An Introduction to the Global Circulation of the Atmosphere

David A. Randall

Contents

Preface **1**

1 Perpetual Motion **3**

2 What Makes It Go? **16**
2.1 The Earth's radiation budget: An "upper boundary condition" on the global circulation . . . 16
2.2 Meridional energy transports by the atmosphere-ocean system . . . 20
2.3 Surface boundary conditions . . . 25
2.3.1 Temperature . . . 25
2.3.2 Wetness . . . 26
2.3.3 Topography . . . 28
2.3.4 Heat capacity . . . 30
2.3.5 Albedo . . . 31
2.3.6 Roughness . . . 31
2.3.7 Vegetation . . . 31
2.3.8 Sea ice . . . 32
2.3.9 Land ice . . . 34
2.4 Energy and moisture budgets of the surface and atmosphere . . . 35
2.5 Segue . . . 42
2.6 Problems . . . 43

3 A quick overview of the circulation **44**
3.1 Introduction . . . 44
3.2 The global distribution of atmospheric mass . . . 44
3.3 Zonal wind . . . 53
3.4 Meridional wind . . . 57
3.5 Vertical velocity and the mean meridional circulation . . . 63
3.6 Temperature . . . 73
3.7 Storm tracks . . . 79
3.8 Moisture . . . 81
3.9 Lots of questions . . . 85

3.10 Problems 86

4 The rules of the game **88**
4.1 Introduction 88
4.2 Conservation of the