{\psi}, \quad (15.9)$$

$$\frac{\partial \hat{b}}{\partial \hat{t}} + \hat{J}(\hat{\psi}, \hat{b}) = \hat{\nabla}^2 \hat{b}, \quad (15.10)$$

and the non-dimensional parameters that govern the behaviour of the system are

$$Ra = \left(\frac{\Delta b L^3}{\nu \kappa} \right), \quad (\text{the Rayleigh number}), \quad (15.11a)$$

$$\sigma = \frac{\nu}{\kappa}, \quad (\text{the Prandtl number}), \quad (15.11b)$$

$$\alpha = \frac{H}{L}, \quad (\text{the aspect ratio}). \quad (15.11c)$$

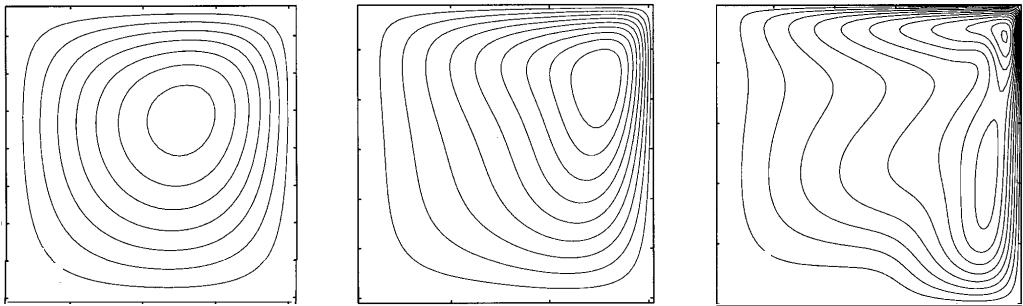


Fig. 15.4 The streamfunction in a numerical simulation of two-dimensional sideways convection.⁷ The circulation is clockwise, and the imposed temperature at the top linearly decreases from left to right, and the other walls are insulating. From left to right the Rayleigh numbers are 10^4 , 10^6 and 10^8 , and the contour interval is 1, 4 and 10 in arbitrary units. The Prandtl number is 10.

The Rayleigh number is a measure of the strength of the buoyancy forcing relative to the viscous term, and in the ocean it will be very large indeed, perhaps $\sim 10^{24}$ if molecular values are used. (Sometimes H is used instead of L in the Rayleigh number definition; we use L here because it is an external parameter.)

For steady non-turbulent flows, and also perhaps for statistically steady flows, then we can demand that the buoyancy term in (15.9) is $\mathcal{O}(1)$. If it is smaller then the flow is not buoyancy driven, and if it is larger there is nothing to balance it. Our demand can be satisfied only if the vertical scale of the motion adjusts appropriately and, for $\sigma = \mathcal{O}(1)$, this suggests the scalings:⁶

$$H = L\sigma^{-1/5}Ra^{-1/5} = \left(\frac{\kappa^2 L^2}{\Delta b} \right)^{1/5}, \quad \Psi = Ra^{1/5}\sigma^{-4/5}\nu = (\kappa^3 L^3 \Delta b)^{1/5}. \quad (15.12a,b)$$

The vertical scale H arises as a consequence of the analysis, and the vertical size of the domain plays no direct role. [For $\sigma \gg 1$ we might expect the nonlinear terms to be small and if the buoyancy term balances the viscous term in (15.9) the right-hand sides of (15.12) are multiplied by $\sigma^{1/5}$ and $\sigma^{-1/5}$. For seawater, $\sigma \approx 7$ using the molecular values of κ and ν . If small scale turbulence exists, then the eddy viscosity will likely be similar to the eddy diffusivity and $\sigma \approx 1$.] Numerical experiments (Figs. 15.4 and 15.5) do provide some support for this scaling, and a few simple and robust points that have relevance to the real ocean emerge, as follows.

- ★ Most of the box fills up with the densest available fluid, with a boundary layer in temperature near the surface required in order to satisfy the top boundary condition. The boundary gets thinner with decreasing diffusivity, consistent with (15.12). This is a diffusive prototype of the oceanic thermocline.
- ★ The horizontal scale of the overturning circulation is large, being at or near the scale of the box.
- ★ The downwelling regions (the regions of convection) are of smaller horizontal scale than the upwelling regions, especially as the Rayleigh number increases.

Let us now try to explain some of the features in a simple and heuristic way.

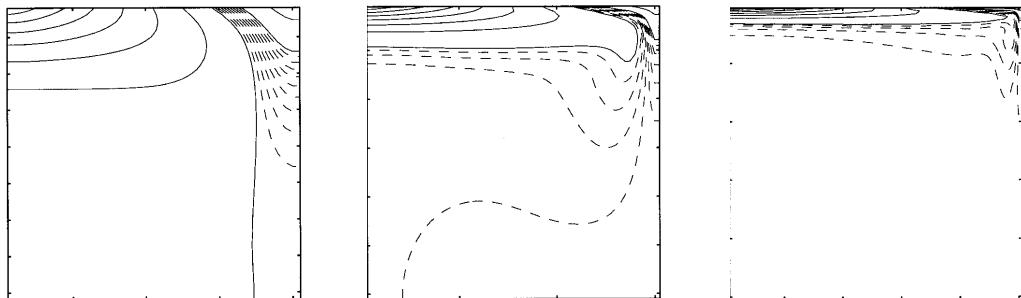


Fig. 15.5 The temperature or buoyancy field corresponding to the streamfunction fields shown in Fig. 15.4. Note an increasingly sharp gradient (a thermocline) near the top as the Rayleigh number increases, and that the bulk of the domain is filled with the densest available fluid.⁸

15.1.2 † Phenomenology of the overturning circulation

No water can be denser (or, more accurately, have a greater potential density) than the densest water at the surface, and if the subsurface water is a little lighter than this the surface water will be convectively unstable and sink in a plume.⁹ The plume slowly entrains the warmer water that surrounds it, and then spreads horizontally when it reaches the bottom or when its density becomes similar to that of its surroundings. The presence of water denser than its surroundings creates a pressure gradient, and the ensuing flow will displace any adjacent lighter fluid, and so the domain fills with the densest available fluid. This process is a continuous one: the plumes take cold water into the interior, where the water slowly warms by diffusion, and the source of cold water at the surface is continuously replenished. If diffusion is small, the end result is that the potential density of the fluid in the interior will be slightly less than that of the densest fluid formed at the surface. (Because diffusion can act only to reduce extrema, no fluid in the interior can be colder than the coldest fluid formed at the surface.) However, the value at the surface is given by the boundary condition $b(x, y, z = 0) = f(x, y)$. Thus, the interior cannot fill all the way to the surface with this cold water and there must be a boundary layer connecting the cold, dense interior with the surface; its thickness δ is given by the height scale of (15.12a); that is $\delta \sim H = (\kappa^2 L^2 / \Delta b)^{1/5}$. Such a strong boundary layer will not necessarily be manifest in the velocity field, however, because the no-normal flow boundary condition on the velocity field is satisfied by setting $\psi = 0$ as a boundary condition to the elliptic problem $\nabla^2 \psi = \zeta$, where ζ is the prognostic variable in (15.5a), and this boundary condition has a global effect on the velocity field.

Why is the horizontal scale of the circulation large? The circulation transfers heat meridionally, and it is far more efficient to do this by a single overturning cell than by a multitude of small cells; hence, although we cannot entirely eliminate the possibility that some instability will produce such small scales of motion, it seems likely the horizontal scale of the mean circulation will be determined by the domain scale. Indeed, at low Rayleigh number we can explicitly calculate an approximate analytic solution for the flow, demonstrating this (problem 15.4). For higher Rayleigh number perturbation approaches fail and we must resort to numerical solutions; these (e.g., Fig. 15.4), do show the circulation

dominated by a single overturning circulation rather than many small convective cells. We cannot rigorously prove that this will be the case for still higher Rayleigh numbers, but a general energetic argument that shows that the flow cannot in fact break up into a succession of ever smaller cells in a turbulent cascade is given in the next section.

Why is the downwelling region narrower than the upwelling? At high Rayleigh number, the extreme efficiency of convection may be responsible. Suppose that the basin is initially filled with water of an intermediate temperature, and that surface boundary conditions of a temperature decreasing linearly from low latitudes to high latitudes are imposed. The deep water will be convectively unstable, and convection at high latitudes (where the surface is coldest) will occur quickly, filling the abyss with dense water as described in the paragraphs above. After this initial adjustment, the deep, dense water at lower latitudes will slowly warmed by diffusion, but at the same time surface forcing will maintain a cold high latitude surface, thus leading to high latitude convection. A steady state or statistically steady state is eventually reached with the deep water having a slightly higher potential density than the surface water at the highest latitudes, and so with continual convection — but convection that is perforce restricted to the highest latitudes. However, it remains to put this heuristic argument on more rigorous or quantitative terms.

Finally, it is important to realize that *even for large diffusion and viscosity there is no stationary solution*: as soon as we impose a temperature gradient at the top the fluid begins to circulate, a manifestation of the dictum that a baroclinic fluid is a moving fluid, encountered in section 4.2. The reason is simply that a temperature gradient leads a density gradient, which in turn leads to a pressure gradient. The pressure gradient leads to motion, for viscosity can have an effect only if the velocity is already non-zero.

15.2 THE MAINTENANCE OF SIDEWAYS CONVECTION

In most conventional convection problems the fluid is heated from below, becomes buoyant and rises, and is cooled at the top. In contrast, in sideways convection the heating and cooling occur at the same level, and the conditions under which a circulation can be maintained are by no means clear; our purpose here to make them clearer. The energetic derivations of this section are but an extension of section 2.4.3, but now with starring roles for friction, diffusion and the boundary conditions, and the reader may wish to review that section first. The derivations below are not difficult, but they lead to powerful and perhaps counter-intuitive results that provide important information about the overturning circulation of the real ocean. To begin, we write down the equations of motion:

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{f} + 2\mathbf{\omega}) \times \mathbf{v} = -\nabla B + b\mathbf{k} + \nu \nabla^2 \mathbf{v}, \quad (15.13a)$$

$$\frac{D b}{D t} = \frac{\partial b}{\partial t} + \nabla \cdot (b\mathbf{v}) = \dot{Q} = J + \kappa \nabla^2 b, \quad (15.13b)$$

$$\nabla \cdot \mathbf{v} = 0, \quad (15.13c)$$

where B is the Bernoulli function and \dot{Q} ($= \dot{b}$) is the total rate of heating (including diffusion, and absorbing constant factors such as heat capacity into its definition) with J its non-diffusive component. The fluid occupies a finite volume, and in a steady state $\langle \dot{Q} \rangle = 0$, where the angle brackets denote a volume and time integration.

15.2.1 The energy budget

To obtain an energy budget we follow the procedure of section 2.4.3. First take the dot product of (15.13a) with \mathbf{v} to give

$$\frac{1}{2} \frac{\partial \mathbf{v}^2}{\partial t} = -\nabla \cdot (\mathbf{v} B) + wb + \mathbf{v} \cdot \nabla^2 \mathbf{v}. \quad (15.14)$$

Integrating over a domain bounded by stress-free rigid walls gives the kinetic energy equation

$$\frac{d}{dt} \left\langle \frac{1}{2} \mathbf{v}^2 \right\rangle = \langle wb \rangle - \varepsilon, \quad (15.15)$$

where angle brackets denote (for the moment) just a volume integration and ε is the total dissipation of kinetic energy ($\varepsilon = -\nu \langle \mathbf{v} \cdot \nabla^2 \mathbf{v} \rangle = \nu \langle \mathbf{w}^2 \rangle$), a positive definite quantity. Thus, in a statistically steady state in which the left-hand side vanishes after time averaging, the dissipation of kinetic energy is maintained by the buoyancy flux; that is, by a release of potential energy with light fluid ascending and dense fluid descending.

We obtain a potential energy budget by using (15.13b) to write

$$\frac{Dbz}{Dt} = z \frac{Db}{Dt} + b \frac{Dz}{Dt} = z \dot{Q} + bw, \quad (15.16)$$

and integrating this over the domain gives the potential energy equation

$$\frac{d}{dt} \langle bz \rangle = \langle z \dot{Q} \rangle + \langle bw \rangle. \quad (15.17)$$

Subtracting (15.17) from (15.15) gives the energy equation

$$\frac{d}{dt} \left\langle \frac{1}{2} \mathbf{v}^2 - bz \right\rangle = -\langle z \dot{Q} \rangle - \varepsilon. \quad (15.18)$$

15.2.2 Conditions for maintaining a thermally-driven circulation

In a statistically steady state the left-hand side of (15.18) vanishes and the kinetic energy dissipation is balanced by the buoyancy source terms; that is

$$\langle z \dot{Q} \rangle = -\varepsilon < 0.$$

(15.19)

The right-hand side (the term $-\varepsilon$) is negative definite, and to balance this the heating must be negatively correlated with height. (Recall that $\langle \dot{Q} \rangle = 0$, and the origin of the z -coordinate is then immaterial.) Thus, *in order to maintain a circulation in which kinetic energy is dissipated, the heating (including diffusive heating) must occur, on average, at lower levels than the cooling*. This result is related to, but not quite the same as, a postulate due to Sandström, discussed below.

In the ocean the non-diffusive heating occurs predominantly at the surface, except for the negligible effects of hydrothermal vents. Thus, $\langle Jz \rangle \approx 0$ and a kinetic-energy-dissipating circulation can *only* be maintained, in the absence of mechanical forcing, if the diffusion is non-zero — in that case heat may be diffused from the surface to depth so effectively providing a deep heat source. In the atmosphere, the heating is mostly at the surface and cooling is mostly in the mid-troposphere so that $\langle z \dot{Q} \rangle < 0$; thus, (15.19) is readily satisfied and the circulation is not restricted.

* *Maintaining a steady baroclinic circulation*

By a rather different method — and one closer to a suggestion of Sandström dating from 1908 — we can obtain a result that is different from but related to (15.19).¹⁰ Using (4.37) and (15.13a), the circulation in a Boussinesq system obeys

$$C = \oint \mathbf{v} \cdot d\mathbf{r}, \quad \frac{DC}{Dt} = \oint b \mathbf{k} \cdot d\mathbf{r} + \oint \mathbf{F} \cdot d\mathbf{r}, \quad (15.20a,b)$$

where C is the circulation, $d\mathbf{r}$ is a path element, and \mathbf{F} represents the frictional terms. (The Coriolis parameter plays no role in this argument because the Coriolis force does no work, and f may be set to zero without loss of generality.) Now we may write the rate of change of circulation in the form (for example problem 2.3):

$$\frac{DC}{Dt} = \oint \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) \cdot d\mathbf{r} = \oint \left(\frac{\partial \mathbf{v}}{\partial t} + \boldsymbol{\omega} \times \mathbf{v} \right) \cdot d\mathbf{r}. \quad (15.21)$$

Let us assume the flow is steady, so that $\partial \mathbf{v} / \partial t$ vanishes. Let us further choose the path of integration to be a streamline, which since the flow is steady is also a parcel trajectory. The second term on the right-most expression of (15.21) then also vanishes and (15.20b) becomes

$$\oint b dz = - \oint \mathbf{F} \cdot d\mathbf{r} = - \oint \frac{\mathbf{F}}{|\mathbf{v}|} \cdot \mathbf{v} dr, \quad (15.22)$$

where the last equality follows because the path is everywhere parallel to the velocity. Let us now assume that the friction retards the flow, and that $\oint \mathbf{F} \cdot \mathbf{v} / |\mathbf{v}| dr < 0$. (One form of friction that has this property is linear drag, $\mathbf{F} = -C\mathbf{v}$ where C is a constant. The property is similar to, but not the same as, the property that the friction dissipates kinetic energy over the circuit.) Making this assumption, if we integrate the term on the left-hand side by parts we obtain

$$\oint z db < 0. \quad (15.23)$$

Now, because the integration circuit in (15.23) is a fluid trajectory, the change in buoyancy db is proportional to the heating of a fluid element as it travels the circuit (in the notation of (15.13b), $db = \dot{Q} dt$, where the heating includes diffusive effects). Thus, the inequality implies that the net heating must be negatively correlated with height: that is, *the heating must occur, on average, at a lower level than the cooling in order that a steady circulation may be maintained against the retarding effects of friction*.

A similar result can be obtained for a compressible fluid. From equations (4.43)–(4.45) on page 173 we write the baroclinic circulation theorem as

$$\frac{DC}{Dt} = \oint p d\alpha + \oint \mathbf{F} \cdot d\mathbf{r} = \oint T d\eta + \oint \mathbf{F} \cdot d\mathbf{r}, \quad (15.24)$$

where η is the specific entropy. Then, by precisely the same arguments as led to (15.23), we are led to the requirements that

$$\oint T d\eta > 0 \quad \text{or equivalently} \quad \oint p d\alpha > 0. \quad (15.25a,b)$$

Equation (15.25a) means that parcels must gain entropy at high temperatures and lose

entropy at low temperatures; similarly, from (15.23b), a parcel must expand ($d\alpha > 0$) at high pressures and contract at low pressures.

For an ideal gas we can put these statements into a form analogous to (15.23) by noting that $d\eta = c_p(d\theta/\theta)$, where θ is potential temperature, and using the definition of potential temperature, (1.105). With these we have

$$\oint T d\eta = \oint c_p \frac{T}{\theta} d\theta = \oint c_p \left(\frac{p}{p_R} \right)^\kappa d\theta, \quad (15.26)$$

and (15.25a) becomes

$$\boxed{\oint c_p \left(\frac{p}{p_R} \right)^\kappa d\theta > 0}. \quad (15.27)$$

Because the path of integration is a fluid trajectory, $d\theta$ is proportional to the heating of a fluid element. Thus [and analogous to the Boussinesq result (15.23)], (15.27) implies that *the heating (the potential temperature increase) must occur at a higher pressure than the cooling in order that a steady circulation may be maintained against the retarding effects of friction*.

These results may be understood by noting that the heating must occur at a higher pressure than the cooling in order that work may be done, the work being necessary to convert potential energy into kinetic energy to maintain a circulation against friction. More informally, if the heating is below the cooling, then the heated fluid will expand and become buoyant and rise, and a steady circulation between heat source and heat sink can readily be imagined. On the other hand, if the heating is above the cooling there is no obvious pathway between source and sink.

Intuitive as these results may be, it is important to realize that the conditions required to prove (15.23) and (15.27) are much more restrictive than those needed to prove (15.19). To prove the former, we must *assume* that the flow is absolutely steady, *and* that streamlines form a closed path, *and* that the friction has retarding properties. The second of these conditions is not generally satisfied in three dimensions, even when the flow is steady. Furthermore, one cannot prove that Newtonian viscosity ($\nu \nabla^2 \mathbf{v}$) will always act to retard the flow. On the other hand, (15.19) provides a condition for the maintenance of a statistically steady circulation, assuming only that the friction acts to dissipate kinetic energy. In any case, it is clear from all of the above results that the overturning circulation is greatly affected by the relative pressures at the locations of the heating and cooling, and this is called *Sandström's effect*. In all of these cases, the 'heating' must be taken to include diffusive effects; if the molecular diffusivity is small and the heating is at the surface we can further constrain the flow, as we now see.

15.2.3 Surface fluxes and non-turbulent flow at small diffusivities

Suppose that the only heating to the fluid is via diffusion through the upper surface; that is $J = 0$ in (15.13b). Is a circulation possible? If the diffusivity is finite then heat can diffuse into the fluid, and thereby potentially provide an difference in altitude between the heating and the cooling. However, as $\kappa \rightarrow 0$ this mechanism ceases to operate and we therefore expect that the left-hand sides of (15.19) and (15.23) will go to zero, and the circulation will cease. In what follows we put this argument on a more rigorous footing: we will show

that as $\kappa \rightarrow 0$ the kinetic energy dissipation also goes to zero, and therefore the flow is ‘non-turbulent’, the meaning of which will be made clearer below.

Assuming a statistically steady state, integrating (15.13b) horizontally gives

$$\frac{\partial \overline{bw}}{\partial z} = \kappa \frac{\partial^2 \overline{b}}{\partial z^2}, \quad (15.28)$$

where an overbar indicates a horizontal and time average. Integrating this equation up from the bottom (where there is no flux) to a level z gives

$$\overline{wb} - \kappa \overline{b}_z = 0 \quad (15.29)$$

at every level in the fluid. The two terms on the left-hand side together comprise the total buoyancy flux through the level z , and this vanishes because there is no buoyancy input except at the surface. If we integrate this vertically we have

$$\langle wb \rangle = H^{-1} \kappa \left[\overline{b}(0) - \overline{b}(-H) \right], \quad (15.30)$$

where the angle brackets denote an average over the entire volume. In the limit $\kappa \rightarrow 0$, the integrated advective buoyancy flux will vanish, because the term $\overline{b}(0) - \overline{b}(-H)$ remains finite. (This follows because b is conserved on parcels, except for the effects of diffusion, which can only act to reduce the value of extrema in the fluid. Thus, $\overline{b}(0) - \overline{b}(-H)$ can only be as large as the temperature difference at the surface, which is set by the boundary conditions.)

Now consider the kinetic energy budget. Using (15.15) and (15.30) we have in a statistically steady state

$$\varepsilon = H^{-1} \kappa \left[\overline{b}(0) - \overline{b}(-H) \right]. \quad (15.31)$$

Because, as noted above, the buoyancy difference on the right-hand side is bounded, the kinetic energy dissipation must go to zero if the thermal diffusivity goes to zero; that is, $\varepsilon \rightarrow 0$ as $\kappa \rightarrow 0$ and in particular $\varepsilon < \kappa \Delta b / H$ where Δb is the maximum buoyancy difference at the surface. We may also consider the limit $(\kappa, \nu) \rightarrow 0$ with a fixed Prandtl number, $\sigma \equiv \nu / \kappa$, and in this limit the energy dissipation also vanishes with κ .

Finally, let us see how the surface buoyancy is related to the buoyancy flux, for any value of κ . Multiplying (15.13b) (with $J = 0$) by b and integrating over the domain gives the buoyancy variance equation

$$\frac{1}{2} \frac{d\langle b^2 \rangle}{dt} = \kappa \left[\overline{b} \frac{\partial b}{\partial z} \Big|_{z=0} - \langle |\nabla b|^2 \rangle \right]. \quad (15.32)$$

We have assumed that the normal derivative of b vanishes on all surfaces except the top one ($z = 0$) and an overbar denotes a horizontal integral. In a statistically steady state,

$$\overline{b} \frac{\partial b}{\partial z} \Big|_{z=0} = \langle |\nabla b|^2 \rangle, \quad (15.33)$$

where the overbar and angle brackets now also imply a time average. The right-hand side is positive definite, and thus there must be a positive correlation between b and $\partial b / \partial z$, meaning that there is a heat flux into the fluid where it is hot, and a heat flux out of the fluid where it is cold. This result holds no matter whether the upper boundary condition is a condition on b or on $\partial b / \partial z$.

Interpretation

The result encapsulated by (15.31) means that, for a fluid forced only at the surface by buoyancy forcing, as the diffusivity goes to zero so does the energy dissipation. For a fluid of finite viscosity the vorticity in the fluid must then go to zero, because $\varepsilon = v \langle \omega^2 \rangle$; this in turn means that the flow cannot be baroclinic, because baroclinicity generates vorticity, even in the presence of viscosity (section 4.2). An even more interesting result follows for a fluid with small viscosity. In turbulent flow, the energy dissipation at high Reynolds number is not a function of the viscosity; if the viscosity is reduced, the cascade of energy to smaller scales merely continues to still smaller scale, generating vorticity at these smaller scales, and the energy dissipation is unaltered, remaining finite even in the limit $v \rightarrow 0$. In contrast, for a fluid heated and cooled only at the upper surface, the energy dissipation *tends to zero* as $\kappa \rightarrow 0$, whether or not one is in the high-Reynolds number limit. This means that vorticity cannot be generated at the viscous scales by the action of a turbulent cascade, as that would lead to energy dissipation. Effectively, the result prohibits an ocean that is forced only at the surface by a buoyancy flux from having an ‘eddy viscosity’ that would enable the fluid to efficiently dissipate energy, and if there is no small scale motion producing an eddy viscosity there can be no eddy diffusivity either. Thus, such an ocean is *non-turbulent*. This is a rather different picture from that of the real ocean, where there is some dissipation of energy in the interior because of breaking gravity waves, and dissipation at the boundary in Ekman layers, and the eddy diffusivity is needed for there to be a non-negligible buoyancy-driven meridional overturning circulation.

Of course, thermal forcing in the ocean is in part an imposed flux, coming from radiation among other things, and this penetrates below the surface. However, this makes little physical difference to the argument, provided that this forcing remains confined to the upper ocean. If so, then for any level below this forcing we still have the result (15.29), and the final result (15.31) holds, assuming that the range of temperatures produced by the forcing is still finite.

15.2.4 The importance of mechanical forcing

The results of (15.19) and (15.31) do not, strictly speaking, prohibit there from being a thermal circulation, with fluid sinking at high latitudes and rising at low, even for zero diffusivity. However, in the absence of any mechanical forcing, this circulation must be laminar, even at high Rayleigh number, meaning that flow is not allowed to break in such a way that energy can be dissipated — a very severe constraint that most flows cannot satisfy. The scalings (15.12) further suggest that the magnitude of the circulation in fact scales (albeit nonlinearly) with the size molecular diffusivity, and if these scalings are correct the circulation will in fact diminish as $\kappa \rightarrow 0$. For small diffusivity, the solution most likely to be adopted by the fluid is for the flow to become confined to a very thin layer at the surface, with no abyssal motion at all, which is completely unrealistic vis-à-vis the observed ocean. Thus, the deep circulation of the ocean cannot be considered to be wholly forced by buoyancy gradients at the surface.

Suppose we add a mechanical forcing, \mathbf{F} , to the right-hand side of (15.13a); this might represent wind forcing at the surface, or tides. The kinetic energy budget becomes

$$\varepsilon = \langle wb \rangle + \langle \mathbf{F} \cdot \mathbf{v} \rangle = H^{-1} \kappa [\bar{b}(0) - \bar{b}(-H)] + \langle \mathbf{F} \cdot \mathbf{v} \rangle. \quad (15.34)$$

In this case, even for $\kappa = 0$, there is a source of energy and therefore turbulence (i.e., a dissipative circulation) can be maintained. The turbulent motion at small scales then provides a mechanism of mixing and so can effectively generate an ‘eddy diffusivity’ of buoyancy. *Given* such an eddy diffusivity, wind forcing is no longer necessary for there to be an overturning circulation. Therefore, it is useful to think of mechanical forcing as having two distinct effects.

- (i) The wind provides a stress on the surface that may directly drive the large-scale circulation, including the overturning circulation. (An explicit example of this is discussed in section 16.5.)
- (ii) Both tides and the wind provide a mechanical source of energy to the system that allows the flow to become turbulent and so provides a source for an eddy diffusivity and eddy viscosity.

In either case, we may conclude that the presence of mechanical forcing is necessary for there to be an overturning circulation in the world’s oceans of the kind observed. In the remainder of this chapter we ignore the effects of item (i), and suppose that the most important effect of the wind is that it enables there to be an eddy diffusivity that is much larger than the molecular one; the eddy diffusivity enables large volumes of the ocean to become mixed, so allowing a buoyancy-driven overturning circulation (a ‘thermohaline circulation’) to exist. We first consider extremely simple models of this circulation, so-called box models.

15.3 SIMPLE BOX MODELS

Even though they are far simpler than the real ocean, the fluid dynamical models of the previous section are still quite daunting. The analysis that can be performed is either very specific and of little generality, for example the construction of solutions at low Rayleigh number, or it is a very general form, being of the form of scaling or energetic arguments at high Rayleigh number. Models based on the fluid dynamical equations do not easily allow for the construction of explicit solutions in the parameter regime — high Rayleigh and Reynolds numbers — of interest. It is therefore useful to consider an extreme simplification of the overturning circulation, *box models*. These are very simple caricatures of the circulation, constructed by dividing the ocean into a very small number of boxes with simple rules determining the transport of fluid properties between them.¹¹ The purist may consider this section a diversion away from a consideration of the fluid dynamical properties of the ocean, and the content implied by the title of this book, but such box models have been quite fecund and an evident source of qualitative understanding, and thus find a place in our discussion.

15.3.1 A two-box model

Consider two boxes as illustrated in Fig. 15.6. Each box is well-mixed and has a uniform temperature and salinity, T_1 , T_2 and S_1 , S_2 . The boxes are connected with a capillary tube at the bottom along which the flow is viscous, obeying Stokes’ Law. That is, the flow along the tube is proportional to the pressure gradient which, because the flow is hydrostatic, is proportional to the density difference between the two boxes. An overflow at the top keeps

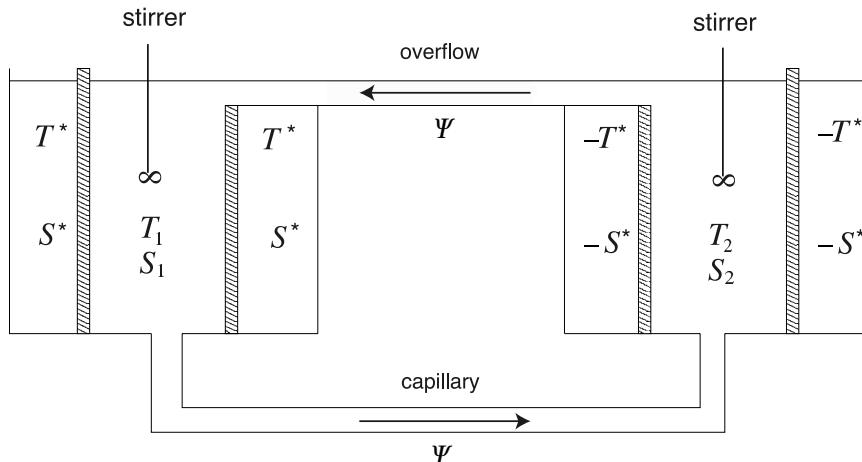


Fig. 15.6 A two-box model of relevance to the overturning circulation of the ocean. The shaded walls are porous, and each box is well mixed by its stirrer. Temperature and salinity evolve by way of fluid exchange between the boxes via the capillary tube and the overflow, and by way of relaxation with the two infinite reservoirs at $(+T^*, +S^*)$ and $(-T^*, -S^*)$.

the upper surfaces of the two boxes at the same level. Thus, the circulation is given by

$$\Psi = A(\rho_1 - \rho_2), \quad (15.35)$$

where Ψ is proportional to the flow in the pipe, ρ_1 and ρ_2 are the densities of the fluids in the two boxes and A is a constant. The boxes are enclosed by porous walls beyond which are reservoirs of constant temperature and salinity, and we are at liberty to choose the origin of the temperature scale such that the two reservoirs are at $+T^*$ and $-T^*$, and similarly for salinity. Thus, heat and salt are transferred into and out of the boxes as represented by simple linear laws and we have

$$\begin{aligned} \frac{dT_1}{dt} &= c(T^* - T_1) - 2|\Psi|(T_1 - T_2), & \frac{dT_2}{dt} &= c(-T^* - T_2) - 2|\Psi|(T_2 - T_1), \\ \frac{dS_1}{dt} &= d(S^* - S_1) - 2|\Psi|(S_1 - S_2), & \frac{dS_2}{dt} &= d(-S^* - S_2) - 2|\Psi|(S_2 - S_1). \end{aligned} \quad (15.36)$$

Note that the advective transfer is independent of the sign the circulation, because it occurs through both the capillary tube and the overflow. From these equations it is easy to show that the sum of the temperatures, $T_1 + T_2$ decays to zero and is uncoupled from the difference, and similarly for salinity. Defining $\hat{T} = (T_1 - T_2)/(2T^*)$ and $\hat{S} = (S_1 - S_2)/(2S^*)$, we obtain

$$\begin{aligned} \frac{d\hat{T}}{dt} &= c(1 - \hat{T}) - 4|\Psi|\hat{T}, \\ \frac{d\hat{S}}{dt} &= d(1 - \hat{S}) - 4|\Psi|\hat{S}. \end{aligned} \quad (15.37)$$

Using a linear equation of state of the form $\rho = \rho_0(1 - \beta_T T + \beta_S S)$ (where the variables are dimensional) the circulation (15.35) becomes

$$\Psi = 2\rho_0 T^* \beta_T A \left(-\hat{T} + \frac{\beta_S S^*}{\beta_T T^*} \hat{S} \right). \quad (15.38)$$

Finally, non-dimensionalizing time using $\tau = ct$, the equations of motion become

$$\frac{d\hat{T}}{d\tau} = (1 - \hat{T}) - |\Phi| \hat{T}, \quad (15.39a)$$

$$\frac{d\hat{S}}{d\tau} = \delta(1 - \hat{S}) - |\Phi| \hat{S}, \quad (15.39b)$$

$$\Phi = -\gamma(\hat{T} - \mu\hat{S}), \quad (15.39c)$$

where $\Phi = 4\Psi/c$ and the three parameters that determine the behaviour of the system are

$$\gamma = \frac{8\rho_0 T^* \beta_T A}{c}, \quad \delta = \frac{d}{c}, \quad \mu = \frac{\beta_S S^*}{\beta_T T^*}. \quad (15.40)$$

The parameter γ measures the overall strength of the forcing in determining the strength of the circulation, and is the ratio of a relaxation time scale to an advective time scale. The parameter δ is the ratio of the reciprocal time constants of temperature and salinity relaxation, and μ is a measure of the ratio of the effect of the salinity and temperature forcings on the density. Salinity transfer will normally be much slower than heat transfer so that $\delta \ll 1$, whereas if salinity and temperature are both to play a role in the dynamics we need $\mu = \mathcal{O}(1)$. We might also expect both advection and relaxation to be important if $\gamma = \mathcal{O}(1)$, and this will depend on the properties of the capillary tube.

Interpretation

Although the above model describes a potentially real system, one that might be constructed in the laboratory, it is the analogy to aspects of the ocean circulation that interests us here. To make the analogy, we suppose that one box represents the entire high-latitude ocean and the other the entire low-latitude ocean, and the capillary tube and the overflow carry the overturning circulation between them. The reservoirs at $\pm T^*$ and $\pm S^*$ represent the atmosphere. Typically, we would choose the low latitudes to be both heated and salted (the latter because of the low rainfall and high evaporation in the subtropics) and the high latitudes to be cooled and freshened by rainfall. Thus, T^* and S^* have the same sign, and they force the circulation in opposite directions.

It is a common fluid-dynamical experience that the behaviour of a highly-truncated system has little resemblance to that of the corresponding continuous system, and so we expect the model to be only be a cartoon of the ocean circulation. For example, we have restricted the circulation to be of basin scale, and the parameterization of the intensity of the overturning circulation by (15.39c) must be regarded with caution, because it represents a frictionally controlled flow rather than a nearly inviscid geostrophic flow. Nevertheless, observations and numerical simulations do indicate that the overturning circulation does have a relatively simple vertical and horizontal structure: the circulation in the North Atlantic is similar to that of a single cell, for example, indicating that an appropriate low-order model may be useful.

One might also question the oceanic appropriateness of the linear relaxation terms. For temperature, the bulk aerodynamic formulae often used to parameterize air-sea fluxes do have a similar form, but the freshening of seawater by rainfall is more akin to an imposed negative flux of salinity, and evaporation is a function of temperature. An alternative might be to impose an effective salt flux so that

$$\frac{d}{dt}(S_1 - S_2) = 2E - 2|\Psi|(S_1 - S_2), \quad (15.41)$$

where E is an imposed, constant, effective rate of salt exchange with the atmosphere. After non-dimensionalization, using E/c to non-dimensionalize salt, (15.39b) is replaced by

$$\frac{dS}{d\tau} = 1 - |\Phi|S. \quad (15.42)$$

Another aspect of the model that is oceanographically questionable is that the model assumes that the water masses can be mixed below the surface. Thus, when water enters one box from the other it immediately mixes with its surroundings. Without the stirrer this would not occur and the equations of the box model would not represent a real system. In the real ocean, most of the mixing of water masses happens near the surface (in the mixed layer) and near lateral boundaries or regions of steep topography. Elsewhere in the ocean, mixing is quite small, and probably far from sufficient to mix a large volume of water in the advective or relaxation times of the box model. Having noted all these objections, we will put them aside and continue with an analysis of the model.

Solutions

Equilibria occur when the time-derivatives vanish, and the circulation then satisfies

$$\Phi = g(\Phi) \equiv \gamma \left(\frac{-1}{1 + |\Phi|} + \frac{\mu}{1 + |\Phi|/\delta} \right). \quad (15.43)$$

A graphical solution of this is obtained as the intercept of the right-hand side with the left-hand side, the latter being a straight line through the origin at an angle of 45° , and this is plotted in Fig. 15.7. Perhaps the most interesting aspect of the solutions is that they exhibit *multiple equilibria*; that is, there are multiple steady solutions with the same parameters.

Evidently, for a range of parameters three solutions are possible, whereas for others only one solution exists. Although a fairly complete analysis of the nature of the steady solutions is possible, it is instructive to consider the special case with $\gamma \gg 1$ and $\delta \ll 1$. This corresponds to the situation in which the advective time scale is shorter than the diffusive one and temperature relaxation is much faster than salt relaxation. Using the graphical solution as a guide, two of the solutions are then close to the origin, with $\Phi \ll 1$, and satisfy

$$\Phi \approx \gamma \left(-1 + \frac{\mu\delta}{\delta + |\Phi|} \right), \quad (15.44)$$

giving, for small $|\Phi|$ and $\mu > 1$,

$$\Phi \approx \pm[\delta(\mu - 1)]. \quad (15.45)$$

The positive solution, with flow in the capillary tube from box 1 to box 2, is salinity driven — driven by the density gradient of the same sign as that caused by the salinity, with the

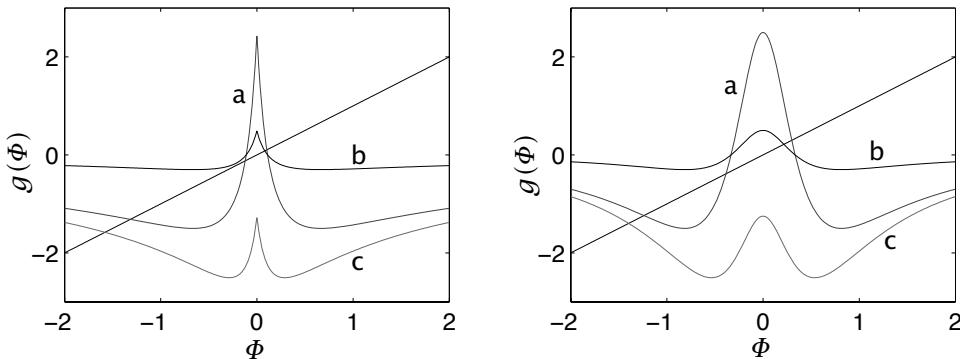


Fig. 15.7 Left panel: graphical solution of the two-box model. The straight line has unit slope and passes through the origin, and the curved lines plot the function $f(\Phi)$ as given by the right-hand side of (15.43). The intercepts of the two are solutions to the equation. The parameters for the three curves are: a, $\gamma = 5$, $\delta = 1/6$, $\mu = 1.5$; b, $\gamma = 1$, $\delta = 1/6$, $\mu = 1.5$; c, $\gamma = 5$, $\delta = 1/6$, $\mu = 0.75$. Right panel: the same except with Φ^2 in place of $|\Phi|$ on the rhs of (15.43).

density gradient due to temperature opposing the motion. That is, box 1 is denser than box 2 because it is more saline, even though it is also warmer. The negative solution is thermally driven, with the flow in the capillary tube going from the denser (cold and fresh) box 2 to the lighter (warm and salty) box 1. However, this solution is unstable, and any small perturbation will amplify and the system will move away from this solution. Solving for temperature and salinity we find that $T \approx 1$ (i.e., it is close to its relaxation value and hardly altered by advection), and $S \approx 1/\mu$. (A more complete analysis of a very similar system is set as problem 15.7.)

The third solution has a circulation far from the origin, and the balance in (15.43) is between the left-hand side and the first term on the right. In the limiting case we find

$$\Phi \approx -\sqrt{\gamma}. \quad (15.46)$$

This solution has a density gradient dominated by the temperature effect: the temperature difference is $T \approx 1/\sqrt{\gamma}$ whereas the salinity difference is $S \approx \delta/\sqrt{\gamma}$, and thus its effect on density is much smaller.

15.3.2 * More boxes

More boxes can be added in a variety of ways and, now forgoing an easy relevance to a laboratory apparatus, one such is illustrated in Fig. 15.8. The three boxes represent the mid- and high-latitude Northern Hemisphere, the mid- and high-latitude Southern Hemisphere, and the equatorial regions. Each of the three boxes can exchange fluid with its neighbour, and each is also in contact with a reservoir and subject to a relaxation to a fixed value of temperature and salinity, (T_s^*, S_s^*) , (T_e^*, S_e^*) and (T_n^*, S_n^*) . (A variation on this theme allows direct communication between the two poleward boxes.) Then, with obvious notation, we

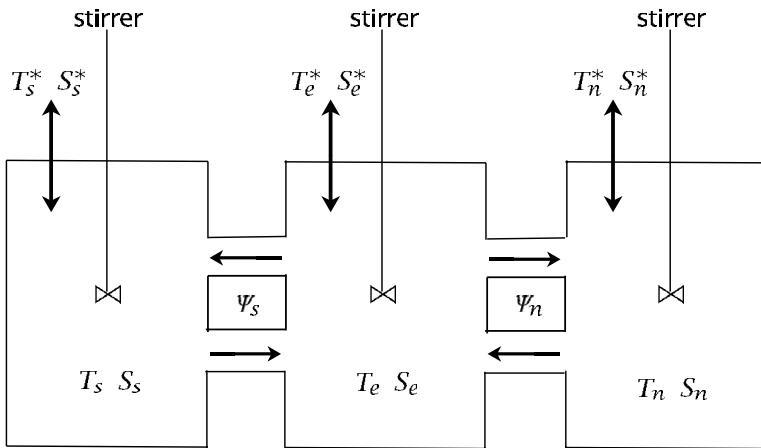


Fig. 15.8 A three-box model. Each box contains fluid with uniform values of temperature and salinity, each exchanges fluid with its neighbour, and in each the temperature and salinity are relaxed toward fixed atmospheric values.

infer the equations of motion:

$$\begin{aligned} \frac{dT_s}{dt} &= c(T_s^* - T_s) - 2|\Psi_s|(T_s - T_e), & \frac{dT_n}{dt} &= c(T_n^* - T_n) - 2|\Psi_n|(T_n - T_e), \\ \frac{dT_e}{dt} &= c(T_e^* - T_e) - 2|\Psi_s|(T_e - T_s) - 2|\Psi_n|(T_e - T_n), \end{aligned} \quad (15.47a)$$

$$\begin{aligned} \frac{dS_s}{dt} &= d(S_s^* - S_s) - 2|\Psi_s|(S_s - S_e), & \frac{dS_n}{dt} &= d(S_n^* - S_n) - 2|\Psi_n|(S_n - S_e), \\ \frac{dS_e}{dt} &= d(S_e^* - S_e) - 2|\Psi_s|(S_e - S_s) - 2|\Psi_n|(S_e - S_n), \end{aligned} \quad (15.47b)$$

with flow rates given by the density differences

$$\Psi_s = A\rho_0[-\beta_T(T_s - T_e) + \beta_S(S_s - S_e)], \quad \Psi_n = A\rho_0[-\beta_T(T_n - T_e) + \beta_S(S_n - S_e)]. \quad (15.48)$$

These equations may be non-dimensionalized and reduced to four prognostic equations for the quantities $T_e - T_n$, $T_e - T_s$, $S_e - S_n$, $S_e - S_s$. Not surprisingly, multiple equilibria can again be found. One interesting aspect is that stable asymmetric solutions arise with symmetric forcing ($T_s^* = T_n^*$, $S_s^* = S_n^*$). These effectively have a pole-to-pole circulation, illustrated in the upper row of Fig. 15.9. Such a circulation can be thought of as the superposition of a thermal circulation in one hemisphere and a salinity-driven circulation in the other.

The box models are useful because they are suggestive of behaviour that might occur in real fluid systems, and because they provide a means of interpreting behaviour that does occur in more complete numerical models and perhaps in the real world. However, they are by no means good approximations of the real equations of motion and without other supporting evidence the solutions found in box models should not be regarded as representing real solutions of the fluid equations for the world's oceans.¹³

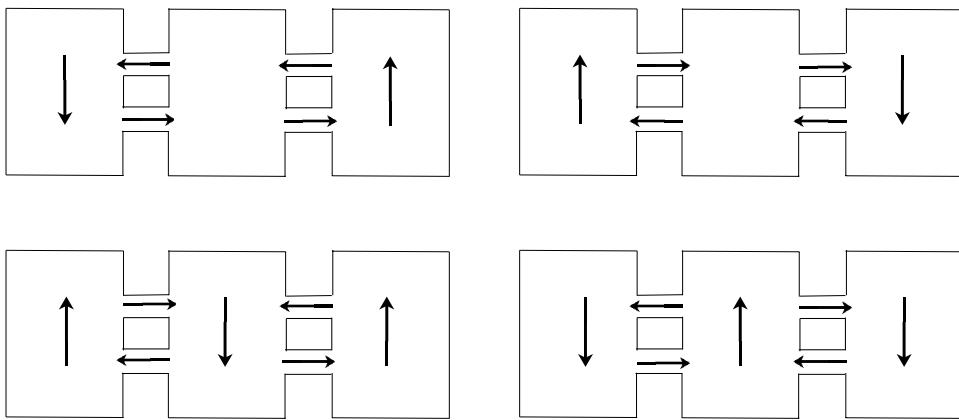


Fig. 15.9 Schematic of four solutions to the three box model with the symmetric forcing $S_s^* = S_n^*$ and $T_n^* = T_s^*$. The two solutions on the top row have an asymmetric, ‘pole-to-pole’, circulation whereas the solutions on the bottom row are symmetric.¹²

15.4 A LABORATORY MODEL OF THE ABYSSAL CIRCULATION

We now return to a more fluid dynamical description of the deep ocean circulation, and consider two simple, closely related, models that are relevant to aspects of the deep circulation. The first, which we consider in this section, is a laboratory model, originally envisioned as being a prototype for the deep circulation. The second model, considered in the following sections, is explicitly a model of the deep circulation. Both models are severe idealizations that describe only limited aspects of the circulation, but they are both very helpful tools that enable us to understand more complete models and, in part, the real circulation itself.

15.4.1 Set-up of the laboratory model

Let us consider flow in a rotating tank, as illustrated in Fig. 15.10. The fluid is confined by vertical walls to occupy a sector, and the entire tank rotates anticlockwise when viewed from above, like the Northern Hemisphere. When the fluid is stationary in the rotating frame, the fluid slopes up toward the outer edge of the tank and the balance of forces in the rotating frame is between a centrifugal force pointing outwards and the pressure gradient due to the sloping fluid pointing inwards. In the inertial frame of the laboratory itself, the pressure gradient pointing inwards provides a centripetal force that causes the fluid to accelerate toward the centre of the tank, resulting in a circular motion. (Recall that steady circular motion is always accompanied by an acceleration toward the centre of the circle.)

This set-up, and the accompanying theory, has become known as the *Stommel-Arons-Faller* model.¹⁴ The motivation of the construct is clear, in that the sector represents an ocean basin. However, rather than driving the fluid with wind or by differential heating, we drive it with localized mass sources and sinks, for example from small pipes inserted into the tank. (If an oceanic analogy is desired, the mass source might be thought of as representing a source of deep abyssal water due to deep convection. However the oceanic

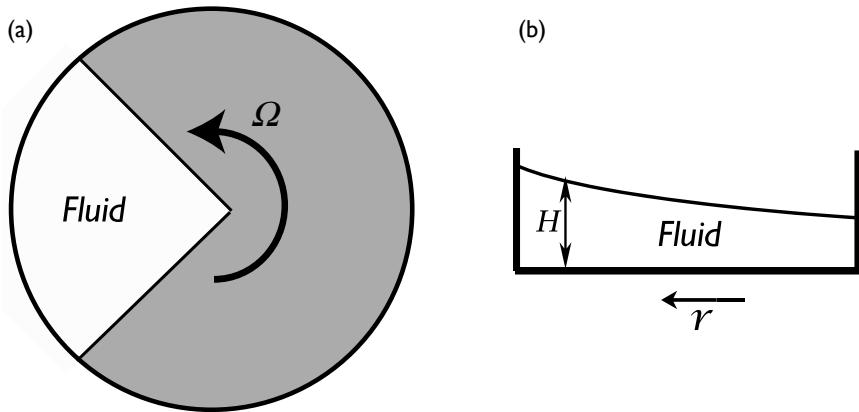


Fig. 15.10 The experimental set-up in the Stommel–Arons–Faller rotating tank experiment. (a) A plan view of the apparatus. The fluid is contained in the sector at left. (b) Side view. The free surface of the fluid slopes up with increasing radius, giving a balance (in the rotating frame) between the centrifugal force pointing outwards and the pressure force pointing inwards. Small pipes may be introduced into the fluid to provide mass sources and sinks.

analogy is not perfect, and the model is best thought of as a purely laboratory one, but one that will build intuition for consideration of the real ocean.)

15.4.2 Dynamics of flow in the tank

Let us assume that the motion of the fluid in the tank is sufficiently weak that its Rossby number is small, and that it obeys the shallow water planetary-geostrophic equations, namely

$$f_0 \times \mathbf{u} = -g \nabla h + \Omega^2 r \hat{\mathbf{r}} + \mathbf{F}, \quad (15.49a)$$

$$\frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{u} h) = S, \quad (15.49b)$$

where $\mathbf{u} = (v^r, v^\theta)$ is the horizontal velocity in cylindrical (r, θ) coordinates, $\hat{\mathbf{r}}$ is a unit vector in the direction of increasing r , \mathbf{F} represents frictional terms (which we will suppose are small except in boundary layers) and S represents mass sources. These two equations yield the potential vorticity equation,

$$\frac{D}{Dt} \left(\frac{f_0}{h} \right) = \frac{\text{curl}_z \mathbf{F}}{h} - \frac{f_0 S}{h^2}. \quad (15.50)$$

Let us write the height field as

$$h = H(r, t) + \eta(r, \theta, t) \quad (15.51)$$

where $H(r, t)$ is the height field corresponding to the rest state of the fluid (in the rotating frame) and η the perturbation. Thus, from (15.49a)

$$0 = -g \nabla H + \Omega^2 r \hat{\mathbf{r}}, \quad (15.52)$$

which gives

$$H = \frac{\Omega^2 r^2}{2g} + \hat{H}(t), \quad (15.53)$$

where \hat{H} is a measure of the overall mass of the fluid. Its rate of change is determined by the mass source

$$\frac{d\hat{H}}{dt} = \langle S \rangle, \quad (15.54)$$

the angle brackets indicating a domain average. The equations of motion (15.49a), and (15.49b) become

$$\mathbf{f}_0 \times \mathbf{u} = -g \nabla \eta + \mathbf{F}, \quad (15.55a)$$

$$\frac{\partial}{\partial t}(\eta + H) + \nabla \cdot [\mathbf{u}(\eta + H)] = 0. \quad (15.55b)$$

Equation (15.55a) tells us that, away from frictional regions, the velocity is in geostrophic balance with the pressure field due to the perturbation height η .

Let us now suppose $|\eta| \ll H$, which holds if the mass source is small and gentle enough. Then (15.55b) may be written

$$\frac{\partial H}{\partial t} + \nabla \cdot (\mathbf{u}H) = 0. \quad (15.56)$$

In this approximation, the potential vorticity equation (15.50) becomes, away from friction and mass sources,

$$\frac{D}{Dt} \left(\frac{f_0}{H} \right) = 0 \quad \text{or} \quad \frac{DH}{Dt} = 0, \quad (15.57a,b)$$

where the second equation follows because f_0 is a constant. [This equation also follows directly from (15.56), because the velocity is geostrophic and divergence-free where friction is absent; however, it is better thought of as a potential vorticity equation, not a mass conservation equation.] Equation (15.57b) means that fluid columns change position in order to keep the same value of H . Further, because H only varies with r , (15.57b) becomes

$$\frac{\partial H}{\partial t} + v^r \frac{\partial H}{\partial r} = 0, \quad (15.58)$$

which, using (15.53) and (15.54), gives

$$v^r = -\frac{g}{\Omega^2 r} \langle S \rangle. \quad (15.59)$$

This is a remarkable result, for it implies that, if $\langle S \rangle$ is positive, the flow is *toward* the apex of the dish, except at the location of the mass sources and in frictional boundary layers, *no matter where the mass source is actually located*. The explanation of this counter-intuitive result is simple enough. If $\langle S \rangle > 0$ the overall height of the fluid is increases with time; thus, in order that a given material column of fluid keep its height fixed, it must move toward the apex of the dish. The full velocity field may be obtained, away from the frictional regions, using the divergence-free nature of the velocity:

$$\nabla \cdot \mathbf{u} = \frac{1}{r} \left[\frac{\partial(rv^r)}{\partial r} + \frac{\partial v^\theta}{\partial \theta} \right] = 0. \quad (15.60)$$

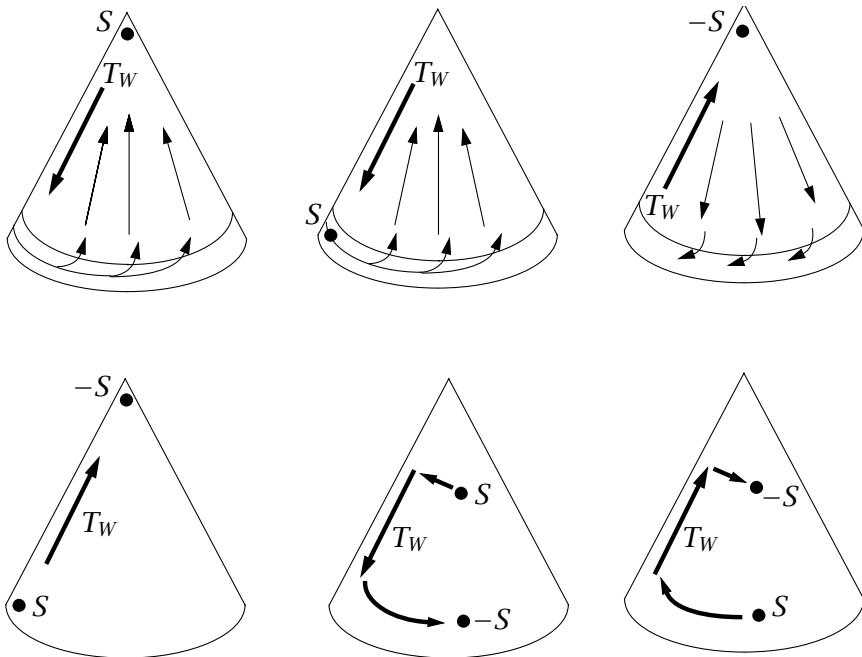


Fig. 15.11 Idealized examples of the flow in the rotating sector experiments, with various locations of a source (S) or sink ($-S$) of mass.

Then, using (15.59), $\partial v^\theta / \partial \theta = 0$ except at a source or sink, or in a frictional boundary layer. Assuming there is only one frictional boundary layer, $v^\theta = 0$ except at those latitudes (i.e., values of r) that contain a mass source or sink.

To balance the flow toward the apex there must, then, be a *boundary layer* in which the flow has the opposite sense, and therefore in which frictional effects are important. To determine where the boundary layer is — on the east or west side of the domain — we need some vorticity dynamics. Away from the mass source, but including friction, the potential vorticity equation is

$$\frac{D}{Dt} \left(\frac{f_0}{H} \right) = \frac{\text{curl}_z \mathbf{F}}{H} \quad \text{or} \quad - \frac{f_0}{H^2} \frac{DH}{Dt} = \text{curl}_z \mathbf{F}, \quad (15.61a,b)$$

and the free surface of the water slopes downwards toward the apex, as illustrated in Fig. 15.10. Now, suppose that there are mass source and a sink of equal magnitudes, with the source further from the apex than the sink, as in the panel at the bottom right of Fig. 15.11. The flow from source to sink must be along either the left or right boundary of the container. To see which, note that the flow is toward smaller values of H , and therefore the left-hand side of (15.61a) is positive. To balance this, the friction in the boundary current must import a positive vorticity to the flow (i.e., $\text{curl}_z \mathbf{F} > 0$), which means in general that the flow itself must have negative vorticity, and the flow is clockwise. (For example, if $\mathbf{F} = -\lambda \mathbf{u}$ the right hand side of (15.61) is $-(\lambda/H) \text{curl}_z \mathbf{u}$ and this is positive if the flow is clockwise.) Clockwise flow implies a *western* boundary layer, on the left of the container. A western boundary layer is a general feature, not dependent on the placement of mass sources or sinks. For suppose

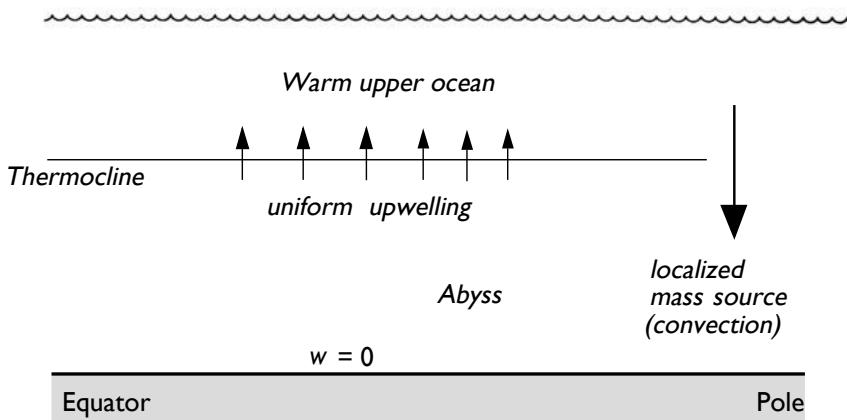


Fig. 15.12 The structure of simple Stommel-Arons ocean model of the abyssal circulation. Convection at high latitudes provides a localized mass-source to the lower layer, and upwelling through the thermocline provides a more uniform mass sink.

there is a single source of mass, as for example in the upper left example of Fig. 15.11; the interior mass flow will then be toward the apex and the flow in the boundary layer away from the apex. The left-hand side of (15.61a) is then negative, and so $\text{curl}_z \mathbf{F}$ must be negative. The flow must then have an anticlockwise sense, again requiring a western boundary layer to achieve a balance in the potential vorticity equation.

The flow is in some ways analogous to flow on the β -plane, and in particular:

- (i) the r -dependence of the height field provides a background potential vorticity gradient, analogous to the β -effect;
- (ii) the time-dependence of H is analogous to a wind curl, since it is this that ultimately drives the fluid motion.

The analogies are drawn out explicitly in the shaded box on the next page; the box also includes a column for abyssal flow in the ocean, discussed in the next two sections.

15.5 A MODEL FOR OCEANIC ABYSSAL FLOW

We will now extend the reasoning applied to the rotating tank to the rotating sphere, and so construct a model — the *Stommel-Arons model* — of the abyssal flow in the ocean.¹⁵ The basic idea is simple: we will model the deep ocean as a single layer of homogeneous fluid in which there is a localized injection of mass at high latitudes, representing convection (Fig. 15.12). However, unlike the rotating dish, mass is extracted from this layer by upwelling into the warmer waters above it, keeping the average thickness of the abyssal layer constant. We assume that this upwelling is nearly uniform, that the ocean is flat-bottomed, and that a passive western boundary current may be invoked to satisfy mass conservation, and which does not affect the interior flow. Obviously, these assumptions are very severe and the model can at best be a conceptual model of the real ocean. Given that, we will work in Cartesian coordinates on the β -plane, and use the planetary-geostrophic approximation. Our

Analogies Between a Rotating Dish, Wind-driven and Abyssal flows

Consider homogeneous models of (i) rotating dish, (ii) wind-driven flow on the β -plane, and (iii) abyssal flow on the β -plane. We model them all with a single layer of homogeneous fluid satisfying the planetary geostrophic equations. In (i) the mass source, $\langle S \rangle$, is localized and the total depth of the fluid layer changes with time; fluid columns move to keep their depth constant. In (ii) there is no mass source and the depth of the fluid layer is constant; the fluid motion is determined by the wind-stress curl, $\text{curl}_z \boldsymbol{\tau}$, and by β . In (iii) the fluid source (convection) is localized at high latitudes and exactly balanced by a mass loss, S_u , due to upwelling everywhere else, so that layer depth is constant and S_u is uniform and negative nearly everywhere. The equations below then apply away from frictional boundary layers and localized mass sources.

(i) Rotating dish

(ii) *Wind-driven flow*

(iii) Abyssal flow

PV Conservation

$$\frac{D}{Dt} \left(\frac{f_0}{H} \right) = 0 \quad \frac{D}{Dt} \left(\frac{f}{H_0} \right) = \frac{1}{H_0} \operatorname{curl}_z \boldsymbol{\tau} \quad \frac{D}{Dt} \left(\frac{f}{h} \right) = - \frac{f S_u}{h^2}$$

This leads to

$$\nu r \frac{\partial H}{\partial r} = - \frac{\partial H}{\partial t} \quad \quad \quad \frac{\nu}{H_0} \frac{\partial f}{\partial \gamma} = \frac{1}{H_0} \operatorname{curl}_z \boldsymbol{\tau} \quad \quad \quad \frac{\nu}{h} \frac{\partial f}{\partial \gamma} = - \frac{f S_u}{h^2}$$

and

$$v^r = -\frac{g}{\Omega^2 r} \langle S \rangle \quad v = \frac{1}{\beta} \text{curl}_z \boldsymbol{\tau} \quad v = -\frac{f S_u h}{\beta}$$

$\langle S \rangle$ is localized mass source

$\text{curl}_z \tau$ is wind stress
curl

S_u is upwelling
mass loss

Meridional mass flow away from boundaries is thus determined by:

sign (and not location) of localized mass source, $\langle S \rangle$, sign of wind-stress curl, $\text{curl}_z \mathbf{\tau}$, upwelling and sign of f , so polewards if $S_u < 0$ (upwelling).

treatment in this section is physically based but quite heuristic; in section 15.6 we are a little more mathematical and a little more formal.

The momentum and mass continuity equations are

$$\mathbf{f} \times \mathbf{u} = -\nabla_z \phi \quad \text{and} \quad \nabla_z \cdot \mathbf{u} = -\frac{\partial w}{\partial z}, \quad (15.62a,b)$$

where $\mathbf{f} = (f_0 + \beta y)\mathbf{k}$. On elimination of ϕ , (15.62) yields the now-familiar balance,

$$\beta v = f \frac{\partial w}{\partial z}. \quad (15.63)$$

Except in the localized regions of convection, the vertical velocity is, by assumption, positive and uniform at the top of the lower layer, and zero at the bottom. Thus (15.63) becomes

$$v = \frac{f}{\beta} \frac{w_0}{H}, \quad (15.64)$$

where w_0 is the uniform upwelling velocity and H the layer thickness. Thus, the flow is *polewards* everywhere (including the Southern Hemisphere), vanishing at the equator.

15.5.1 Completing the solution

Since $v = f^{-1}(\partial \phi / \partial x)$, the pressure is given by

$$\phi = \int_{x_0}^x \left(\frac{f^2 w_0}{\beta H} \right) dx', \quad (15.65)$$

where x_0 is a constant of integration, to be determined by the boundary conditions. Because there no flow into the Eastern boundary, x_E , we set $\phi = \text{constant}$ at $x = x_E$, and because this is a one-layer model we are at liberty to set that constant equal to zero. Thus,

$$\phi(x) = - \int_x^{x_E} \left(\frac{f^2 w_0}{\beta H} \right) dx' = - \frac{f^2}{\beta H} w_0 (x_E - x). \quad (15.66)$$

The zonal velocity follows using geostrophic balance,

$$u = \frac{1}{f} \frac{\partial \phi}{\partial y} = \frac{2}{H} w_0 (x_E - x), \quad (15.67)$$

where we have also used $\partial f / \partial y = \beta$ and $\partial \beta / \partial y = 0$. Thus the velocity is eastwards in the interior, and independent of f and latitude, provided x_E is not a function of y .

Using (15.64) and (15.67) we can confirm mass conservation is indeed satisfied:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = -\frac{2w_0}{H} + \frac{w_0}{H} + \frac{w_0}{H} = 0. \quad (15.68)$$

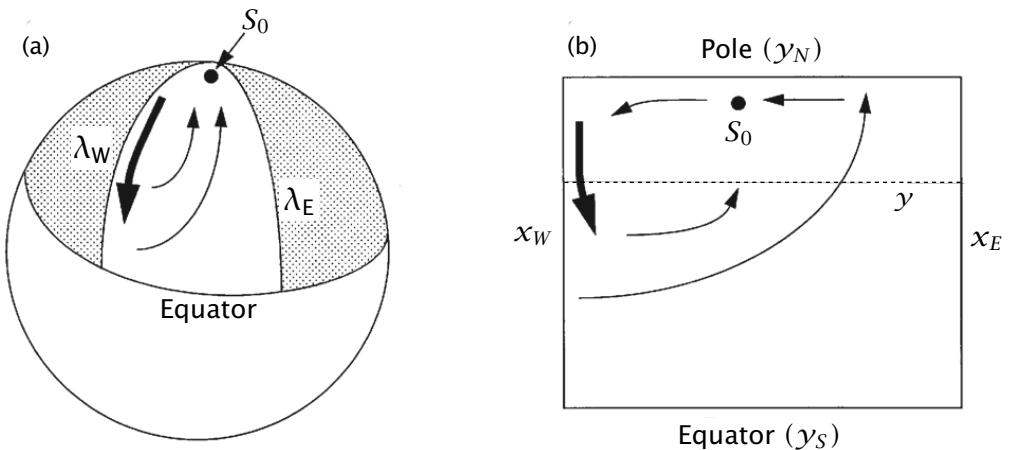


Fig. 15.13 Abyssal circulation in a spherical sector (left) and in a corresponding Cartesian rectangle (right).

15.5.2 Application to the ocean

Let us consider a rectangular ocean with a mass source at the northern boundary, balanced by uniform upwelling (see Figs. 15.13 and 15.14). Since the interior flow will be northwards, we anticipate a southwards flowing western boundary current to balance mass. Conservation of mass in the area polewards of the latitude y demands that

$$S_0 + T_I(y) = T_W(y) + U(y), \quad (15.69)$$

where S_0 is the strength of the source, T_W the equatorwards transport in the western boundary current, T_I the poleward transport in the interior, and U is the integrated loss due to upwelling polewards of y . Then, using (15.64),

$$T_I = \int_{x_W}^{x_E} v H dx = \int_{x_W}^{x_E} \frac{f w_0}{\beta} dx = \frac{f}{\beta} w (x_E - x_W). \quad (15.70)$$

The upwelling loss is given by

$$U = \int_{x_W}^{x_E} \int_{y}^{y_N} w dx dy = w_0 (x_E - x_W) (y_N - y), \quad (15.71)$$

where y_N denotes the northern (polar) boundary. Assuming the source term is known, then using (15.69) we obtain the strength of the western boundary current,

$$T_W(y) = S_0 + T_I - U = S_0 + \frac{f}{\beta} w (x_E - x_W) - w_0 (x_E - x_W) (y_N - y). \quad (15.72)$$

To close the problem we note that over the entire basin mass must be balanced, which gives a relationship between w_0 and S_0 ,

$$S_0 = w_0 \Delta x \Delta y, \quad (15.73)$$

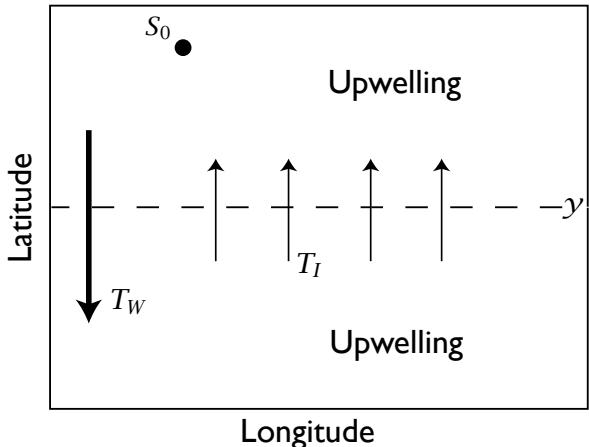


Fig. 15.14 Mass budget in an idealized abyssal ocean. Polewards of some latitude y , the mass source (S_0) plus the poleward mass flux across y (T_I) are equal to the sum of the equatorward mass flux in the western boundary current (T_W) and the integrated loss due to upwelling (U) polewards of y . See (15.69).

where $\Delta x = x_E - x_W$ and $\Delta y = y_N - y_S$, with y_S being the southern boundary of the domain. Using (15.73), (15.72) becomes

$$T_W(y) = -w_0 \left(\Delta x(y_N - y) - \frac{f}{\beta} \Delta x - \Delta x \Delta y \right) = w_0 \Delta x \left(y - y_S + \frac{f}{\beta} \right). \quad (15.74)$$

With no loss of generality we will take $y_S = 0$ and $f = f_0 + \beta y$. Then (15.74) becomes

$$T_W(y) = w_0 \Delta x (2y + f_0/\beta) \quad (15.75)$$

or, using $S_0 = w_0 \Delta x y_N$,

$$T_W(y) = \frac{S_0}{y_N} \left(2y - \frac{f_0}{\beta} \right). \quad (15.76)$$

With a slight loss of generality (but consistent with the spirit of the planetary-geostrophic approximation) we take $f_0 = 0$, which is equivalent to supposing that the equatorial boundary of the domain is at the equator, and finally obtain

$$T_W(y) = 2S_0 \frac{y}{y_N}. \quad (15.77)$$

At the northern boundary this becomes

$$T_W(y) = 2S_0, \quad (15.78)$$

which means that the flow southwards from the source is *twice* the strength of the source itself! We also see that:

- (i) the western boundary current is equatorwards everywhere;
- (ii) at the northern boundary the equatorward transport in the western boundary current is equal to *twice* the strength of the source;
- (iii) the northward mass flux at the northern boundary is equal to the strength of the source itself.

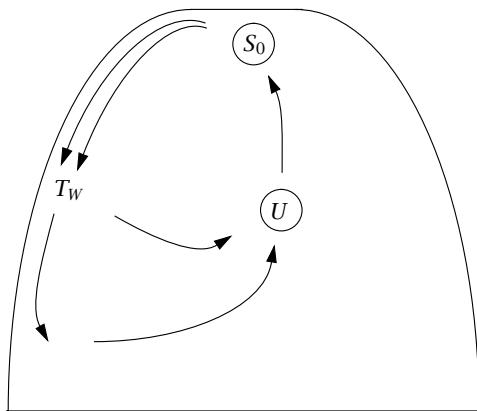


Fig. 15.15 Schematic of a Stommel-Arons circulation in a single sector. The transport of the western boundary current is greater than that provided by the source at the apex, illustrating the property of *recirculation*. The transport in the western boundary current T_W decreases in intensity equatorwards, as it loses mass to the polewards interior flow, and thence to upwelling. The integrated sink, due to upwelling, U , exactly matches the strength of the source, S .

We may check this last point directly: from (15.70)

$$T_I(\gamma_N) = \frac{\beta \gamma_N}{\beta} w_0 \Delta x = S. \quad (15.79)$$

The fact that convergence at the pole balances T_W and S_0 does not of course depend on the particular choice we made for f and γ_S .

The flow pattern evidently has the property of *recirculation* (see Fig. 15.15): this is one of the most important properties of the solution, and one that is likely to transcend all the limitations inherent in the model. This single-hemisphere model may be thought of as a crude model for aspects of the abyssal circulation in the North Atlantic, in which convection at high latitudes near Greenland is at least partially associated with the abyssal circulation. In the North Pacific there is, in contrast, little if any deep convection to act as a mass source. Rather, the deep circulation is driven by mass sources in the opposite hemisphere, and we now consider a simple model of this.

15.5.3 A two-hemisphere model

Our treatment now is even more obviously heuristic, since our domain crosses the equator yet we continue to use the planetary-geostrophic equations, invalid at the equator. We also persist with Cartesian geometry, even for these global-scale flows. In our defense, we note that the value of the solutions lies in their qualitative structure, not in their quantitative predictions. Let us consider a situation with a source in the Southern Hemisphere but none in the Northern Hemisphere. For later convenience we take the Southern Hemisphere source to be of strength $2S_0$, and we suppose the two hemispheres have equal area. As before, the upwelling is uniform, so that to satisfy global mass balance

$$S_0 = w_0 \Delta x \Delta y, \quad (15.80)$$

where $\Delta x \Delta y$ is the area of each hemisphere. Then, given w_0 , the zonally integrated poleward interior flow in each hemisphere, away from the equator, follows from Sverdrup balance,

$$T_I(\gamma) = \frac{f}{\beta} w_0 (x_E - x_W) = S_0 \frac{\gamma}{\gamma_p}, \quad (15.81)$$

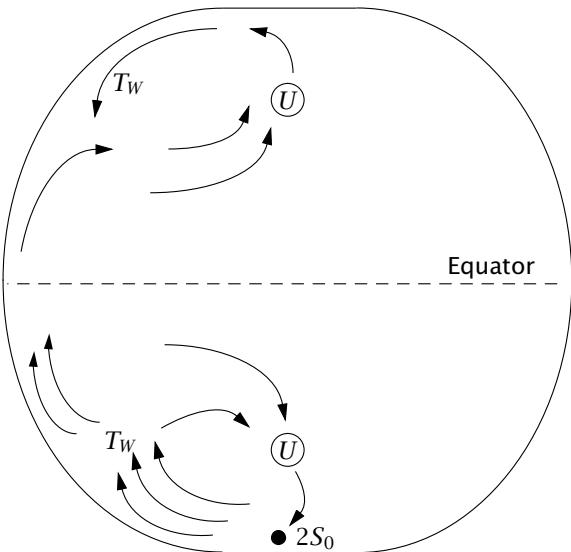


Fig. 15.16 Schematic of a Stommel-Arons circulation in a two-hemisphere basin. There is only one mass source, and this is in the Southern Hemisphere and for convenience it has a strength of 2. Although there is no source in the Northern Hemisphere, there is still a western boundary current and a recirculation. The integrated sinks due to upwelling exactly match the strength of the source.

where γ_p is either γ_N (the northern boundary) or γ_S . The western boundary current is assumed to ‘take up the slack’, that is to be able to adjust its strength to satisfy mass conservation. Thus, since $T_I(\gamma_N) = S_0$, where S_0 is half the strength of the source in the Southern Hemisphere, it is plain that there must be a southwards flowing western boundary current near the northern end of the Northern Hemisphere, even in the absence of any deep water formation there!

In the northern hemisphere, the total loss due to upwelling polewards of a latitude γ is given by

$$U(\gamma) = w_0 \Delta x |\gamma_N - \gamma|. \quad (15.82)$$

The strength of the western boundary current is then given by, with southward flow positive,

$$T_W(\gamma) = T_I - U = \frac{f}{\beta} w_0 \Delta x - w_0 \Delta x (\gamma_N - \gamma) = -w_0 \Delta x (\gamma_N - 2\gamma), \quad (15.83)$$

using $f = \beta\gamma$. The boundary current thus changes sign halfway between equator and North Pole, at $\gamma = \gamma_N/2$. [In spherical coordinates, the analogous latitude turns out to be at $\theta = \sin^{-1}(1/2)$.] At the North Pole $\gamma = \gamma_N$ and we have

$$T_W(\gamma_N) = w_0 \Delta x \gamma_N = S. \quad (15.84)$$

The solution is illustrated schematically in Fig. 15.16. We can (rather fancifully) imagine this to represent the abyssal circulation in the Pacific Ocean, with no source of deep water at high northern latitudes.

15.6 * A SHALLOW WATER MODEL OF THE ABYSSAL FLOW

We can obtain a more complete solution of the flow that explicitly includes the western boundary current by constructing a shallow water Stommel-Arons type model of the abyssal

circulation.¹⁶ However, the essential dynamics is the same as that of the previous section. The model is similar to that illustrated in Fig. 15.12 and comprises a single moving layer of homogeneous fluid lying underneath a lighter, stationary layer. The lower fluid is forced by a mass source at its poleward end that represents convection, and by a uniform mass sink everywhere else that represents diffusive upwelling into the upper layer. The mass source and sink are specified — that is, they are not functions of the flow — and are equal and opposite, so that there is no net mass source. The motion of the lower layer is then governed by the planetary geostrophic reduced-gravity shallow water equations (sections 3.2 and 5.2), to wit:

$$-fv = -g' \frac{\partial h}{\partial x} - ru, \quad fu = -g' \frac{\partial h}{\partial y} - rv, \quad \frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{u}h) = S. \quad (15.85a,b,c)$$

Here h is the thickness of the lower layer (below the thermocline of Fig. 15.12), $\mathbf{u} = (u, v)$ is the two-dimensional velocity, g' is the reduced gravity, and r is a constant frictional coefficient; we will assume this is small, and in particular that $r \ll f$. The mass source term S on the right-hand side of the mass continuity equation represents both upwelling and a localized convective source at the poleward end of the domain. We will assume the upwelling is uniform and that, when integrated over area, it exactly balances the convective mass source. Thus, we write $S = S_0 + S_u$ where S_u is uniform and negative, S_0 is the localized convective source, and $\int_A S \, dA = 0$.

15.6.1 Potential vorticity and poleward interior flow

Straightforward manipulation of (15.85) gives the potential vorticity equation

$$\frac{D}{Dt} \left(\frac{f}{h} \right) = -\frac{r}{h} \text{curl}_z \mathbf{u} - \frac{fS}{h^2}. \quad (15.86)$$

Away from boundaries the first term on the right-hand side is negligible, and so, away from the convective source and in a steady state (15.86) becomes

$$\frac{\beta v}{h} + \frac{f}{h^2} \mathbf{u} \cdot \nabla h = -\frac{fS_u}{h^2}. \quad (15.87)$$

However, the second term on the left-hand side is small if friction is small, for then the flow is nearly in geostrophic balance and, using (15.85a) with $r = 0$, it follows that

$$f\mathbf{u} \cdot \nabla h = -g' \frac{\partial h}{\partial y} \frac{\partial h}{\partial x} + g' \frac{\partial h}{\partial x} \frac{\partial h}{\partial y} = 0. \quad (15.88)$$

The potential vorticity equation is then just

$$\beta v = -fS_u/h, \quad (15.89)$$

and because $S_u < 0$ the flow is polewards, regardless of the location of the convective mass source. From the perspective of the continuously stratified equations, the corresponding potential vorticity equation is just

$$\beta v = f \frac{\partial w}{\partial z}. \quad (15.90)$$

Upwelling from the abyss into the upper ocean corresponds to a positive value of the stretching term $f\partial w/\partial z$, and again the interior flow is polewards.

15.6.2 The solution

It is convenient to deal with mass transports rather than velocities and we define $\mathbf{U} \equiv U\mathbf{i} + V\mathbf{j} \equiv uh\mathbf{i} + vh\mathbf{j}$. Away from the convective source equations of motion, (15.85), become, in steady state,

$$-fV = -\frac{\partial\Phi}{\partial x} - rU, \quad fU = -\frac{\partial\Phi}{\partial y} - rV, \quad \nabla \cdot \mathbf{U} = S_u, \quad (15.91\text{a,b,c})$$

where $\Phi = g'h^2/2$. For small friction ($r \ll f$) we may write (15.91a,b) as

$$V = \frac{1}{f} \frac{\partial\Phi}{\partial x} - \frac{r}{f^2} \frac{\partial\Phi}{\partial y}, \quad U = -\frac{1}{f} \frac{\partial\Phi}{\partial y} - \frac{r}{f^2} \frac{\partial\Phi}{\partial x}, \quad (15.92\text{a,b})$$

which, after differentiation and use of (15.91c), combine to give

$$\frac{\beta}{f} \frac{\partial\Phi}{\partial x} = -fS_u - f\nabla \cdot \left(\frac{r}{f^2} \nabla\Phi \right). \quad (15.93)$$

In the interior where the effects of friction are small (15.93) becomes $\beta V = -fS$, so recovering (15.89) and implying poleward interior flow. However, by mass conservation, the flow cannot be polewards at all longitudes, and for reasons similar to those articulated in section 14.1.1 we expect there to be a frictional western boundary current. We thus let $\Phi = \Phi_I + \Phi_b$, where Φ_I is the interior field and Φ_b the boundary layer correction. The interior field obeys $\partial\Phi_I/\partial x = -(f^2/\beta)S$, and is therefore given by, for constant S ,

$$\Phi_I(x, y) = \frac{f^2}{\beta} S_u (x_E - x) + \Phi_E \quad (15.94)$$

where Φ_E is the value of Φ_I at the eastern boundary, x_E . The value of Φ_E does not affect the solution and may be set to zero.

In the western boundary layer the dominant balance in (15.93) is

$$\beta \frac{\partial\Phi_b}{\partial x} = -r \frac{\partial^2\Phi_b}{\partial x^2}, \quad (15.95)$$

with solution

$$\Phi_b = P(y) \exp(-\beta x/r), \quad (15.96)$$

where $P(y)$ is to be determined by mass conservation: the southward mass flux at a latitude y in the western boundary current must be equal to the sum of the poleward mass flux in the interior at that latitude plus the total mass lost to upwelling equatorwards of y (Fig. 15.14). These fluxes are:

$$\text{upwelling flux} = \int_0^{x_E} \int_0^y S_u \, dx \, dy = S_u x_E y, \quad (15.97)$$

using the definition of S ;

$$\text{interior flux} = \int_0^{x_E} V \, dx = \frac{1}{f} \int_0^{x_E} \frac{\partial\Phi_I}{\partial x} \, dx = \frac{1}{f} [\Phi_E - \Phi_I(0, y)] = -\frac{f}{\beta} S x_E, \quad (15.98)$$

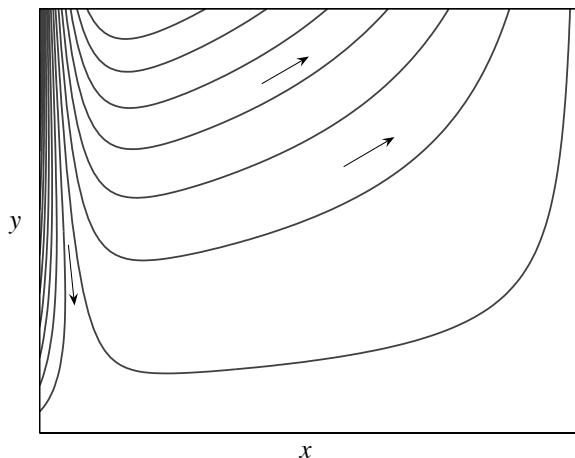


Fig. 15.17 The pressure field Φ for the shallow water Stommel–Arons model, as given by (15.102) with $r/\beta = 0.04x_E$, $f = \beta y$ and $y = 0$ at the equatorward edge of the domain. The arrows indicate the flow direction, with the western boundary current diminishing in intensity as it moves equatorwards. The convective mass source is, implicitly, just polewards of the domain. (Note that the pressure field is not exactly a streamline.)

using (15.94); and

$$\text{boundary flux} = \int_0^\infty \frac{1}{f} \frac{\partial \Phi_b}{\partial x} dx = P(y), \quad (15.99)$$

using (15.92a), neglecting the very small term $r\partial\Phi/\partial y$, and (15.96). By mass conservation we have

$$P(y) = S_u x_E y + \frac{f}{\beta} S_u x_E, \quad (15.100)$$

and if $f = \beta y$ and the domain goes from $y = 0$ to $y = y_N$ this becomes

$$P(y) = 2S_u x_E y = -2S_0 \frac{y}{y_N}, \quad (15.101)$$

where the second equality follows because the integrated upwelling balances the mass source. Consistent with our earlier heuristic treatment, the western boundary current:

- (i) is equatorwards, away from the convective mass source;
- (ii) diminishes in intensity as it moves equatorwards, as it feeds the interior;
- (iii) has a maximum mass flux of twice that of the convective source; that is, the flow recirculates.

Using (15.94) and (15.101) the complete (western boundary layer plus interior) solution is thus given by

$$\Phi = \Phi_I + \Phi_b = \frac{f^2}{\beta} S_u (x_E - x) + 2S_u x_E y e^{-\beta x/r}, \quad (15.102)$$

and this is plotted in Fig. 15.17. This solution does not, of course, include the flow in the neighbourhood of the convective source itself, nor does it satisfy no-normal flow conditions at the poleward edge of the domain.

15.7 SCALING FOR THE BUOYANCY-DRIVEN CIRCULATION

Thus far, we have taken the strength of the upwelling as a given. In reality, this is a consequence of the presence of diapycnal diffusion, because in its absence the flow is along

isopycnals, and not across isopycnals as in the flow we have envisioned in the sections above. In section 15.1 we estimated the upwelling for a non-rotating model; we now do the same for a fluid obeying the steady planetary-geostrophic equations, namely

$$\mathbf{v} \cdot \nabla b = \kappa \nabla^2 b, \quad \nabla \cdot \mathbf{u} + \frac{\partial w}{\partial z} = 0 \quad (15.103a,b)$$

$$\mathbf{f} \times \mathbf{u} = -\nabla \phi, \quad b = \frac{\partial \phi}{\partial z}. \quad (15.104a,b)$$

On the first line we have the thermodynamic equation and the mass continuity equation, and on the second line we have the momentum equations, that is geostrophic and hydrostatic balance, respectively. We use the continuously stratified equations rather than the corresponding shallow water equations because the former allow for a straightforward representation of diapycnal diffusion.

The momentum and mass continuity equations yield the linear geostrophic vorticity equation and the thermal wind equation, and associated scales, as follows:

$$\beta v = f \frac{\partial w}{\partial z} \rightarrow \beta U = \frac{f_0 W}{\delta}, \quad (15.105a)$$

$$f \frac{\partial \mathbf{u}}{\partial z} = \mathbf{k} \times \nabla b \rightarrow f \frac{U}{\delta} = \frac{\Delta b}{L}. \quad (15.105b)$$

We use uppercase to denote scaling variables, except that the vertical scale is denoted δ and the scaling for buoyancy variations is denoted Δb . We will take Δb as given, and equal to the buoyancy difference at the surface that is ultimately driving the motion. We also assume that the horizontal scales are isotropic, with $U = V$ and $X = Y = L$. The thermodynamic relation gives a relationship between W , κ and δ if we assume a broad upwelling region with a balance between upward advection and diffusion:

$$w \frac{\partial b}{\partial z} = \kappa \frac{\partial^2 b}{\partial z^2} \rightarrow W = \frac{\kappa}{\delta}. \quad (15.106)$$

Eliminating U from (15.105) gives

$$W = \frac{\delta^2 \beta \Delta b}{f^2 L}, \quad (15.107)$$

and using this with (15.106) gives scalings for the vertical velocity and δ :

$$W = \kappa^{2/3} \left(\frac{\beta \Delta b}{f^2 L} \right)^{1/3}, \quad \delta = \kappa^{1/3} \left(\frac{f^2 L}{\beta \Delta b} \right)^{1/3}.$$

(15.108a,b)

These scalings mean that, in so far as (15.104) describes the flow, the upwelling strength, and the circulation more generally, are dependent on a finite value of the diffusivity and scale as the two-thirds power of that diffusivity. The upwelling water is cold but the water at the surface is warm, and (15.108b) is a measure of the depth of the transition region — that is, the thickness of the thermocline, as illustrated in Fig. 15.2. This is an important topic that we return to in more detail in the next chapter, but we will make one important point now: In order for the scales given in (15.108) to be at all representative of those observed in the real ocean, we must use an eddy diffusivity for κ . Using $f = 10^{-4} \text{ s}^{-1}$, $\beta = 10^{-11} \text{ m}^{-1} \text{ s}^{-1}$,

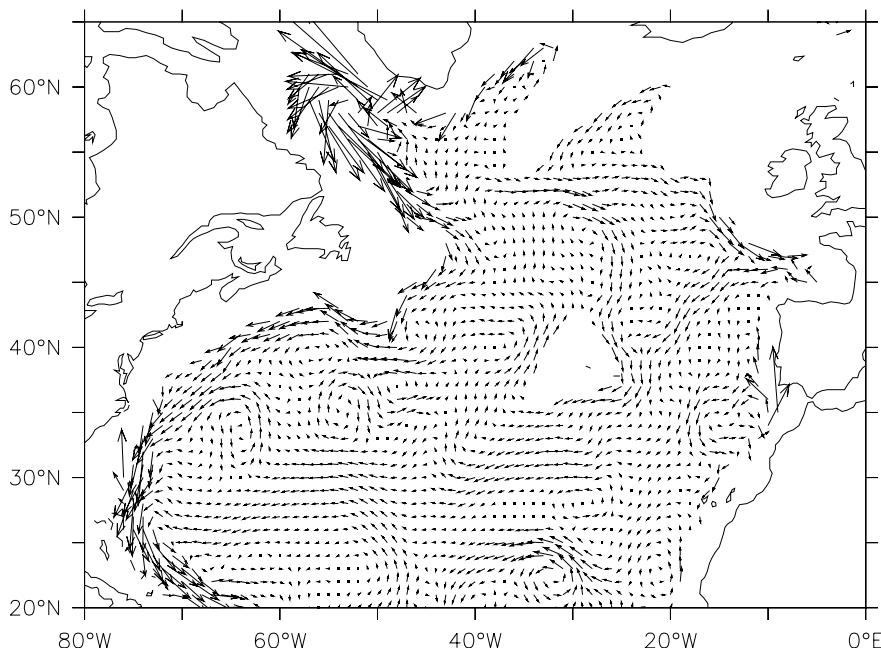


Fig. 15.18 The ocean currents at a depth of 2500 m in the North Atlantic, obtained using a combination of observations and model (as in Fig. 14.2). Note the southwards flowing *deep western boundary current*.

$L = 5 \times 10^6$ m, $g = 10$ m s $^{-2}$, $\kappa = 10^{-5}$ m 2 s $^{-1}$, $\Delta b = -g\Delta\rho/\rho_0 = g\beta_T\Delta T$ and $\Delta T = 10$ K we find the not unreasonable values of $\delta \approx 150$ m and $W = 10^{-7}$ m s $^{-1}$, albeit δ is rather smaller than the thickness of the observed thermocline. However, if we take the molecular value of $\kappa \approx 10^{-7}$ m 2 the values of W and δ are unrealistically small (although non-zero). Evidently, if the deep circulation of the ocean is buoyancy driven, it must take advantage of turbulence that enhances the small scale mixing and produces an eddy diffusivity.

15.7.1 Summary remarks on the Stommel–Arons model

If we were given the location and strength of the sources of deep water in the real ocean, the Stommel–Arons model could give us a global solution for the abyssal circulation. The solution for the Atlantic, for example, resembles a superposition of Fig. 15.15 and Fig. 15.16 (with deep water sources in the Weddell Sea and near Greenland), and that for the Pacific resembles Fig. 15.16 (with a deep water source emanating from the Antarctic Circumpolar Current). Perhaps the greatest success of the model is that it introduces the notions of deep western boundary currents and recirculation — enduring concepts of the deep circulation that remain with us today. For example, the North Atlantic ocean does have a well-defined deep western boundary current running south along the eastern seaboard of Canada and the United States, as seen in Fig. 15.18.¹⁷ However, in other important aspects the model is found to be in error, in particular it is found that there is little upwelling through the main thermocline — much of the water formed by deep convection in the North Atlantic in fact

upwells in the Southern Hemisphere.¹⁸ Are there fundamental problems with the model, or just discrepancies in details that might be corrected with a slight reformulation? To help answer that we summarize the assumptions and corresponding predictions of the model, and distinguish the essential aspects from what is merely convenient.

- (i) A foundational assumption is that of linear geostrophic vorticity balance in the ocean abyss, represented by $\beta v \approx f \partial w / \partial z$, or its shallow water analogue.
- The effects of mesoscale eddies are thereby neglected. As discussed in chapter 9, in their mature phase mesoscale eddies seek to barotropize the flow, and so create deep eddying motion that might dominate the deep flow.
- (ii) A second important assumption is that of uniform upwelling, across isopycnals, into the upper ocean, and that $w = 0$ at the ocean bottom. When combined with (i) this gives rise to a poleward interior flow, and by mass conservation a deep western boundary current. The upwelling is a consequence of a finite diffusion, which in turn leads to deep convection as in the model of sideways convection of section 15.1.
 - The uniform-upwelling assumption might be partially relaxed, while remaining in the Stommel-Arons framework, by supposing (for example) that the upwelling occurs near boundaries, or intermittently, with corresponding detailed changes to the interior flow.
 - If bottom topography is important, then $w \neq 0$ at the ocean bottom. This effect may be most important if mesoscale eddies are present, for then in an attempt to maintain its value of potential vorticity the abyssal flow will have a tendency to meander nearly inviscidly along contours of constant topography. In the presence of a mid-ocean ridge, some of the deep western boundary current might travel meridionally along the eastern edge of the ridge instead of along the coast.
 - In reality, the deep water might not upwell across isopycnals at all, but might move along isopycnals that intersect the surface (or are connected to the surface by convection). If so, then in the presence of mechanical forcing a deep circulation could be maintained even in the absence of a diapycnal diffusivity. The circulation might then be qualitatively different from the Stommel-Arons model, although a linear vorticity balance might still hold, with deep western boundary currents. This is discussed in section 16.5.

Note that, even if the Stommel-Arons picture were to be essentially correct, we should not consider the deep flow as being driven by deep convection at the source regions. It is a *convenience* to specify the strength of these regions for the calculations but, just as in the models of sideways convection considered in section 15.1, the overall strength of the circulation (insofar as it is buoyancy driven) is a function of the size of the diffusivity and the meridional buoyancy gradient at the surface.

Notes

- 1 From Wunsch (2002).
- 2 Figure kindly prepared by Neven Stjepan Fučkar, using the climatology of Conkright *et al.* (2001).
- 3 See, for example, Warren (1981) who provides a review and historical background and Schmitz (1995) who surveys the observations and provides an interpretation of the deep global circulation.
- 4 Adapted from Paparella & Young (2002).
- 5 As in Haney (1971). The value of C , which is not necessarily related to that of κ , is often taken to be such that the heat flux is of order $30 \text{ W m}^{-2} \text{ K}^{-1}$, but it is certainly not a universal constant.
- 6 Rossby (1965).
- 7 Adapted from Rossby (1998).
- 8 Adapted from Rossby (1998).
- 9 Ocean convection is reviewed by Marshall & Schott (1999).
- 10 Sandström (1908, 1916). Sandström's discussion was rather qualitative and generally thermodynamic in nature, with friction playing only an implicit role. Since then a number of related statements with varying degrees of generality and precision have been given (see e.g., Dutton 1986, Huang 1999, Paparella & Young 2002). Section 15.2.3 follows Paparella and Young.
- 11 The original box model is due to Stommel (1961), and many studies with variations around this have followed. Rooth (1982) developed the idea of a buoyancy-driven pole-to-pole overturning circulation, and Welander (1986) discussed, among other things, the role of boundary conditions on temperature and salinity at the ocean surface. Thual & McWilliams (1992) systematically explored how box models compare with two-dimensional fluid models of sideways convection, Quon & Ghil (1992) explored how multiple equilibria arise in related fluid models, and Dewar & Huang (1995) discussed the problem of flow in loops. Cessi & Young (1992) tried to derive simple models systematically from the equations of motion, obtaining various nonlinear amplitude equations. Our discussion is just a fraction of all this; see also Whitehead (1995), Cessi (2001) and Dijkstra (2002) for reviews and more.
- 12 Adapted from Welander (1986) and Dijkstra (2002).
- 13 Having said this, Bryan (1986), Manabe & Stouffer (1988) and Marotzke (1989) did find evidence of multiple equilibria in various three-dimensional numerical models, motivated in part by the solutions of box models.
- 14 After Stommel *et al.* (1958).
- 15 Following Stommel & Arons (1960).
- 16 Loosely following Cessi (2001).
- 17 A global Stommel-Arons-like solution was presented by Stommel (1958). The discovery of deep western boundary currents, by Swallow & Worthington (1961), was motivated by the theoretical model. Using neutrally-buoyant floats underneath the Gulf Stream they found a robust equatorwards-flowing undercurrent with typical speeds of $9\text{--}18 \text{ m s}^{-1}$. Some relevant observations of the deep circulation are summarized by Hogg (2001).
- 18 For example, Toggweiler & Samuels (1995).

Problems

- 15.1 (a) Obtain an expression analogous to (15.19) for the anelastic equations.
 (b) ♦ Obtain, if possible, an expression analogous to (15.19) for a compressible gas (which you may suppose to be ideal, if needed). Interpret the result.
- 15.2 (a) Consider an ocean with a flat surface, forced only by radiation. Suppose that the net radiation at the surface falls from about $+100 \text{ W m}^{-2}$ at the equator to -100 W m^{-2} at the pole, and that associated heating \dot{Q} varies in vertical with an e-folding scale of 10 metres with a volume integral that is zero. Is such a forcing sufficient to account for the observed dissipation of kinetic energy in the ocean, which is about 1 kW km^{-3} ?
 (b) Now let a wind blow over the ocean, providing an average stress of 0.05 N m^{-2} . Estimate the kinetic energy dissipation that this will lead to.
 (c) Estimate the rate of heating (in degrees per day) associated with a kinetic energy dissipation rate of 1 kW km^{-3} ?
- 15.3 ♦ Consider a model of sideways convection in which the boundary condition at the top is the relaxation, or Haney, condition $b_z = A(b^*(\gamma) - b)$ where A (a constant) and $b^*(\gamma)$ are given, b is the buoyancy, proportional to the temperature, the other boundaries are insulating, and the flow is statistically steady.
 (a) Is it still the case that on average the fluid is heated (i.e., there is a heat flux into the fluid through the upper surface) where it is already warm? If so, how may this be reconciled with the intuition that if the ocean surface is anomalously warm it will cool by way of a heat flux from the ocean to the atmosphere?
 (b) Show that if $A < 0$ a statistically steady state cannot be reached.
 (c) Suppose that b^* varies monotonically with latitude. Show that if κ is non-zero, the average surface temperature gradient will be less than that of b^* .
- 15.4 Define $\mathcal{R} \equiv Ra\alpha^5$ and expand
- $$\hat{\psi} = \mathcal{R}\hat{\psi}_1 + \mathcal{R}^2\hat{\psi}_2 + \mathcal{O}(\mathcal{R}^3), \quad \hat{b} = \hat{b}_0 + \mathcal{R}\hat{b}_1 + \mathcal{O}(\mathcal{R}^2). \quad (\text{P15.1})$$
- For small aspect ratio flow show that (15.9) gives at zeroth order
- $$\frac{\partial^2 \hat{b}_0}{\partial \hat{z}^2} = 0, \quad (\text{P15.2})$$
- giving a solution that satisfies the boundary conditions at both top and bottom. Show that at next order
- $$\frac{\partial^4 \hat{\psi}_1}{\partial \hat{z}^4} = -\frac{\partial f}{\partial \hat{y}}, \quad J(\hat{\psi}_1, f) = \frac{\partial^2 \hat{b}_1}{\partial \hat{z}^2}. \quad (\text{P15.3a,b})$$
- Show that the solutions are of the general form $\hat{\psi}_1(\hat{y}, \hat{z}) = (\partial f / \partial \hat{y})A(\hat{z})$ and $b_1(\hat{y}, \hat{z}) = (\partial f / \partial \hat{y})^2B(\hat{z})$, where A and B are functions only of \hat{z} . Thus deduce that the horizontal form of the solution is determined by the surface forcing, provided there are no meridional walls that force $\hat{\psi}$ to zero where $\partial f / \partial \hat{y} \neq 0$. (From notes by W. R. Young.)
- 15.5 ♦ Obtain the explicit solution to (P15.3) at lowest order, and at one higher order, and show that the sinking region is narrower than the upwelling region.
- 15.6 Write a code (e.g., in Matlab, Fortran, or C) that will timestep the Stommel two box model, (15.39), or the modified box model of problem 15.7. Integrate the model to equilibrium for a variety of initial conditions and parameters. Verify some of the solutions against the graphical/analytic solutions. In cases in which three solutions are present, vary the initial conditions to try to converge on all three.

- 15.7 Consider a variation of the Stommel box model in which the equation of motion (15.39) is replaced by

$$\frac{dT}{d\tau} = (1 - T) - \Phi^2 T, \quad \frac{dS}{d\tau} = \delta(1 - S) - \Phi^2 S, \quad \Phi = -\gamma(T - \mu S). \quad (\text{P15.4})$$

The interbox flow equations now depends on the density difference squared, so allowing differentiable equations. Show that multiple equilibria are possible, and that one is thermally driven and one salinity driven. Obtain approximate expressions for the flow (i.e., Φ), the temperature and salinity for these equilibria, in the limit of large γ and small δ if you wish. If δ is large, what changes about these solutions?

- 15.8 (a) ♦ In the Stommel two-box problem (or the modified box model of problem 15.7, which is easier) show physically that when three solutions are present, the middle one is generally unstable. One way to do this is to suppose that the system is perturbed slightly from that equilibrium, and argue qualitatively that the forces on the system will then take it farther from that state. By use of similar arguments, show that the other solutions are generally stable.
- (b) ♦ Alternatively, linearize the equations about the equilibrium points and show that small perturbations will grow if the solution is that of the middle equilibrium state, but will be damped in the other two cases.
- 15.9 ♦ Suppose that seawater were transparent to solar radiation, but nearly opaque to infrared radiation, with its other properties being the same as real seawater. Estimate the strength of the oceanic meridional overturning circulation, stating clearly any assumptions you make. You may neglect the effects of wind. Does the addition of wind make a difference to your solution?

No fairer destiny [has] any physical theory, than that it should of itself point out the way to the introduction of a more comprehensive theory, in which it lives on as a limiting case.

Albert Einstein, *Relativity, the Special and the General Theory*, 1916.

CHAPTER SIXTEEN

The Wind- and Buoyancy-Driven Ocean Circulation

N THIS CHAPTER we try to understand the combined effect of wind and buoyancy forcing in setting the three-dimensional structure of the ocean. There are three main topics we will consider.

- (i) The *main thermocline*, the region in the upper 1 km or so of the ocean where density and temperature change most rapidly.
- (ii) The ‘wind-driven’ overturning circulation. That is, we look into whether and how the ocean might maintain a deep overturning circulation that owes its existence to the effects of wind at the surface, as well as buoyancy forcing at the surface, and that persists even as the diapycnal diffusivity in the ocean interior goes to zero.
- (iii) The circulation of the flow in a channel, as a model of the Antarctic Circumpolar Current (ACC).

16.1 THE MAIN THERMOCLINE: AN INTRODUCTION

In the previous chapter we saw that a fluid that is differentially heated from above will develop both an overturning circulation and a region near the surface where the temperature changes rapidly. To examine this in more detail, we consider the circulation in a closed, single hemispheric basin, and again suppose that there is a net surface heating at low latitudes and a net cooling at high latitudes that maintains a meridional temperature gradient at the surface. We presume, *ab initio*, that there is a single overturning cell, with water rising at low latitudes before returning to polar regions, illustrated schematically in Fig. 16.1.

At lower latitudes the surface water is warmer than the cold water in the abyss. Thus there

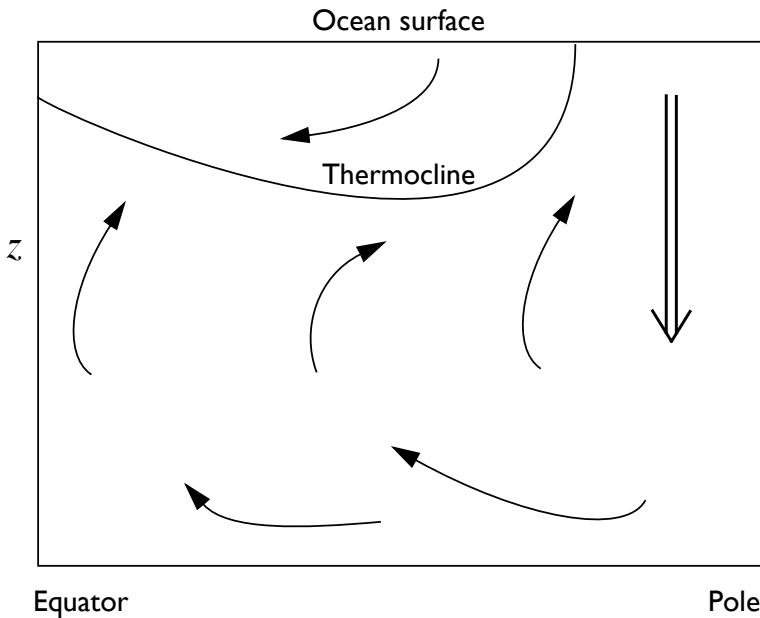


Fig. 16.1 Cartoon of a single-celled meridional overturning circulation, with a wall at the equator. Sinking is concentrated at high latitudes and upwelling spread out over lower latitudes. The thermocline is the boundary between the cold abyssal waters, with polar origins, and the warmer near-surface subtropical water. Wind forcing in the subtropical gyre mechanically pushes the warm water down, increasing the depth of the thermocline.

must be a vertical temperature gradient everywhere except possibly at the highest latitudes where the cold dense water sinks. This temperature gradient is called the *thermocline*, and is illustrated in Fig. 16.2. In purely buoyancy-driven flows the thickness of the thermocline is determined by way of an advective-diffusive balance, and proportional to some power of the thermal diffusivity as we considered in section 15.7. Let us revisit this issue, first by way of a simple kinematic model.

16.1.1 A simple kinematic model

The fact that cold water with polar origins upwells into a region of warmer water suggests that we consider the simple one-dimensional advective-diffusive balance,

$$w \frac{\partial T}{\partial z} = \kappa \frac{\partial^2 T}{\partial z^2}, \quad (16.1)$$

where w is the vertical velocity, κ is a diffusivity and T is temperature. In mid-latitudes, where this might hold, w is positive and the equation represents a balance between the upwelling of cold water and the downward diffusion of heat. If w and κ are given constants, and if T is specified at the top ($T = T_T$ at $z = 0$) and if $\partial T / \partial z = 0$ at great depth ($z = -\infty$)

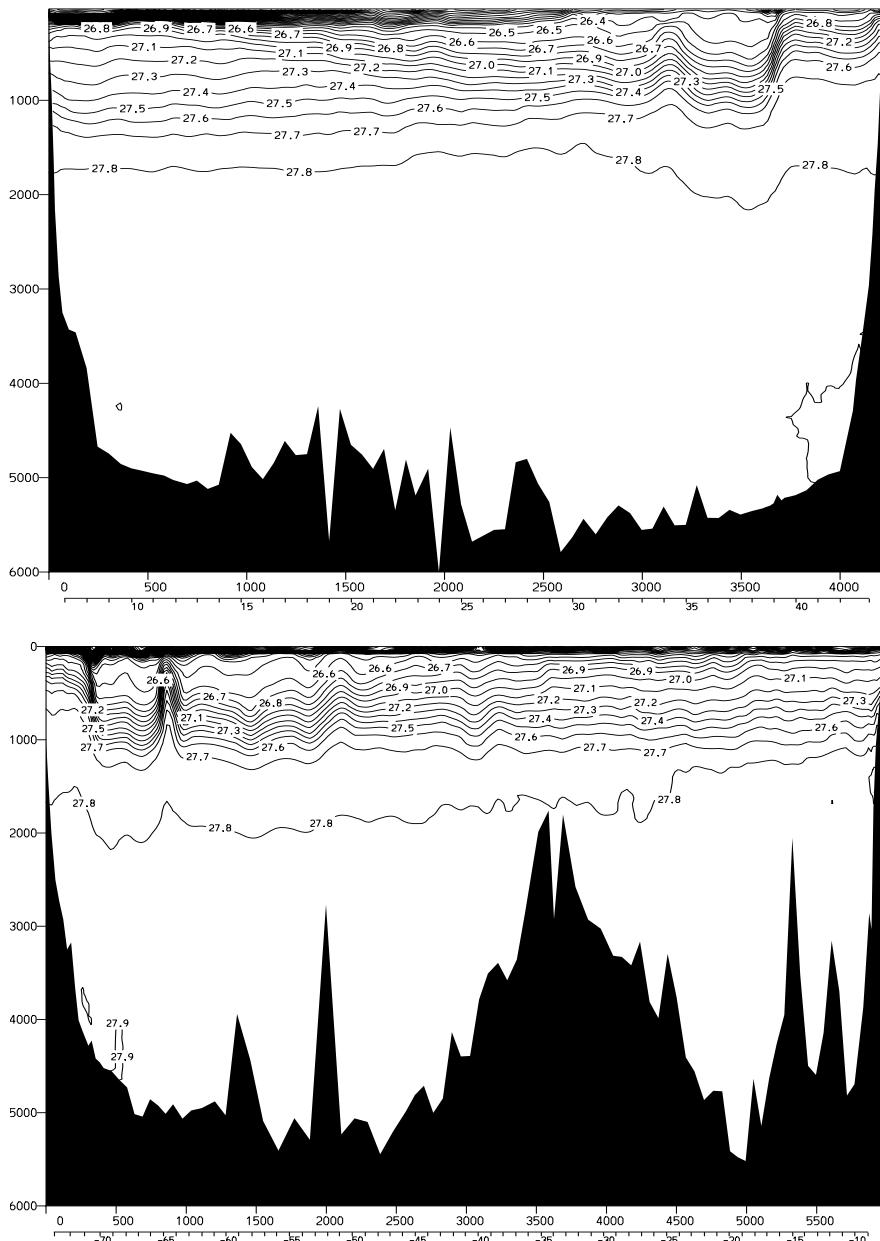


Fig. 16.2 Sections of potential density (σ_θ) in the North Atlantic. Upper panel: meridional section at 53°W , from 5°N to 45°N , across the subtropical gyre. Lower panel: zonal section at 36°N , from about 75°W to 10°W . A front is associated with the western boundary current and its departure from the coast near 40°N . In the upper northwestern region of the subtropical thermocline there is a region of low stratification known as MODE water: isopycnals above this outcrop in the subtropical gyre and are 'ventilated'; isopycnals below the MODE water outcrop in the subpolar gyre or ACC.¹

then the temperature falls exponentially away the surface according to

$$T = (T_T - T_B) e^{wz/\kappa} + T_B, \quad (16.2)$$

where T_B is a constant. This expression cannot be used to estimate how the thermocline depth scales with either w or κ , because the magnitude of the overturning circulation depends on κ (section 15.7). However, it is reasonable to see if the observed ocean is broadly consistent with this expression. The diffusivity κ can be measured; it is an eddy diffusivity, maintained by small-scale turbulence, and measurements produce values that range between $10^{-5} \text{ m}^2 \text{ s}^{-1}$ in the main thermocline and $10^{-4} \text{ m}^2 \text{ s}^{-1}$ in abyssal regions over rough topography and in and near continental margins, with still higher values locally.² The vertical velocity is too small to be measured directly, but various estimates based on deep water production suggest a value of about 10^{-7} m s^{-1} . Using this and the smaller value of κ in (16.2) gives an e-folding vertical scale, κ/w , of just 100 m, beneath which the stratification is predicted to be very small (i.e., nearly uniform potential density). Using the larger value of κ increases the vertical scale to 1000 m, which is probably closer to the observed value for the total thickness of the thermocline (look at Fig. 16.2), but using such a large value of κ in the main thermocline is not supported by the observations. Similarly, the deep stratification of the ocean is rather larger than that given by (16.1), except with values of diffusivity on the large side of those observed.³ Thus, there are two conclusions to be drawn.

- (i) The observed thickness of the thermocline is somewhat larger than what one might infer from observed values of the diffusivity and overturning circulation.
- (ii) The observed deep stratification is somewhat larger than what one might infer from the advective-diffusive balance (16.1) with observed values of diffusivity and overturning circulation.

Of course the model itself, (16.1), is overly simple but these conclusions suggest that additional physical factors may play a role in thermocline dynamics. Mechanical forcing, and in particular the wind, is one such: the wind-stress curl forces water to converge in the subtropical Ekman layer, thereby forcing relatively warm water to downwell and meet the upwelling colder abyssal water at some finite depth, thus deepening the thermocline from its purely diffusive value. Indeed, in so far as we can separate the two effects of wind and diffusion, we can say that the strength of the wind influences the *depth* at which the thermocline occurs, whereas the strength of the diffusivity influences the *thickness* of the thermocline. The influence of the wind on the abyssal circulation is not quite as direct, but we will find in section 16.5 that it will enable both a circulation and deep stratification to persist even in the absence of diffusion.

16.2 SCALING AND SIMPLE DYNAMICS OF THE MAIN THERMOCLINE

We now begin to consider the dynamics that produce an overturning circulation and a thermocline. The Rossby number of the large-scale circulation is small and the scale of the motion large, and the flow obeys the planetary-geostrophic equations:

$$\mathbf{f} \times \mathbf{u} = -\nabla \phi, \quad \frac{\partial \phi}{\partial z} = b, \quad (16.3a,b)$$

$$\nabla \cdot \mathbf{v} = 0, \quad \frac{Db}{Dt} = \kappa \frac{\partial^2 b}{\partial z^2}. \quad (16.4a,b)$$

We suppose that these equations hold below an Ekman layer, so that the effects of a wind stress may be included by specifying a vertical velocity at the top of the domain. The diffusivity, κ , is, as we noted above, an eddy diffusivity, but since its precise form and magnitude are uncertain we must proceed with due caution, and a useful practical philosophy is to try to ignore dissipation and viscosity where possible, and to invoke them only if there is no other way out. Let us therefore scale the equations in two ways, with and without diffusion; these scalings will be central to our theory.

16.2.1 An advective scale

As usual we denote (with one or two exceptions) scaling values with capital letters and non-dimensional values with a hat, so that, for example, $u = U\hat{u}$ and $u = \mathcal{O}(U)$. Let us ignore the diffusive term in (16.4b) and try to construct a scaling estimate for the depth of the wind's influence.

If there is upwelling ($w > 0$) from the abyss, and Ekman downwelling ($w < 0$) at the surface, there is some depth D_a at which $w = 0$. By cross-differentiating (16.3a) we obtain $\beta v = -f\nabla_z \cdot \mathbf{u}$, and combining this with (16.4a) gives the familiar geostrophic vorticity equation and corresponding scaling

$$\beta v = f \frac{\partial w}{\partial z} \quad \rightarrow \quad \beta V = f \frac{W}{D_a}. \quad (16.5)$$

Here, D_a is the unknown depth scale of the motion, L is the horizontal scale of the motion, which we take as the gyre or basin scale, and V is a horizontal velocity scale. (It is reasonable to suppose that $V \sim U$, where U is the zonal velocity scale, and henceforth we will denote both by U .) The appropriate vertical velocity to use is that due to Ekman pumping, W_E ; we will assume (and demonstrate later) that this is much larger than the abyssal upwelling velocity, which in any case is zero by assumption at $z = -D_a$. With this, (16.5) yields the Sverdrup-balance estimate

$$U = \frac{f}{\beta} \frac{W_E}{D_a}. \quad (16.6)$$

We may determine an appropriate value of U using the thermal wind relation, which from (16.3) is

$$\mathbf{f} \times \frac{\partial \mathbf{u}}{\partial z} = -\nabla b \quad \rightarrow \quad \frac{U}{D_a} = \frac{1}{f} \frac{\Delta b}{L}, \quad (16.7)$$

where Δb is the scaling value of variations of buoyancy in the horizontal. Assuming the vertical scales are the same in (16.6) and (16.7) then eliminating U gives

$$D_a = W_E^{1/2} \left(\frac{f^2 L}{\beta \Delta b} \right)^{1/2}.$$

(16.8)

[This is essentially the same as the estimate (14.130).] If we relate U and W_E using mass conservation, $U/L = W_E/D_a$, instead of using Sverdrup balance, then we write L in place of f/β and (16.8) becomes $D_a = (W_E f L^2 / \Delta b)^{1/2}$, which is not qualitatively different for large scales. The important aspect of these equations is that the depth of the wind-influenced region increases with the magnitude of the wind stress (because $W_E \propto \text{curl}_z \tau$) and decreases

with the meridional temperature gradient. The former dependence is reasonably intuitive, and the latter arises because as the temperature gradient increases the associated thermal wind-shear U/D_a correspondingly increases. But the horizontal transport (the product UD_a) is fixed by mass conservation; the only way that these two can remain consistent is for the vertical scale to decrease. Taking $W_E = 10^{-6} \text{ m s}^{-1}$, $\Delta b = g\Delta\rho/\rho_0 = g\beta_T\Delta T \sim 10^{-2} \text{ m s}^{-2}$, $L = 5000 \text{ km}$ and $f = 10^{-4} \text{ s}^{-1}$ gives $D_a = 500 \text{ m}$. Such a scaling argument cannot be expected to give more than an estimate of the depth of the wind-influenced region; nevertheless, because D_a is much less than the ocean depth, the estimate does suggest that the wind-driven circulation is predominantly an upper-ocean phenomenon.

16.2.2 A diffusive scale

The estimate (16.8) cares nothing about the thermodynamic equation; if we do take into account thermodynamics, with non-zero diffusivity, we recover the model of section 15.7. Thus, briefly, the scaling follows from advective-diffusive balance in the thermodynamic equation, the linear geostrophic vorticity equation, and thermal wind balance:

$$w \frac{\partial b}{\partial z} = \kappa \frac{\partial^2 b}{\partial z^2}, \quad \beta v = f \frac{\partial w}{\partial z}, \quad f \frac{\partial \mathbf{u}}{\partial z} = \mathbf{k} \times \nabla b, \quad (16.9a,b,c)$$

with corresponding scales

$$\frac{W}{\delta} = \frac{\kappa}{\delta^2}, \quad \beta U = \frac{fW}{\delta}, \quad \frac{U}{\delta} = \frac{\Delta b}{fL}, \quad (16.10a,b,c)$$

where δ is the vertical scale. Because there is now one more equation than in the advective scaling theory we cannot take the vertical velocity as a given, otherwise the equations would be overdetermined. We therefore take it to be the abyssal upwelling velocity, which then becomes part of the *solution*, rather than being imposed. From (16.10) we obtain the diffusive vertical scale,

$$\delta = \left(\frac{\kappa f^2 L}{\beta \Delta b} \right)^{1/3}.$$

(16.11)

With $\kappa = 10^{-5} \text{ m}^2 \text{ s}^{-2}$ and with the other parameters taking the values given following (16.8), (16.11) gives $\delta \approx 150 \text{ m}$ and, using (16.10a), $W \approx 10^{-7} \text{ m s}^{-1}$, which is an order of magnitude smaller than the Ekman pumping velocity W_E .

A wind-influenced diffusive scaling

The scaling above assumes that the length scale over which thermal wind balance holds is the gyre scale itself. In fact, there is another length scale that is more appropriate, and this leads to a slightly different scaling for the thickness of the thermocline. To obtain this scaling, we first note that the depth of the subtropical thermocline is not constant: it shoals up to the east because of Sverdrup balance, and it may shoal up polewards as the curl of the wind stress falls (and is zero at the poleward edge of the gyre). Thus, referring to Fig. 16.3, the appropriate horizontal length scale \tilde{L} is given by

$$\tilde{L} = \delta \frac{L}{D_a}. \quad (16.12)$$

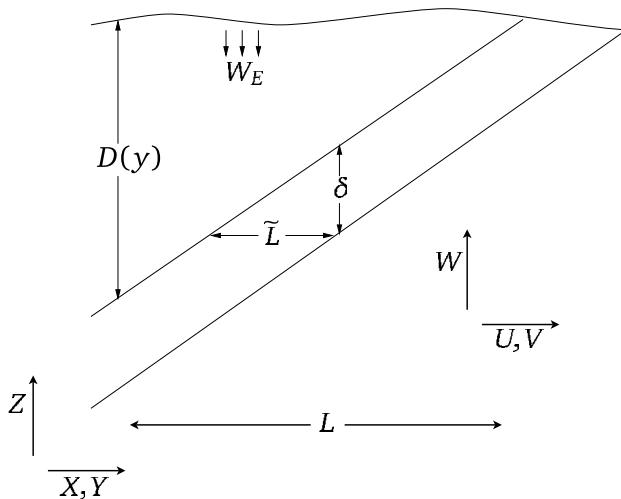


Fig. 16.3 Scaling the thermocline. The diagonal lines mark the diffusive thermocline of thickness δ and depth $D(y)$. The advective scaling for $D(y)$, i.e., D_a , is given by (16.8), and the diffusive scaling for δ is given by (16.13).

This is no longer an externally imposed parameter, but must be determined as part of the solution. Using \tilde{L} instead of L as the length scale in the thermal wind equation (16.10c) gives, using (16.8), the modified diffusive scale

$$\delta = \kappa^{1/2} \left(\frac{f^2 \tilde{L}}{\Delta b \beta D_a} \right)^{1/2} = \kappa^{1/2} \left(\frac{f^2 L}{\Delta b \beta W_E} \right)^{1/4}. \quad (16.13)$$

Substituting values of the various parameters results in a thickness of about 100–200 m. The thermocline thickness now scales as $\kappa^{1/2}$. The interpretation of this scale and that of (16.11) is that the thickness of the thermocline scales as $\kappa^{1/3}$ in the absence of a wind stress, but scales as $\kappa^{1/2}$ if a wind stress is present that can provide a finite slope to the base of the thermocline that is independent of κ , and this is confirmed by numerical simulations.⁴ From (16.9a) the vertical velocity, and hence the meridional overturning circulation, no longer scales as $\kappa^{2/3}$ but as

$$W = \frac{\kappa}{\delta} \propto \kappa^{1/2}. \quad (16.14)$$

16.2.3 Summary of the physical picture

What do the vertical scales derived above represent? The wind-influenced scaling, D_a , is the depth to which the directly wind-driven circulation can be expected to penetrate. Thus, over this depth we can expect to see wind-driven gyres and associated phenomena. At greater depths lies the abyssal circulation, and this is not wind-driven in the same sense. Now, in general, the water at the base of the wind-driven layer will not have the same thermodynamic properties as the upwelling abyssal water — this being cold and dense, whereas the water in the wind-driven layer is warm and subtropical (look again at Fig. 16.1). The thickness δ characterizes the diffusive transition region between these two water masses and in the limit of very small diffusivity this becomes a *front*. One might say that D_a is the *depth* of the thermocline, while δ is the *thickness* of the thermocline. In the diffusive region, no

matter how small the diffusivity κ is in the thermodynamic equation, the diffusive term is important. Of course if the diffusion is sufficiently large, the thickness will be as large or larger than the depth, and the two regions will blur into each other, and this may indeed be the case in the real ocean. (The real world is also complicated by the effects of mesoscale eddies.) Nevertheless, these scales are a useful foundation on which to build.⁵

16.3 THE INTERNAL THERMOCLINE

We now try to go beyond simple scaling arguments and investigate in more detail the dynamics of the thermocline. In this section we consider the diffusive, or internal, thermocline and in section 16.4 we consider the advective, or ventilated, thermocline. Such an investigation requires trying to actually solve the equations of motion, but the advective term in the thermodynamic equation makes this extremely difficult; indeed this prevents us from constructing wholly analytic models, but not from constructing informative models. We begin by expressing the planetary-geostrophic equations as an equation in a single unknown.

16.3.1 The M equation

The planetary-geostrophic equations can be written as a single partial differential equation in a single variable, although the resulting equation is of quite high order and is nonlinear. We write the equations of motion as

$$-fv = -\frac{\partial \phi}{\partial x}, \quad fu = -\frac{\partial \phi}{\partial y}, \quad b = \frac{\partial \phi}{\partial z}, \quad (16.15a,b,c)$$

$$\nabla \cdot \mathbf{v} = 0, \quad \frac{\partial b}{\partial t} + \mathbf{v} \cdot \nabla b = \kappa \nabla^2 b, \quad (16.16a,b)$$

where we take $f = \beta y$. Cross-differentiating the horizontal momentum equations and using (16.16a) gives the linear geostrophic vorticity relation $\beta v = f \partial w / \partial z$ which, using (16.15a) again, may be written as

$$\frac{\partial \phi}{\partial x} + \frac{\partial}{\partial z} \left(-\frac{f^2}{\beta} w \right) = 0. \quad (16.17)$$

This equation is the divergence in (x, z) of $(\phi, -f^2 w / \beta)$ and is automatically satisfied if

$$\phi = M_z \quad \text{and} \quad \frac{f^2 w}{\beta} = M_x. \quad (16.18a,b)$$

where the subscripts on M denote derivatives. Then straightforwardly

$$u = -\frac{\partial_y \phi}{f} = -\frac{M_{zy}}{f}, \quad v = \frac{\partial_x \phi}{f} = \frac{M_{zx}}{f}, \quad b = \partial_z \phi = M_{zz}. \quad (16.19a,b,c)$$

The thermodynamic equation, (16.16b) becomes

$$\frac{\partial M_{zz}}{\partial t} + \left(\frac{-M_{zy}}{f} M_{zzx} + \frac{M_{zx}}{f} M_{zzy} \right) + \frac{\beta}{f^2} M_x M_{zzz} = \kappa M_{zzzz} \quad (16.20)$$

or

$$\frac{\partial M_{zz}}{\partial t} + \frac{1}{f} J(M_z, M_{zz}) + \frac{\beta}{f^2} M_x M_{zzz} = \kappa M_{zzzz}. \quad (16.21)$$

where J is the usual horizontal Jacobian. This is the M *equation*,⁶ somewhat analogous to the potential vorticity equation in quasi-geostrophic theory in that it expresses the entire dynamics of the system in a single, nonlinear, advective-diffusive partial differential equation, although note that M_{zz} is materially conserved (in the absence of diabatic effects) by the three-dimensional flow. Because of the high differential order and nonlinearity of the system analytic solutions of (16.21) are very hard to find, and from a numerical perspective it is easier to integrate the equations in the form (16.15) and (16.16) than in the form (16.21). Nevertheless, it is possible to move forward by approximating the equation to one or two dimensions, or by a priori assuming a boundary-layer structure.

A one-dimensional model

Let us consider an illustrative one-dimensional model (in z) of the thermocline.⁷ Merely setting all horizontal derivatives in (16.21) to zero is not very useful, for then all the advective terms on the left-hand side vanish. Rather, we look for steady solutions of the form $M = M(x, z)$, and the M equation then becomes

$$\frac{\beta}{f^2} M_x M_{zzz} = \kappa M_{zzzz}, \quad (16.22)$$

which represents the advective-diffusive balance

$$w \frac{\partial b}{\partial z} = \kappa \frac{\partial^2 b}{\partial z^2}. \quad (16.23)$$

(We must also suppose that the value of κ varies meridionally in the same manner as does β/f^2 ; without this technicality M would be a function of y , violating our premise.) If the ocean surface is warm and the abyss is cold, then (16.22) represents a balance between the upward advection of cold water and the downward diffusion of warm water. The horizontal advection terms vanish because the zonal velocity, u , and the meridional buoyancy gradient, b_y , are each zero. Let us further consider the special case

$$M = (x - x_e)W(z), \quad (16.24)$$

where the domain extends from $0 \leq x \leq x_e$, so satisfying $M = 0$ on the eastern boundary. Equation (16.22) becomes the ordinary differential equation

$$\frac{\beta}{f^2} WW_{zzz} = \kappa W_{zzzz}, \quad (16.25)$$

where W has the dimensions of velocity squared. We non-dimensionalize this by setting

$$z = H\hat{z}, \quad \kappa = \hat{\kappa}(HW_S), \quad W = \left(\frac{f^2 W_S}{\beta} \right) \hat{W}, \quad (16.26a,b,c)$$

where the hatted variables are non-dimensional and W_S is a scaling value of the dimensional vertical velocity, w (e.g., the magnitude of the Ekman pumping velocity W_E). Equation (16.25) becomes

$$\hat{W}\hat{W}_{\hat{z}\hat{z}\hat{z}} = \hat{\kappa}\hat{W}_{\hat{z}\hat{z}\hat{z}\hat{z}}, \quad (16.27)$$

The parameter $\hat{\kappa}$ is a non-dimensional measure of the strength of diffusion in the interior, and the interesting case occurs when $\hat{\kappa} \ll 1$; in the ocean, typical values are $H = 1$ km,

$\kappa = 10^{-5} \text{ m s}^{-2}$ and $W_S = W_E = 10^{-6} \text{ m s}^{-1}$ so that $\hat{\kappa} \approx 10^{-2}$, which is indeed small. (It might appear that we could completely scale away the value of κ in (16.25) by scaling W appropriately, and if so there would be no meaningful way that one could say that κ was small. However, this is a chimera, because the value of κ would still appear in the boundary conditions. See also problem 16.5.)

The time-dependent form of (16.27), namely $\hat{W}_{\hat{z}\hat{z}t} + \hat{W}\hat{W}_{\hat{z}\hat{z}\hat{z}} = \hat{\kappa}\hat{W}_{\hat{z}\hat{z}\hat{z}\hat{z}}$, is similar to Burger's equation, $V_t + VV_z = \nu V_{zz}$, which is known to develop fronts. (In the inviscid Burger's equation, $DV/Dt = 0$, where the advective derivative is one-dimensional, and therefore the velocity of a given fluid parcel is preserved on the line. Suppose that the velocity of the fluid is positive but diminishes in the positive z -direction, so that a fluid parcel will catch-up with the fluid parcel in front of it. But since the velocity of a fluid parcel is fixed, there are two values of velocity at the same point, so a singularity must form. In the presence of viscosity, the singularity is tamed to a front.) Thus, we might similarly expect (16.27) to produce a front, but because of the extra derivatives the argument is not as straightforward and it is simplest to obtain solutions numerically.

Equation (16.27) is fourth order, so four boundary conditions are needed, two at each boundary. Appropriate ones are a prescribed buoyancy and a prescribed vertical velocity at each boundary, for example

$$\begin{aligned} \hat{W} &= \hat{W}_E, & -\hat{W}_{\hat{z}\hat{z}} &= B_0, & \text{at top} \\ \hat{W} &= 0, & -\hat{W}_{\hat{z}\hat{z}} &= 0, & \text{at bottom,} \end{aligned} \quad (16.28)$$

where \hat{W}_E is the (non-dimensional) vertical velocity at the base of the top Ekman layer, which is negative for Ekman pumping in the subtropical gyre, and B_0 is a constant, proportional to the buoyancy difference across the domain. We obtain solutions numerically by Newton's method,⁸ and these are shown in Figs. 16.4 and 16.5. The solutions do indeed display fronts, or boundary layers, for small diffusivity. If the wind forcing is zero (Fig. 16.5), the boundary layer is at the top of the fluid. If the wind forcing is non-zero, an internal boundary layer — a front — forms in the fluid interior with an adiabatic layer above and below. In the real ocean, where wind forcing is of course non-zero, the frontal region is known as the *internal thermocline*.

16.3.2 * Boundary-layer analysis

The reasoning and the numerical solutions of the above sections suggest that the internal thermocline has a boundary-layer structure whose thickness decreases with κ . If the Ekman pumping at the top of the ocean is non-zero, the boundary layer is internal to the fluid. To learn more, let us perform a boundary layer analysis, much as we did when investigating western boundary currents in section 14.1.3. The nonlinearity precludes a complete solution of the equation, but we can nevertheless obtain some useful information.

One-dimensional model

Let us now *assume* a steady two-layer structure of the form illustrated in Fig. 16.6, and that the dynamics are governed by (16.27) in a domain that extends from 0 to -1 . The buoyancy thus varies rapidly only in an internal boundary layer of non-dimensional thickness $\hat{\delta}$ located at $\hat{z} = -h$; above and below this the buoyancy is assumed to be only very slowly varying.

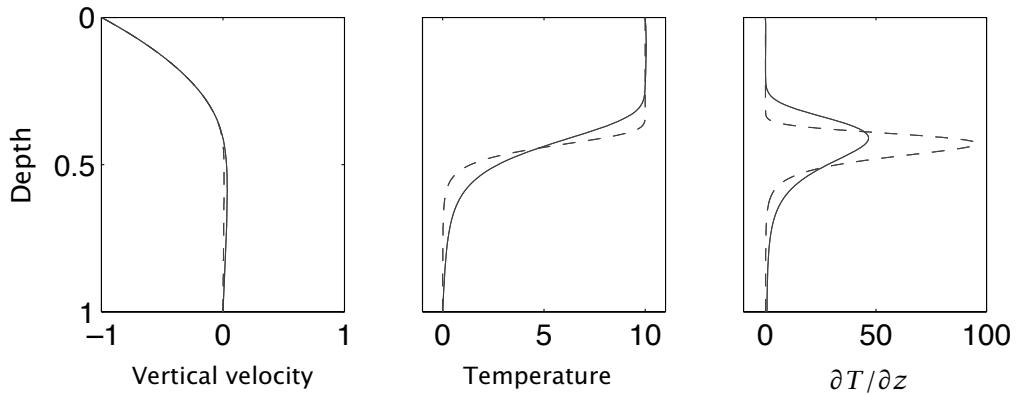


Fig. 16.4 Solution of the one-dimensional thermocline equation, (16.27), with boundary conditions (16.28), for two different values of the diffusivity: $\hat{\kappa} = 3.2 \times 10^{-3}$ (solid line) and $\hat{\kappa} = 0.4 \times 10^{-3}$ (dashed line), in the domain $0 \leq \hat{z} \leq -1$. ‘Vertical velocity’ is W , ‘temperature’ is $-W_{zz}$, and all units are the non-dimensional ones of the equation itself. A negative vertical velocity, $\hat{W}_E = -1$, is imposed at the surface (representing Ekman pumping) and $B_0 = 10$. The internal boundary layer thickness increases as $\hat{\kappa}^{1/3}$, so doubling in thickness for an eightfold increase in $\hat{\kappa}$. The upwelling velocity also increases with $\hat{\kappa}$ (as $\hat{\kappa}^{2/3}$), but this is barely noticeable on the graph because the downwelling velocity, above the internal boundary layer, is much larger and almost independent of $\hat{\kappa}$. The depth of the boundary layer increases as $\hat{W}_E^{1/2}$, so if $\hat{W}_E = 0$ the boundary layer is at the surface, as in Fig. 16.5.

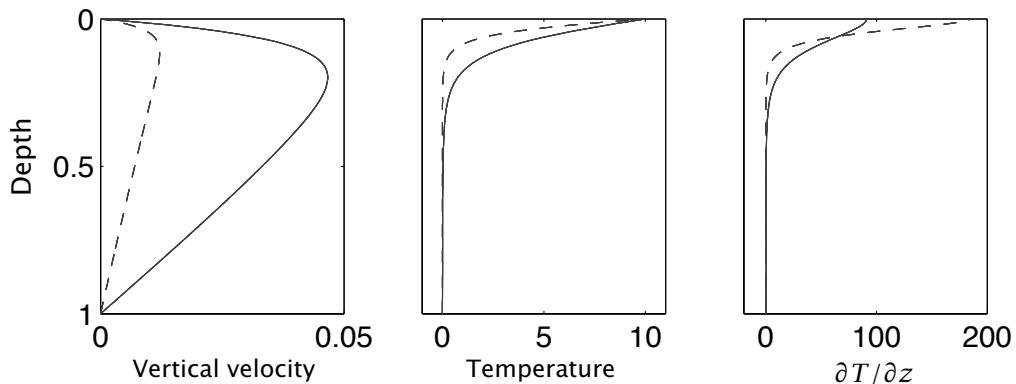


Fig. 16.5 As for Fig. 16.4, but with no imposed Ekman pumping velocity at the upper boundary ($\hat{W}_E = 0$), again for two different values of the diffusivity: $\hat{\kappa} = 3.2 \times 10^{-3}$ (solid line) and $\hat{\kappa} = 0.4 \times 10^{-3}$ (dashed line). The boundary layer now forms at the upper surface. The boundary thickness again increases with diffusivity and, even more noticeably, so does the upwelling velocity — this scales as $\hat{\kappa}^{2/3}$, and so increases fourfold for an eightfold increase in $\hat{\kappa}$.

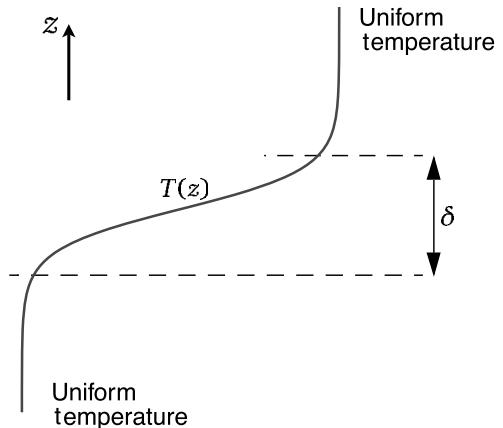


Fig. 16.6 The simplified boundary-layer structure of the internal thermocline. In the limit of small diffusivity the internal thermocline forms a boundary layer, of thickness δ in the figure, in which the temperature and buoyancy change rapidly.

Following standard boundary layer procedure, we introduce a stretched boundary layer coordinate ζ , where

$$\hat{\delta}\zeta = \hat{z} + h. \quad (16.29)$$

That is, ζ is the distance from $\hat{z} = -h$, scaled by the boundary layer thickness $\hat{\delta}$, and within the boundary layer ζ is an order-one quantity. We also let

$$\hat{W}(\hat{z}) = \hat{W}_I(\hat{z}) + \widetilde{W}(\zeta), \quad (16.30)$$

where \hat{W}_I is the solution away from the boundary layer and \widetilde{W} is the boundary layer correction. Because the boundary layer is presumptively thin, \hat{W}_I is effectively constant through it and, furthermore, for $\hat{z} < -h$, \hat{W} vanishes in the limit as $\kappa = 0$. We thus take $\hat{W}_I = 0$ throughout the boundary layer. (The small diffusively-driven upwelling below the boundary layer is part of the boundary layer solution, not the interior solution.) Now, buoyancy varies rapidly in the boundary layer but it remains an order-one quantity throughout. To satisfy this we explicitly scale \widetilde{W} in the boundary layer by writing

$$\widetilde{W}(\zeta) = \hat{\delta}^2 B_0 A(\zeta), \quad (16.31)$$

where B_0 is defined by (16.28) and A is an order-one field. The derivatives of W are

$$\frac{\partial \hat{W}}{\partial \hat{z}} = \frac{1}{\hat{\delta}} \frac{\partial \widetilde{W}}{\partial \zeta} = \hat{\delta} B_0 \frac{\partial A}{\partial \zeta}, \quad \frac{\partial^2 \hat{W}}{\partial \hat{z}^2} = B_0 \frac{\partial^2 A}{\partial \zeta^2}, \quad (16.32)$$

so that $\hat{W}_{\hat{z}\hat{z}}$ is an order-one quantity. Far from the boundary layer the solution must be able to match the external conditions on temperature and velocity, (16.28); the buoyancy condition on $W_{\hat{z}\hat{z}}$ is satisfied if

$$A_{\zeta\zeta} \rightarrow \begin{cases} 1 & \text{as } \zeta \rightarrow +\infty \\ 0 & \text{as } \zeta \rightarrow -\infty. \end{cases} \quad (16.33)$$

On vertical velocity we require that $W \rightarrow (\hat{z}/h + 1)W_E$ as $\zeta \rightarrow +\infty$, and $W \rightarrow \text{constant}$ as $\zeta \rightarrow -\infty$. The first matches the Ekman pumping velocity above the boundary layer, and

the second condition produces the abyssal upwelling velocity, which as noted vanishes for $\kappa \rightarrow 0$.

Substituting (16.30) and (16.31) into (16.27) we obtain

$$B_0 A A_{\zeta\zeta\zeta} = \frac{\hat{\kappa}}{\hat{\delta}^3} A_{\zeta\zeta\zeta\zeta}. \quad (16.34)$$

Because all quantities are presumptively $\mathcal{O}(1)$, (16.34) implies that $\hat{\delta} \sim (\hat{\kappa}/B_0)^{1/3}$. We restore the dimensions of δ by using $\kappa = \hat{\kappa}(HW_S)$ and $\Delta b = B_0 L f^2 W_S / (\beta H^2)$, where Δb is the dimensional buoyancy difference across the boundary layer [note that $b = M_{zz} = (x-1)W_{zz} \sim LW_{zz} \sim LB_0 f^2 W_S / (\beta H^2)$ using (16.26)]. The dimensional boundary layer thickness, δ , is then given by

$$\delta \sim \left(\frac{\kappa f^2 L}{\Delta b \beta} \right)^{1/3}, \quad (16.35)$$

which is the same as the heuristic estimate (16.11). The dimensional vertical velocity scales as

$$W \sim \frac{\kappa}{\delta} \sim \kappa^{2/3} \left(\frac{\Delta b \beta}{\kappa k^2 L} \right)^{1/3}, \quad (16.36)$$

this being an estimate of strength of the upwelling velocity at the base of the thermocline and, more generally, the strength of the diffusively-driven component of meridional overturning circulation of the ocean.

Somewhat different one-dimensional thermocline models may be constructed (see problem 16.3), and these have slightly different scaling properties. However, their qualitative features transcend their detailed construction, and in particular:

- ★ the thickness of the internal thermocline increases with increasing diffusivity, and decreases with increasing buoyancy difference across it, and as the diffusivity tends to zero the thickness of the internal thermocline tends to zero.
- ★ the strength of the upwelling velocity, and hence the strength of the meridional overturning circulation, increases with increasing diffusivity and increasing buoyancy difference.

* The three-dimensional equations

We now apply boundary layer techniques to the three-dimensional M equation.⁹ The main difference is that the depth of the boundary layer is now a function of x and y , so that the stretched coordinate ζ is given by

$$\hat{\delta}\zeta = z + h(x, y). \quad (16.37)$$

[The coordinates (x, y, z) in this subsection are non-dimensional, but we omit their hats to avoid too cluttered a notation.] Just as in the one-dimensional case we rescale M in the boundary layer and write

$$M = B_0 \hat{\delta}^2 \hat{A}(x, y, \zeta), \quad (16.38)$$

where the scaling factor $\hat{\delta}^2$ again ensures that the temperature remains an order-one quantity. In the boundary layer the derivatives of M become

$$\frac{\partial M}{\partial z} = \frac{1}{\hat{\delta}} \frac{\partial A}{\partial \zeta}, \quad (16.39)$$

and

$$\frac{\partial M}{\partial x} = \hat{\delta}^2 B_0 \left(\frac{\partial A}{\partial \zeta} \frac{\partial \zeta}{\partial x} + \frac{\partial A}{\partial x} \right) = \hat{\delta}^2 B_0 \left(\frac{\partial A}{\partial \zeta} \frac{1}{\hat{\delta}} \frac{\partial h}{\partial x} + \frac{\partial A}{\partial x} \right). \quad (16.40)$$

Substituting these into (16.20) we obtain, omitting the time-derivative,

$$\begin{aligned} \hat{\delta} \left[\frac{1}{f} \left(A_{\zeta x} A_{\zeta \zeta y} - A_{\zeta y} A_{\zeta \zeta x} \right) + \frac{\beta}{f^2} A_x A_{\zeta \zeta \zeta} \right] + \frac{\beta}{f^2} h_x A_{\zeta} A_{\zeta \zeta \zeta} \\ + \frac{1}{f} \left[h_x \left(A_{\zeta \zeta} A_{\zeta \zeta y} - A_{\zeta y} A_{\zeta \zeta \zeta} \right) + h_y \left(A_{\zeta x} A_{\zeta \zeta \zeta} - A_{\zeta \zeta} A_{\zeta \zeta x} \right) \right] \\ = \frac{\kappa}{B_0 \hat{\delta}^2} A_{\zeta \zeta \zeta \zeta}, \end{aligned} \quad (16.41)$$

where the subscripts on A and h denote derivatives. If $h_x = h_y = 0$, that is if the base of the thermocline is flat, then (16.41) becomes

$$\frac{1}{f} \left[A_{\zeta x} A_{\zeta \zeta y} - A_{\zeta y} A_{\zeta \zeta x} \right] + \frac{\beta}{f^2} A_x A_{\zeta \zeta \zeta} = \frac{\kappa}{B_0 \hat{\delta}^3} A_{\zeta \zeta \zeta \zeta}. \quad (16.42)$$

Since all the terms in this equation are, by construction, order one, we immediately see that the non-dimensional boundary layer thickness $\hat{\delta}$ scales as

$$\hat{\delta} \sim \left(\frac{\kappa}{B_0} \right)^{1/3}, \quad (16.43)$$

just as in the one-dimensional model. On the other hand, if h_x and h_y are order-one quantities then the dominant balance in (16.41) is

$$\frac{1}{f} \left[h_x \left(A_{\zeta \zeta} A_{\zeta \zeta y} - A_{\zeta y} A_{\zeta \zeta \zeta} \right) + h_y \left(A_{\zeta x} A_{\zeta \zeta \zeta} - A_{\zeta \zeta} A_{\zeta \zeta x} \right) \right] = \frac{\kappa}{B_0 \hat{\delta}^2} A_{\zeta \zeta \zeta \zeta} \quad (16.44)$$

and

$$\hat{\delta} \sim \left(\frac{\kappa}{B_0} \right)^{1/2}, \quad (16.45)$$

confirming the heuristic scaling arguments. Thus, if the isotherm slopes are fixed independently of κ (for example, by the wind stress), then as $\kappa \rightarrow 0$ an internal boundary layer will form whose thickness is proportional to $\kappa^{1/2}$. We expect this to occur at the base of the main thermocline, with purely advective dynamics being dominant in the upper part of the thermocline, and determining the slope of the isotherms (i.e., the form of h_x and h_y), as in Fig. 16.3. Interestingly, the balance in the three-dimensional boundary layer equation does not in general correspond locally to $w T_z \approx \kappa T_{zz}$. Both at $\mathcal{O}(1)$ and $\mathcal{O}(\delta)$ the horizontal advective terms in (16.41) are of the same asymptotic size as the vertical advection terms. In the boundary layer the thermodynamic balance is thus $\mathbf{u} \cdot \nabla_z T + w T_z \approx \kappa T_{zz}$, whether the isotherms are sloping or flat. We might have anticipated this, because the vertical velocity passes through zero within the boundary layer.

What are the dynamics above the diffusive layer, presuming that it does not extend all the way to the surface? Answering this leads us into our next topic, the ‘ventilated thermocline’.

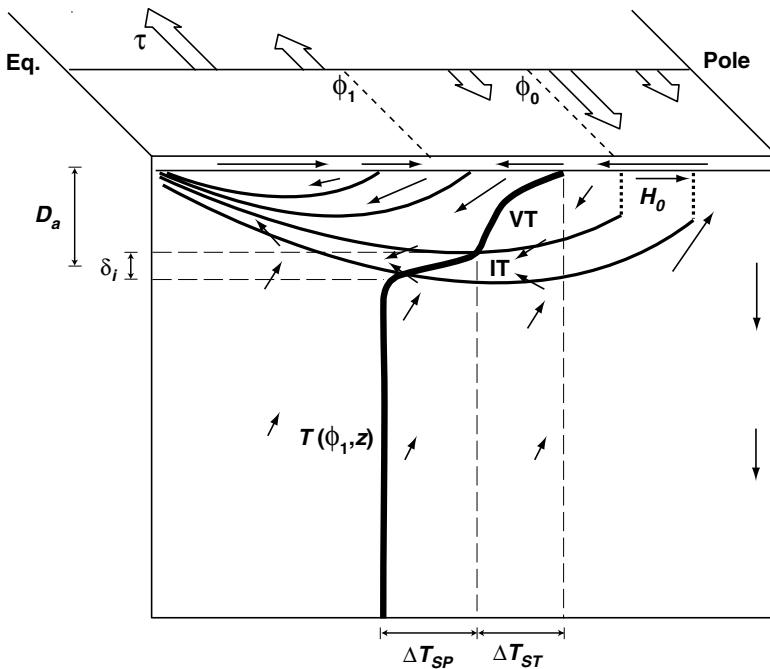


Fig. 16.7 Schema of the large-scale circulation and structure of the main thermocline, in a single-hemisphere ocean driven by wind stress (broad arrows) and a meridional gradient of heating at the surface. The subtropical–subpolar gyre boundary is a constant latitude ϕ_0 , where the wind-stress curl changes sign. ‘VT’ denotes the ventilated thermocline, an advective regime of thickness D_a , and ‘IT’ denotes the internal thermocline, a diffusive internal boundary layer of thickness δ_I . The thin arrows indicate the meridional overturning circulation and the flow in the Ekman layer near the ocean surface. The thick line is a temperature profile at latitude ϕ_1 : the temperature drop across the internal thermocline is ΔT_{SP} , equal to the meridional temperature difference across the subpolar gyre; the temperature drop across the ventilated thermocline is ΔT_{ST} , the temperature difference across the subtropical gyre.¹⁰

16.4 THE VENTILATED THERMOCLINE

We now consider the nature of the dynamics *above* the diffusive layer, presuming that the diffusivity is sufficiently small that there is a meaningful separation of the internal boundary layer and the advective dynamics above. In the advective region there is no general reason that the temperature profile should be uniform, and we envision an essentially adiabatic region that is both wind-driven and stratified. This region of the thermocline has become known, for reasons that will become apparent, as the *ventilated thermocline*. The main thermocline is composed of the internal thermocline plus a ventilated region, and to set our bearings it may be useful to refer now to the overall picture sketched in Fig. 16.7. To elucidate the structure of the ventilated thermocline we will assume the following.¹¹

- (i) The motion satisfies the ideal, steady, planetary-geostrophic equations.
- (ii) The surface temperature, and the vertical velocity due to Ekman pumping, are given. (These surface conditions are, in reality, influenced by the ocean’s dynamics, but we

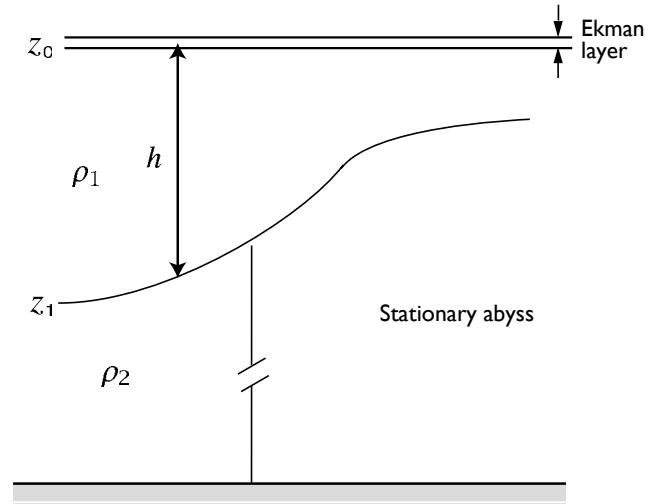


Fig. 16.8 A reduced gravity, single-layer model. A single moving layer lies above a deep, stationary layer of higher density. The upper surface is rigid. A thin Ekman layer may be envisioned to lie on top of the moving layer, providing a vertical velocity boundary condition.

assume that we can calculate a solution with specified surface conditions.) At the base of the wind-influenced region we will impose $w = 0$.

- (iii) Rather than use the continuously stratified equations, we will assume that the solution can be adequately represented by a small number of layers, each of constant density. The abyss is represented by a single stationary layer.
- (iv) We will not take into account the possible effects of a western boundary current. In that sense the model is an extension of the Sverdrup interior of homogeneous models.

The model is thus not a complete one, yet we may hope that it is revealing about the structure of the real ocean.

16.4.1 A reduced gravity, single-layer model

The simplest possible model along these lines is to suppose the ocean is composed of just two layers, as illustrated in Fig. 16.8. The upper layer of density ρ_1 is wind-driven, whereas the lower layer of density ρ_2 is assumed to be stationary; this is called a 'one-and-a-half-layer' model or a 'reduced gravity single-layer' model. Pertinent questions are: how deep is the upper layer? What is the velocity field in it?

In the planetary-geostrophic approximation, the momentum and mass conservation equations of the reduced gravity shallow water model may be written as:

$$\mathbf{f} \times \mathbf{u} = -g' \nabla h, \quad \nabla \cdot \mathbf{u} = -\frac{\partial w}{\partial z}, \quad (16.46a,b)$$

where ∇ is a two-dimensional operator (as it will be for the rest of this section) and $g' = g(\rho_2 - \rho_1)/\rho_0$ is the *reduced gravity*. Taking the curl of (16.46a) gives the geostrophic vorticity equation, $\beta v + f \nabla \cdot \mathbf{u} = 0$, and integrating this over the depth of the layer and using mass conservation gives

$$h\beta v = f(w_E - w_b), \quad (16.47)$$

where w_E is the velocity at the top of the layer, due mainly to Ekman pumping, and w_b is the vertical velocity at the layer base. If the flow is steady, w_b is zero for then

$$w_b = \mathbf{u} \cdot \nabla h = -\frac{g'}{f} \frac{\partial h}{\partial y} \frac{\partial h}{\partial x} + \frac{g'}{f} \frac{\partial h}{\partial x} \frac{\partial h}{\partial y} = 0. \quad (16.48)$$

Using this result and geostrophic balance, (16.47) becomes

$$\frac{g'}{f} \beta h \frac{\partial h}{\partial x} = f w_E \quad (16.49)$$

which integrates to

$$h^2 = -2 \frac{f^2}{g' \beta} \int_x^{x_e} w_E \, dx' + H_e^2, \quad (16.50)$$

where H_e is the (unknown) value of h at the eastern boundary x_e , and it is a constant to satisfy the no-normal flow condition. This apart, the equation contains complete information about the solution. We note that:

- ★ the depth of the moving layer scales as the magnitude of the wind stress (or Ekman pumping velocity) to the one-half power;
- ★ the horizontal solution is similar to the simpler Sverdrup interior solution previously obtained in section 14.1.3;
- ★ there is no solution if w_E is positive; that is, if there is Ekman upwelling;
- ★ the solution depends on the unknown parameter H_e , the layer depth at the eastern boundary. (That the eastern boundary depth is undetermined is perhaps the main incomplete aspect of the theory as presented here.¹²⁾

16.4.2 A two-layer model

If there is a meridional buoyancy gradient at the surface then isopycnals *outcrop*, or intersect the surface. Thus, at some latitude (say $y = y_2$, which for simplicity we assume not to be a function of longitude) layer 2 passes underneath layer 1, which is of lower density, as sketched in Fig. 16.9. Thus, polewards of y_2 the dynamics are just those of a single layer discussed above, whereas equatorwards of y_2 layer 2 does not feel the wind directly, and its dynamics are governed by two principles.

- (i) Sverdrup balance. This still applies to the vertically integrated motion, and thus to the sum of layer 1 and layer 2.
- (ii) Conservation of potential vorticity. The motion in layer 2 is shielded from the wind forcing, and the effects of dissipation are assumed to be negligible. Thus, the fluid parcels in the layer will conserve their potential vorticity.

We first use potential vorticity conservation to obtain an expression for the depths of each layer in terms of the total depth of the moving fluid, h , and then use Sverdrup balance to obtain h .

Use of potential vorticity conservation

Conservation of potential vorticity in the region equatorwards of y_2 is, for steady flow,

$$\mathbf{u}_2 \cdot \nabla q_2 = 0 \quad \text{for } y < y_2, \quad (16.51)$$

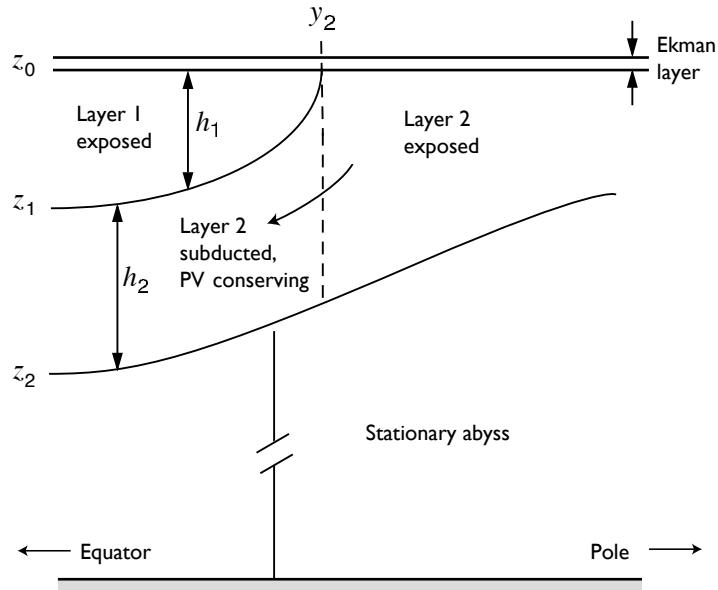


Fig. 16.9 A two-layer model of the ventilated thermocline. Two moving layers lie above an infinitely deep, stationary layer of higher density. Models with more moving layers may be constructed by straightforward extension.

where $q_2 = f/h_2$. Now, the velocity field in layer 2 is given by $\mathbf{u}_2 = (g'_2/f)\mathbf{k} \times \nabla h$, where $h = h_1 + h_2$ is the total depth of the moving fluid (see the appendix to this chapter). Thus, (16.51) becomes

$$-\frac{g'_2}{f} \frac{\partial h}{\partial y} \frac{\partial}{\partial x} \left(\frac{f}{h_2} \right) + \frac{g'_2}{f} \frac{\partial h}{\partial x} \frac{\partial}{\partial y} \left(\frac{f}{h_2} \right) = \frac{g'_2}{f} J \left(\frac{f}{h_2}, h \right) = 0. \quad (16.52)$$

This is an equation relating h and h_2 and it has the general solution

$$q_2 = \frac{f}{h_2} = G_2(h), \quad (16.53)$$

where G_2 is an *arbitrary* function of its argument. However, we *know* what the potential vorticity of layer 2 is at the moment it is subducted; it is just

$$q_2(y_2) = \frac{f(y_2)}{h_2} = \frac{f_2}{h}, \quad (16.54)$$

where $f_2 \equiv f(y_2)$, and $h_2 = h$ because $h_1 = 0$. This relationship must therefore hold everywhere in layer 2, equatorwards of y_2 ; that is,

$$G_2(h) = \frac{f_2}{h} \quad (16.55)$$

Thus, in the subducted region, and taking $z_0 = 0$,

$$\frac{f}{h_2} = \frac{f_2}{h}, \quad \text{or} \quad \frac{f}{z_1 - z_2} = -\frac{f_2}{z_2}. \quad (16.56)$$

From this we easily obtain expressions for the depths of each layer as functions of the total depth, h , namely

$$h_2 = z_1 - z_2 = \frac{f}{f_2} h \quad \text{and} \quad h_1 = -z_1 = \left(1 - \frac{f}{f_2}\right) h. \quad (16.57)$$

It remains only to find an expression for the total depth of the moving fluid, h , and this we do using Sverdrup balance. Note that because potential vorticity, f/h_2 , is conserved, as the subducted fluid column moves equatorwards its height must decrease.

Using Sverdrup balance to find the total depth

Equations (16.57) contain the unknown total depth h , and we now use Sverdrup balance to find this and close the problem. The linear vorticity equation is $\beta v = f \partial w / \partial z$, where the velocity at the top of layer one is that due to Ekman layer and the velocity at the base of layer two is zero. Given this, we may write the Sverdrup balance as

$$\beta(h_1 v_1 + h_2 v_2) = f w_E \quad (16.58)$$

where, using (16.120), the velocities in each layer are given by

$$f v_1 = \frac{\partial}{\partial x} (g'_2 h + g'_1 h_1) \quad \text{and} \quad f v_2 = \frac{\partial}{\partial x} (g'_2 h). \quad (16.59)$$

Using these, Sverdrup balance becomes,

$$\beta h_1 \frac{\partial}{\partial x} (g'_2 h + g'_1 h_1) + \beta(h - h_1) g'_2 \frac{\partial h}{\partial x} = f^2 w_E, \quad (16.60)$$

or

$$\frac{\partial}{\partial x} (g'_2 h^2 + g'_1 h_1^2) = \frac{2f^2}{\beta} w_E. \quad (16.61)$$

On integrating this gives

$$\left(h^2 + \frac{g'_1}{g'_2} h_1^2\right) = D_0^2 + C, \quad (16.62)$$

where

$$D_0^2(x, y) = -\frac{2f^2}{\beta g'_2} \int_x^{x_e} w_E(x', y) \, dx', \quad (16.63)$$

which by construction vanishes at the eastern wall ($x = x_e$). The constant of integration C may be interpreted as follows. We write $C = H_e^2 + (g'_1/g'_2)H_1^2$ where H_e is the (unknown) total depth of layers 1 and 2 at the eastern boundary, and H_1 is the depth of layer 1. These must both be constants in order to satisfy the no-normal flow condition. However, H_1 must be zero, because at the outcrop line $h_1 = 0$. Thus, H_1 is zero at $y = y_2$, and therefore zero everywhere, and $C = H_e^2$.

Using (16.57) and (16.62) we obtain a closed expression for h , namely

$$h = -z_2 = \frac{(D_0^2 + H_e^2)^{1/2}}{\left[1 + (g'_1/g'_2)(1 - f/f_2)^2\right]^{1/2}}. \quad (16.64)$$

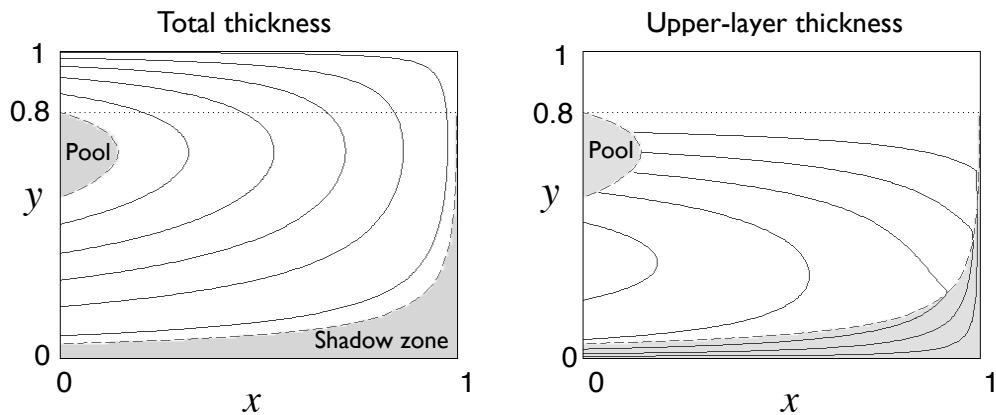


Fig. 16.10 Contour plots of total thickness and upper layer thickness in a two-layer model of the ventilated thermocline. The thickness generally increases westwards, and the flow is clockwise. The shadow zone and the western pool are shaded, and no contours are drawn in the latter. The outcrop latitude, $y_2 = 0.8$, is marked with a dotted line. [Parameters are $g'_1 = g'_2 = 1$, $\beta = 1$, $f_0 = 0.5$, $H_e = 0.5$, and $w_E = -\sin(\pi y)$.]

Using (16.57) the depths in each layer, and the corresponding geostrophic velocities, can readily be obtained.

A typical solution is shown in Fig. 16.10. The upper layer exists only equatorwards of the outcrop latitude, $y_2 = 0.8$, and isolines of total thickness correspond to streamlines of the lower layer. We see, as expected, the overall shape of a subtropical gyre, with the circulation being closed by an implicit western boundary current that is not part of the calculation. Two regions are shaded in the figure, the 'pool' region in the west and the 'shadow zone' in the south-east. The solutions above do not apply to these, and they require some special attention.

16.4.3 The shadow zone

In the fluid interior the potential vorticity of a parcel in layer 2 is determined by tracing its trajectory back to its outcrop latitude where the potential vorticity is given. That trajectory is determined by its velocity, and this turn is determined by inverting the potential vorticity. Now, parcels subducted at y_2 sweep equatorwards and westwards, so that a parcel, labelled '**a**' say, subducted at the eastern boundary will in general leave the eastern boundary tracing a southwestern trajectory. Consider another parcel, '**b**' say, in the interior that lies eastwards of the subducted position of **a**, in the shaded region of Fig. 16.11. It is impossible to trace **b** back to the outcrop line without trajectories crossing, and this is forbidden in steady flow. Rather, it seems as if the trajectory of **b** would emanate from the eastern wall. What is the potential vorticity there?

At the eastern boundary the condition of no normal flow at the boundary demands that h be constant (so that $u_2 = 0$), and h_1 be constant (so that $u_1 = 0$). But if a parcel in layer 2 moves *along* the boundary potential vorticity conservation demands that f/h_2

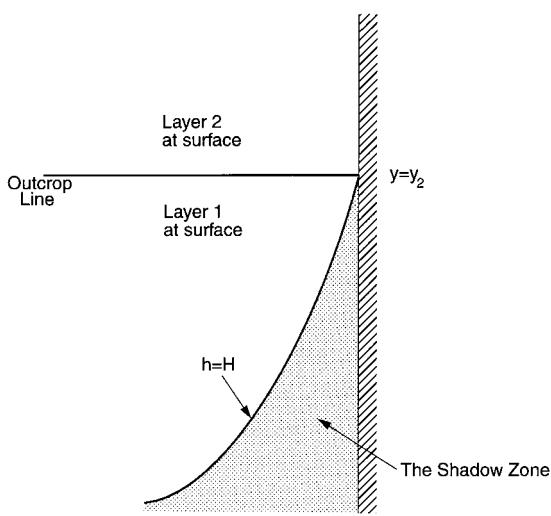


Fig. 16.11 The shadow zone in the ventilated thermocline. Layer 2 outcrops at $y = y_2$. A column moving equatorwards along the eastern boundary in layer 2 is subducted at y_2 . It cannot remain against the eastern wall and both preserve its potential vorticity, which implies the column shrinks, at the same time that the no-normal flow condition is satisfied, as by geostrophy this implies the layer depth is constant. Thus, the column must move westwards, along the boundary of a 'shadow zone' within which there is no motion. The streamline it follows is the isoline of constant total thickness of the two moving layers [see (16.120) or (16.122c)].

is constant, and therefore h_2 must change, contradicting the no-normal flow requirement. Thus, the velocity at the boundary can have neither a normal component nor a tangential component, and so we cannot trace parcels in the shaded region back to the wall. Rather, in the absence of closed trajectories (for example, eddying motion), we may assume that the shaded region is stagnant, and h is constant. Of course, potential vorticity is everywhere given by f/h_2 , which varies spatially, but since there is no motion potential vorticity is still, rather trivially, conserved along trajectories. This region is aptly called the *shadow zone*, since the region falls under the shadow of the eastern boundary; an analogous region arose in the quasi-geostrophic discussion of section 14.7.

To obtain an expression for the fields within the shadow zone, first note that because h is constant, its value is equal to that on the eastern wall; that is, $h = H_e$. The wind forcing must then all be taken up by the upper layer, and Sverdrup balance then implies

$$\beta v_1 h_1 = f w_E, \quad (16.65)$$

and using (16.59) we obtain an expression for h_1 , to wit

$$h_1^2 = -\frac{2f^2}{\beta g'_1} \int_x^{x_e} w_E(x', y) dx' = \frac{g'_2}{g'_1} D_0^2, \quad (16.66)$$

which is zero at the eastern wall. In the lower layer the thickness is just $h_2 = H_e - h_1$. The boundary of the shadow zone is given by the trajectory of a fluid parcel in layer 2 that emanates from the eastern boundary at the outcrop line where $h_1 = 0$ and $h = h_2 = H_e$. Since the flow is steady, the trajectory is an isoline of h . Thus, from (16.64) we have

$$h^2 = \frac{(D_0^2(x_s, y_s) + H_e^2)}{[1 + g'_1/g'_2 (1 - f/f_2)^2]} = H_e^2, \quad (16.67)$$

where (x_s, y_s) denotes the boundary of the shadow zone. (Note that $x_s = x_e$ at $y = y_2$.)

From this

$$D_0^2(x_s, y_s) = H_e^2 \left[\frac{g'_1}{g'_2} \left(1 - \frac{f}{f_2} \right)^2 \right] \quad (16.68)$$

which, given the wind stress, determines the shadow zone boundary x_s as a function of y .

16.4.4 † The western pool

Polewards of the outcrop latitude the fluid of layer 2 feels the wind directly and the layer thickness is determined by Sverdrup balance. Equatorwards of the outcrop latitude the properties of this layer are determined by potential vorticity conservation, with the potential vorticity being determined by the layer thickness at the outcrop. However, just as there is a region in the east where trajectories cannot be traced back to the outcrop, there is a ‘pool’ region in the west that is bounded by the trajectory that emerges from the western boundary at the outcrop latitude. Within the pool, trajectories cannot be traced back to the outcrop (Fig. 16.10), and one might imagine that they emerge from the western boundary current. In the context of ventilated thermocline theory, there are two plausible hypotheses for determining the layer depths within this region:

- (i) within layer 2, potential vorticity is homogenized;
- (ii) because there is no source for layer-2 water, layer-2 water does not exist and the pool consists solely of ventilated, layer-1 water.

There is a loose similarity between these two hypotheses and those for the stratification of the troposphere summarized on page 535.

(i) Potential vorticity homogenization

The pool region is a region of recirculation, receiving water from and depositing water into the western boundary current. Thus, following the ideas described in chapter 10 and employed in section 14.7, we hypothesize that the potential vorticity within this region becomes homogenized. The value of potential vorticity within the pool is just the value of potential vorticity at its boundary, and this is given by $f_2/h_2(w)$, where $h_2(w)$ is the thickness of layer 2 at the western boundary at the outcrop latitude. This is given using (16.50) with $f = f_2$ and $g' = g'_2$, and thus the potential vorticity in the pool is given by

$$q_{pool} = \frac{f_2}{D_w^2 + H_e^2}, \quad (16.69)$$

where $D_w^2 = -2(f_2^2/g'_2\beta) \int_{x_w}^{x_e} w_E(x', y_2) dx'$. The thickness of layer 2 in the pool must be consistent with this, and so is given by

$$h_2 = \frac{f}{q_{pool}}. \quad (16.70)$$

The thickness of layer 1 is determined by using Sverdrup balance, (16.58), which, given h_2 and geostrophy, reduces to an equation for h_1 .

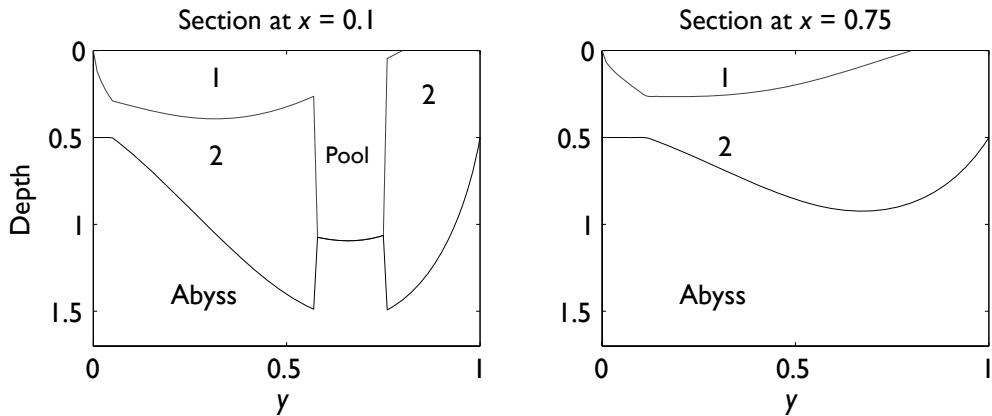


Fig. 16.12 Two north-south section of layer thickness, at different longitudes, from the same solution as Fig. 16.10 and assuming a ventilated western pool. The numbers refer to the fluid layer. The section on the left passes through the western ventilated pool region, where all the Sverdrup transport is taken up by the top layer. The region near $y = 0$ in both plots where the total depth of the thermocline is constant is the shadow zone.

(ii) The ventilated pool

'I came for the waters.' 'What waters?' 'I was misinformed.'

From *Casablanca* (1942).

The homogenization hypothesis, although plausible, depends on the assumption of down-gradient diffusion of potential vorticity by eddies. Also, because there is no source of layer-2 water in the pool, we must suppose that it is ventilated by eddy pathways that meander down from the surface. An alternative hypothesis, and one that does not rely on the properties of mesoscale eddies, is to suppose that the western pool is filled with water that is directly ventilated from the surface. That is, if there is no surface source for a water mass, we simply suppose that that water mass does not exist. In the two-layer model, this means that the western pool is filled entirely with layer-1 fluid. If a non-ventilated (e.g., layer-2) fluid is present initially, then we hypothesize that it is slowly expunged by the continuous downwards Ekman pumping of layer-1 water into the pool.¹³

Because the layer-2 fluid is absent, the layer-1 fluid extends all the way down to the stagnant abyss; it takes up all the Sverdrup transport, and this determines the depth of the ventilated pool. Thus, rounding up the usual equations, we set $h = h_1$ in (16.62) to give

$$h_1^2 = D_1^2 + g'_2 H_e^2, \quad (16.71)$$

where

$$D_1^2(x, y) = -\frac{2f^2}{\beta g'_{1a}} \int_x^{x_e} w_E(x', y) dx', \quad (16.72)$$

with $g'_{1a} = g'_1 + g'_2$ being the reduced gravity between layer 1 and the abyss, and H_e , as before, being the thickness of layer 2 at the eastern boundary. Because $g'_{1a} > g'_2$ this pool

Thermocline Dynamics — a Summary and Overview

The model of the main thermocline that we have constructed in sections 16.1–16.4 is illustrated schematically in Fig. 16.7. Some of the features, and limitations, of this model are listed below.

- ★ The main subtropical thermocline consists of an advective upper region overlying a diffusive base.
 - The diffusive base forms the *internal thermocline*, and in the limit of small diffusivity this is an internal boundary layer. The advective region forms the *ventilated thermocline*. The separation of the two regions may, in reality, not be sharp.
 - The relative thicknesses of these layers is a function of various parameters, notably the strength of the wind and the magnitude of the diffusivity.
- ★ Above the ventilated thermocline there may be a mixed layer with a seasonally varying depth. In certain regions, for example at the poleward edge of the subtropical gyre, convection may deepen the mixed layer as far as the base of the thermocline.
- ★ A western boundary layer is needed to close the circulation and the heat budget.
- ★ The single-hemisphere model assumes that the water that sinks at high latitude either upwells through the main thermocline or returns to the subpolar gyre beneath the main thermocline. In reality much of this water may cross into the other hemisphere, so that, for example, the deep water in the North Atlantic may upwell in the Antarctic Circumpolar Current.
 - In this case, the diffusion-dependent overturning circulation represents only part of the overall meridional overturning circulation.
 - Nevertheless, there would remain a diffusive internal thermocline, because there is still a boundary between the warm subtropical water and cold abyssal water.
- ★ Within the ventilated thermocline there are two regions — the shadow zone and the western pool — whose dynamics are not determined without additional assumptions. A plausible assumption for the western pool, namely that all the water within it is ventilated, leads to a model of mode water.
 - Ventilated thermocline theory contains an unknown parameter, H_e , the depth of the thermocline on the eastern boundary.
- ★ Mesoscale eddies may play an important role in thermocline dynamics. Numerical simulations and some observations suggest that eddies may be particularly important in the internal thermocline, and they may provide a tendency for potential vorticity to homogenize.

will generally be shallower than the total depth of moving fluid ($h_1 + h_2$) just outside, but the depth of layer-1 fluid alone will be much greater; that is, there will be discontinuities in layer depths at the pool boundary. A section through the pool region is shown in Fig. 16.12.

Although we may be shocked by the appearance of discontinuities in layer depths in a fluid model, the model does provide a simple mechanism for the appearance of *mode water*. This is a distinct mass of weakly stratified, low potential vorticity water appearing in the north-west corner of the North Atlantic subtropical gyre (where it is sometimes called '18 degree water'), with analogues in the other gyres of the world's oceans; it is so-called because it appears as a distinct mode in a census of water properties. The proximate mechanism for mode water formation is convection in winter, but for such convection to occur the large-scale ocean circulation must maintain a weakly stratified region, and it is the ventilated pool that enables this, and sets the formation in the context of thermocline structure. In reality, the vertical isopycnals predicted by the simple model will, of course, be highly baroclinically unstable, and the ensuing mesoscale eddies will erode the pool interface and cause the isopycnals to slump, so that the discontinuities in layer depths will be manifest only as rapid changes or fronts.

We conclude by emphasizing that both the ventilated pool theory and the potential vorticity homogenization theory rely on hypotheses that cannot be derived from the governing equations of motion without making additional physical assumptions that are neither *a priori* true nor obvious. For an overview of the entire main thermocline, refer to the shaded box on the facing page.

16.5 † A MODEL OF DEEP WIND-DRIVEN OVERTURNING

The sciences do not try to explain... they mainly make models.

John von Neumann (1903–1957).

In sections 15.7 and 16.2 we made estimates for the strength of the overturning circulation assuming that it was buoyancy driven. Using values of the diapycnal diffusivity that are measured in the main thermocline these estimates tend to be rather smaller than the overturning circulation that is observed. Relatedly, and as we noted in section 16.1, the deep stratification of the ocean, although small, is non-zero and this can only be maintained, if the flow is across isopycnals, by a somewhat larger value of diffusivity than is generally measured in the upper ocean. There are two possible resolutions to these problems. One is to note that the measured diapycnal diffusivity is in fact large in some parts of the ocean (e.g., in the abyss over steep topography), and if this were sufficient to produce the measured overturning and stratification the issue would be resolved. However, such a calculation would likely be fraught with uncertainty. A second and more straightforward resolution would arise if a deep circulation, and deep stratification, could be maintained by a mechanism that was *independent* of the diapycnal diffusivity. This second approach is the one we shall take in this section. Specifically, our goal is to construct a model that illustrates that the overturning circulation of the ocean, and a concomitant deep stratification, may have a 'wind-driven' component that persists even as the diffusivity goes to zero.¹⁴

In the absence of a diapycnal diffusivity no upwelling can occur *through* the stratification, because that is a diabatic process. Rather, if there is deep stratification, the deep water must be directly connected to the surface along isopycnals or via a convective pathway, for

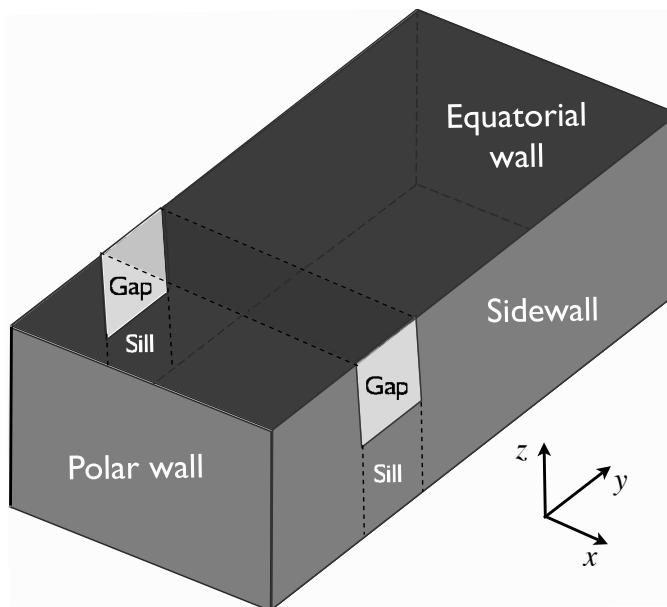


Fig. 16.13 Idealized geometry of the Southern Ocean: a re-entrant channel, partially blocked by a sill, is embedded within a closed rectangular basin; thus, the channel has periodic boundary conditions. The channel is a crude model of the Antarctic Circumpolar Current, with the area over the sill analogous to the Drake Passage.

convection, although diabatic, does not rely on a finite eddy diapycnal diffusivity. Let us recall two *de facto* principles that were useful in our discussion of the overturning circulation and the thermocline.

- (i) A basin will, in the absence of mechanical forcing, tend to fill with the densest available fluid.
- (ii) Light fluid forced down by wind may displace the cold fluid, so producing stratification.

A completely closed ocean thus fills completely with dense polar water, except in the upper several hundred metres where the main thermocline forms, and whose thickness is the sum of wind-driven and diffusive components. However, suppose that the polewards part of the basin is not fully enclosed but is periodic, as illustrated in Fig. 16.13, with a sill across it at mid-depth, and suppose too that the surface boundary conditions are such that the surface temperature decreases monotonically polewards. A fully enclosed basin exists only beneath the level of the sill, and we may expect the densest water in the basin, formed at the polewards edge of the domain, to fill the basin only below the level of the sill, and that above this may lie warmer water with origins at lower latitudes. Furthermore, suppose that an eastward wind blows over the channel that produces an equatorial flow in the Ekman layer. Then mass conservation demands that there must exist a *subsurface* return flow, and thus a meridional overturning circulation is set up. Note the essential role of the channel in this: if the gap were closed, then the return flow could take place at the surface via a western boundary current, as in a conventional subpolar gyre, and no overturning circulation need be set up.

16.5.1 A single-hemisphere model

We consider first a single-hemisphere basin with a periodic channel near its poleward edge. We suppose it to be in the southern hemisphere, so the channel represents the Antarctic Circumpolar Current (ACC), and that the dynamics are Boussinesq and planetary-geostrophic. We will choose extremely simple forms of wind and buoyancy forcing to allow us to obtain an analytic solution, and then later discuss how the qualitative forms of these solutions might more apply generally.

Wind and buoyancy forcing

Thermodynamic forcing is imposed by fixing the surface buoyancy, b_s . (In the discussion following salinity is absent, and buoyancy is virtually equivalent to temperature.) South of the gap we suppose the buoyancy to be constant, then that it increases linearly across the gap, and is constant again polewards of the gap. Thus, there is no temperature gradient across the subtropical gyre, focusing attention on the influence of the channel. Thus, referring to Fig. 16.14 or Fig. 16.15 for the definitions of the geometric factors,

$$b_s = \begin{cases} b_1, & y_0 \leq y \leq y_1 \\ b_1 + \frac{(b_2 - b_1)(y - y_1)}{y_2 - y_1}, & y_1 \leq y \leq y_2 \\ b_2, & y \geq y_2, \end{cases} \quad (16.73)$$

where $b_2 > b_1$, and both are constants, and we may take $b_1 = 0$ and $y_0 = 0$.

The wind forcing is purely zonal, and it is convenient to express this in terms of the Ekman transport and associated pumping (refer to section 2.12). In the channel the Ekman transport is chosen to be (realistically) equatorwards and (less realistically) constant, a simplification that avoids complications of wind-driven upwelling in the channel. South (polewards) of the channel there is a conventional subpolar gyre, with an Ekman upwelling and an equatorward Ekman transport that joins smoothly to that of the channel. Equatorwards of the channel there is conventional subtropical gyre, with Ekman downwelling. All this may be achieved by specifying:

$$v_E = \begin{cases} \frac{V}{2} \left[1 - \cos \left(\frac{\pi y}{\Delta y_1} \right) \right] & 0 \leq y < y_1 \\ V & y_1 \leq y < y_2 \\ \frac{V}{2} \left[1 + \cos \left(\frac{\pi(y - y_2)}{\Delta y_2} \right) \right] & y_2 \leq y < y_3, \end{cases} \quad w_E = \begin{cases} W_1 \sin \left(\frac{\pi y}{\Delta y_1} \right) & 0 \leq y < y_1 \\ 0 & y_1 \leq y < y_2 \\ -W_2 \sin \left(\frac{\pi(y - y_2)}{\Delta y_2} \right) & y_2 \leq y < y_3, \end{cases} \quad (16.74a,b)$$

where $\Delta y_1 = y_1$, $\Delta y_2 = y_3 - y_2$, and V is a constant that determines the magnitude of the meridional Ekman flow. The meridional Ekman transport, v_E , is related to the Ekman pumping by $w_E/\delta_E = \partial v_E/\partial y$, so that $W_i = \delta_E \pi V / (2\Delta y_i)$, where δ_E is the Ekman layer thickness. If f were constant, the wind-stress curl would be proportional to the w_E field above. The precise details of the forcing do not affect the qualitative form of the solution — they merely allow an analytic solution to be obtained — but there are two essential aspects to it.

- (i) The surface is cold south of the channel, warm north of the channel, and there is a temperature gradient across the channel.

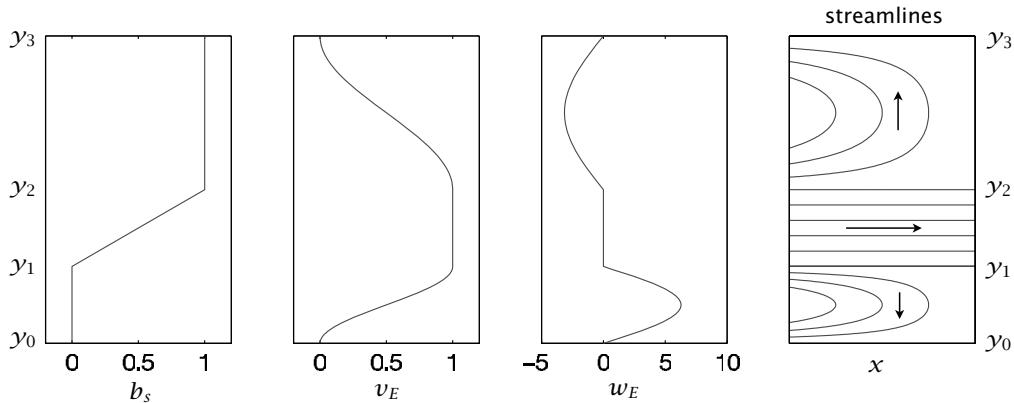


Fig. 16.14 The surface buoyancy b_s , meridional Ekman velocity v_E , vertical Ekman velocity w_E and the solution streamlines for the geostrophic horizontal flow, omitting the western boundary currents. The ordinate in all plots is latitude, with the pole at the bottom, and the four fields are given by, respectively, (16.73), (16.74a), (16.74b) and (16.75), with purely zonal flow given by (16.78) in the channel.

- (ii) The Ekman flow is equatorwards within the channel, with conventional gyres to either side.

The meridional extent of the region south of the channel and the wind forcing within it are relatively unimportant, and the region could be shrunk to nearly zero.

Solution in the gyres

Below the depth of the sill the basin is fully enclosed, and therefore up to that level the basin will fill with the densest available water (much as described in section 15.1), except where it may be displaced by warmer fluid equatorwards of the gap that is pumped down below the level of the sill by the wind (Fig. 16.15). Thus, all of the domain south of the channel, and nearly everywhere below the sill, the water has buoyancy b_1 . Polewards of the channel, then, the fluid is barotropic and its vertically integrated horizontal circulation is given by Sverdrup balance, $\beta V = f w_E / H$, where V is the vertically integrated flow. With the wind stress of (16.74) we get a conventional barotropic subpolar gyre (and associated western boundary current) by the same methods that we employed in chapter 14.

Above the sill, net meridional geostrophic transfer is forbidden in the channel region, because by geostrophic balance $f \bar{v}_g = \partial \bar{\phi} / \partial x = 0$, where ϕ is the pressure and the overbar denotes a zonal average. Equatorwards of the channel the region above the sill will therefore tend to fill with the densest water available to it, and this is water with buoyancy equal to b_2 (which is the buoyancy of the water as it emerges from the channel). However because of the presence of wind forcing, the base of this layer is not flat; rather, this fluid obeys the dynamics of the reduced-gravity single-layer ventilated thermocline model discussed in section 16.4.1. In such a model the depth of the fluid on the eastern boundary is constant, and this must be specified. Here, this is given by the height of the sill, and therefore $h(x = x_e, y) = h_e = H - \eta_{sill}$, where H is the total depth of the basin and η_{sill} is the sill height. Then, using (16.50) and (16.74), the thickness of the moving layer equatorwards of

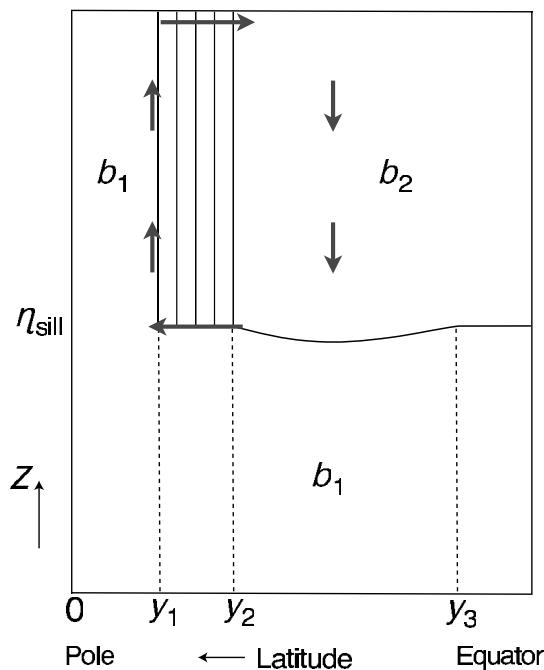


Fig. 16.15 Cross-section of the structure of the single-hemisphere ocean model described in section 16.5.1. The domain is zonally closed equatorwards of y_2 and polewards of y_1 , with a zonally periodic channel between latitudes y_1 and y_2 and above the sill, which has height η_{sill} . The arrows indicate the fluid flow driven by the equatorward Ekman transport in the channel, and the solid lines are isopycnals.

the sill is given by, for $y_2 < y < y_3$,

$$h^2 = D^2(x, y) + h_e^2 \quad (16.75)$$

where

$$D^2 = -\frac{2f^2}{g'\beta} \int_x^{x_e} w_E \, dx' = \frac{2f^2}{g'\beta} W_2(x_e - x) \sin\left(\frac{\pi(y - y_2)}{\Delta y_2}\right) \quad (16.76)$$

with $g' = b_2 - b_1$. The solution is closed by the addition of a western boundary current. Note that because $h > h_e$, the light fluid is pushed below the level of the sill in the subtropical gyre.

Solution in the channel

In the channel, the fluid in the Ekman layer flows equatorwards, and therefore there must be a compensating poleward flow at depth. This will occur at and just below the level of the sill: it cannot be deeper, because here the basin is full of denser, b_1 fluid, and in the absence of eddying or ageostrophic flow it cannot be shallower because of the geostrophic constraint. Now, because of the temperature gradient across the channel the polewards flowing fluid is warmer than the fluid at the surface, and therefore convectively unstable. Convection ensues, the result of which is the entire column of fluid between the top of the sill and the surface mixes and takes on the temperature of the surface. Thermal wind demands that there be a zonal flow associated with this meridional temperature gradient, so this temperature distribution is advected eastwards into the interior of the channel. Because the interior is presumed to be adiabatic, this temperature field extends zonally throughout the channel. Thus, in steady state, the temperature *everywhere* in the channel above the

level of the sill is given by

$$b(x, y, z) = b_s(y) = b_1 + \frac{(b_2 - b_1)(y - y_1)}{y_2 - y_1}, \quad y_1 \leq y \leq y_2, \quad z > \eta_{sill}. \quad (16.77)$$

Note that convective mixing does not rely on a diapycnal diffusivity other than a molecular one: convective plumes are generally turbulent, generating small scales in the fluid interior where mixing and entrainment may occur; failing that, the lighter fluid is displaced to the surface where it cools by way of interaction with the atmosphere. The zonal velocity within the channel is then given by thermal wind balance, so that

$$u(x, y, z) = -\frac{1}{f} \left(\frac{b_2 - b_1}{y_2 - y_1} \right) (z - \eta_{sill}). \quad (16.78)$$

(Note that $f < 0$ so that the shear is positive.)

Regarding the depth-integrated zonal momentum budget, the wind stress at the surface is balanced by a pressure force against the sill walls. This pressure gradient arises through the meridional circulation, as the southward return flow just below the level of the sill is associated with a zonal pressure gradient that is exactly equal, but opposite, to the stress exerted by the wind. That is to say, in the Ekman layer the wind stress is balanced by the Coriolis force on the equatorward flow in the Ekman layer, which by mass conservation is equal and opposite to the Coriolis force on the deep poleward flow, which by geostrophy is equal to the net pressure force on the sill walls. The wind stress plays no role in determining the zonal transport of the channel: if the wind increases, the meridional overturning and the pressure force increase but with no change to the transport. This is a somewhat unrealistic feature of the model, for in reality the form stress induced by the flow over bottom topography (and that balances the wind stress) is likely to be a function of the zonal transport as well as the meridional transport.

A qualitative summary

The circulation of the model may be described as follows. The entire basin polewards of the channel fills with dense, b_1 , water. Below the sill this fluid extends equatorwards, filling the lower part of the channel and subtropical basin, up to the level of the sill. Now, Ekman pumping in the channel forces near-surface fluid equatorwards, which warms as it goes, entering the subtropical basin with buoyancy b_2 . This fluid fills the basin down to the level of the sill, where it encounters the dense, b_1 , fluid. The subtropical basin is wind-driven, and it forms a subtropical gyre with a single moving layer. Its dynamics are completely determined by specifying the wind, the reduced gravity ($g' = b_2 - b_1$), and the depth of the fluid at the eastern boundary (the sill depth). Because of the requirements of mass conservation, there must be a poleward return flow at depth, and so at the level of the sill warm water flows polewards. This flow is convectively unstable (because the water is lighter than that at the surface), and so the entire column of fluid mixes and its density takes on the value at the surface. The meridional temperature gradient gives rise to an eastward flow, and this temperature field is advected zonally, and in steady state the temperature distribution is zonally symmetric and given by (16.77). The overturning circulation within the ACC is known as the Deacon Cell, and this is a crude model of it. It is considered further in section 16.6.

If the diapycnal diffusivity were non-zero, the sharp boundary between the two fluid

masses at the sill height would be diffused to a front of finite thickness, with some upwelling and water mass transformation occurring across the front. This diffusive loss of dense fluid would be compensated by water-mass formation at the surface, polewards of the channel, leading to a deep, diffusively-driven circulation. That is, the deep water mass of b_1 fluid would circulate: this is a crude model of the ‘Antarctic Bottom Water’ cell.

Suppose now that the wind were everywhere zero, and the diffusivity small but non-zero. The cold, b_1 fluid would quickly completely fill the basin polewards of the channel, and would also fill the basin equatorwards of the channel up to the level of the sill. However, with no wind to drive an overturning circulation dense b_1 water would slowly drift ageostrophically across the channel, displacing any warmer water until the *entire basin* were filled with the dense, b_1 , fluid, except for a thin boundary layer at the top needed to satisfy the upper boundary condition. The final state would be one of no motion, and no stratification, below this boundary layer.

The important overall conclusion to be drawn is the following: *a deep meridional circulation and a deep stratification can be maintained, even as the diapycnal diffusivity goes to zero, in the presence of a wind forcing and a circumpolar channel.* Of course there are a number of idealized or unrealistic aspects to this model, perhaps the most egregious being the ones following.

- ★ The vertical isopycnals in the channel will be highly baroclinically unstable. This will cause the isopycnals to slump and will potentially set up an eddy-induced circulation. We consider this in section 16.6.
- ★ The model has no surface temperature gradient across the subtropical gyre. If one were present, it would lead to the formation of a ‘main’ subtropical thermocline, a full treatment of which would require determining its eastern boundary conditions. This would, however, be unlikely to qualitatively affect the presence of a deep, wind-driven overturning circulation.
- ★ The wind stress in the model channel is chosen so that the meridional Ekman transport is constant. (This means the wind stress is chosen to vary in the same fashion as the Coriolis parameter, and if f were constant, the wind-stress curl would vanish.) Thus, there is no wind-driven downwelling or upwelling in the channel, and this simplifies the solution. Numerical simulations suggest that this choice does not affect the qualitative nature of the overturning circulation or temperature distribution.

16.5.2 A cross-equatorial wind-driven deep circulation

Let us now qualitatively and heuristically extend the above model to consider flow across the equator. Thus, we suppose that the ocean basin extends to high northern latitudes, where there is, potentially, another source of cold deep water. To keep the model simple and tractable we will assume a very simple buoyancy structure:

$$b_s = \begin{cases} b_1, & 0 \leq y \leq y_1 \\ b_1 + \frac{(b_2 - b_1)(y - y_1)}{y_2 - y_1}, & y_1 \leq y \leq y_2 \\ b_2, & y_2 \leq y \leq y_4 \\ b_3 & y > y_4, \end{cases} \quad (16.79)$$

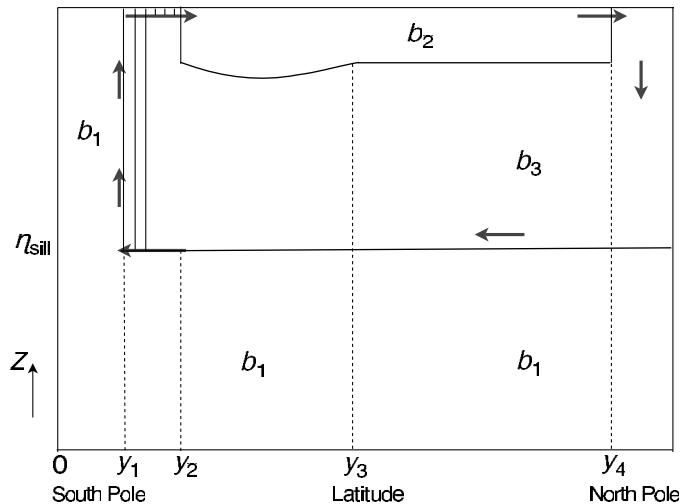


Fig. 16.16 As for Fig. 16.15, but now for a two-hemisphere ocean with a source of dense water, b_3 , at high northern latitudes. The solid lines are isopycnals, and here the wind is zero in the Northern Hemisphere.

where the geometry is illustrated in Fig. 16.16. Given that $b_2 > b_1$, there are three cases to consider.

- (i) $b_3 > b_2$. This is not oceanographically relevant to today's climate, nor does it provide another potential deep water source.
- (ii) $b_3 < b_1$. The northern water is now the densest in the ocean, and would fill up the entire basin north of the channel (except near the surface in regions where some b_2 water is pushed down by the wind), and so provide no mid-depth stratification.
- (iii) $b_1 < b_3 < b_2$. This is the most interesting and relevant case, and the only one we explore further.

As regards the wind, we will assume that south of the equator this is given by (16.74). North of the equator the wind forcing does not affect the qualitative nature of the overturning circulation, and may be taken to be zero.

Descriptive solution

In case (iii), the entire basin below the sill fills with b_1 water, except where wind forcing forces warmer fluid below the sill level, as before. However, unlike the earlier case, the fluid above the sill is predominantly b_3 water from high northern latitudes. This forms in high polar latitudes and fills most of the basin above the sill, from the basin boundary in the north to the channel in the south (as discussed more below). However, except at latitudes where the b_3 is formed, it does not reach the surface because of the presence of b_2 water. That water is pushed down by the wind in the southern hemisphere to some as yet undetermined depth (discussed below), the boundary between b_2 and b_3 water then forming the upper ocean thermocline.

These water masses circulate because of the wind forcing in the channel. As in the single-hemisphere case, northwards flowing water emerges from the channel with buoyancy b_2 . This emerges into a region of Ekman downwelling, with a northward transport carried by a western boundary current. This transport crosses the equator finally reaching the latitudes where b_3 water is formed where it sinks and returns equatorwards, again in a

western boundary current. (Away from the western boundary layer there is no meridional flow in the absence of diffusion, because the flow satisfies $\beta v = f \partial w / \partial z$ and there is no upwelling.) This water then crosses the sill. However, unlike the single-hemisphere case, in the northern part of the sill this water is denser than the surface water; no convection occurs and so the b_3 water extends upwards to the surface, where it warms by contact with the atmosphere and advected equatorwards to become b_2 water. Further south the surface buoyancy in the channel is less than b_3 , and the column now mixes convectively, much as in the single-hemisphere case. The solution is completed by specifying the thickness of the layer of b_2 water at the surface. Now, if the circulation is in steady state, the meridional transport between the gyres must equal that of the northward Ekman flow at the northern edge of the circumpolar channel, and given the wind forcing, this is determined by the depth of the layer at the eastern boundary, a constant. Thus, in this model, global constraints determine the depth of the eastern boundary of the thermocline.

Suppose that the wind were everywhere zero. Then, as in the single-hemisphere case, the circulation would eventually die. Again, though, slow ageostrophic motion across the channel would first allow the entire basin, within and on both sides of the channel, to fill with the densest available water, and in the final steady state there would be no stratification (and no motion) below a thin surface layer.

Suppose, on the other hand, that a small amount of diffusion were added to the wind-forced model above. Then there would be mass exchange between the layers and, in particular, the deep cell of b_1 water would begin to circulate diffusively. In addition, the mid-depth cell would begin to upwell through the b_2 - b_3 interface, and develop a diffusively driven circulation, in much the same way as is illustrated in Fig. 16.7.

Summary remarks

The key result of this model is that, even as the diffusivity falls and the interior of the ocean becomes more and more adiabatic, a meridional cross-hemispheric circulation can be maintained, provided that the wind across the circumpolar channel remains finite. The diabatic water mass transformations all occur at the surface or in convection: these processes require a non-zero diffusivity, but this can be the molecular diffusivity because the associated mixing involves turbulence, which can generate arbitrarily small scales. (Note also that the convection that occurs in the circumpolar channel reduces the potential energy of the column, and requires no mechanical input of energy.) Aside from the region of the ACC, the meridional transport will occur (in this model) in western boundary layers. Indeed, we may still expect to see a southwards flowing deep western boundary current south of γ_4 and below the b_2 water in Fig. 16.16, just as in the implicitly diffusive Stommel-Arons model. In the ACC itself, the meridional transport occurs in a subsurface current, nestled against the sill. Although the overturning circulation in this model is ‘wind-driven’, the possibility that it may be cross-equatorial depends upon the thermodynamic forcing; in particular, if there is no source of dense water in the northern hemisphere, then the basin above the sill simply fills with b_2 water, as in the model of section 16.5.1, and there need be little or no interhemispheric flow. We emphasize, too, that our model of interhemispheric flow is quite heuristic: we have essentially *posited* that b_2 water may continuously flow across the equator, without examining the equatorial dynamics in any detail.

Letting our reach exceed our grasp, with a little imagination we can combine the above model of the deep circulation with the models of the thermocline discussed in sections

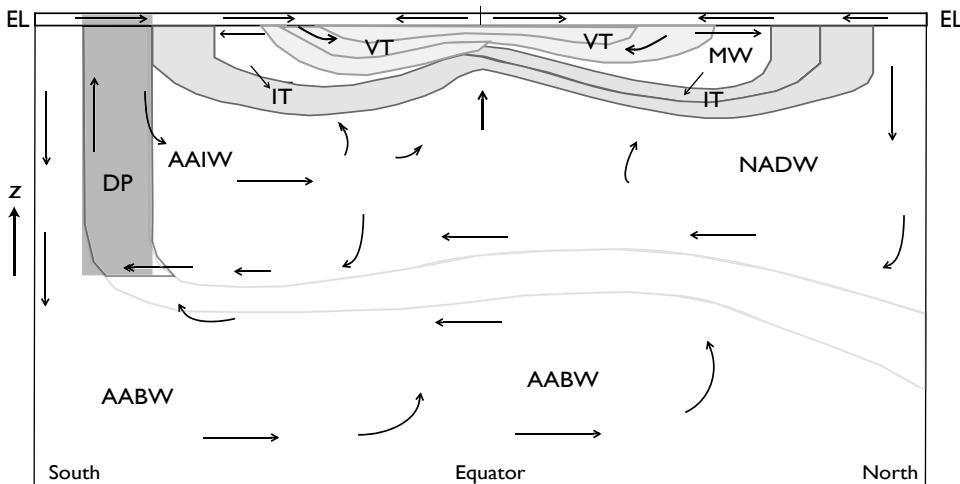


Fig. 16.17 A schema of the stratification and overturning circulation obtained by combining the thermocline models of sections 16.1–16.4 with the model of deep overturning of section 16.5. Key: DP – Drake Passage; EL – Ekman layer; VT – ventilated thermocline; IT – internal thermocline; AABW – Antarctic Bottom Water; AAIW – Antarctic Intermediate Water; NADW – North Atlantic Deep Water; MW – Mode Water. The shaded regions mark the main regions of stratification and the Drake Passage. The real ocean is more complex; see text.

16.1–16.4 and this gives us an overturning circulation as illustrated in Fig. 16.17, with plausible identification of the observed main water masses as indicated. For illustrative clarity we have separated the various dynamical regimes, but the real ocean's overturning circulation defies such simple characterizations. The dynamically different regimes merge into one another via complex transition zones — for example, the outcropping region of the southern hemisphere internal thermocline merges with the ACC, and the deep stratification provided by the difference in density between North Atlantic Deep Water and Antarctic Bottom Water is not nearly as vertically confined as indicated, and topography also plays a role. The ACC itself is a complex beast, for the nearly vertical isopycnals of the model above will be highly baroclinically unstable, and this provides a segue into our next topic.

16.6 † FLOW IN A CHANNEL AND THE ANTARCTIC CIRCUMPOLAR CURRENT

We now take a closer look at the Antarctic Circumpolar Current (ACC) itself, with less of a focus on how the ACC connects the rest of the world's ocean and more of a focus on its own internal dynamics. This current system, sketched in Fig. 16.18, differs from other oceanic regimes primarily in that the flow is, like that of the atmosphere, predominantly zonal and re-entrant. The two obvious influences on the circulation are the strong, eastward winds (the 'roaring forties' and the 'furious fifties') and the buoyancy forcing associated with the meridional gradient of atmospheric temperature and radiative effects that cause ocean cooling at high latitudes and warming at low ones. Providing a detailed description of the resulting flow is properly the province of numerical models, and here our goals are

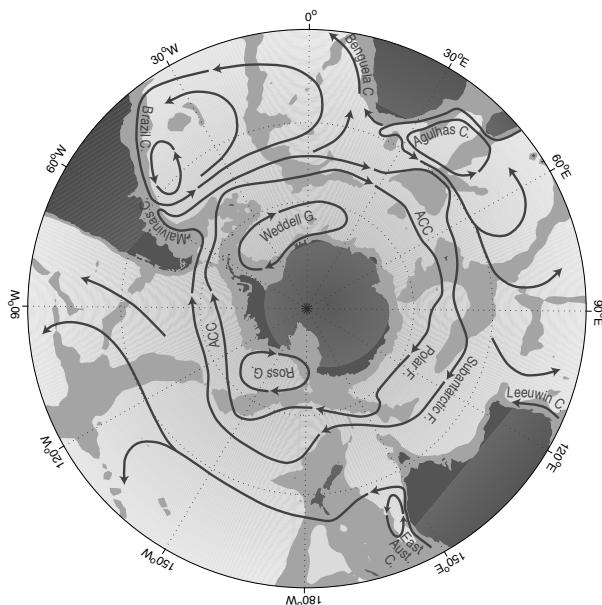


Fig. 16.18 Schema of the major currents in the Southern Ocean. Shown are the South Atlantic subtropical gyre, and the two main cores of the ACC, associated with the Polar front and the sub-Antarctic front.¹⁵

much more modest, namely to describe and understand some of the fundamental dynamical mechanisms that determine the structure and transport of the system.¹⁶

16.6.1 Steady and eddying flow

Consider again the simplified geometry of the Southern Ocean as sketched in Fig. 16.13. The ocean floor is flat, except for a ridge (or 'sill') at the same longitude as the gyre walls; this is a crude representation of the topography across the Drake Passage, that part of the ACC between the tip of South America and the Antarctic Peninsula. In the planetary-geostrophic approximation, the steady response is that of nearly vertical isopycnals in the area above the sill, as illustrated in Fig. 16.15. Below the sill a meridional flow can be supported and the isotherms spread polewards, as illustrated in numerical solutions using the primitive equations (Fig. 16.19).¹⁷

The stratification of the non-eddying simulation is similar to that predicted by the idealized model illustrated in Fig. 16.15. However, the steep isotherms within the channel contain a huge amount of available potential energy (APE), and the flow is highly baroclinically unstable. If baroclinic eddies are allowed to form, the solution is dramatically different: the isotherms slump, releasing that APE and generating mesoscale eddies that exercise control over much of the circulation. An important conclusion is that *baroclinic eddies are of leading-order importance in the dynamics of the ACC*. A dynamical description of the ACC without eddies would be *qualitatively* in error, in much the same way as would a similar description of the mid-latitude troposphere (i.e., the Ferrel Cell). (However, it is less clear whether these eddies are important in the interaction of the ACC with the rest of the world's oceans.) These eddies transfer both heat and momentum, and much of the rest of our description will focus on their effects.

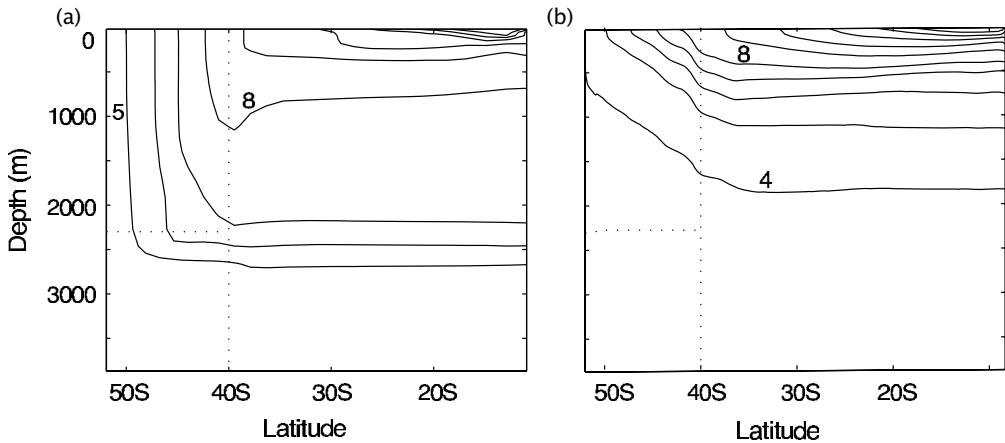


Fig. 16.19 The zonally averaged temperature field in numerical solutions of the primitive equations in a domain similar to that of Fig. 16.13 (except that here the channel and sill are nestled against the poleward boundary). Panel (a) shows the steady solution of a diffusive model with no baroclinic eddies, and (b) shows the time averaged solution in a higher-resolution model that allows baroclinic eddies to develop. Two contour values in each panel are labelled. The dotted lines show the channel boundaries and the sill.¹⁸

16.6.2 Vertically integrated momentum balance

The momentum supplied by the strong eastward winds must somehow be removed. Presuming that lateral transfers of momentum are small the momentum must be removed by fluid contact with the solid Earth at the bottom of the channel. Thus, let us first consider the vertically integrated momentum balance in a channel, without regard to how the momentum might be vertically transferred. We begin with the frictional-geostrophic balance, namely

$$\mathbf{f} \times \mathbf{u} = -\nabla \phi + \frac{\partial \tilde{\tau}}{\partial z}, \quad (16.80)$$

where $\tilde{\tau}$ is the kinematic stress, τ/ρ_0 , and $\phi = p/\rho_0$. Integrating over the depth of the ocean and using Leibnitz's rule [cf. (14.95)] gives

$$\mathbf{f} \times \hat{\mathbf{u}} = -\nabla \hat{\phi} - \phi_b \nabla \eta_b + \tilde{\tau}_w - \tilde{\tau}_f, \quad (16.81)$$

where $\tilde{\tau}_w$ is the stress at the surface (due mainly to the wind) and $\tilde{\tau}_f$ is the frictional stress at the bottom, a hat denotes a vertical integral and ϕ_b is the pressure at $z = \eta_b$, where η_b is the z -coordinate of the bottom topography. The x -component of (16.81) is

$$-f \hat{v} = -\frac{\partial \hat{\phi}}{\partial x} - \phi_b \frac{\partial \eta_b}{\partial x} + \tilde{\tau}_w^x - \tilde{\tau}_f^x, \quad (16.82)$$

and on integrating around a line of latitude the term on the left-hand side vanishes by mass conservation and we are left with

$$-\overline{\phi_b \frac{\partial \eta_b}{\partial x}} + \overline{\tilde{\tau}_w^x} - \overline{\tilde{\tau}_f^x} = 0, \quad (16.83)$$

where overbars denote zonal averages. The first term is the bottom, or topographic, form drag, encountered in sections 3.5 and 14.6, and observations and numerical simulations indicate that it is this, rather than the frictional term $\tilde{\tau}_f^x$, that predominantly balances the wind stress.¹⁹ We address the question of *why* this should be so in section 16.6.5.

The vorticity balance is similarly dominated by a balance between bottom form-stress curl and wind-stress curl. Taking the curl of (16.81), noting that $\nabla \cdot \hat{\mathbf{u}} = 0$, gives

$$\beta \hat{v} = -\mathbf{k} \cdot \nabla \phi_b \times \nabla \eta_b + \text{curl}_z \tilde{\mathbf{\tau}}_w - \text{curl}_z \tilde{\mathbf{\tau}}_f. \quad (16.84)$$

Now, on integrating over an area bounded by two latitude circles and applying Stokes' theorem the β -term vanishes by mass conservation and we regain (16.83). This means that Sverdrup balance, in the usual sense of $\beta v \approx \text{curl}_z \mathbf{\tau}_w$, cannot hold in the zonal average: the left-hand side vanishes but the right-hand side does not. The same could be said for the zonal integral of (16.84) across a gyre, but the two cases do differ: In a gyre Sverdrup balance can (in principle) hold over most of the interior, with mass balance being satisfied by the presence of an intense western boundary current. In contrast, in a channel where the dynamics are zonally homogeneous then v must be, on average, zero at *all* longitudes and form drag and/or frictional terms must balance the wind-stress curl in a given water column. Sverdrup balance is thus a less useful foundation for channel dynamics — at least zonally homogeneous ones — than it is for gyres. Of course, the real ACC is *not* zonally homogeneous, and may contain regions of poleward Sverdrup flow balanced by equatorward flow in boundary currents along the eastern edges of sills and continents, and the extent to which Sverdrup flow is a leading-order descriptor of its dynamics is partly a matter of geography.

We cannot in general completely neglect non-conservative frictional terms, for two reasons. First, they are the means whereby kinetic energy is dissipated. Second, if there is a contour of constant orographic height encircling the domain (i.e., encircling Antarctica) then the form drag will vanish when integrated along it. However, the same integral of the wind stress will not vanish, and therefore must be balanced by something else. To see this explicitly, write the vertically integrated vorticity equation, (16.84), in the form

$$\beta \hat{v} + J(\phi_b, \eta_b) = \text{curl}_z \tilde{\mathbf{\tau}}_w - \text{curl}_z \tilde{\mathbf{\tau}}_f. \quad (16.85)$$

If we integrate over an area bounded by a contour of constant orographic height (i.e., constant η_b) then both terms on the left-hand side vanish, and the wind stress along that line must be balanced by friction. In the real ocean there may be no such contour that is confined to the ACC — rather, any such contour would meander through the rest of the ocean; indeed, no such confined contour exists in the idealized geometry of Fig. 16.13.

16.6.3 Form drag and baroclinic eddies

How does the momentum put in at the surface by the wind stress make its way to the bottom of the ocean where it may be removed by form drag? We saw in section 16.5.1 that one mechanism is by way of a mean meridional overturning circulation, with an upper branch in the Ferrel Cell and a lower branch at the level of the sill, with no meridional flow between. However, the presence of baroclinic eddies changes things in two related ways:

- (i) eddy form drag can pass momentum vertically within the fluid;

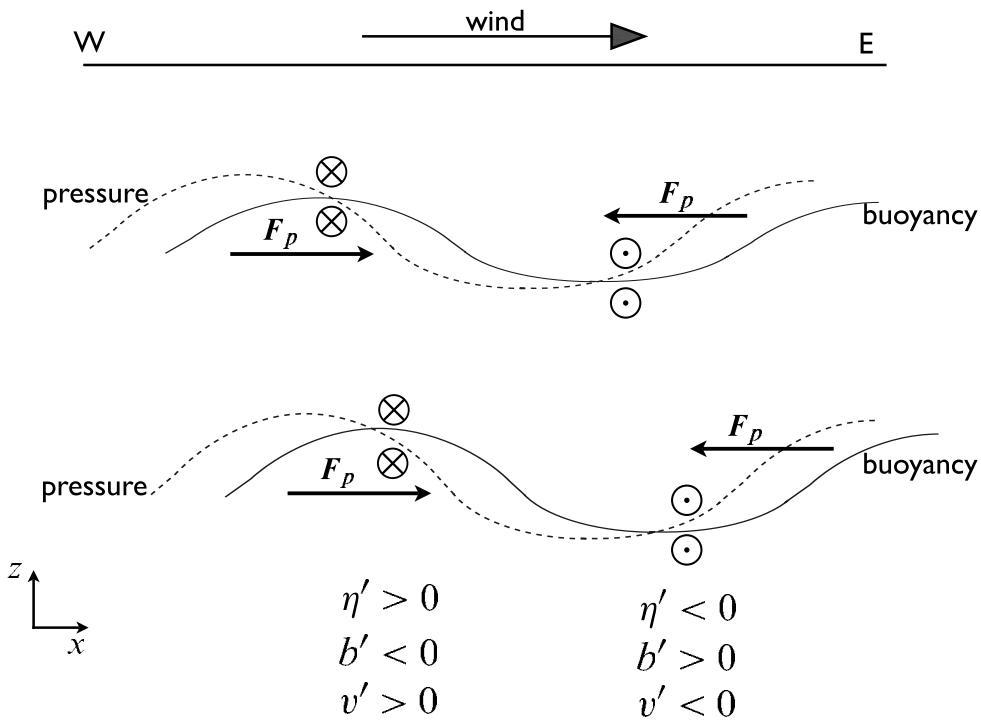


Fig. 16.20 Eddy fluxes and form drag in a Southern Hemisphere channel, viewed from the south. In this example, cold (less buoyant) water flows equatorwards and warm water polewards, so that $\overline{v' b'} < 0$. The pressure field associated with this flow (dashed lines) provides a form drag on the successive layers, F_p , shown. At the ocean bottom the westward form drag on the fluid arising through its interaction with the orography of the sea-floor is equal and opposite to that of the eastward wind stress at the top. The mass fluxes in each layer are given by $\overline{v' h'} \approx -\partial_z(\overline{v' b'}/N^2)$. If the magnitude of buoyancy displacement increases with depth then $\overline{v' h'} < 0$.

- (ii) eddies can allow a net meridional thickness flux, or equivalently a mass flux in a given density class.

Momentum dynamics of layers

Let us model the channel as a finite number of fluid layers, each of constant density and lying one on top of the other — a ‘stacked shallow water’ model, and one equivalent to a model expressed in isopycnal coordinates. The wind provides a stress on the upper layer, which sets it into motion, and this in turn, via the mechanism of form drag, provides a stress to the layer below, and so on until the bottom is reached. The lowest layer then equilibrates via form drag with the bottom topography or via Ekman friction, and the general mechanism is illustrated in Fig. 16.20.

Recalling the results of section 3.5, the zonal form drag at a layer interface is given by

$$\tau_i = -\eta_i \overline{\frac{\partial p_i}{\partial x}} = -\rho_0 f \overline{\eta_i v_i}, \quad (16.86)$$

where p_i is the pressure and η_i is the displacement at the i th interface (i.e., between the i th and $(i+1)$ th layer as in Fig. 16.21), and the overbar denotes a zonal average. If we define the averaged meridional transport in each layer by

$$V_i = \int_{\eta_i}^{\eta_{i-1}} \rho_0 v \, dz, \quad (16.87)$$

then, neglecting the meridional momentum flux divergence (for reasons given in the next subsection), the time and zonally averaged zonal momentum balance for each layer of fluid are:

$$-f\bar{V}_1 = \tau_w - \tau_1 = \eta_1 \overline{\frac{\partial p_1}{\partial x}} + \tau_w, \quad (16.88a)$$

$$-f\bar{V}_i = \tau_{i-1} - \tau_i = -\eta_{i-1} \overline{\frac{\partial p_{i-1}}{\partial x}} + \eta_i \overline{\frac{\partial p_i}{\partial x}}, \quad (16.88b)$$

$$-f\bar{V}_N = \tau_{N-1} - \tau_N = -\eta_{N-1} \overline{\frac{\partial p_{N-1}}{\partial x}} + \eta_b \overline{\frac{\partial p_b}{\partial x}} - \tau_f, \quad (16.88c)$$

where the subscripts 1, i and N refer to the top layer, an interior layer, and the bottom layer, respectively. Also, η_b is the height of bottom topography and τ_w is the zonal stress imparted by the wind which, we assume, is confined to the uppermost layer. The term τ_f represents drag at the bottom due to Ekman friction, but we have neglected any other viscous terms or friction between the layers.

The vertically integrated meridional mass transport must vanish, and thus summing over all the layers (16.88) becomes

$$0 = \tau_w - \tau_f - \tau_N, \quad (16.89)$$

or, noting that $\tau_N = -\overline{\eta_b \partial p_b / \partial x}$,

$$\boxed{\tau_w = \tau_f - \eta_b \overline{\frac{\partial p_b}{\partial x}}}. \quad (16.90)$$

Thus, the stress imparted by the wind (τ_w) may be communicated vertically through the fluid by form drag, and ultimately balanced by the sum of the bottom form stress (τ_N) and the bottom friction (τ_f).

Momentum dynamics in height coordinates

We now look at these same dynamics in height coordinates, using the quasi-geostrophic TEM formalism, and it may be helpful to review section 7.3 before proceeding. As in (7.66), we write the zonally averaged momentum equation in the form

$$-f_0 \bar{v}^* = \nabla_m \cdot \mathcal{F} + \frac{\partial \tilde{\tau}}{\partial z} \quad (16.91)$$

where $\bar{v}^* = \bar{v} - \partial_z (\overline{v' b'}) / \bar{b}_z$ is the residual meridional velocity, $\tilde{\tau}$ is the zonal component of the kinematic stress (wind-induced and frictional, and typically important only in an Ekman layer at the surface and in a frictional layer at the bottom) and \mathcal{F} is the Eliassen-Palm flux, which satisfies

$$\nabla_m \cdot \mathcal{F} = -\frac{\partial}{\partial y} \overline{u' v'} + \frac{\partial}{\partial z} \left(\frac{f_0}{N^2} \overline{v' b'} \right) = \overline{v' q'}. \quad (16.92)$$

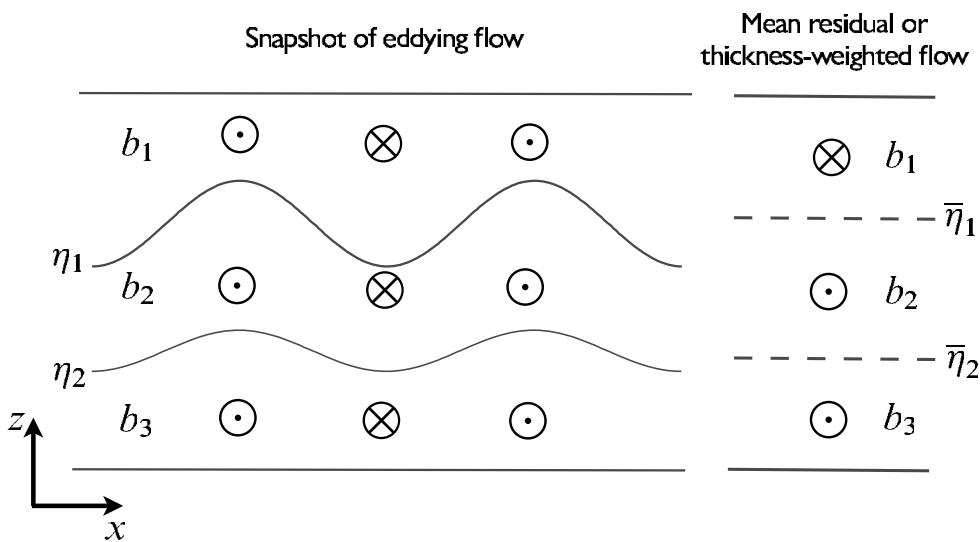


Fig. 16.21 An example of the meridional flow in an eddying channel. The eddying flow may be organized such that, even though at any given level the Eulerian meridional flow may be small, there is a net flow in a given isopycnal layer. The residual (\bar{v}^*) and Eulerian (\bar{v}) flows are related by $\bar{v}^* = \bar{v} + \bar{v}'h'/\bar{h}$; thus, the thickness-weighted average of the eddying flow on the left gives rise to the residual flow on the right, where $\bar{\eta}_i$ denotes the mean elevation of the isopycnal interface η_i .

Now, if the horizontal velocity and buoyancy perturbations are related by $v' \sim b'/N$ (meaning roughly similar available potential energy and kinetic energy, see also section 9.4), then the two terms comprising the potential vorticity flux scale as

$$\frac{\partial}{\partial y} \overline{u'v'} \sim \frac{v'^2}{L_e}, \quad \frac{\partial}{\partial z} \left(f_0 \frac{\overline{v' b'}}{\overline{b}_z} \right) \sim \frac{v'^2}{L_d}, \quad (16.93)$$

where L_e is the scale of the eddies and L_d is the deformation radius. If the former is much larger than the latter, as we might expect in a field of developed geostrophic turbulence (and as is observed in the ACC), then the potential vorticity flux is dominated by the buoyancy flux [so also justifying the neglect of the lateral momentum fluxes in (16.88)] and (16.91) becomes

$$-f_0 \bar{v}^* \approx \frac{\partial \tilde{\tau}}{\partial z} + \frac{\partial}{\partial z} \left(f_0 \frac{\bar{v}' \bar{b}'}{\bar{b}_z} \right). \quad (16.94)$$

In the ocean interior the frictional terms, $\partial \tilde{\tau} / \partial z$, are small, and (16.94) represents a balance between the Coriolis force on the residual flow and the form stress associated with the vertical component of the EP flux (an association explained in section 7.4.3).

If we integrate (16.94) equation over the depth of the channel the term on the left-hand side vanishes and we have

$$\tilde{\tau}_w = \tilde{\tau}_f - \left[f_0 \frac{v' b'}{b_z} \right]_u^0, \quad (16.95)$$

where $\tilde{\tau}_w$ is the wind stress and $\tilde{\tau}_f$ is the frictional stress at the bottom (both divided by

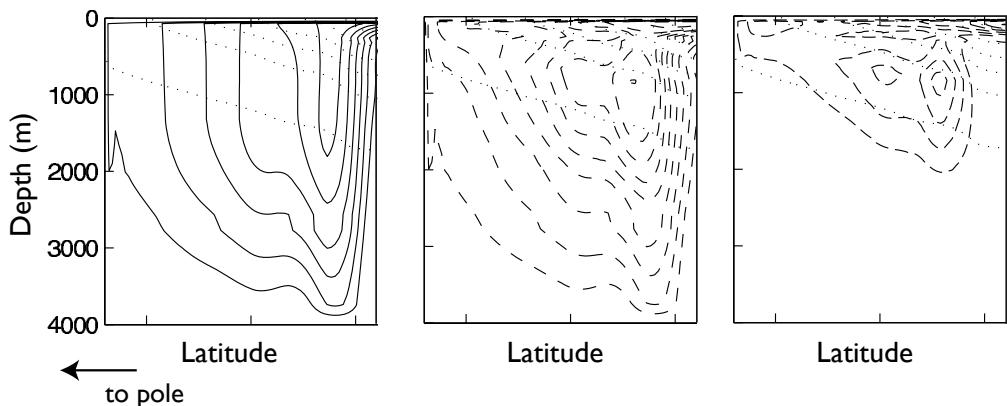


Fig. 16.22 The meridional circulation in the re-entrant channel of an idealized, eddying numerical model of the ACC (as in Fig. 16.19, but showing only the region south of 40°S). Left panel, the zonally averaged Eulerian circulation. Middle panel, the eddy-induced circulation. Right panel, the residual circulation. Solid lines represent a clockwise circulation and dashed lines represent anticlockwise circulation. The faint dotted lines are the mean isopycnals. The overturning circulation of the ACC is called the Deacon Cell.

ρ_0). Equation (16.95) expresses essentially the same momentum balance as (16.90). Thus, the EP flux expresses the passage of momentum vertically through the water column, the momentum being removed at the bottom through frictional stresses and/or form drag with the orography.

Mass fluxes and thermodynamics

Associated with the form drag is a meridional mass flux in each layer, which in the layered model appears as V_i (a thickness flux) in each layer. The satisfaction of the momentum balance at a particular latitude goes hand-in-hand with the satisfaction of the mass balance. Above any topography the Eulerian mean momentum equation is, with quasi-geostrophic scaling and neglecting eddy momentum fluxes,

$$f_0 \bar{v} = \frac{\partial \tilde{\tau}}{\partial z} \quad (16.96)$$

where \bar{v} is the zonally averaged meridional velocity and $\tilde{\tau}$ is the zonal component of the kinematic stress. The zonally averaged meridional flow is ageostrophic and is non-zero only near the surface (i.e., equatorward Ekman flow) and near the ocean bottom, where the flow can be supported by friction and/or form drag. Even in an eddying flow, the Eulerian circulation is primarily confined to the upper Ekman layer and a frictional or topographically interrupted layer at the bottom, as illustrated in Fig. 16.22. This is a perfectly acceptable description of the flow, and is not an artifact in any way. However, and analogously to the atmospheric Ferrel Cell (sections 11.7 and 12.2.2), if the flow is unsteady this circulation does not necessarily represent the flow of water parcels, nor does it imply that water parcels cross isopycnals, as might be suggested by the leftmost panel of Fig. 16.22. The flow of

parcels is better represented by the residual flow, or the thickness-weighted flow, and as sketched in Fig. 16.21 there can be a net meridional flow in a given layer (i.e., of a given water mass type) *even when the net meridional Eulerian flow at the level of mean height of the layer is zero*.

[As an aside, we note that rather than being a mass flux, the residual flow is less ambiguously considered as a thickness flux (or equivalently a buoyancy flux) because in an incompressible fluid, and so in the Boussinesq approximation, the eddies do not carry net mass, *per se*: mass flux is a volume flux and $\bar{v}' = 0$. In any given isopycnal layer the residual circulation does represent a flux of mass of fluid of a particular buoyancy class, and this will be non-zero if there is a correlation between the thickness of the layer and the velocity (or between velocity and buoyancy) so that $\bar{v}'\bar{h}' \neq 0$ and $\bar{v}'\bar{b}' \neq 0$.]

The vertically integrated residual mass flux must vanish, and even though one component of this — the equatorward Ekman flow — is determined mechanically, the overall sense of the residual circulation cannot in general be determined by the momentum balance alone: thermodynamic effects play a role. As in (7.101), the zonally averaged thermodynamic equation may be written in TEM form as

$$\frac{\partial \bar{b}}{\partial t} + J(\psi^*, \bar{b}) = Q[b] \quad (16.97)$$

where $J(\psi^*, \bar{b}) = (\partial_y \psi^*)(\partial_z b) - (\partial_z \psi^*)(\partial_y b) = \bar{v}^* \partial_y b + \bar{w}^* \partial_z b$, ψ^* is the streamfunction of the residual flow and $Q[b]$ represents heating and cooling, which occurs mainly at the surface. In the ocean interior and in a statistically steady state we therefore have

$$J(\psi^*, \bar{b}) = 0, \quad (16.98)$$

the general solution of which is $\psi^* = G(\bar{b})$, where G is an arbitrary function. That is, the interior residual flow is along isopycnals, and this is approximately satisfied by the numerical solution shown in the rightmost panel of Fig. 16.22. At the surface, however, the flow is generally not adiabatic, because of heat exchange with the atmosphere, and so the residual flow can be across isopycnals. At the edge of the channel, the flow is also diapycnal. Note that the sense of the subsurface circulation determines how the form drag varies with depth; if the residual flow were zero, for example, then, either from (16.88) or from (16.94), we see that the form drag must be constant with depth.

16.6.4 † An idealized adiabatic model

We finally consider a simple model of the ACC that, although very idealized, provides a starting point for understanding more complete numerical models and the system itself.²⁰ The simplifying assumption we make is that the flow is adiabatic everywhere; it then follows that the net overturning, as given by the residual circulation, is zero. We can see this by first noting that in a statistically steady state the flow satisfies (16.98), implying that the residual flow is along isopycnals. However, if there is a meridional buoyancy gradient at the surface there can be no surface residual flow (because this would be cross-isopycnal); it then follows that there can be no net flow along isopycnals in the interior, because if these outcrop there would be a net fluid source, and hence diapycnal flow, at the surface. This idealized limit has thus led to the ‘vanishing of the Deacon Cell’. In reality the flow is not

completely adiabatic and the residual flow will not vanish, but it is likely to be weaker than either the Eulerian or the eddy-induced flow (as in Fig. 16.22).

The zonal momentum equation in this limit follows from (16.94), which with $\bar{v}^* = 0$ gives

$$\frac{\partial \tilde{\tau}}{\partial z} \approx -\frac{\partial}{\partial z} \left(f_0 \frac{\bar{v}' b'}{b_z} \right). \quad (16.99)$$

The equivalent balance for the Eulerian flow is, using the definition of \bar{v}^* ,

$$-f_0 \bar{v} = \frac{\partial \tilde{\tau}}{\partial z} \quad \rightarrow \quad f_0 \bar{v} = f_0 \frac{\partial}{\partial z} \left(\frac{\bar{v}' b'}{b_z} \right). \quad (16.100a,b)$$

These equations represent *dynamical* balances; they do not follow from the momentum equation without making addition assumptions, in this case that $\bar{v}^* = 0$. In the residual equation, (16.99), the wind stress is balanced by the divergence of the Eliassen-Palm flux, which is dominated by the contribution from the buoyancy flux, and in the ocean interior where the stress is small the meridional buoyancy flux will be constant with height. Equation (16.100b) represents a balance between the Coriolis force on the equatorward flow and form drag. If the frictional stress is small in the interior then the form drag does not vary in the vertical and the right-hand side of (16.100) is small. The zonally averaged meridional flow in the interior is then also small, and the equatorial flow in the top Ekman layer is balanced by a return flow at the bottom of the ocean involving topographic form stress or a bottom Ekman layer.

Integrating (16.99) from the surface (where $\tilde{\tau} = \tilde{\tau}_w$) to a stress-free level in the interior (where $\tilde{\tau} = 0$) gives

$$\tilde{\tau}_w = f_0 \frac{\bar{v}' b'}{b_z}, \quad (16.101)$$

if the buoyancy flux at the surface is small. If we are now willing to parameterize the eddy fluxes in terms of the mean flow, then we can predict the stratification. Thus, let $\bar{v}' b' = -\kappa \partial \bar{b} / \partial y$, where κ is an eddy diffusivity, and noting that $s = -\bar{b}_y / \bar{b}_z$ is the slope of the isopycnals, we find

$$\tilde{\tau}_w = \kappa f_0 s = \kappa \frac{f_0^2}{b_z} \frac{\partial \bar{u}}{\partial z}, \quad (16.102)$$

where the second equality uses thermal wind balance. Thus, given κ , we can predict the isopycnal slope [$s = \tilde{\tau}_w / (\kappa f_0)$] and, potentially, the total baroclinic transport of the ACC as a function of the wind stress. Further progress then depends on making a specific choice for κ , as described for example in chapter 10, but our reach has once more exceeded our grasp. In a more realistic diabatic model, the sense of the residual circulation can be inferred if the diabatic fluxes at the surface are known, but at the same time these fluxes depend in a complicated way on both the lateral eddy fluxes and the general circulation itself.

16.6.5 Form stress and Ekman stress at the ocean bottom

Earlier, we noted that the stress at the ocean bottom is observed to be dominated by form stress, rather than Ekman friction, in the ACC. A simple scaling argument helps understand why this should be. The form stress scales like

$$\tau_{form} \sim \eta_b \frac{\partial p_b}{\partial x} \sim \eta_b \rho_0 U f, \quad (16.103)$$

where we have used geostrophic balance and U is a scaling for the horizontal velocity. The frictional stress due to an Ekman layer (section 2.12) scales like

$$\tau_{Ekman} \sim \rho_0 A \frac{\partial u}{\partial z} \sim \frac{\rho_0 A U}{d} \sim \rho_0 d U f \quad (16.104)$$

where A is the eddy kinematic viscosity and $d = \sqrt{A/f}$ is the Ekman layer thickness. The ratio of these two stresses thus scales as

$$\frac{\tau_{form}}{\tau_{Ekman}} \sim \frac{\eta_b}{d}. \quad (16.105)$$

We therefore expect the form stress to dominate the Ekman stress if the variations in topography are greater than the Ekman layer thickness, and if the flow goes *over* the topography rather than around it. In the ACC the topography is hundreds or even thousands of metres high whereas the bottom Ekman layer may be of order tens of metres, and furthermore the predominantly eastward flow must (unlike the situation in gyre circulations) go *over* the topography. Thus, form stress dominates the frictional, Ekman layer, stress at the bottom of the ACC.

16.6.6 Differences between gyres and channels

In the dynamics of the ACC, the wind stress itself seems to play an important role, whereas in our discussion of gyres in chapter 14 the wind stress *curl* was dominant. What is the root of this difference?²¹ Suppose that we change the wind stress, but not its curl, in a closed basin. The vertically integrated gyral flow, as given for example by the Stommel solution (14.39) or its two-gyre counterpart, does not change at all. However, the vertical structure of this flow will in general change; for example, if the wind is made uniformly more eastward, there will be a corresponding increase in the equatorward flux in the Ekman layer that must return polewards at depth (assuming that the western boundary current balances only the Sverdrup flow). At the same time, the added force from the wind must be balanced by an increased pressure difference between the western and eastern boundaries. This may be achieved if the sea-surface tilts upwards to the east, so producing a net (vertically integrated) poleward geostrophic flow. The subsurface isopycnal slopes may then adjust in order to reduce this flow to near zero in the abyss. The added force provided by the basin walls on the fluid in the basin is a kind of form drag (rather like the force provided by the sill in section 16.5), and integrated around the basin this force must be equal and opposite to the force supplied by the wind. In contrast, in a channel adding a constant wind produces a direct change in its zonal transport. This is because the wind stress is balanced by form drag and bottom friction (with the former probably dominating), and both of these, in practice, depend on the zonal flow at the channel bottom.

APPENDIX: MISCELLANEOUS RELATIONSHIPS IN A LAYERED MODEL

Here we collect various expressions relating pressure, density and velocity in a geostrophic and Boussinesq layered model. The layers and the interfaces are numbered, increasing downwards, as in Fig. 16.23, and the bottom layer is stationary.

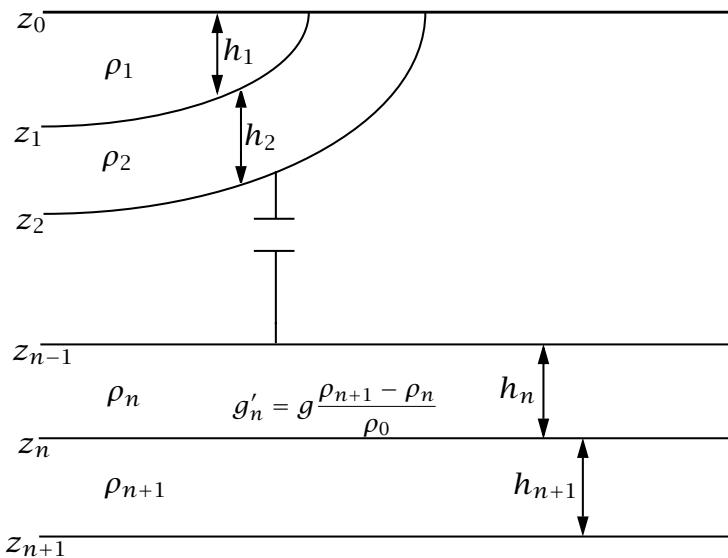


Fig. 16.23 Structure and notational conventions used for a multi-layered model.

16.A.1 Hydrostatic balance

Hydrostatic balance is $\partial p / \partial z = -\rho g$. We can integrate this to give, in layers n and $n-1$,

$$p_n = -\rho_n g z + p'_n(x, y), \quad p_{n-1} = -\rho_{n-1} g z + p'_{n-1}(x, y). \quad (16.106)$$

Since pressure is continuous, at $z = z_{n-1}$ these two expressions are equal so that

$$-\rho_n g z_{n-1} + p'_n = p_{n-1} = -\rho_{n-1} g z_{n-1} + p'_{n-1} \quad (16.107)$$

whence

$$g'_{n-1} z_{n-1} = \frac{1}{\rho_0} (p'_n - p'_{n-1}) \quad (16.108)$$

where $g'_n = g(\rho_{n+1} - \rho_n)/\rho_0$ is the *reduced gravity*, and ρ_0 is the constant, reference, value of the density used in the Boussinesq approximation, taken to be equal to ρ_1 .

16.A.2 Geostrophic and thermal wind balance

In the Boussinesq approximation, geostrophic balance is:

$$\rho_0 f \mathbf{u}_n = \mathbf{k} \times \nabla p_n. \quad (16.109)$$

Using (16.109) with (16.108) gives

$$\mathbf{u}_{n+1} - \mathbf{u}_n = \frac{g'_n}{f} \mathbf{k} \times \nabla z_n, \quad (16.110)$$

which is the appropriate form of thermal wind balance for this system. Let us suppose that at sufficient depth there is no motion, and in particular that layer $N + 1$ is stationary and contains no pressure gradients. That is, $p'_{N+1} = 0$ and so, from (16.107) and (16.109)

$$p'_N = -(\rho_{N+1} - \rho_N)g z_N = -g'_N \rho_0 z_N. \quad (16.111)$$

Integrating upwards we obtain the pressure in each layer,

$$p'_n = -\rho_0 \sum_{i=n}^{i=N} g'_i z_i \quad (16.112)$$

where $n \leq N$. Thus, the geostrophic velocities in each layer are given by

$$f \mathbf{u}_n = -\mathbf{k} \times \nabla \left(\sum_{i=n}^{i=N} g'_i z_i \right). \quad (16.113)$$

The quantity in brackets on the right-hand side is not a velocity streamfunction because it is $f \mathbf{u}_n$, and not \mathbf{u}_n that is given by its curl. Nevertheless, the velocity is normal to its gradient, and therefore its isolines define streamlines.

The upper surface of the ocean is assumed to be fixed; this is the 'rigid-lid' approximation. Thus, $z_0 = 0$ and $h_1 = -z_1$. More generally, the layer thicknesses and the interfaces between the layers are related by

$$z_n = - \sum_{i=M}^{i=n} h_i \quad (16.114)$$

where M is the index of the uppermost layer, and $n \geq M$. If there is no outcropping, then $M = 1$.

The geostrophic velocity in the lowest moving layer is given by

$$f \mathbf{u}_N = -g'_N \mathbf{k} \times \nabla z_N = g'_N \mathbf{k} \times \nabla h. \quad (16.115)$$

This means that lines of constant depth of the lowest layer are also streamlines; the velocity moves parallel to the depth contours. The vertical velocity at the base of the lowest layer is given by, for steady flow,

$$w(z = -h) = \mathbf{u}_N \cdot \nabla h = \frac{1}{f} g'_N (\mathbf{k} \times \nabla h) \cdot \nabla h = 0. \quad (16.116)$$

That is, there is no vertical motion at the base of the moving layers.

16.A.3 Explicit cases

A one-layer reduced-gravity model

The perturbation pressure in the moving layer (layer 1) is

$$p'_1 = -\rho_0 g'_1 z_1 = \rho_0 g'_1 h_1. \quad (16.117)$$

The geostrophic velocity is given by

$$f \mathbf{u}_1 = \frac{1}{\rho_0} \mathbf{k} \times \nabla p'_1 = g'_1 \mathbf{k} \times \nabla h_1. \quad (16.118)$$

(In a single-layer model, the subscripts are often omitted.)

A two-layer model

The perturbation pressure in the upper and lower moving layers are given by

$$p'_1 = -\rho_0(g'_2 z_2 + g'_1 z_1) = \rho_0(g'_2 h + g'_1 h_1), \quad (16.119a)$$

$$p'_2 = -\rho_0 g'_2 z_2 = \rho_0 g'_2 (h_1 + h_2) = \rho_0 g'_2 h, \quad (16.119b)$$

[cf. (16.111)], where $h = h_1 + h_2 = -z_2$.

The corresponding geostrophic velocities are

$$f\mathbf{u}_1 = \mathbf{k} \times \nabla(g'_2 h + g'_1 h_1), \quad (16.120a)$$

$$f\mathbf{u}_2 = \mathbf{k} \times \nabla(g'_2 h). \quad (16.120b)$$

A three-layer model

The perturbation pressures in the three moving layers are

$$p_1 = -\rho_0[g'_3 z_3 + g'_2 z_2 + g'_1 z_1] = \rho_0[g'_3 h + g'_2 (h_1 + h_2) + g'_1 h_1], \quad (16.121a)$$

$$p_2 = -\rho_0[g'_2 z_2 + g'_3 z_3] = \rho_0[g'_2 (h_1 + h_2) + g'_3 h], \quad (16.121b)$$

$$p_3 = -\rho_0 g'_3 z_3 = \rho_0 g'_3 h, \quad (16.121c)$$

where $h = h_1 + h_2 + h_3 = -z_3$. The corresponding geostrophic velocities are:

$$f\mathbf{u}_1 = \mathbf{k} \times \nabla[g'_3 h + g'_2 (h_1 + h_2) + g'_1 h_1] \quad (16.122a)$$

$$f\mathbf{u}_2 = \mathbf{k} \times \nabla[g'_2 (h_1 + h_2) + g'_3 h] \quad (16.122b)$$

$$f\mathbf{u}_3 = \mathbf{k} \times \nabla[g'_3 h]. \quad (16.122c)$$

Notes

- 1 Sections courtesy of L. Talley.
- 2 See, among others, Toole *et al.* (1994), Polzin *et al.* (1997), Gregg (1998) and Ledwell *et al.* (1998).
- 3 For estimates of the strength of the overturning circulation in the ocean, and its relation to diapycnal diffusivity and the observed stratification, see Munk (1966), revisited by Munk & Wunsch (1998). See also Huang (1999) and, for a review, Wunsch & Ferrari (2004). If the abyssal flow is along rather than across isopycnals, smaller values of diffusivity suffice to maintain deep stratification. See section 16.5.
- 4 Vallis (2000).
- 5 The modern development of the theory of the thermocline began with two back-to-back papers in 1959 in the journal *Tellus*. Welander (1959) suggested an adiabatic model, based on the ideal-fluid thermocline equations (i.e., the planetary-geostrophic equations, with no diffusion terms in the buoyancy equation), whereas Robinson & Stommel (1959) proposed a model that is intrinsically diffusive. In this model [developed further by Stommel & Webster (1963), Salmon (1990), and others] the thermocline is an internal boundary layer or front that forms at the convergence of two different homogeneous water types, warm surface fluid above and cold abyssal fluid below. Meanwhile, the adiabatic model continued its own development (see Veronis 1969), culminating in the ventilated thermocline model of Luyten

et al. (1983) and its continuous extensions (Killworth 1987, Huang 1988). Signs that the two classes of theory might not be wholly incompatible came from Welander (1971b) and Colin de Verdière (1989) who noted that the diffusion might become important below an adiabatic near-surface flow, and Samelson & Vallis (1997) eventually suggested a model in which the upper thermocline is adiabatic, as in the ventilated thermocline model, but has a diffusive base, constituting an internal boundary layer. The role of mesoscale eddies in all of this is still unclear; however, numerical simulations suggest they do play a role in determining the structure of the thermocline, at least in its diffusive base and perhaps everywhere (e.g. Henning & Vallis 2004, Cessi & Fantini 2004). The near-certainty that some of the deep overturning circulation is wind-driven, and inter-hemispheric, will also affect the dynamics of the diffusive layer but will not obviate the need for it.

- 6 Welander (1971a).
- 7 Drawing from Salmon (1990).
- 8 Newton's method is an iterative way to numerically solve certain types of differential equations. The solutions here are obtained using about 1000 uniformly spaced grid points to span the domain, taking just a few seconds of computer time. Because of the boundary layer structure of the solutions employing a non-uniform grid would be even more efficient for this problem, but there is little point in designing a streamlined hat to reduce the effort of walking.
- 9 Following Samelson (1999b).
- 10 Adapted from Samelson & Vallis (1997).
- 11 Following Luyten *et al.* (1983).
- 12 It may be that the eastern boundary depth is determined by global thermodynamic and/or mass constraints; see, for example, Boccaletti *et al.* (2004). There must also be a poleward transport across the boundary of the subtropical–subpolar gyre to balance the equatorward transport in the Ekman layer and that of the meridional overturning circulation, and this balance requirement may influence the boundary depth.
- 13 Dewar *et al.* (2005). This paper also discusses the nature of discontinuities at the pool boundaries, and their treatment via shock conditions.
In the shadow zone, layer-2 fluid also has no direct surface source, so one may wonder why this is also not expunged. However, the shadow zone is not a recirculating regime and the Ekman induced displacement will be much less efficient. More directly, the eastern boundary condition $h_1(x = x_e) = 0$ precludes the vanishing of layer-2 water there, and this boundary condition is propagated westwards into the interior.
- 14 Drawing from the numerical and conceptual models of Toggweiler & Samuels (1995, 1998), Döös & Coward (1997), Gnanadesikan (1999), Vallis (2000), Webb & Sugino (2001), Nof (2003), and the analytic model of Samelson (1999a, 2004). The notion of a deep interhemispheric circulation driven by winds in the ACC was earlier proposed by Eady (1957).
- 15 From Rintoul *et al.* (2001).
- 16 See Rintoul *et al.* (2001) and Olbers *et al.* (2004) for reviews.
- 17 These simulations, described in Henning & Vallis (2005), solve the primitive equations in a domain similar to Fig. 16.13. The wind forcing produces a poleward Ekman drift across the channel, as well as a subtropical gyre, and there is a meridional temperature gradient across the whole domain, so giving rise to a subtropical thermocline.
- 18 Adapted from Henning & Vallis (2005).

- 19 Munk & Palmén (1951), Gille (1997) and Stevens & Ivchenko (1997).
- 20 Models of this ilk stem from Johnson & Bryden (1989), Straub (1993), Hallberg & Gnanadesikan (2001), Karsten *et al.* (2002), Marshall & Radko (2003), Henning & Vallis (2005), and others, consider related issues and extensions.
- 21 See also Warren *et al.* (1996) and Hughes (2002).

Further reading

Siedler, G., Church, J. & Gould, J., Eds., 2001. *Ocean Circulation and Climate*.

A comprehensive collection of review articles, with a general emphasis on observations and modelling of the large-scale circulation, and the ocean's role in the climate system.

Problems

- 16.1 Show that the height of the upper layer in a model of the ventilated thermocline is continuous at the edge of the shadow zone, although the gradient normal to the shadow zone boundary need not be. [Hint: In the ventilated region $h_1 = (1 - f/f_2)h$, and in the shadow zone $g'_1/g'_2h_1^2 = D_0^2$. Show that these lead to a shadow zone boundary that is the same as (16.68).]
- 16.2 Obtain an expression for the boundary of the western pool in a two-layer model of the ventilated thermocline.
- 16.3 ♦ A one-dimensional equation similar to (16.27) was put forward by Stommel & Webster (1963) as a simple model of the thermocline. The equation is

$$\left(2W - z \frac{\partial W}{\partial z}\right) \frac{\partial^3 W}{\partial z^3} = \kappa \frac{\partial^4 W}{\partial z^4}. \quad (\text{P16.1})$$

Perform a boundary layer analysis on this equation, and show that if the downward vertical velocity at the surface $z = 0$ is non-zero then the thickness of an internal boundary layer scales as $\kappa^{1/2}$. How does the vertical velocity scale with κ in this case? Why do these differ from the corresponding scalings for (16.27)? How does the boundary layer thickness scale if it is at the surface ($z = 0$)?

- 16.4 Construct a one-dimensional thermocline boundary layer equation by simply ignoring the variations in x and y in A , so by substituting $A(x, y, \zeta) = B(\zeta)$ in (16.41). Show that this leads to an equation of the form

$$C \frac{dB}{d\zeta} \frac{d^3 B}{d\zeta^3} = \frac{\kappa}{\delta^2} \frac{d^4 B}{d\zeta^4}, \quad (\text{P16.2})$$

where C depends on β , f , h_x and T_0 . If C were independent of x and y , show that this equation is essentially the same as the boundary layer equation that emerges from (P16.1). Is C in fact constant (i.e., not a function of x) in general?

- 16.5 (a) What differentiates the top of the ocean from the bottom in (16.27)? That is, why are the solutions not vertically symmetric, especially when boundary conditions of zero vertical velocity at top and bottom are imposed?
 (b) By appropriately scaling (16.26), show that a possible non-dimensional form is

$$\widehat{W} \widehat{W}_{\widehat{z}\widehat{z}\widehat{z}} = \widehat{W}_{\widehat{z}\widehat{z}\widehat{z}\widehat{z}}. \quad (\text{P16.3})$$

- Does the value of κ play any role in the solution? (Hint: consider the boundary conditions.)
 (c) If the Ekman pumping velocity at the surface, W_E , is zero then we cannot use it to non-dimensionalize W in (16.27). In this case what would be an appropriate choice for the scaling value W_S ? Does this choice affect whether $\widehat{\kappa}$ is 'small' and whether \widehat{W} is $\mathcal{O}(1)$ in (16.27)?

References

- Abarbanel, H. D. I. & Young, W. R., Eds., 1987. *General Circulation of the Ocean*. Springer-Verlag, 291 pp.
- Allen, J. S., 1993. Iterated geostrophic intermediate models. *J. Phys. Oceanogr.*, **23**, 2447–2461.
- Allen, J. S., Barth, J. A. & Newberger, P. A., 1990a. On intermediate models for barotropic continental shelf and slope flow fields. Part I: formulation and comparison of exact solutions. *J. Phys. Oceanogr.*, **20**, 1017–1042.
- Allen, J. S., Barth, J. A. & Newberger, P. A., 1990b. On intermediate models for barotropic continental shelf and slope flow fields. Part II: comparison of numerical model solutions in doubly periodic domains. *J. Phys. Oceanogr.*, **20**, 1043–1082.
- Allen, J. S., Holm, D. D. & Newberger, P. A., 2002. Extended-geostrophic Euler–Poincaré models for mesoscale oceanographic flow. In J. Norbury & I. Roulstone, Eds., *Large-scale Atmosphere–Ocean Dynamics I*. Cambridge University Press.
- Andrews, D. G., 1987. On the interpretation of the Eliassen–Palm flux divergence. *Quart. J. Roy. Meteor. Soc.*, **113**, 323–338.
- Andrews, D. G., Holton, J. R. & Leovy, C. B., 1987. *Middle Atmosphere Dynamics*. Academic Press, 489 pp.
- Andrews, D. G. & McIntyre, M. E., 1976. Planetary waves in horizontal and vertical shear: the generalized Eliassen–Palm relation and the mean zonal acceleration. *J. Atmos. Sci.*, **33**, 2031–2048.
- Andrews, D. G. & McIntyre, M. E., 1978. Generalized Eliassen–Palm and Charney–Drazin theorems for waves on axisymmetric mean flows in compressible atmospheres. *J. Atmos. Sci.*, **35**, 175–185.
- Arbic, B., Flierl, G. & Scott, R., 2005. Cascade inequalities for forced-dissipated geostrophic turbulence. (In review).
- Aris, R., 1962. *Vectors, Tensors and the Basic Equations of Fluid Mechanics*. Dover, 286 pp. First published by Prentice-Hall.
- Arnold, V. I., 1965. Conditions for nonlinear stability of stationary plane curvilinear flows of an ideal fluid. *Dokl. Akad. Nauk SSSR*, **162**, 975–978. Engl. transl.: *Sov. Math.* **6**, 773–777 (1965).
- Arnold, V. I., 1966. On an a priori estimate in the theory of hydrodynamic stability. *Izv. Vyssh. Uchebn. Zaved. Math.*, **54**, 3–5. Engl. transl.: *Am Mat. Soc. Transl. Ser.*, **2**, **79**, 267–289 (1969).

- Assmann, R., 1902. Über die Existenz eines wärmeren Luftstromes in der Höhe von 10 bis 15 km. (On the existence of a warmer airflow at heights from 10 to 15 km). *Sitzber. Königl. Preuss. Akad. Wiss. Berlin*, **24**, 495–504.
- Aubin, D. & Dahan Dalmedico, A., 2002. Writing the history of dynamical systems and chaos: *longue durée* and revolution, disciplines and cultures. *Historia Mathematica*, **29**, 1–67.
- Ball, J. M. & James, R. D., 2002. The scientific life and influence of Clifford Ambrose Truesdell III. *Arch. Rational Mech. Anal.*, **161**, 1–26.
- Bannon, P. R., 1995. Potential vorticity conservation, hydrostatic adjustment, and the anelastic approximation. *J. Atmos. Sci.*, **52**, 2301–2312.
- Bannon, P. R., 1996. On the anelastic equation for a compressible atmosphere. *J. Atmos. Sci.*, **53**, 3618–3628.
- Barnier, B., Hua, B. L. & Provost, C. L., 1991. On the catalytic role of high baroclinic modes in eddy-driven large-scale circulations. *J. Phys. Oceanogr.*, **21**, 976–997.
- Bartello, P. & Warn, T., 1996. Self-similarity of decaying two-dimensional turbulence. *J. Fluid Mech.*, **326**, 357–372.
- Batchelor, G. K., 1953a. The conditions for dynamical similarity of motions of a frictionless perfect-gas atmosphere. *Quart. J. Roy. Meteor. Soc.*, **79**, 224–235.
- Batchelor, G. K., 1953b. *The Theory of Homogeneous Turbulence*. Cambridge University Press, 197 pp.
- Batchelor, G. K., 1959. Small-scale variation of convected quantities like temperature in turbulent fluid. Part 1: general discussion and the case of small conductivity. *J. Fluid Mech.*, **5**, 113–133.
- Batchelor, G. K., 1967. *An Introduction to Fluid Dynamics*. Cambridge University Press, 615 pp.
- Batchelor, G. K., 1969. Computation of the energy spectrum in homogeneous two-dimensional turbulence. *Phys. Fluids Suppl.*, **12**, II–233–II–239.
- Batchelor, G. K., Howells, I. D. & Townsend, A. A., 1959. Small-scale variation of convected quantities like temperature in turbulent fluid. Part 2: the case of large conductivity. *J. Fluid Mech.*, **5**, 134–139.
- Batchelor, G. K. & Townsend, A. A., 1956. Turbulent diffusion. In G. K. Batchelor & R. M. Davies, Eds., *Surveys in Mechanics*, pp. 352–399. Cambridge University Press.
- Baumert, H. Z., Simpson, J. & Sündermann, J., Eds., 2005. *Marine Turbulence: Theories, Observations and Models*. Cambridge University Press, 630 pp.
- Bender, C. M. & Orszag, S. A., 1978. *Advanced Mathematical Methods for Scientists and Engineers*. McGraw-Hill, 593 pp.
- Berrisford, P., Marshall, J. C. & White, A. A., 1993. Quasi-geostrophic potential vorticity in isentropic co-ordinates. *Quart. J. Roy. Meteor. Soc.*, **50**, 778–782.
- Birner, T., 2006. Fine-scale structure of the extratropical tropopause region. *J. Geophys. Res.*, **111**, D04104. doi:10.1029/2005JD006301.
- Birner, T., Dörnbrack, A. & Schumann, U., 2002. How sharp is the tropopause at midlatitudes? *Geophys. Res. Lett.*, **29**, 1700. doi:10.1029/2002GL015142.
- Bjerknes, J., 1919. On the structure of moving cyclones. *Geofys. Publ.*, **1** (2), 1–8.
- Bjerknes, J., 1959. Atlantic air-sea interaction. In *Advances in Geophysics*, Vol. 10, pp. 1–82. Academic Press.
- Bjerknes, J., 1969. Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.*, **97**, 163–172.
- Bjerknes, V., 1898a. Über die Bildung von Cirkulationsbewegungen und Wirbeln in reibunglosen Flüssigkeiten (On the generation of circulation and vortices in inviscid fluids). *Skr. Nor. Vidensk.-Akad. 1: Mat.-Naturvidensk. Kl.*, **5**, 3–29.
- Bjerknes, V., 1898b. Über einen hydrodynamischen Fundamentalsatz und seine Anwendung beson-

- ders auf die Mechanik der Atmosphäre und des Weltmeeres (On a fundamental principle of hydrodynamics and its application particularly to the mechanics of the atmosphere and the world's oceans). *Kongl. Sven. Vetensk. Akad. Handlingar*, **31**, 1–35.
- Bjerknes, V., 1902. Cirkulation relativ zu der Erde (Circulation relative to the Earth). *Meteor. Z.*, **37**, 97–108.
- Bjerknes, V., 1904. Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physic (The problem of weather forecasting as a problem in mathematics and physics). *Meteor. Z.*, January, 1–7. Engl. transl.: Y. Mintz, in Shapiro and Grønås (1999), pp. 1–7.
- Blumen, W., 1968. On the stability of quasi-geostrophic flow. *J. Atmos. Sci.*, **25**, 929–933.
- Boccaletti, G., Pacanowski, R. C., Philander, S. G. H. & Fedorov, A. V., 2004. The thermal structure of the upper ocean. *J. Phys. Oceanogr.*, **34**, 888–902.
- Boer, G. J. & Shepherd, T. G., 1983. Large-scale two-dimensional turbulence in the atmosphere. *J. Atmos. Sci.*, **40**, 164–184.
- Boussinesq, J., 1903. Théorie analytique de la chaleur (Analytic theory of heat). *Tome, Paris, Gauthier-Villars*, **II**, 170–172.
- Boyd, J. P., 1976. The noninteraction of waves with the zonally averaged flow on a spherical Earth and the interrelationships of eddy fluxes of energy, heat and momentum. *J. Atmos. Sci.*, **33**, 2285–2291.
- Branscome, L. E., 1983. The Charney baroclinic stability problem: approximate solutions and modal structures. *J. Atmos. Sci.*, **40**, 1393–1409.
- Bretherton, C. S. & Schär, C., 1993. Flux of potential vorticity substance: a simple derivation and a uniqueness property. *J. Atmos. Sci.*, **50**, 1834–1836.
- Bretherton, F. P., 1966. Critical layer instability in baroclinic flows. *Quart. J. Roy. Meteor. Soc.*, **92**, 325–334.
- Brewer, A. W., 1949. Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Quart. J. Roy. Meteor. Soc.*, **75**, 251–363.
- Brillouin, L., 1926. La mecanique ondulatoire de Schrödinger; une methode generale de resolution par approximations successives (The wave mechanics of Schrödinger: a general method of solution by successive approximation). *Comptes Rendus*, **183**, 24–26.
- Bryan, F., 1986. High-latitude salinity effects and interhemispheric thermohaline circulations. *Nature*, **323**, 301–304.
- Bryden, H. L., 1973. New polynomials for thermal expansion, adiabatic temperature gradient and potential temperature gradient of sea water. *Deep Sea Res.*, **20**, 401–408.
- Burger, A., 1958. Scale considerations of planetary motions of the atmosphere. *Tellus*, **10**, 195–205.
- Callen, H. B., 1985. *Thermodynamics and an Introduction to Thermostatistics*. John Wiley & Sons, 493 pp.
- Cessi, P., 2001. Thermohaline circulation variability. In *Conceptual Models of the Climate, Woods Hole Program in Geophysical Fluid Dynamics* (2001). Also available from <http://gfd.whoi.edu/proceedings/2001/PDFvol2001.html>.
- Cessi, P. & Fantini, M., 2004. The eddy-driven thermocline. *J. Phys. Oceanogr.*, **34**, 2642–2658.
- Cessi, P. & Young, W. R., 1992. Multiple equilibria in two-dimensional thermohaline circulation. *J. Fluid Mech.*, **241**, 291–309.
- Chandrasekhar, S., 1961. *Hydrodynamic and Hydromagnetic Stability*. Oxford University Press, 652 pp. Reprinted by Dover Publications, 1981.
- Chapman, S. & Lindzen, R. S., 1970. *Atmospheric Tides*. Gordon and Breach, 200 pp.
- Charney, J. G., 1947. The dynamics of long waves in a baroclinic westerly current. *J. Meteor.*, **4**, 135–162.
- Charney, J. G., 1948. On the scale of atmospheric motion. *Geofys. Publ. Oslo*, **17**(2), 1–17.

- Charney, J. G., 1955. The Gulf Stream as an inertial boundary layer. *Proc. Nat. Acad. Sci.*, **41**, 731–740.
- Charney, J. G., 1971. Geostrophic turbulence. *J. Atmos. Sci.*, **28**, 1087–1095.
- Charney, J. G. & Drazin, P. G., 1961. Propagation of planetary scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.*, **66**, 83–109.
- Charney, J. G. & Eliassen, A., 1949. A numerical method for predicting the perturbations of the mid-latitude westerlies. *Tellus*, **1**, 38–54.
- Charney, J. G. & Eliassen, A., 1964. On the growth of the hurricane depression. *J. Atmos. Sci.*, **21**, 68–75.
- Charney, J. G., Fjørtoft, R. & Neumann, J. V., 1950. Numerical integration of the barotropic vorticity equation. *Tellus*, **2**, 237–254.
- Charney, J. G. & Stern, M. E., 1962. On the stability of internal baroclinic jets in a rotating atmosphere. *J. Atmos. Sci.*, **19**, 159–172.
- Charnock, H., Green, J., Ludlam, F., Scorer, R. & Sheppard, P., 1966. Dr. E. T. Eady, B. A. (Obituary). *Quart. J. Roy. Meteor. Soc.*, **92**, 591–592.
- Chasnov, J. R., 1991. Simulation of the inertial-conductive subrange. *Phys. Fluids A*, **3**, 1164–1168.
- Chelton, D. B., de Szoeke, R. A., Schlax, M. G., Naggar, K. E. & Siwertz, N., 1998. Geographical variability of the first-baroclinic Rossby radius of deformation. *J. Phys. Oceanogr.*, **28**, 433–460.
- Colin de Verdrière, A., 1980. Quasi-geostrophic turbulence in a rotating homogeneous fluid. *Geophys. Astrophys. Fluid Dyn.*, **15**, 213–251.
- Colin de Verdrière, A., 1989. On the interaction of wind and buoyancy driven gyres. *J. Mar. Res.*, **47**, 595–633.
- Conkright, M. E., Antonov, J., Baranova, O., Boyer, T. P. *et al.*, 2001. World ocean database 2001, vol. 1. In S. Levitus, Ed., *NOAA Atlas NESDIS 42*, pp. 167. US Government Printing Office, Washington D. C.
- Coriolis, G. G., 1832. Mémoire sur le principe des forces vives dans les mouvements relatifs des machines (On the principle of kinetic energy in the relative movement of machines). *J. Ec. Polytech.*, **13**, 268–301.
- Coriolis, G. G., 1835. Mémoire sur les équations du mouvement relatif des systèmes de corps (On the equations of relative motion of a system of bodies). *J. Ec. Polytech.*, **15**, 142–154.
- Corrsin, S., 1951. On the spectrum of isotropic temperature fluctuations in an isotropic turbulence. *J. Appl. Phys.*, **22**, 469–473. Erratum: *J. Appl. Phys.* **22**, 1292, (1951).
- Cressman, G. P., 1996. The origin and rise of numerical weather prediction. In J. R. Fleming, Ed., *Historical Essays on Meteorology 1919–1995*, pp. 617. American Meteorological Society.
- Cushman-Roisin, B., 1994. *Introduction to Geophysical Fluid Dynamics*. Prentice Hall, 320 pp.
- Danielsen, E. F., 1990. In defense of Ertel's potential vorticity and its general applicability as a meteorological tracer. *J. Atmos. Sci.*, **47**, 2353–2361.
- Danilov, S. & Gryankin, V., 2002. Rhines scale and spectra of the β -plane turbulence with bottom drag. *Phys. Rev. E*, **65**, 067301–1–067301–3.
- Danilov, S. & Gurarie, D., 2001. Quasi-two-dimensional turbulence. *Usp. Fiz. Nauk.*, **170**, 921–968.
- Davies-Jones, R., 2003. Comments on “A generalization of Bernoulli's theorem”. *J. Atmos. Sci.*, **60**, 2039–2041.
- Davis, R. E., de Szoeke, R., Halpern, D. & Niiler, P., 1981. Variability in the upper ocean during MILE. Part I: The heat and momentum balances. *Deep-Sea Res.*, **28**, 1427–1452.
- De Morgan, A., 1872. *A Budget of Paradoxes*. Thoemmes Continuum, 814 pp.
- de Szoeke, R., 2004. An effect of the thermobaric nonlinearity of the equation of state: A mechanism for sustaining solitary Rossby waves. *J. Phys. Oceanogr.*, **34**, 2042–2056.

- de Szoëke, R. & Bennett, A. F., 1993. Microstructure fluxes across density surfaces. *J. Phys. Oceanogr.*, **24**, 2254–2264.
- Defant, A., 1921. Die Zirkulation der Atmosphäre in den gemäßigten Breiten der Erde. Grundzüge einer Theorie der Klimaschwankungen (The circulation of the atmosphere in the Earth's midlatitudes. Basic features of a theory of climate fluctuations). *Geograf. Ann.*, **3**, 209–266.
- Dewar, W. K. & Huang, R. X., 1995. Fluid flow in loops driven by freshwater and heat fluxes. *J. Fluid Mech.*, **297**, 153–191.
- Dewar, W. K., Samelson, R. S. & Vallis, G. K., 2005. The ventilated pool: a model of subtropical mode water. *J. Phys. Oceanogr.*, **35**, 137–150.
- Dickinson, R. E., 1968. Planetary Rossby waves propagating vertically through weak westerly wind wave guides. *J. Atmos. Sci.*, **25**, 984–1002.
- Dickinson, R. E., 1969. Theory of planetary wave–zonal flow interaction. *J. Atmos. Sci.*, **26**, 73–81.
- Dickinson, R. E., 1980. Planetary waves: theory and observation. In *Orographic Effects on Planetary Flows*, Number 23 in GARP Publication Series. World Meteorological Organization.
- Dijkstra, H. A., 2002. *Nonlinear Physical Oceanography*. Kluwer, 456 pp.
- Dima, I. & Wallace, J. M., 2003. On the seasonality of the Hadley Cell. *J. Atmos. Sci.*, **60**, 1522–1527.
- Dobson, G. M. B., 1956. Origin and distribution of the polyatomic molecules in the atmosphere. *Proc. Roy. Soc. Lond. A*, **236**, 187–193.
- Döös, K. & Coward, A., 1997. The Southern Ocean as the major upwelling zone of the North Atlantic. *Int. WOCE Newsletter*, No. 27, 3–4.
- Drazin, P. G. & Reid, W. H., 1981. *Hydrodynamic Stability*. Cambridge University Press, 527 pp.
- Drijfhout, S. S. & Hazeleger, W., 2001. Eddy mixing of potential vorticity versus temperature in an isopycnic ocean model. *J. Phys. Oceanogr.*, **31**, 481–505.
- Dunkerton, T. J., 1980. A Lagrangian-mean theory of wave, mean-flow interaction with applications to non-acceleration and its breakdown. *Rev. Geophys. Space Phys.*, **18**, 387–400.
- Durran, D. R., 1989. Improving the anelastic approximation. *J. Atmos. Sci.*, **46**, 1453–1461.
- Durran, D. R., 1993. Is the Coriolis force really responsible for the inertial oscillations? *Bull. Am. Meteor. Soc.*, **74**, 2179–2184.
- Durst, C. S. & Sutcliffe, R. C., 1938. The effect of vertical motion on the “geostrophic departure” of the wind. *Quart. J. Roy. Meteor. Soc.*, **64**, 240.
- Dutton, J. A., 1986. *The Ceaseless Wind: An Introduction to the Theory of Atmospheric Motion*. Dover Publications, 617 pp.
- Eady, E. T., 1949. Long waves and cyclone waves. *Tellus*, **1**, 33–52.
- Eady, E. T., 1950. The cause of the general circulation of the atmosphere. In *Cent. Proc. Roy. Meteor. Soc.* (1950), pp. 156–172.
- Eady, E. T., 1954. The maintenance of the mean zonal surface currents. *Proc. Toronto Meteor. Conf. 1953*, **138**, 124–128. Royal Meteorological Society.
- Eady, E. T., 1957. The general circulation of the atmosphere and oceans. In D. R. Bates, Ed., *The Earth and its Atmosphere*, pp. 130–151. New York, Basic Books.
- Eady, E. T. & Sawyer, J. S., 1951. Dynamics of flow patterns in extra-tropical regions. *Quart. J. Roy. Meteor. Soc.*, **77**, 531–551.
- Edmon, H. J., Hoskins, B. J. & McIntyre, M. E., 1980. Eliassen–Palm cross sections for the troposphere. *J. Atmos. Sci.*, **37**, 2600–2616.
- Egger, J., 1976. Linear response of a two-level primitive equation model to forcing by topography. *Mon. Wea. Rev.*, **104**, 351–364.
- Egger, J., 1999. Inertial oscillations revisited. *J. Atmos. Sci.*, **56**, 2951–2954.
- Einstein, A., 1916. *Relativity: The Special and the General Theory*. Random House, 188 pp.

- Ekman, V. W., 1905. On the influence of the Earth's rotation on ocean currents. *Arch. Math. Astron. Phys.*, 2, 1-52.
- Eliassen, A. & Palm, E., 1961. On the transfer of energy in stationary mountain waves. *Geophys. Publ.*, 22, 1-23.
- Eluszkiewicz, J., Crisp, D., Grainger, R. G., Lambert, A. et al., 1997. Sensitivity of the residual circulation diagnosed from the UARS data to the uncertainties in the input fields and to the inclusion of aerosols. *J. Atmos. Sci.*, 54, 1739-1757.
- Er-El, J. & Peskin, R., 1981. Relative diffusion of constant-level balloons in the Southern Hemisphere. *J. Atmos. Sci.*, 38, 2264-2274.
- Ertel, H., 1942a. Ein neuer hydrodynamischer Wirbelsatz (A new hydrodynamic eddy theorem). *Meteorol. Z.*, 59, 277-281.
- Ertel, H., 1942b. Über des Verhältnis des neuen hydrodynamischen Wirbelsatzes zum Zirculationssatz von V. Bjerknes (On the relationship of the new hydrodynamic eddy theorem to the circulation theorem of V. Bjerknes). *Meteorol. Z.*, 59, 385-387.
- Ertel, H. & Rossby, C.-G., 1949a. Ein neuer Erhaltungs-satz der Hydrodynamik (A new conservation theorem of hydrodynamics). *Sitzungsber. d. Deutschen Akad. Wissenschaften Berlin*, 1, 3-11.
- Ertel, H. & Rossby, C.-G., 1949b. A new conservation theorem of hydrodynamics. *Geofis. Pura Appl.*, 14, 189-193.
- Fang, M. & Tung, K. K., 1996. A simple model of nonlinear Hadley circulation with an ITCZ: analytic and numerical solutions. *J. Atmos. Sci.*, 53, 1241-1261.
- Fang, M. & Tung, K. K., 1999. Time-dependent nonlinear Hadley circulation. *J. Atmos. Sci.*, 56, 1797-1807.
- Farrell, B., 1984. Modal and non-modal baroclinic waves. *J. Atmos. Sci.*, 41, 668-673.
- Farrell, B. F. & Ioannou, P. J., 1996. Generalized stability theory. Part I: autonomous operators. *J. Atmos. Sci.*, 53, 2025-2040.
- Feistel, R., 2003. A new extended thermodynamic potential of seawater. *Prog. Oceanogr.*, 58, 41-114.
- Feistel, R. & Hagen, E., 1995. On the Gibbs thermodynamic potential of seawater. *Prog. Oceanogr.*, 36, 249-347.
- Fels, S., 1985. Radiative-dynamical interactions in the middle atmosphere. In S. Manabe, Ed., *Adv. Geophys. Vol 28, Part A, Issues in Atmospheric and Oceanic Modeling*, pp. 27-300. Academic Press.
- Ferrari, R. & McWilliams, J. C., 2006. Parameterization of eddy fluxes near ocean boundaries. *Ocean Modelling*. (In review).
- Ferrari, R. & Plumb, R. A., 2003. The residual circulation in the ocean. In *Proc. 'Aha Huliko'a Winter Workshop, Honolulu, Hawaii* (2003), pp. 1-10.
- Ferrel, W., 1856a. An essay on the winds and currents of the ocean. *Nashville J. Med. & Surg.*, 11, 287-301.
- Ferrel, W., 1856b. The problem of the tides. *Astron. J.*, 4, 173-176.
- Ferrel, W., 1858. The influence of the Earth's rotation upon the relative motion of bodies near its surface. *Astron. J.*, V, No. 13 (109), 97-100.
- Ferrel, W., 1859. The motion of fluids and solids relative to the Earth's surface. *Math. Monthly*, 1, 140-148, 210-216, 300-307, 366-373, 397-406.
- Fjørtoft, R., 1950. Application of integral theorems in deriving criteria for laminar flows and for the baroclinic circular vortex. *Geophys. Publ.*, 17, 1-52.
- Fjørtoft, R., 1953. On the changes in the spectral distribution of kinetic energy for two-dimensional nondivergent flow. *Tellus*, 5, 225-230.
- Fleming, E. L., Chandra, S., Schoeberl, M. R. & Barnett, J. J., 1988. Monthly mean global climatology

- of temperature, wind, geopotential height, and pressure for 0-120 km. Technical report, NASA/Goddard Space Flight Center, Greenbelt, MD. NASA Tech. Memo. 100697.
- Fofonoff, N. P., 1954. Steady flow in a frictionless homogeneous ocean. *J. Mar. Res.*, **13**, 254-262.
- Fofonoff, N. P., 1985. Physical properties of seawater: a new salinity scale and equation of state for seawater. *J. Geophys. Res.*, **90**(C), 3332-3342.
- Foster, T. D., 1972. An analysis of the cabbeling instability in sea water. *J. Phys. Oceanogr.*, **2**, 294-301.
- Fox-Kemper, B. & Pedlosky, J., 2004. Wind-driven barotropic gyre I: circulation control by eddy vorticity fluxes to an enhanced removal region. *J. Mar. Res.*, **62**, 169-193.
- Franklin, W. S., 1898. Review of P. Duhamel, *Traité Elementaire de Méchanique Chimique fondée sur la Thermodynamique*, Two volumes. Paris, 1897. *Phys. Rev.*, **6**, 170-175.
- Friedman, R. M., 1989. *Appropriating the Weather: Vilhelm Bjerknes and the Construction of a Modern Meteorology*. Cornell University Press, 251 pp.
- Frisch, U., 1995. *Turbulence: The Legacy of A. N. Kolmogorov*. Cambridge University Press, 296 pp.
- Fu, L. L. & Flierl, G. R., 1980. Nonlinear energy and enstrophy transfers in a realistically stratified ocean. *Dyn. Atmos. Oceans*, **4**, 219-246.
- Gage, K. S. & Nastrom, G. D., 1986. Theoretical interpretation of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft during GASP. *J. Atmos. Sci.*, **43**, 729-740.
- Galperin, B., Sukoriansky, S., Dikovskaya, N., Read, P. et al., 2006. Anisotropic turbulence and zonal jets in rotating flows with a β -effect. *Nonlinear Proc. Geophys.*, **13**, 83-98.
- Gardiner, C. W., 1985. *Handbook of Stochastic Methods*. Springer-Verlag, 442 pp.
- Gent, P. R. & McWilliams, J. C., 1990. Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.*, **20**, 150-155.
- Gent, P. R., Willebrand, J., McDougall, T. J. & McWilliams, J. C., 1995. Parameterizing eddy-induced transports in ocean circulation models. *J. Phys. Oceanogr.*, **25**, 463-474.
- Gierasch, P. J., 1975. Meridional circulation and the maintenance of the Venus atmospheric rotation. *J. Atmos. Sci.*, **32**, 1038-1044.
- Gill, A. E., 1982. *Atmosphere-Ocean Dynamics*. Academic Press, 662 pp.
- Gill, A. E., Green, J. S. A. & Simmons, A. J., 1974. Energy partition in the large-scale ocean circulation and the production of mid-ocean eddies. *Deep-Sea Res.*, **21**, 499-528.
- Gille, S. T., 1997. The Southern Ocean momentum balance: evidence for topographic effects from numerical model output and altimeter data. *J. Phys. Oceanogr.*, **27**, 2219-2232.
- Gilman, P. A. & Glatzmaier, G. A., 1981. Compressible convection in a rotating spherical shell. I. Anelastic equations. *Astrophys. J. Suppl. Ser.*, **45**, 335-349.
- Gnanadesikan, A., 1999. A simple predictive model for the structure of the oceanic pycnocline. *Science*, **283**, 2077-2099.
- Godske, C. L., Bergeron, T., Bjerknes, J. & Bundgaard, R. C., 1957. *Dynamic Meteorology and Weather Forecasting*. American Meteorological Society, 864 pp.
- Gough, D. O., 1969. The anelastic approximation for thermal convection. *J. Atmos. Sci.*, **216**, 448-456.
- Grant, H. L., Hughes, B. A., Vogel, W. M. & Moilliet, A., 1968. The spectrum of temperature fluctuation in turbulent flow. *J. Fluid Mech.*, **344**, 423-442.
- Grant, H. L., Stewart, R. W. & Moilliet, A., 1962. Turbulent spectra from a tidal channel. *J. Fluid Mech.*, **12**, 241-268.
- Greatbatch, R. J., 1998. Exploring the relationship between eddy-induced transport velocity, vertical momentum transfer, and the isopycnal flux of potential vorticity. *J. Phys. Oceanogr.*, **28**, 422-432.

- Green, J. S. A., 1960. A problem in baroclinic stability. *Quart. J. Roy. Meteor. Soc.*, **86**, 237–251.
- Green, J. S. A., 1970. Transfer properties of the large-scale eddies and the general circulation of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **96**, 157–185.
- Green, J. S. A., 1977. The weather during July 1976: some dynamical considerations of the drought. *Weather*, **32**, 120–128.
- Green, J. S. A., 1999. *Atmospheric Dynamics*. Cambridge University Press, 213 pp.
- Greenspan, H., 1962. A criterion for the existence of inertial boundary layers in oceanic circulation. *Proc. Nat. Acad. Sci.*, **48**, 2034–2039.
- Gregg, M. C., 1998. Estimation and geography of diapycnal mixing in the stratified ocean. In J. Imberger, Ed., *Physical Processes in Lakes and Oceans*, pp. 305–338. American Geophysical Union.
- Griannik, N., Held, I., Smith, K. & Vallis, G. K., 2004. Effect of nonlinear drag on the inverse cascade. *Phys. Fluids*, **16**, 73–78.
- Griffies, S. M., 1998. The Gent–McWilliams skew flux. *J. Phys. Oceanogr.*, **28**, 831–841.
- Griffies, S. M., 2004. *Fundamentals of Ocean Climate Models*. Princeton University Press, 518 pp.
- Grose, W. & Hoskins, B., 1979. On the influence of orography on large-scale atmospheric flow. *J. Atmos. Sci.*, **36**, 223–234.
- Hadley, G., 1735. Concerning the cause of the general trade-winds. *Phil. Trans. Roy. Soc.*, **29**, 58–62.
- Hallberg, R. & Gnanadesikan, A., 2001. An exploration of the role of transient eddies in determining the transport of a zonally reentrant current. *J. Phys. Oceanogr.*, **31**, 3312–3330.
- Hamilton, K. P., 1998. Dynamics of the tropical middle atmosphere: a tutorial review. *Atmosphere–Ocean*, **36**, 319–354.
- Haney, R. L., 1971. Surface thermal boundary condition for ocean circulation models. *J. Phys. Oceanogr.*, **1**, 241–248.
- Hayes, M., 1977. A note on group velocity. *Proc. Roy. Soc. Lond. A*, **354**, 533–535.
- Haynes, P. H., Marks, C. J., McIntyre, M. E., Shepherd, T. G. & Shine, K. P., 1991. On the “downward control” of extratropical diabatic circulations by eddy-induced mean zonal forces. *J. Atmos. Sci.*, **48**, 651–678.
- Haynes, P. H. & McIntyre, M. E., 1987. On the evolution of vorticity and potential vorticity in the presence of diabatic heating and frictional or other forces. *J. Atmos. Sci.*, **44**, 828–841.
- Haynes, P. H. & McIntyre, M. E., 1990. On the conservation and impermeability theorem for potential vorticity. *J. Atmos. Sci.*, **47**, 2021–2031.
- Held, I. M., 1982. On the height of the tropopause and the static stability of the troposphere. *J. Atmos. Sci.*, **39**, 412–417.
- Held, I. M., 1983. Stationary and quasi-stationary eddies in the extratropical troposphere: theory. In B. Hoskins & R. P. Pearce, Eds., *Large-Scale Dynamical Processes in the Atmosphere*, pp. 127–168. Academic Press.
- Held, I. M., 1985. Pseudomomentum and the orthogonality of modes in shear flows. *J. Atmos. Sci.*, **42**, 2280–2288.
- Held, I. M., 2000. The general circulation of the atmosphere. In *Woods Hole Program in Geophysical Fluid Dynamics* (2000), pp. 66.
- Held, I. M. & Hou, A. Y., 1980. Nonlinear axially symmetric circulations in a nearly inviscid atmosphere. *J. Atmos. Sci.*, **37**, 515–533.
- Held, I. M. & Larichev, V. D., 1996. A scaling theory for horizontally homogeneous, baroclinically unstable flow on a beta-plane. *J. Atmos. Sci.*, **53**, 946–952.
- Held, I. M. & Schneider, T., 1999. The surface branch of the zonally averaged mass transport circulation in the troposphere. *J. Atmos. Sci.*, **56**, 1688–1697.

- Held, I. M., Tang, M. & Wang, H., 2002. Northern winter stationary waves: theory and modeling. *J. Climate*, **15**, 2125–2144.
- Helmholtz, H., 1858. Über Integrale der hydrodynamischen Gleichungen welche den Wirbelbewegungen entsprechen (On the integrals of the hydrodynamic equations that correspond to eddy motion). *J. Reine Angew. Math*, **25**, 25–55. Engl. transl.: C. Abbe, *Smithson. Misc. Collect.*, no. 34, pp. 78–93, Smithsonian Institution, Washington D. C., 1893.
- Helmholtz, H., 1868. Über discontinuirliche Flüssigkeitsbewegungen (On discontinuous liquid motion). *Monats. Königl. Preuss. Akad. Wiss. Berlin*, **23**, 215–228. Engl. trans.: F. Guthrie: On discontinuous movements of fluids. *Phil. Mag.*, **36**, 337–346 (1868).
- Hendershott, M., 1987. Single layer models of the general circulation. In H. Abarbanel & W. R. Young, Eds., *General Circulation of the Ocean*, pp. 202–267. Springer-Verlag.
- Henning, C. C. & Vallis, G. K., 2004. The effect of mesoscale eddies on the main subtropical thermocline. *J. Phys. Oceanogr.*, **34**, 2428–2443.
- Henning, C. C. & Vallis, G. K., 2005. The effects of mesoscale eddies on the stratification and transport of an ocean with a circumpolar channel. *J. Phys. Oceanogr.*, **35**, 880–896.
- Hide, R., 1969. Dynamics of the atmospheres of major planets with an appendix on the viscous boundary layer at the rigid boundary surface of an electrically conducting rotating fluid in the presence of a magnetic field. *J. Atmos. Sci.*, **26**, 841–853.
- Hilborn, R. C., 2004. Sea gulls, butterflies, and grasshoppers: a brief history of the butterfly effect in nonlinear dynamics. *Am. J. Phys.*, **72**, 425–427.
- Hockney, R., 1970. The potential calculation and some applications. In *Methods of Computational Physics*, Vol. 9, pp. 135–211. Academic Press.
- Hogg, N., 2001. Quantification of the deep circulation. In G. Siedler, J. Church, & J. Gould, Eds., *Ocean Circulation and Climate: Observing and Modelling the Global Ocean*, pp. 259–270. Academic Press.
- Hoinka, K. P., 1997. The tropopause: discovery, definition and demarcation. *Meteorol. Z.*, **6**, 281–303.
- Holland, W. R., Keffer, T. & Rhines, P. B., 1984. Dynamics of the oceanic circulation: the potential vorticity field. *Nature*, **308**, 698–705.
- Holloway, G. & Hendershott, M. C., 1977. Stochastic closure for nonlinear Rossby waves. *J. Fluid Mech.*, **82**, 747–765.
- Holm, D. D., Marsden, J. E., Ratiu, T. & Weinstein, A., 1985. Nonlinear stability of fluid and plasma equilibria. *Phys. Rep.*, **123**, 1–116.
- Holton, J. R., 1974. Forcing of mean flows by stationary waves. *J. Atmos. Sci.*, **31**, 942–945.
- Holton, J. R., 1992. *An Introduction to Dynamic Meteorology*. 3rd edn. Academic Press, 507 pp.
- Hoskins, B. J. & Karoly, D. J., 1981. The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179–1196.
- Hua, B. L. & Haidvogel, D. B., 1986. Numerical simulations of the vertical structure of quasi-geostrophic turbulence. *J. Atmos. Sci.*, **43**, 2923–2936.
- Huang, R. X., 1988. On boundary value problems of the ideal-fluid thermocline. *J. Phys. Oceanogr.*, **18**, 619–641.
- Huang, R. X., 1998. Mixing and available potential energy in a Boussinesq ocean. *J. Phys. Oceanogr.*, **28**, 669–678.
- Huang, R. X., 1999. Mixing and energetics of the oceanic thermohaline circulation. *J. Phys. Oceanogr.*, **29**, 727–746.
- Hughes, C. W., 2002. Sverdrup-like theories of the Antarctic Circumpolar Current. *J. Mar. Res.*, **60**, 1–17.

- Hughes, C. W. & de Cuevas, B., 2001. Why western boundary currents in realistic oceans are inviscid: a link between form stress and bottom pressure torques. *J. Phys. Oceanogr.*, **31**, 2871–2885.
- Ierley, G. R. & Ruehr, O. G., 1986. Analytic and numerical solutions of a nonlinear boundary value problem. *Stud. Appl. Math.*, **75**, 1–36.
- Il'in, A. M. & Kamenkovich, V. M., 1964. The structure of the boundary layer in the two-dimensional theory of ocean currents (in Russian). *Okeanologiya*, **4** (5), 756–769.
- Ingersoll, A. P., 2005. Boussinesq and anelastic approximations revisited: potential energy release during thermobaric instability. *J. Phys. Oceanogr.*, **35**, 1359–1369.
- Iwayama, T., Shepherd, T. G. & Watanabe, T., 2002. An ‘ideal’ form of decaying two-dimensional turbulence. *J. Fluid Mech.*, **456**, 183–198.
- Jackett, J. & McDougall, T., 1995. Minimal adjustment of hydrographic profiles to achieve static stability. *J. Atmos. Ocean. Tech.*, **12**, 381–389.
- Jackson, L., Hughes, C. W. & Williams, R. G., 2006. Topographic control of basin and channel flows: the role of bottom pressure torques and friction. *J. Phys. Oceanogr.*, **36**, 1786–1805.
- James, I. N., 1994. *Introduction to Circulating Atmospheres*. Cambridge University Press, 422 pp.
- Jeffreys, H., 1924. On certain approximate solutions of linear differential equations of the second order. *Proc. London Math. Soc.* (2), **23**, 428–436.
- Jeffreys, H., 1926. On the dynamics of geostrophic winds. *Quart. J. Roy. Meteor. Soc.*, **51**, 85–104.
- Johnson, G. C. & Bryden, H. L., 1989. On the size of the Antarctic Circumpolar Current. *Deep-Sea Res.*, **36**, 39–53.
- Juckes, M. N., 2000. The static stability of the midlatitude troposphere: the relevance of moisture. *J. Atmos. Sci.*, **57**, 3050–3057.
- Juckes, M. N., 2001. A generalization of the transformed Eulerian-mean meridional circulation. *Quart. J. Roy. Meteor. Soc.*, **127**, 147–160.
- Kalnay, E., 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kalnay, E., 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.
- Karsten, R., Jones, H. & Marshall, J., 2002. The role of eddy transfer in setting the stratification and transport of a circumpolar current. *J. Phys. Oceanogr.*, **32**, 39–54.
- Keffer, T., 1985. The ventilation of the world’s oceans: maps of potential vorticity. *J. Phys. Oceanogr.*, **15**, 509–523.
- Kibel, I., 1940. Prilozhenie k meteorogi uravnenii mekhaniki baroklinnoi zhidkosti (Application of baroclinic fluid dynamic equations to meteorology). *SSSR Ser. Geogr. Geofiz.*, **5**, 627–637.
- Killworth, P. D., 1987. A continuously stratified nonlinear ventilated thermocline. *J. Phys. Oceanogr.*, **17**, 1925–1943.
- Killworth, P. D., 1997. On the parameterization of eddy transfer. Part I: theory. *J. Marine Res.*, **55**, 1171–1197.
- Kimoto, M. & Ghil, M., 1993. Multiple flow regimes in the northern hemisphere winter. Part 1: methodology and hemispheric regimes. *J. Atmos. Sci.*, **50**, 2625–2643.
- Kolmogorov, A. N., 1941. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Dokl. Acad. Sci. USSR*, **30**, 299–303.
- Kolmogorov, A. N., 1962. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds numbers. *J. Fluid Mech.*, **13**, 82–85.
- Kraichnan, R., 1967. Inertial ranges in two-dimensional turbulence. *Phys. Fluids*, **10**, 1417–1423.
- Kraichnan, R. & Montgomery, D., 1980. Two-dimensional turbulence. *Rep. Prog. Phys.*, **43**, 547–619.
- Kramers, H. A., 1926. Wellenmechanik und halbzahlige Quantisierung (Wave mechanics and semi-integral quantization). *Zeit. fur Physik A*, **39**, 828–840.

- Kundu, P., Allen, J. S. & Smith, R. L., 1975. Modal decomposition of the velocity field near the Oregon coast. *J. Phys. Oceanogr.*, **5**, 683–704.
- Kundu, P. & Cohen, I. M., 2002. *Fluid Mechanics*. Academic Press, 730 pp.
- Kuo, H.-I., 1949. Dynamic instability of two-dimensional nondivergent flow in a barotropic atmosphere. *J. Meteorol.*, **6**, 105–122.
- Kuo, H.-I., 1951. Vorticity transfer as related to the development of the general circulation. *J. Meteorol.*, **8**, 307–315.
- Kushnir, Y., Robinson, W. A., Bladé, I., Hall, N. M. J. *et al.*, 2002. Atmospheric GCM response to extratropical SST anomalies: synthesis and evaluation. *J. Climate*, **15**, 2233–2256.
- LaCasce, J. H. & Ohlmann, C., 2003. Relative dispersion at the surface of the Gulf of Mexico. *J. Mar. Res.*, **65**, 285–312.
- Lait, L. R., 1994. An alternative form for potential vorticity. *J. Atmos. Sci.*, **51**, 1754–1759.
- Lanczos, C., 1970. *The Variational Principles of Mechanics*. University of Toronto Press, Reprinted by Dover Publications 1980, 418 pp.
- Landau, L. D., 1944. On the problem of turbulence. *Dokl. Akad. Nauk SSSR*, **44**, 311–314.
- Landau, L. D. & Lifshitz, E. M., 1987. *Fluid Mechanics* (Course of Theoretical Physics, v. 6). 2nd edn. Pergamon Press, 539 pp.
- Larichev, V. D. & Held, I. M., 1995. Eddy amplitudes and fluxes in a homogeneous model of fully developed baroclinic instability. *J. Phys. Oceanogr.*, **25**, 2285–2297.
- Ledwell, J. R., Watson, A. J. & Law, C. S., 1998. Mixing of a tracer in the pycnocline. *J. Geophys. Res.*, **103**, 21 499–21 529.
- Lee, M.-M., Marshall, D. P. & Williams, R. G., 1997. On the eddy transfer of tracers: advective or diffusive? *J. Mar. Res.*, **55**, 483–595.
- Lee, T. D., 1951. Difference between turbulence in a two-dimensional fluid and in a three-dimensional fluid. *J. Appl. Phys.*, **22**, 524.
- Leetmaa, A., Niiler, P. & Stommel, H., 1977. Does the Sverdrup relation account for the mid-Atlantic circulation? *J. Mar. Res.*, **35**, 1–10.
- Leith, C. E., 1968. Diffusion approximation for two-dimensional turbulence. *Phys. Fluids*, **11**, 671–672.
- Lesieur, M., 1997. *Turbulence in Fluids: Third Revised and Enlarged Edition*. Kluwer, 515 pp.
- Levinson, N., 1950. The 1st boundary value problem for $\epsilon \Delta U + A(x, y)U_x + B(x, y)U_y + C(x, y)U = D(x, t)$ for small epsilon. *Ann. Math.*, **51**, 429–445.
- Lewis, R., Ed., 1991. *Meteorological Glossary*. 6th edn. Her Majesty's Stationery Office, 335 pp.
- Lighthill, J., 1978. *Waves in Fluids*. Cambridge University Press, 504 pp.
- Lighthill, M. J., 1965. Group velocity. *J. Inst. Math. Appl.*, **1**, 1–28.
- Lilly, D. K., 1969. Numerical simulation of two-dimensional turbulence. *Phys. Fluid Suppl. II*, **12**, 240–249.
- Lilly, D. K., 1996. A comparison of incompressible, anelastic and Boussinesq dynamics. *Atmos. Res.*, **40**, 143–151.
- Lindborg, E., 1999. Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence? *J. Fluid Mech.*, **388**, 259–288.
- Lindborg, E. & Alvelius, K., 2000. The kinetic energy spectrum of the two-dimensional enstrophy turbulence cascade. *Phys. Fluids*, **12**, 945–947.
- Lindzen, R. S. & Farrell, B., 1980. The role of the polar regions in global climate, and a new parameterization of global heat transport. *Mon. Wea. Rev.*, **108**, 2064–2079.
- Lindzen, R. S. & Hou, A. Y., 1988. Hadley circulation for zonally averaged heating centered off the equator. *J. Atmos. Sci.*, **45**, 2416–2427.

- Lindzen, R. S., Lorenz, E. N. & Plazman, G. W., Eds., 1990. *The Atmosphere — a Challenge: the Science of Jule Gregory Charney*. American Meteorological Society, 321 pp.
- Lipps, F. B. & Hemler, R. S., 1982. A scale analysis of deep moist convection and some related numerical calculations. *J. Atmos. Sci.*, **39**, 2192–2210.
- Longuet-Higgins, M. S., 1964. Planetary waves on a rotating sphere, I. *Proc. Roy. Soc. Lond. A*, **279**, 446–473.
- Lorenz, E. N., 1955. Available potential energy and the maintenance of the general circulation. *Tellus*, **7**, 157–167.
- Lorenz, E. N., 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.
- Lorenz, E. N., 1967. *The Nature and the Theory of the General Circulation of the Atmosphere*. WMO Publications, Vol. 218, World Meteorological Organization.
- Lozier, S., Owens, W. B. & Curry, R. G., 1996. The climatology of the North Atlantic. *Prog. Oceanogr.*, **36**, 1–44.
- Luyten, J. R., Pedlosky, J. & Stommel, H., 1983. The ventilated thermocline. *J. Phys. Oceanogr.*, **13**, 292–309.
- Mahrt, L., 1986. On the shallow motion approximations. *J. Atmos. Sci.*, **43**, 1036–1044.
- Majda, A. J., 2003. *Introduction to PDEs and Waves for the Atmosphere and Ocean*. American Mathematical Society, 234 pp.
- Maltrud, M. E. & Vallis, G. K., 1991. Energy spectra and coherent structures in forced two-dimensional and beta-plane turbulence. *J. Fluid Mech.*, **228**, 321–342.
- Manabe, S. & Stouffer, R. J., 1988. Two stable equilibria of a coupled ocean-atmosphere model. *J. Climate*, **1**, 841–866.
- Margules, M., 1903. Über die Energie der Stürme (On the energy of storms). *Jahrb. Kais.-kön Zent. für Met. und Geodynamik, Vienna*, 26 pp. Engl. transl.: C. Abbe, *Smithson. Misc. Collect.*, no. 51, pp. 533–595, Smithsonian Institution, Washington D. C., 1910.
- Marotzke, J., 1989. Instabilities and multiple steady states of the thermohaline circulation. In D. L. T. Anderson & J. Willebrand, Eds., *Oceanic Circulation Models: Combining Data and Dynamics*, pp. 501–511. NATO ASI Series, Kluwer.
- Marshall, D. P., Williams, R. G. & Lee, M.-M., 1999. The relation between eddy-induced transport and isopycnic gradients of potential vorticity. *J. Phys. Oceanogr.*, **29**, 1571–1578.
- Marshall, J. C., 1981. On the parameterization of geostrophic eddies in the ocean. *J. Phys. Oceanogr.*, **11**, 257–271.
- Marshall, J. C. & Nurser, A. J. G., 1992. Fluid dynamics of oceanic thermocline ventilation. *J. Phys. Oceanogr.*, **22**, 583–595.
- Marshall, J. C. & Radko, T., 2003. Residual-mean solutions for the Antarctic Circumpolar Current and its associated overturning circulation. *J. Phys. Oceanogr.*, **22**, 2341–2354.
- Marshall, J. C. & Schott, F., 1999. Open-ocean convection: observations, theory, and models. *Rev. Geophys.*, **37**, 1–64.
- McComb, W. D., 1990. *The Physics of Fluid Turbulence*. Clarendon Press, 572 pp.
- McDougall, T. J., 1998. Three-dimensional residual mean theory. In E. P. Chassignet & J. Verron, Eds., *Ocean Modeling and Parameterization*, pp. 269–302. Kluwer Academic.
- McDougall, T. J., 2003. Potential enthalphy: a conservative oceanic variable for evaluating heat content and heat fluxes. *J. Phys. Oceanogr.*, **33**, 945–963.
- McDougall, T. J., Jackett, D. R., Wright, D. G. & Feistel, R., 2002. Accurate and computationally efficient formulae for potential temperature and density of seawater. *J. Atmos. Ocean. Tech.*, **20**, 730–741.
- McIntosh, P. C. & McDougall, T. J., 1996. Isopycnal averaging and the residual mean circulation. *J. Phys. Oceanogr.*, **26**, 1655–1660.

- McIntyre, M. E., 1997. Lucidity and science I: writing skills and the pattern perception hypothesis. *Interdisc. Sci. Rev.*, **22**, 199–216.
- McIntyre, M. E. & Norton, W. A., 1990. Dissipative wave-mean interactions and the transport of vorticity or potential vorticity. *J. Fluid Mech.*, **212**, 403–435.
- McIntyre, M. E. & Norton, W. A., 2000. Potential vorticity inversion on a hemisphere. *J. Atmos. Sci.*, **57**, 1214–1235.
- McIntyre, M. E. & Shepherd, T. G., 1987. An exact local conservation theorem for finite-amplitude disturbances to nonparallel shear flows, with remarks on Hamiltonian structure and on Arnol'd's stability theorems. *J. Fluid Mech.*, **181**, 527–565.
- McWilliams, J. C., 1984. The emergence of isolated coherent vortices in turbulent flow. *J. Fluid Mech.*, **146**, 21–43.
- Mellor, G. L., 1991. An equation of state for numerical models of oceans and estuaries. *J. Atmos. Ocean. Tech.*, **8**, 609–611.
- Mihaljan, J. M., 1962. A rigorous exposition of the Boussinesq approximations applicable to a thin layer of fluid. *Astrophysical J.*, **136**, 1126–1133.
- Morel, P. & Larcheveque, M., 1974. Relative dispersion of constant-level balloons in the 200 mb general circulation. *J. Atmos. Sci.*, **31**, 2189–2196.
- Mundt, M., Vallis, G. K. & Wang, J., 1997. Balanced models for the large- and meso-scale circulation. *J. Phys. Oceanogr.*, **27**, 1133–1152.
- Munk, W. H., 1950. On the wind-driven ocean circulation. *J. Meteorol.*, **7**, 79–93.
- Munk, W. H., 1966. Abyssal recipes. *Deep-Sea Res.*, **13**, 707–730.
- Munk, W. H., Groves, G. W. & Carrier, G. F., 1950. Note on the dynamics of the Gulf Stream. *J. Mar. Res.*, **9**, 218–238.
- Munk, W. H. & Palmén, E., 1951. Note on dynamics of the Antarctic Circumpolar Current. *Tellus*, **3**, 53–55.
- Munk, W. H. & Wunsch, C., 1998. Abyssal recipes II: energetics of tidal and wind mixing. *Deep-Sea Res.*, **45**, 1976–2009.
- Namias, J., 1959. Recent seasonal interaction between North Pacific waters and the overlying atmospheric circulation. *J. Geophys. Res.*, **64**, 631–646.
- Newell, A. C., 1969. Rossby wave packet interactions. *J. Fluid Mech.*, **35**, 255–271.
- Nicholls, S., 1985. Aircraft observations of the Ekman layer during the joint air-sea interaction experiment. *Quart. J. Roy. Meteor. Soc.*, **111**, 391–426.
- Nof, D., 2003. The Southern Ocean's grip on the northward meridional flow. In *Progress in Oceanography*, Vol. 56, pp. 223–247. Pergamon.
- Novikov, E. A., 1959. Contributions to the problem of the predictability of synoptic processes. *Izv. An. SSSR Ser. Geophys.*, **11**, 1721. Eng. transl.: *Am. Geophys. U. Transl.*, 1209–1211.
- Nurser, A. G. & Lee, M.-M., 2004. Isopycnal averaging at constant height. Part I: the formulation and a case study. *J. Phys. Oceanogr.*, **34**, 2721–2739.
- Oberbeck, A., 1879. Über die Wärmeleitung der Flüssigkeiten bei Berücksichtigung der Strömungen infolge vor Temperaturdifferenzen (On the thermal conduction of liquids taking into account flows due to temperature differences). *Ann. Phys. Chem., Neue Folge*, **7**, 271–292.
- Oberbeck, A., 1888. Über die Bewegungserscheinungen der Atmosphäre (On the phenomena of motion in the atmosphere). *Sitzb. K. Preuss. Akad. Wiss.*, **7**, 383–395 and 1129–1138. Engl trans.: in B. Saltzman, Ed., *Theory of Thermal Convection*, Dover, 162–183.
- Obukhov, A. M., 1941. Energy distribution in the spectrum of turbulent flow. *Izv. Akad. Nauk. SSR Ser. Geogr. Geofiz.*, **5**, 453–466.
- Obukhov, A. M., 1949. Structure of the temperature field in turbulent flows. *Izv. Akad. Nauk. SSR Ser. Geogr. Geofiz.*, **13**, 58–63.

- Obukhov, A. M., 1962. On the dynamics of a stratified liquid. *Dokl. Akad. Nauk SSSR*, **145**, 1239–1242. Engl. transl.: *Soviet Physics-Dokl.* **7**, 682–684.
- Oetzel, K. & Vallis, G. K., 1997. Strain, vortices, and the enstrophy inertial range in two-dimensional turbulence. *Phys. Fluids*, **9**, 2991–3004.
- O’Gorman, P. A. & Pullin, D. I., 2005. Effect of Schmidt number of the velocity-scaler cospectrum in isotropic turbulence with a mean scalar gradient. *J. Fluid Mech.*, **532**, 111–140.
- Ogura, Y. & Phillips, N. A., 1962. Scale analysis of deep and shallow convection in the atmosphere. *J. Atmos. Sci.*, **19**, 173–179.
- Olbers, D., Borowski, D., Völker, C. & Wolff, J.-O., 2004. The dynamical balance, transport and circulation of the Antarctic Circumpolar Current. *Antarctic Science*, **16**, 439–470.
- Ollitrault, M., Gabet, C. & Colin de Verdière, A., 2005. Open ocean regimes of relative dispersion. *J. Fluid Mech.*, **533**, 381–407.
- Onsager, L., 1949. Statistical hydrodynamics. *Nuovo Cim. (Suppl.)*, **6**, 279–287.
- Palmer, T. N., 1997. A nonlinear dynamical perspective on climate prediction. *J. Climate*, **12**, 575–591.
- Paparella, F. & Young, W. R., 2002. Horizontal convection is non-turbulent. *J. Fluid Mech.*, **466**, 205–214.
- Pedlosky, J., 1964. The stability of currents in the atmosphere and ocean. Part I. *J. Atmos. Sci.*, **21**, 201–219.
- Pedlosky, J., 1987. *Geophysical Fluid Dynamics*. 2nd edn. Springer-Verlag, 710 pp.
- Pedlosky, J., 1996. *Ocean Circulation Theory*. Springer-Verlag, 453 pp.
- Peixoto, J. P. & Oort, A. H., 1992. *Physics of Climate*. American Institute of Physics, 520 pp.
- Peltier, W. R. & Stuhne, G., 2002. The upscale turbulence cascade: shear layers, cyclones and gas giant bands. In R. P. Pearce, Ed., *Meteorology at the Millennium*, pp. 43–61. Academic Press.
- Persson, A., 1998. How do we understand the Coriolis force? *Bull. Am. Meteor. Soc.*, **79**, 1373–1385.
- Philander, S. G., 1990. *El Niño, La Niña, and the Southern Oscillation*. Academic Press, 289 pp.
- Phillips, N. A., 1954. Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level, quasi-geostrophic model. *Tellus*, **6**, 273–286.
- Phillips, N. A., 1956. The general circulation of the atmosphere: a numerical experiment. *Quart. J. Roy. Meteor. Soc.*, **82**, 123–164.
- Phillips, N. A., 1963. Geostrophic motion. *Rev. Geophys.*, **1**, 123–176.
- Phillips, N. A., 1966. The equations of motion for a shallow rotating atmosphere and the traditional approximation. *J. Atmos. Sci.*, **23**, 626–630.
- Phillips, N. A., 1973. Principles of large-scale numerical weather prediction. In P. Morel, Ed., *Dynamic Meteorology*, pp. 1–96. Riedel.
- Phillips, N. A., 2000. Explication of the Coriolis effect. *Bull. Am. Meteor. Soc.*, **81**, 299–303.
- Pierini, S. & Vulpiani, A., 1981. Nonlinear stability analysis in multi-layer quasigeostrophic system. *J. Phys. A*, **14**, L203–L207.
- Pierrehumbert, R. T. & Swanson, K. L., 1995. Baroclinic instability. *Ann. Rev. Fluid Mech.*, **27**, 419–467.
- Plumb, R. A., 1979. Eddy fluxes of conserved quantities by small-amplitude waves. *J. Atmos. Sci.*, **36**, 1699–1704.
- Plumb, R. A., 1990. A nonacceleration theorem for transient quasi-geostrophic eddies on a three-dimensional time-mean flow. *J. Atmos. Sci.*, **47**, 1825–1836.
- Plumb, R. A., 2002. Stratospheric transport. *J. Meteor. Soc. Japan*, **80**, 793–809.
- Plumb, R. A. & Ferrari, R., 2005. Transformed Eulerian-mean theory. Part I: Non-quasigeostrophic theory for eddies on a zonal mean flow. *J. Phys. Oceanogr.*, **35**, 165–174.

- Poincaré, H., 1893. *Théorie des Tourbillons (Theory of Vortices [literally, Swirls])*. Georges Carré, Éditeur. Reprinted by Éditions Jacques Gabay, 1990, 211 pp.
- Poincaré, H., 1908. *Science and Method*. T. Nelson and Sons. Engl. transl.: F. Maitland. Reprinted in *The Value of Science: Essential Writings of Henri Poincaré*, Ed. S. J. Gould, Random House, 584 pp.
- Polzin, K. L., Toole, J. M., Ledwell, J. R. & Schmidt, R. W., 1997. Spatial variability of turbulent mixing in the abyssal ocean. *Science*, **276**, 93–96.
- Price, J. F., Weller, R. A. & Schudlich, R. R., 1987. Wind-driven ocean currents and Ekman transport. *Science*, **238**, 1534–1538.
- Proudman, J., 1916. On the motion of solids in liquids. *Proc. Roy. Soc. Lond. A*, **92**, 408–424.
- Quon, C. & Ghil, M., 1992. Multiple equilibria in thermosaline convection due to salt-flux boundary conditions. *J. Fluid Mech.*, **245**, 449–484.
- Rayleigh, Lord, 1880. On the stability, or instability, of certain fluid motions. *Proc. London Math. Soc.*, **11**, 57–70.
- Rayleigh, Lord, 1894. *The Theory of Sound, Volume II*. 2nd edn. Macmillan, 522 pp. Reprinted by Dover Publications, 1945.
- Rayleigh, Lord, 1912. On the propagation of waves through a stratified medium, with special reference to the question of reflection. *Proc. Roy. Soc. Lond. A*, **86**, 207–226.
- Read, P. L., 2001. Transition to geostrophic turbulence in the laboratory, and as a paradigm in atmospheres and oceans. *Surveys Geophys.*, **33**, 265–317.
- Reif, F., 1965. *Fundamentals of Statistical and Thermal Physics*. McGraw-Hill, 651 pp.
- Rhines, P. B., 1975. Waves and turbulence on a β -plane. *J. Fluid. Mech.*, **69**, 417–443.
- Rhines, P. B., 1977. The dynamics of unsteady currents. In E. A. Goldberg, I. N. McCane, J. J. O'Brien, & J. H. Steele, Eds., *The Sea*, Vol. 6, pp. 189–318. J. Wiley and Sons.
- Rhines, P. B., 1979. Geostrophic turbulence. *Ann. Rev. Fluid Mech.*, **11**, 401–441.
- Rhines, P. B. & Holland, W. R., 1979. A theoretical discussion of eddy-driven mean flows. *Dyn. Atmos. Oceans*, **3**, 289–325.
- Rhines, P. B. & Young, W. R., 1982a. Homogenization of potential vorticity in planetary gyres. *J. Fluid Mech.*, **122**, 347–367.
- Rhines, P. B. & Young, W. R., 1982b. A theory of wind-driven circulation. I. Mid-ocean gyres. *J. Mar. Res. (Suppl)*, **40**, 559–596.
- Richardson, L. F., 1920. The supply of energy from and to atmospheric eddies. *Proc. Roy. Soc. Lond. A*, **97**, 354–373.
- Richardson, L. F., 1922. *Weather Prediction by Numerical Process*. Cambridge University Press, 236 pp. Reprinted by Dover Publications.
- Richardson, L. F., 1926. Atmospheric diffusion on a distance-neighbour graph. *Proc. Roy. Soc. Lond. A*, **110**, 709–737.
- Richardson, P. L., 1983. Eddy kinetic-energy in the North Atlantic from surface drifters. *J. Geophys. Res.*, **88**, NC7, 4355–4367.
- Riehl, H. & Fultz, D., 1957. Jet stream and long waves in a steady rotating-dishpan experiment: structure of the circulation. *Quart. J. Roy. Meteor. Soc.*, **82**, 215–231.
- Rintoul, S. R., Hughes, C. & Olbers, D., 2001. The Antarctic Circumpolar Current system. In G. Siedler, J. Church, & J. Gould, Eds., *Ocean Circulation and Climate*, pp. 271–302. Academic Press.
- Ripa, P., 1981. On the theory of nonlinear wave-wave interactions among geophysical waves. *J. Fluid Mech.*, **103**, 87–115.
- Robinson, A. R., Ed., 1984. *Eddies in Marine Science*. Springer-Verlag, 609 pp.
- Robinson, A. R. & McWilliams, J. C., 1974. The baroclinic instability of the open ocean. *J. Phys. Oceanogr.*, **4**, 281–294.

- Robinson, A. R. & Stommel, H., 1959. The oceanic thermocline and the associated thermohaline circulation. *Tellus*, **11**, 295–308.
- Rooth, C., 1982. Hydrology and ocean circulation. *Prog. Oceanogr.*, **11**, 131–149.
- Rossby, C.-G., 1936. Dynamics of steady ocean currents in the light of experimental fluid dynamics. *Papers Phys. Oceanogr. Meteor.*, **5**, 1–43.
- Rossby, C.-G., 1938. On the mutual adjustment of pressure and velocity distributions in certain simple current systems, II. *J. Mar. Res.*, **5**, 239–263.
- Rossby, C.-G., 1940. Planetary flow patterns in the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **66**, suppl., 68–87.
- Rossby, C.-G., 1949. On the nature of the general circulation of the lower atmosphere. In G. P. Kuiper, Ed., *The Atmospheres of the Earth and Planets*, pp. 16–48. University of Chicago Press.
- Rossby, H. T., 1965. On thermal convection driven by non-uniform heating from below: an experimental study. *Deep-Sea Res.*, **12**, 9–16.
- Rossby, H. T., 1998. Numerical experiments with a fluid heated non-uniformly from below. *Tellus A*, **50**, 242–257.
- Rudnick, D. L. & Weller, R. A., 1993. Observations of superinertial and near-inertial wind-driven flow. *J. Phys. Oceanogr.*, **23**, 2351–2359.
- Ruelle, D. & Takens, F., 1971. On the nature of turbulence. *Commun. Math. Phys.*, **20**, 167–192.
- Salmon, R., 1980. Baroclinic instability and geostrophic turbulence. *Geophys. Astrophys. Fluid Dyn.*, **10**, 25–52.
- Salmon, R., 1983. Practical use of Hamilton's principle. *J. Fluid Mech.*, **132**, 431–444.
- Salmon, R., 1990. The thermocline as an internal boundary layer. *J. Mar. Res.*, **48**, 437–469.
- Salmon, R., 1998. *Lectures on Geophysical Fluid Dynamics*. Oxford University Press, 378 pp.
- Saltzman, B., 1962. Finite amplitude free convection as an initial value problem. *J. Atmos. Sci.*, **19**, 329–341.
- Samelson, R. M., 1999a. Geostrophic circulation in a rectangular basin with a circumpolar connection. *J. Phys. Oceanogr.*, **29**, 3175–3184.
- Samelson, R. M., 1999b. Internal boundary layer scaling in 'two-layer' solutions of the thermocline equations. *J. Phys. Oceanogr.*, **29**, 2099–2102.
- Samelson, R. M., 2004. Simple mechanistic models of middepth meridional overturning. *J. Phys. Oceanogr.*, **34**, 2096–2103.
- Samelson, R. M. & Vallis, G. K., 1997. Large-scale circulation with small diapycnal diffusion: the two-thermocline limit. *J. Mar. Res.*, **55**, 223–275.
- Sandström, J. W., 1908. Dynamische Versuche mit Meerwasser (Dynamical experiments with seawater). *Annal. Hydrogr. Marit. Meteorol.*, **36**, 6–23.
- Sandström, J. W., 1916. Meteorologische Studien im Schwedischen Hochgebirge (Meteorological studies in the Swedish high mountains). *Göteborgs Kungl. Vetenskaps-och Vitterhets-Samhälles, Handlingar*, **27**, 1–48.
- Sarkisyan, A. & Ivanov, I., 1971. Joint effect of baroclinicity and relief as an important factor in the dynamics of sea currents. *Izv. Akad. Nauk Atmos. Ocean. Phys.*, **7**, 116–124.
- Schär, C., 1993. A generalization of Bernoulli's theorem. *J. Atmos. Sci.*, **50**, 1437–1443.
- Schmitz, W. J., 1995. On the interbasin-scale thermohaline circulation. *Rev. Geophysics*, **33**, 151–173.
- Schneider, E. K., 1977. Axially symmetric steady-state models of the basic state for instability and climate studies. Part II: nonlinear calculations. *J. Atmos. Sci.*, **34**, 280–297.
- Schneider, E. K. & Lindzen, R. S., 1977. Axially symmetric steady-state models of the basic state for instability and climate studies. Part I: linearized calculations. *J. Atmos. Sci.*, **34**, 263–279.
- Schneider, T., Held, I. & Garner, S. T., 2003. Boundary effects in potential vorticity dynamics. *J. Atmos. Sci.*, **60**, 1024–1040.

- Schneider, T. & Sobel, A., Eds., 2007. *The Global Circulation of the Atmosphere: Phenomena, Theory, Challenges*. Princeton University Press. To appear.
- Schneider, T. & Walker, C. C., 2006. Self-organization of atmospheric macroturbulence into critical states of weak nonlinearity. *J. Atmos. Sci.*, **63**, 1569–1586.
- Schubert, W. H., Hausman, S. A., Garcia, M., Ooyama, K. V. & Kuo, H.-C., 2001. Potential vorticity in a moist atmosphere. *J. Atmos. Sci.*, **58**, 3148–3157.
- Schubert, W. H., Ruprecht, E., Hertenstein, R., Nieto Ferreira, R. *et al.*, 2004. English translations of twenty-one of Ertel's papers on geophysical fluid dynamics. *Meteor. Z.*, **13**, 527–576.
- Scinocca, J. F. & Shepherd, T. G., 1992. Nonlinear wave-activity conservation laws and Hamiltonian structure for the two-dimensional anelastic equations. *J. Atmos. Sci.*, **49**, 5–28.
- Scott, R. B., 2001. Evolution of energy and enstrophy containing scales in decaying, two-dimensional turbulence with friction. *Phys. Fluids*, **13**, 2739–2742.
- Shapiro, M. & Grønås, S., Eds., 1999. *The Life Cycles of Extratropical Cyclones*. American Meteorological Society, 359 pp.
- Shepherd, T. G., 1983. Mean motions induced by baroclinic instability in a jet. *Geophys. Astrophys. Fluid Dyn.*, **27**, 35–72.
- Shepherd, T. G., 1987. A spectral view of nonlinear fluxes and stationary-transient interaction in the atmosphere. *J. Atmos. Sci.*, **44**, 1166–1179.
- Shepherd, T. G., 1990. Symmetries, conservation laws, and Hamiltonian structure in geophysical fluid dynamics. *Adv. Geophys.*, **32**, 287–338.
- Shepherd, T. G., 1993. A unified theory of available potential energy. *Atmosphere–Ocean*, **31**, 1–26.
- Shepherd, T. G., 2002. Issues in stratosphere–troposphere coupling. *J. Meteor. Soc. Japan*, **80**, 769–792.
- Shutts, G. J., 1983. Propagation of eddies in diffluent jet streams: eddy vorticity forcing of blocking flow fields. *Quart. J. Roy. Meteor. Soc.*, **109**, 737–761.
- Siedler, G., Church, J. & Gould, J., 2001. *Ocean Circulation and Climate: Observing and Modelling the Global Ocean*. Academic Press, 715 pp.
- Silberstein, L., 1896. O tworzeniu sie wirow, w plynne doskonalym (On the creation of eddies in an ideal fluid). *W Krakowie Nakladem Akademii Umiejetnosci (Proc. Cracow Acad. Sci.)*, **31**, 325–335. Also published as: Über die Entstehung von Wirbelbewegungen in einer reibungslosen Flüssigkeit in *Bull. Int. l'Acad. Sci. Cracovie, Compte Rendue Scéances Année*, 280–290, 1896. Engl. transl. by M. Ziemiánski available from the author.
- Simmons, A. & Hoskins, B., 1978. The life-cycles of some nonlinear baroclinic waves. *J. Atmos. Sci.*, **35**, 414–432.
- Smagorinsky, J., 1953. The dynamical influences of large-scale heat sources and sinks on the quasi-stationary mean motions of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **79**, 342–366.
- Smagorinsky, J., 1969. Problems and promises of deterministic extended range forecasting. *Bull. Am. Meteor. Soc.*, **50**, 286–311.
- Smith, K. S., Boccaletti, G., Henning, C. C., Marinov, I. *et al.*, 2002. Turbulent diffusion in the geostrophic inverse cascade. *J. Fluid Mech.*, **469**, 13–48.
- Smith, K. S. & Vallis, G. K., 1998. Linear wave and instability properties of extended range geostrophic models. *J. Atmos. Sci.*, **56**, 1579–1593.
- Smith, K. S. & Vallis, G. K., 2001. The scales and equilibration of mid-ocean eddies: freely evolving flow. *J. Phys. Oceanogr.*, **31**, 554–571.
- Smith, K. S. & Vallis, G. K., 2002. The scales and equilibration of mid-ocean eddies: forced-dissipative flow. *J. Phys. Oceanogr.*, **32**, 1669–1721.
- Smith, L. M. & Waleffe, F., 1999. Transfer of energy to two-dimensional large scales in forced, rotating three-dimensional turbulence. *Phys. Fluids*, **11**, 1608–1622.

- Smith, R. B., 1979. The influence of mountains on the atmosphere. In B. Saltzman, Ed., *Advances in Geophysics*, vol. 21, pp. 87–230. Academic Press.
- Spall, M. A., 2000. Generation of strong mesoscale eddies by weak ocean gyres. *J. Mar. Res.*, **58**, 97–116.
- Spiegel, E. A. & Veronis, G., 1960. On the Boussinesq approximation for a compressible fluid. *Astrophys. J.*, **131**, 442–447. (Correction: *Astrophys. J.*, **135**, 655–656).
- Squire, H., 1933. On the stability of three-dimensional disturbances of viscous flow between parallel walls. *Proc. Roy. Soc. Lond. A*, **142**, 621–628.
- Stammer, D., 1997. Global characteristics of ocean variability estimated from regional TOPEX/Poseidon altimeter measurements. *J. Phys. Oceanogr.*, **27**, 1743–1769.
- Starr, V. P., 1948. An essay on the general circulation of the Earth's atmosphere. *J. Meteor.*, **78**, 39–43.
- Starr, V. P., 1968. *Physics of Negative Viscosity Phenomena*. McGraw-Hill, 256 pp.
- Steers, J. A., 1962. *An Introduction to the Study of Map Projections*. Univ. of London Press, 288 pp.
- Stevens, D. P. & Ivchenko, V. O., 1997. The zonal momentum balance in an eddy-resolving general-circulation model of the southern ocean. *Quart. J. Roy. Meteor. Soc.*, **123**, 929–951.
- Stewart, G. R., 1941. *Storm*. Random House, 349 pp.
- Stips, A., 2005. Dissipation measurement: theory. In H. Z. Baumert, J. Simpson, & J. Sündermann, Eds., *Marine Turbulence*. Cambridge University Press.
- Stommel, H., 1958. The abyssal circulation. *Deep-Sea Res.*, **5**, 80–82.
- Stommel, H., 1961. Thermohaline convection with two stable regimes of flow. *Tellus*, **13**, 224–230.
- Stommel, H. & Arons, A. B., 1960. On the abyssal circulation of the world ocean—I. Stationary planetary flow patterns on a sphere. *Deep-Sea Res.*, **6**, 140–154.
- Stommel, H., Arons, A. B. & Faller, A. J., 1958. Some examples of stationary planetary flow patterns in bounded basins. *Tellus*, **10**, 179–187.
- Stommel, H. & Moore, D. W., 1989. *An Introduction to the Coriolis Force*. Columbia University Press, 297 pp.
- Stommel, H. & Webster, J., 1963. Some properties of the thermocline equations in a subtropical gyre. *J. Mar. Res.*, **44**, 695–711.
- Stone, P. H., 1972. A simplified radiative-dynamical model for the static stability of rotating atmospheres. *J. Atmos. Sci.*, **29**, 405–418.
- Stone, P. H., 1978. Baroclinic adjustment. *J. Atmos. Sci.*, **35**, 561–571.
- Stone, P. H. & Nemet, B., 1996. Baroclinic adjustment: a comparison between theory, observations, and models. *J. Atmos. Sci.*, **53**, 1663–1674.
- Straub, D. N., 1993. On the transport and angular momentum balance of channel models of the Antarctic Circumpolar Current. *J. Phys. Oceanogr.*, **23**, 776–782.
- Sutcliffe, R. C., 1939. Cyclonic and anticyclonic development. *Quart. J. Roy. Meteor. Soc.*, **65**, 518–524.
- Sutcliffe, R. C., 1947. A contribution to the problem of development. *Quart. J. Roy. Meteor. Soc.*, **73**, 370–383.
- Swallow, J. C. & Worthington, V., 1961. An observation of a deep countercurrent in the western North Atlantic. *Deep-Sea Res.*, **8**, 1–19.
- Talley, L., 1988. Potential vorticity distribution in the North Pacific. *J. Phys. Oceanogr.*, **18**, 89–106.
- Taylor, G. I., 1921a. Diffusion by continuous movements. *Proc. London Math. Soc.*, **2** (20), 196–211.
- Taylor, G. I., 1921b. Experiments with rotating fluids. *Proc. Roy. Soc. Lond. A*, **100**, 114–121.
- Tennekes, H. & Lumley, J. L., 1972. *A First Course in Turbulence*. The MIT Press, 330 pp.
- Tesserenc De Bort, L. P., 1902. Variations de la température de l'air libre dans la zone comprise 8 km et 13 km d'altitude (Variations in the temperature of the free air in the zone between 8 km and 13 km of altitude). *C. R. Hebd. Séances Acad. Sci.*, **134**, 987–989.

- Thompson, A. F. & Young, W. R., 2006. Scaling baroclinic eddy fluxes: vortices and energy balance. *J. Phys. Oceanogr.*, **36**, 720–738.
- Thompson, P. D., 1957. Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus*, **9**, 275–295.
- Thompson, R. O. R. Y., 1971. Why there is an intense eastward current in the North Atlantic but not in the South Atlantic. *J. Phys. Oceanogr.*, **1**, 235–237.
- Thompson, R. O. R. Y., 1980. A prograde jet driven by Rossby waves. *J. Atmos. Sci.*, **37**, 1216–1226.
- Thomson, J., 1892. Bakerian lecture. On the grand currents of atmospheric circulation. *Phil. Trans. Roy. Soc. Lond. A*, **183**, 653–684.
- Thomson, W. (Lord Kelvin), 1869. On vortex motion. *Trans. Roy. Soc. Edinburgh*, **25**, 217–260.
- Thomson, W. (Lord Kelvin), 1871. Hydrokinetic solutions and observations. *Phil. Mag. and J. Science*, **42**, 362–377.
- Thomson, W. (Lord Kelvin), 1879. On gravitational oscillations of rotating water. *Proc. Roy. Soc. Edinburgh*, **10**, 92–100.
- Thorncroft, C. D., Hoskins, B. J. & McIntyre, M. E., 1993. Two paradigms of baroclinic-wave life-cycle behaviour. *Quart. J. Roy. Meteor. Soc.*, **119**, 17–55.
- Thorpe, A. J., Volkert, H. & Ziemienski, M. J., 2003. The Bjerknes' circulation theorem: a historical perspective. *Bull. Am. Meteor. Soc.*, **84**, 471–480.
- Thual, O. & McWilliams, J. C., 1992. The catastrophe structure of thermohaline convection in a two-dimensional fluid model and a comparison with low-order box models. *Geophys. Astrophys. Fluid Dyn.*, **64**, 67–95.
- Thuburn, J. & Craig, G. C., 1997. GCM tests of theories for the height of the tropopause. *J. Atmos. Sci.*, **54**, 869–882.
- Toggweiler, J. R. & Samuels, B., 1995. Effect of Drake Passage on the global thermohaline circulation. *Deep-Sea Res.*, **42**, 477–500.
- Toggweiler, J. R. & Samuels, B., 1998. On the ocean's large-scale circulation in the limit of no vertical mixing. *J. Phys. Oceanogr.*, **28**, 1832–1852.
- Toole, J. M., Polzin, K. L. & Schmitt, R. W., 1994. Estimates of diapycnal mixing in the abyssal ocean. *Science*, **264**, 1120–1123.
- Tréguier, A. M., Held, I. M. & Lariehev, V. D., 1997. Parameterization of quasi-geostrophic eddies in primitive equation ocean models. *J. Phys. Oceanogr.*, **29**, 567–580.
- Trenberth, K. E. & Caron, J. M., 2001. Estimates of meridional atmosphere and ocean heat transports. *J. Climate*, **14**, 3433–3443.
- Tritton, D. J., 1988. *Physical Fluid Dynamics*. Oxford University Press, 519 pp.
- Truesdell, C., 1951. Proof that Ertel's vorticity theorem holds in average for any medium suffering no tangential acceleration on the boundary. *Geofis Pura Appl.*, **19**, 167–169.
- Truesdell, C., 1954. *The Kinematics of Vorticity*. Indiana University Press, 232 pp.
- Tung, K. K., 1979. A theory of stationary long waves. Part III: quasi-normal modes in a singular wave guide. *Mon. Wea. Rev.*, **107**, 751–774.
- UNESCO, 1981. The practical salinity scale 1978 and the international equation of state of seawater 1980. Tenth report of the joint panel on oceanographic tables and standards. Technical report, UNESCO Technical Papers in Marine Science No. 36, Paris.
- Valdes, P. J. & Hoskins, B. J., 1989. Baroclinic instability of the zonally averaged flow with boundary layer damping. *J. Atmos. Sci.*, **45**, 1584–1593.
- Vallis, G. K., 1982. A statistical dynamical climate model with a simple hydrology cycle. *Tellus*, **34**, 211–227.
- Vallis, G. K., 1985. Instability and flow over topography. *Geophys. Astrophys. Fluid Dyn.*, **34**, 1–38.

- Vallis, G. K., 1988. Numerical studies of eddy transport properties in eddy-resolving and parameterized models. *Quart. J. Roy. Meteor. Soc.*, **114**, 183–204.
- Vallis, G. K., 1996. Potential vorticity and balanced equations of motion for rotating and stratified flows. *Quart. J. Roy. Meteor. Soc.*, **122**, 291–322.
- Vallis, G. K., 2000. Large-scale circulation and production of stratification: effects of wind, geometry and diffusion. *J. Phys. Oceanogr.*, **30**, 933–954.
- Vallis, G. K. & Malmstrom, M. E., 1993. Generation of mean flows and jets on a beta plane and over topography. *J. Phys. Oceanogr.*, **23**, 1346–1362.
- Vanneste, J. & Shepherd, T. G., 1998. On the group-velocity property for wave-activity conservation laws. *J. Atmos. Sci.*, **55**, 1063–1068.
- Veronis, G., 1966a. Wind-driven ocean circulation – Part 1: linear theory and perturbation analysis. *Deep-Sea Res.*, **13**, 17–29.
- Veronis, G., 1966b. Wind-driven ocean circulation – Part 2: numerical solutions of the non-linear problem. *Deep-Sea Res.*, **13**, 30–55.
- Veronis, G., 1969. On theoretical models of the thermocline circulation. *Deep-Sea Res.*, **31** Suppl., 301–323.
- Visbeck, M., Marshall, J., Haine, T. & Spall, M., 1997. Specification of eddy transfer coefficients in coarse-resolution ocean circulation models. *J. Phys. Oceanogr.*, **27**, 381–402.
- Walker, C. & Schneider, T., 2005. Response of idealized Hadley circulations to seasonally varying heating. *Geophys. Res. Lett.*, **32**, L06813. doi:10.1029/2004GL022304.
- Wallace, J. M., 1983. The climatological mean stationary waves: observational evidence. In B. Hoskins & R. P. Pearce, Eds., *Large-Scale Dynamical Processes in the Atmosphere*, pp. 27–63. Academic Press.
- Warn, T., Bokhove, O., Shepherd, T. G. & Vallis, G. K., 1995. Rossby number expansions, slaving principles, and balance dynamics. *Quart. J. Roy. Meteor. Soc.*, **121**, 723–739.
- Warren, B. A., 1981. Deep circulation of the world ocean. In B. A. Warren & C. Wunsch, Eds., *Evolution of Physical Oceanography*, pp. 6–41. MIT Press.
- Warren, B. A., LaCasce, J. H. & Robbins, P. E., 1996. On the obscurantist physics of form drag in theorizing about the circumpolar current. *J. Phys. Oceanogr.*, **26**, 2297–2301.
- Wasow, W., 1944. Asymptotic solution of boundary value problems for the differential equation $\Delta U + \lambda(\partial/\partial x)U = \lambda f(x, y)$. *Duke Math J.*, **11**, 405–415.
- Watanabe, T., Iwayama, T. & Fujisaka, H., 1998. Scaling law for coherent vortices in decaying drift Rossby wave turbulence. *Phys. Rev. E*, **57**, 1636–1643.
- Webb, D. J. & Sugihara, N., 2001. Vertical mixing in the ocean. *Nature*, **409**, 37.
- Weinstock, R., 1952. *Calculus of Variations*. McGraw-Hill. Reprinted by Dover Publications, 1980, 328 pp.
- Welander, P., 1959. An advective model of the ocean thermocline. *Tellus*, **11**, 309–318.
- Welander, P., 1968. Wind-driven circulation in one- and two-layer oceans of variable depth. *Tellus*, **20**, 1–15.
- Welander, P., 1971a. Some exact solutions to the equations describing an ideal-fluid thermocline. *J. Mar. Res.*, **29**, 60–68.
- Welander, P., 1971b. The thermocline problem. *Phil. Trans. Roy. Soc. Lond. A*, **270**, 415–421.
- Welander, P., 1973. Lateral friction in the ocean as an effect of potential vorticity mixing. *Geophys. Fluid Dyn.*, **5**, 101–120.
- Welander, P., 1986. Thermohaline effects in the ocean circulation and related simple models. In J. Willebrand & D. L. T. Anderson, Eds., *Large-scale Transport Processes in Oceans and Atmospheres*, pp. 163–200. Reidel.
- Wentzel, G., 1926. Eine Verallgemeinerung der Quantenbedingungen für die Zwecke der Wellen-

- mechanik (A generalization of the quantum conditions for the purposes of wave mechanics). *Zeit. fur Physic A*, **38**, 518–529.
- White, A. A., 1977. Modified quasi-geostrophic equations using geometric height as vertical coordinate. *Quart. J. Roy. Meteor. Soc.*, **103**, 383–396.
- White, A. A., 2002. A view of the equations of meteorological dynamics and various approximations. In J. Norbury & I. Roulstone, Eds., *Large-Scale Atmosphere-Ocean Dynamics I*, pp. 1–100. Cambridge University Press.
- White, A. A., 2003. The primitive equations. In J. Holton, J. Pyle, & J. Curry, Eds., *Encyclopedia of Atmospheric Science*, pp. 694–702. Academic Press.
- Whitehead, J. A., 1975. Mean flow generated by circulation on a beta-plane: An analogy with the moving flame experiment. *Tellus*, **27**, 358–364.
- Whitehead, J. A., 1995. Thermohaline ocean processes and models. *Ann. Rev. Fluid Mech.*, **27**, 89–113.
- Williams, G. P., 1978. Planetary circulations: 1. Barotropic representation of Jovian and terrestrial turbulence. *J. Atmos. Sci.*, **35**, 1399–1426.
- World Meteorological Organization, 1957. Meteorology: a three dimensional science. Second session of the Commission for Aerology. *WMO Bulletin*, **IV**, 134–138.
- Wright, D. G., 1997. An equation of state for use in ocean models: Eckart's formula revisited. *J. Atmos. Ocean. Tech.*, **14**, 735–740.
- Wunsch, C., 2002. What is the thermohaline circulation? *Science*, **298**, 1179–1180.
- Wunsch, C. & Ferrari, R., 2004. Vertical mixing, energy, and the general circulation of the oceans. *Ann. Rev. Fluid Mech.*, **36**, 281–314.
- Wunsch, C. & Roemmich, D., 1985. Is the North Atlantic in Sverdrup balance? *J. Phys. Oceanogr.*, **15**, 1876–1880.
- Wyrki, K., Magaard, L. & Hager, J., 1976. Eddy energy in oceans. *J. Geophys. Res.*, **81**, 15, 2641–2646.
- Yaglom, A. M., 1994. A. N. Kolmogorov as a fluid mechanician and founder of a school in turbulence research. *Ann. Rev. Fluid Mech.*, **26**, 1–22.
- Young, W. R., 2006. Boussinesq notes. Unpublished.
- Young, W. R. & Rhines, P. B., 1982. A theory of the wind-driven circulation II. Gyres with western boundary layers. *J. Mar. Res.*, **40**, 849–872.
- Zdunkowski, W. & Bott, A., 2003. *Dynamics of the Atmosphere: A Course in Theoretical Meteorology*. Cambridge University Press, 719 pp.
- Zhang, R. & Vallis, G. K., 2007. The role of the bottom vortex stretching on the path of the north atlantic western boundary current and on the northern recirculation gyre. *J. Phys. Oceanogr.*, (in press).
- Zurita-Gotor, P. & Lindzen, R., 2007. Theories of baroclinic adjustment and eddy equilibration. In T. Schneider & A. Sobel, Eds., *The Global Circulation of the Atmosphere: Phenomena, Theory, Challenges*. Princeton University Press. (To appear).

Index

Bold face denotes a primary entry or an extended discussion.

Abyssal ocean circulation, **646–662**
wind driven, 691
ACC, 700–710
adiabatic model of, 708
and mesoscale eddies, 701
form drag in, 709
momentum balance, 702, 704, 705
Acoustic-gravity waves, 100–103
Advection derivative, 4
Anelastic approximation, **73–77**
Anelastic equations, 74
energetics of, 76
Angular momentum, 64
spherical coordinates, 65
Antarctic Circumpolar Current, **700–710**
Antisymmetric turbulent diffusivity, 427
APE, 155
Arnold stability conditions, 327
Asymptotic models
conservation properties of, 214
quasi-geostrophy, 216
Atmospheric stratification, 522–536
Auto-barotropic fluid, 15
Available potential energy, **155–159**
Boussinesq fluid, 156
ideal gas, 158

Baroclinic adjustment, 531
Baroclinic circulation theorem, 173
Baroclinic eddies, **395–403**
effect on Hadley Cell, 473
in atmosphere, 397
in ocean, 400
magnitude and scale, 396

Baroclinic eddy diffusivities, 430
Baroclinic Fluid, 15
Baroclinic instability, **247, 261–282**
beta effect in continuous model, 284
beta effect in two-layer model, 274
Eady problem, 265
effect of stratosphere, 288
energetics of, 282
high-wavenumber cut-off, 256, 274, 332, 390
in ocean, 289
interacting edge waves, 277
linear QG equations, 263
mechanism of, 261, 277
minimum shear, 275
necessary conditions for, 264, 330, 331
neutral curve in two-layer problem, 276
non-uniform shear and stratification, 288
sloping convection, 261
two-layer problem, 271
Baroclinic lifecycle, 398
Baroclinic lifecycles
in atmosphere, 397
in ocean, 400, 403
Baroclinic term, 166
Baroclinic triads, 388
Barotropic fluid, 15, 20
Barotropic instability, **247**
Barotropic jet, **486–496**
and Rossby waves, 489
and the EP flux, 494
numerical example, 495
Barotropic triads, 388
Batchelor scale, 368

- Batchelor spectrum, 368
Bernoulli function, 43
Bernoulli's theorem, 43
and potential vorticity flux, 191
Beta effect, 175, 177
in two-dimensional turbulence, 377
Beta plane vorticity equation, 178
Beta scale, 378, 379
Beta-plane approximation, 67
Beta-Rossby number, 508, 588
Bjerknes, Jacob, 193
Bjerknes, Vilhelm, 193
Bjerknes-Silberstein circulation theorem, 173
Bolus velocity, 437, 441
Bottom pressure stress, 611
Boussinesq approximation, 67–73
Boussinesq equations, 68
energetics of, 72
potential vorticity conservation, 186
relation to pressure coordinates, 79
summary, 72
Box ocean models, 640–645
many boxes, 644
two boxes, 640
Bretherton's boundary layer, 220
Brewer–Dobson circulation, 568–571, 579
Brunt–Väisälä frequency, 92
Buoyancy frequency, 71, 92
ideal gas, 94
ocean, 94
Buoyancy-driven ocean circulation, 627, 659
Burger number, 200

Cabbeling, 95
Centrifugal force, 53, 54
Chaos, 362
Charney–Drazin filtering, 557
Charney–Drazin problem, 554
Charney–Eliassen problem, 542
Charney–Green number, 286
Charney–Stern–Pedlosky criterion for instability, 265
Chemical potential, 17, 18
Circulation, 163–174
Circulation theorem
baroclinic, 173
barotropic fluid, 171
hydrostatic flow, 174
Closure problem of turbulence, 338
Compressible flow, 38
Concentration and mixing ratio, 11
Convective instability, 99
Coriolis acceleration, 53
Coriolis force, 53–54
Coriolis, Gaspard Gustave de, 115

Deacon Cell, 696, 707
Deacon cell, 696
Deformation radius, 143
Diffusion
equation of, 408
turbulent, 408
Diffusive thermocline, 632, 674–680
Diffusive transport, 408
Diffusivity tensors, 426
Dispersion relation, 236
Rossby waves, 230
Downward control, 571
Dry adiabatic lapse rate, 95
Dumbbell in beta-plane turbulence, 381

Eady problem, 265–271
eddy effect on mean flow, 319
secondary circulation, 319
Eddy diffusion, 409
two-dimensional, 418
Eddy transport
and the TEM, 443
velocity, 437
Edge waves, 254, 256
Eady problem, 280
in shear flows, 254
Effective gravity, 55
Ekman layers, 104–115
integral properties of, 107
momentum balance, 106
observed, 113
stress in, 104
Ekman number, 105
Ekman spiral, 109, 112
Eliassen–Palm flux, 298–304
and barotropic jets, 494
and form drag, 317
observed, 519
primitive equations, 536
spherical coordinates, 536
Eliassen–Palm relation, 299
Energetics
of quasi-geostrophic equations, 226
Energy budget, 40
constant density fluid, 40
variable density fluid, 42
viscous effects, 43
Energy conservation
Boussinesq equations, 72
primitive equations, 62, 119
shallow water equations, 139
Energy inertial range
in two-dimensional turbulence, 357
Energy transfer in two-dimensional flow, 351
Enstrophy inertial range, 356
Enstrophy transfer in two-dimensional flow, 351
Enthalpy, 19
ideal gas, 21
Entropy, 16
Equation of State, 14–15
Equation of state
fundamental, 16, 20

- ideal gas, 14
seawater, 14, 35
- Equivalent potential temperature, 97
- Equivalent topography, 562
- Euler equations, 32
- Euler, Leonard, 45
- Eulerian derivative, 4
- Eulerian viewpoint, 4
- Exner function, 117, 154
- f-plane approximation, 66
- Ferrel Cell, 480–482, 522
eddy fluxes in, 481
surface flow in, 481
- Ferrel, William, 482
- Field or Eulerian viewpoint, 4
- First law of thermodynamics, 17
- Fjørtoft's criterion for instability, 260, 329
- Fluid element, 4
- Fofonoff model, 606
- Form drag, 135–136
and Eliassen Palm flux, 317
at ocean bottom, 611
in ACC, 703, 709
- Four-thirds law, 417
- Free-slip condition, 594
- Frequency, 236
- frictional-geostrophic balance, 105
- Froude number, 83, 200
- Frozen in property of vorticity, 168
- Fundamental equation of state, 16, 20
ideal gas, 26
- Fundamental postulate of thermodynamics, 16
- Fundamental thermodynamic relation, 18
- Gent-McWilliams scheme, 436
- Geopotential surfaces, 57
- Geostrophic adjustment, 144–152
energetics of, 148
Rossby problem, 146
- Geostrophic and thermal wind balance, 85–91
- Geostrophic balance, 86, 134
a variational perspective, 151
frictional, 105
in shallow water equations, 134
pressure coordinates, 90
- Geostrophic contours, 616
- Geostrophic scaling, 198
in continuously stratified equations, 200
in shallow water equations, 198
- Geostrophic turbulence, 377
Larichev–Held model, 391
stratified, 384
two layers, 385
two-dimensional, beta-plane, 377
- Gibbs function, 19
for seawater, 33
- Gravity waves, 98–99
acoustic, 100
- hydrostatic, 99
- Green and Stone turbulent transport, 430
- Group velocity, 236–241
meaning of, 239
property for wave activity, 300
- Gyres, 583
- Hadley Cell, 457–479
angular-momentum-conserving model, 457
effects of eddies on, 473
effects of moisture on, 465
poleward extent, 458
radiative equilibrium solution, 466
seasonal effects and hemispheric asymmetry, 469
shallow water model of, 468
strength of, 458
- Hadley, George, 482
- Haney boundary condition, 630, 663
- Heat capacity, 21
- Held–Hou model of Hadley Cell, 457
- Hide's theorem, 467, 483
- Homentropic fluid, 20
- Homogenization of a tracer, 423–425
- Horizontal convection, 629–640
maintenance of, 634
- Hydrostasy, 13
- Hydrostatic approximation
accuracy, 84
in deriving primitive equations, 61
- Hydrostatic balance, 13, 80–84
effects of rotation, 91
effects of stratification, 82
scaling and the aspect ratio, 81
- Hydrostatic equations
potential vorticity conservation, 187
- Ideal gas, 20
enthalpy, 21
equation of state, 14
fundamental equation of state, 26
heat capacity, 21
simple and general, 20
thermodynamics of, 23
- Impermeability of potential vorticity, 188
- Incompressible flow, 38–40
conditions for, 39
- Inertia circles, 117
- Inertial oscillations, 117, 143
- Inertial range
energy in 3D, 344
scaling argument for, 348
- Inertial range theory, 342
- Inertial ranges
two-dimensional turbulence, 354
- Inertial waves, 118, 161
- Inertial western boundary currents, 601
- Inertial-diffusive range, 369
- Inflection point criterion, 259

- Instability
baroclinic, 247, 261
barotropic, 247
Kelvin–Helmholtz, 248
necessary conditions in baroclinic flow, 264
necessary conditions in shear flow, 258
parallel shear flow, 250
- Intermediate models, 214
- Intermittency, 349
- Internal thermocline, 674–680
- Inverse cascade, 357
- Inversion, 168
of vorticity, 168
- Inviscid western boundary currents, 608–613
- ISENTROPIC coordinates, 152–155
and quasi-geostrophy, 225
- Boussinesq fluid, 152
- ideal gas, 153
- Isopycnal coordinates, 152
- JEBAR, joint effect of baroclinicity and relief, 612
- Jets, 380–381, 486–496
and the pseudomomentum budget, 492
and the vorticity budget, 487
atmospheric, 486
eddy-driven, 486
in beta-plane turbulence, 380
numerical simulation of, 381
- Joint effect of baroclinicity and relief, 612
- Joint effect of beta and friction, 382
- Jump conditions, 251
- JWKB method, 552
- K41 theory, 341
- Kelvin waves, 143
- Kelvin's circulation theorem, 171
- Kelvin–Helmholtz instability, 248, 253
- Kinematic stress, 105
- Kinematic viscosity, 13
- Kinematics
of waves, 236
- Kolmogorov scale, 346
- Kolmogorov theory, 341–349
- Kolmogorov, A. N., 371
- Lagrange, Joseph-Louis, 45
- Lagrangian derivative, 4
- Lagrangian viewpoint, 4
- Lamb waves, 103
- Lapse rate, 95
adiabatic, of density, 27
adiabatic, of temperature, 29
dry adiabatic, 95
ideal gas, 95
of seawater, 34
saturated, 96
- Larichev–Held model of geostrophic turbulence, 391
- Lifecycle of baroclinic waves, 398
in atmosphere, 397
- in ocean, 400, 403
- Lifting condensation level, 97
- Lindzen–Hou model of Hadley Cell, 469
- Locality in turbulence, 349
- Log-pressure coordinates, 80
- Lorenz equations, 362
- LPS model, 681
- Luyten–Pedlosky–Stommel model, 681
- M equation, 674
one-dimensional model, 675
- Mach number, 40
- Macro-turbulence, 377
- Main thermocline, 667
- Margules relation, 135
- Mass continuity, 7–11
- Mass continuity equation
shallow water, 125
- Material derivative, 4–7
finite volume, 6
fluid property, 5
- Material viewpoint, 4
- Maxwell relations, 19, 20
- Meridional overturning circulation, 627
atmospheric, Eulerian, 456
of atmosphere, 456
of ocean, 628, 691
- Mid-latitude atmospheric circulation, 496–522
- Minimum shear for baroclinic instability, 275
- Mixing length theory, 419–423
- Mixing ratio and concentration, 11
- Moist adiabatic lapse rate, 96
- Moist convection
and tropospheric stratification, 534
- Moisture
effect on potential vorticity conservation, 183
effects on Hadley Cell, 465
- Momentum equation, 11–14
in a rotating frame of reference, 54
shallow water, 124
vector invariant form, 63
- Montgomery potential, 154
- Multi-layer QG equations, 213
- Munk wind-driven model, 593
- Navier Stokes equations, 32
- Navier, Claude, 46
- Necessary conditions for baroclinic instability, 330, 331
- Necessary conditions for instability, 324–333
baroclinic flow, 264
- Charney–Stern–Pedlosky criterion, 264, 325
- Fjørtoft's criterion, 260, 329
- Rayleigh–Kuo criterion, 259, 325
- relation to eddy fluxes, 514
- shear flow, 258
- use of pseudoenergy, 327
- use of pseudomomentum, 325

- No-slip condition, 594
- Non-acceleration theorem, 314–319
- Non-dimensionalization, 43–45
in rotating flow, 198
- Non-homentropic term, 166
- Oblate spheroid, 55, 57
- Observations
- Atlantic Ocean, 629
 - atmospheric meridional overturning circulation, 456, 523
 - atmospheric stratification, mean, 524
 - atmospheric wind and temperature, 453
 - deep ocean circulation, 628
 - deep western boundary current, 661
 - Ekman layers, 113
 - Eliassen–Palm flux, 519, 521
 - Eliassen–Palm flux divergence, 521
 - global ocean currents, 584
 - main thermocline, 669
 - North Atlantic, 661
 - North Atlantic currents, 585
 - ocean stratification, 628
 - oceanic meridional overturning circulation, 628
 - of the atmosphere, 452
 - reanalysis, 482
 - surface winds, 455
 - zonally averaged atmosphere, 456
 - zonally averaged zonal wind, 521
- Ocean circulation
- abyssal, 650
 - laboratory model of, 646
 - scaling for buoyancy-driven, 659
 - wind- and buoyancy-driven, 667
 - wind-driven, 585
 - wind-driven abyssal, 691
- Ocean currents, 584
- Ocean gyres, 583
- Outcropping, 683
- Parcel Method, 91–98
- Passive tracer, 366–370
in three dimensions, 367
in two dimensions, 368
spectra of, 366
- Phillips instability problem, 271
- Piecewise linear flows, 251
- Plane waves, 236
- Planetary waves, 541
- Planetary-geostrophic equations, 203–207
for shallow water flow, 203
for stratified flow, 205
- Planetary-geostrophic potential vorticity equation, 205, 206
shallow water, 205
stratified, 206
- Poincaré waves, 141
- Poincaré, Henri, 160
- Polar vortex, 575
- Polytropic fluid, 15
- Potential density, 24, 31, 33
of liquids, 27
- Potential temperature, 24, 29, 31
equivalent, 97
- ideal gas, 25
 - of liquids, 28
 - seawater, 36
- Potential vorticity, 178–192
and Bernoulli's theorem, 191
and the frozen-in property, 180
- Boussinesq equations, 186
 - concentration, 189
 - conservation of, 178
 - diffusion of, 443
 - for baroclinic fluids, 179
 - for barotropic fluids, 178
 - homogenization of, 688
 - hydrostatic equations, 187
 - impermeability of isentropes, 188
 - mixing, 530, 531
 - on isentropic surfaces, 187
 - planetary-geostrophic, 205
 - quasi-geostrophic, 218
 - relation to circulation, 178
 - salinity effects, 183
 - shallow water, 136, 209
 - shallow water equations, 184
 - substance, 189
- Potential vorticity fluxes, atmospheric, 525
- Potential vorticity transport
and tropospheric stratification, 529
- Prandtl number, 368, 631
- Predictability, 361–366
of Lorenz equations, 362
- of turbulence, 363
 - of weather, 365
- Pressure, 12, 18
- Pressure coordinates, 78
and quasi-geostrophy, 220
relation to Boussinesq equations, 79
- Primitive equations, 61
potential vorticity conservation, 187
- vector form, 62
- Pseduomomentum
and hydrodynamic stability, 325
- Pseudoenergy, 327
and hydrodynamic instability, 327
- and wave activity, 329
- Pseudomomentum, 299
a wave activity, 299
- and zonal jets, 492
- Pseudomomentum equation, 299
- Quasi-geostrophic potential vorticity equation, 218
relation to Ertel PV, 224
- Quasi-geostrophic turbulence, 384

- Quasi-geostrophy, 207–223
asymptotic derivation, 216
buoyancy advection at surface, 219
continuously stratified, 215
energetics, 226
in isentropic coordinates, 225
informal derivation, 222
multi-layer, 213
pressure coordinates, 220
shallow water, 207
sheet at boundary, 220
single layer, 207
stratified equations, 215–221
two-layer, 211–213
two-level, 221
- Radiation condition, 490
- Radiative and dynamical constraints on stratification, 528
- Radiative equilibrium temperature, 452
- Radiative-convective model, 526
- Radiative-equilibrium temperature, 526
- Radius of deformation, 143
- Random walk, 410
- Ray, 239
- Ray tracing, 549, 550
- Rayleigh number, 631
- Rayleigh's criterion for instability, 258
- Rayleigh's equation, 251
- Rayleigh-Kuo criterion, 259, 325
- Reanalysis, 482
- Reduced gravity equations, 129–131
- Relative vorticity, 174
- Residual circulation, 305
and thickness-weighted circulation, 307
atmospheric, mid-latitude, 522
atmospheric, observations of, 523
stratospheric, 569
- Resonance of stationary waves, 543
- Reynolds number, 44
- Reynolds stress, 339
- Reynolds, Osborne, 47
- Rhines length, 379
- Rhines scale, 378
- Rhines–Young model, 613
- Richardson's four-thirds law, 417
- Richardson, Lewis Fry, 371
- Rigid body rotation, 164
- Rigid lid, 127, 130
- Rossby number, 85
- Rossby wave trains, 545
- Rossby waves, 229–236
and barotropic jets, 489
and ray tracing, 550
and turbulence, 378
barotropic, 229, 230
breaking, 569
continuously stratified, 234, 554
dispersion relation, 230
- finite deformation radius, 231
- group velocity property, 300
- mechanism of, 231
- meridional propagation, 549
- momentum transport in, 489
- two layers, 232
- vertical propagation, 554
- Rossby, Carl-Gustav, 241
- Rotating frame, 51–55
- Salinity, 14, 35
effect on potential vorticity conservation, 183
in box models, 640
- Salt, 14
- Sandström's effect, 635, 636
- Scale height
atmosphere, 40
density, 27
temperature, 29
- Scaling, 43–45
geostrophic, 198
in rotating continuously stratified equations, 200
in rotating shallow water equations, 198
- Seawater, 14, 33
adiabatic lapse rate, 34
buoyancy frequency, 120
equation of state, 14, 34, 35
heat capacity, 34
potential temperature, 34
thermodynamic properties, 33
- Shadow zone, 686
- Shallow water
quasi-geostrophic equations, 207
- Shallow water equations
multi-layer, 131
potential vorticity conservation, 184
reduced gravity, 129
rotation effects, 138
- Shallow water model of Hadley Cell, 468
- Shallow water systems, 123–140
conservation properties of, 136
potential vorticity in, 136
- Shallow water waves, 140–144
- Shallow-fluid approximation, 61
- Sideways convection, 629–640
conditions for maintenance, 635
energy budget, 635
limit of small diffusivity, 637
maintenance of, 634
mechanical forcing of, 639
phenomenology, 633
- Single-particle diffusivity, 412
- Skew diffusion, 427
- Skew flux, 427
- Sloping convection, 261
- Solenoidal term, 166
- Solenoids, 166, 173
- Sound waves, 37–38

- Southern Ocean, 700
 Specific heat capacities, 21
 Spectra of passive tracers, 366
 Spherical coordinates, 55–66
 centrifugal force in, 55
 Squire's theorem, 291
 Stacked shallow water equations, 131
 Standard atmosphere, 522, 524
 Static Instability, 91–98
 Stationary waves, 542–554
 adequacy of linear theory, 548
 and ray tracing, 550
 Green's function, 547
 in a single-layer, 542
 meridional propagation, 549
 one-dimensional wave trains, 545
 resonant response, 543
 thermal forcing of, 560
 Stokes, George, 46
 Stommel box models, 640
 Stommel wind-driven model, 586
 boundary layer solution, 589
 properties of, 594
 quasi-geostrophic formulation, 588
 the nonlinear problem, 599
 Stommel, Henry, 624
 Stommel–Arons model, 650–659
 shallow water version, 656
 single-hemisphere, 650
 two-hemisphere, 655
 Stommel–Arons–Faller laboratory model, 646
 Stommel–Arons–Faller model, 646
 Stratification
 of the atmosphere, 522
 Stratified geostrophic turbulence, 384
 Stratosphere, 453, 522, 566
 polar vortex, 575
 sudden warming of, 577
 Stratospheric dynamics, 566–577
 Stress
 Ekman layer, 104
 kinematic, 105
 Stretching, 170
 Sudden warming, 577
 Super-rotation, 467
 Surf zone, 575
 Surface drifters, 418
 Surface westerlies, 486
 Surface winds
 observed, 455
 Sverdrup balance, 589
 Sverdrup interior flow, 590
 Symmetric diffusivity tensor, 426
 Tangent plane, 66
 Taylor–Proudman effect, 88
 TEM equations, 304, 306
 for primitive equations, 309, 536
 Temperature, 18
 Thermal wind
 in shallow water equations, 134, 135
 Thermal wind balance, 89
 pressure coordinates, 90
 Thermobaric parameter, 15
 Thermocline, 667–691
 advection scaling, 671
 boundary-layer analysis, 676
 diffusive, 632
 diffusive scaling, 672
 internal, 674
 kinematic model, 668
 main, 667
 one-dimensional model, 675
 reduced-gravity, single-layer model, 682
 scaling for, 670
 summary and overview, 690
 ventilated, 681
 wind-influenced diffusive scaling, 672
 Thermodynamic equation, 22–31
 Boussinesq equations, 70
 for liquids, 26, 31
 summary table, 30
 Thermodynamic equilibrium, 22
 Thermodynamic relations, 16–21
 Thermodynamics
 first law, 17
 fundamental postulate, 16
 Thermohaline circulation, 627
 Thickness, 80
 Thickness diffusion, 440, 442
 Tilting and tipping, 170
 Topographic effects
 atmospheric stationary waves, 542
 JEBAR, 612
 oceanic western boundary current, 608
 Tracer continuity equation, 11
 Tracer homogenization, 423
 Traditional approximation, 62
 Transformed Eulerian Mean, 304–314
 and eddy transport, 443
 isentropic coordinates, 306
 primitive equations, 309, 536
 quasi-geostrophic form, 304
 spherical coordinates, 536
 Transport by baroclinic eddies, 425
 Triad interactions, 339
 two-layer geostrophic turbulence, 388
 Tropopause, 522–536
 definitions, 524
 Troposphere, 453, 522
 and moist convection, 534
 and potential vorticity transport, 529
 stratification, 522, 528, 529, 535
 ventilation of, 534
 Turbulence, 337
 closure problem, 338
 degrees of freedom, 348

- fundamental problem, 338
predictability of, 361
three-dimensional, 341
two-dimensional, 349
- Turbulent diffusion, 408, 409**
and the TEM, 443
by baroclinic eddies, 425
in the atmosphere and ocean, 430
macroscopic perspective, 422
potential vorticity, 443
requirements for, 421
thickness, 440
two-dimensional, 418
- Turbulent diffusivity, 412**
- Two-box model, 640**
- Two-dimensional turbulence, 349–361**
beta effect, 377
eddy diffusion in, 418
energy and enstrophy transfer, 351
numerical solutions, 360
- Two-dimensional vorticity equation, 167**
- Two-layer instability problem, 271**
- Two-layer model**
of atmospheric mid-latitudes, 503
- Two-layer QG equations, 211**
- Two-level QG equations, 221**
- Two-particle diffusivity, 415, 417**
- Unit vectors, rate of change on sphere, 59**
- Vector invariant momentum equation, 63**
- Ventilated pool, 689**
- Ventilated thermocline, 681–691**
reduced-gravity, single-layer model, 682
two-layer model, 683
- Ventilated troposphere, 534**
- Vertical coordinates, 77–80**
- Vertical vorticity equation, 176**
- Viscosity, 13**
effect on energy budget, 43
- Viscous scale, 346**
- Viscous-advection range, 368**
- Vorticity, 163–174**
equation for a barotropic fluid, 166
equation on beta plane, 178
evolution equation, 165
evolution in a rotating frame, 175
frozen-in property, 168
in two dimensional fluids, 167
stretching and tilting, 170
vertical component, 176
- Vorticity, relative, 174**
- vr vortex, 164**
- Wave activity, 299**
and pseudomomentum, 299
group velocity property, 300
orthogonality of modes, 302
- Wave breaking, 569**
- Wave packet, 240**
- Wave trains, 545**
- Wave–turbulence cross-over, 378**
- Wavelength, 237**
- Waves, 236**
acoustic-gravity, 100
barotropic Rossby, 229
breaking, 569
frequency, 236
gravity, 98
hydrostatic gravity, 99
inertial, 118, 143, 161
Kelvin, 143
kinematics, 236
Lamb, 103
Poincaré, 141
Rossby, 229
Rossby dispersion relation, 230
Rossby wave mechanism, 231
Rossby, continuously stratified, 234
Rossby, single-layer, 229
Rossby, two-layer, 232
rotating shallow water, 141
shallow water, 140
sound, 37
wavevector, 236
- Wavevector, 236**
- Weather predictability, 365**
- Western boundary currents**
topographic and inviscid, 608
- Western boundary layer, 592**
frictional, 591
inertial, 601
- Western intensification, 583**
- Western pool, 688**
- Wind-driven gyres, 583**
- Wind-driven ocean circulation, 585–623**
continuously stratified, 619
homogeneous model, 585
two-layer model, 613
vertical structure, 613
- WKB method, 552**
- Zonal boundary layers in ocean gyres, 597**
- Zonal flow in turbulence, 378**
- Zonal flows in beta-plane turbulence, 380**
- Zonally averaged atmospheric circulation, 485**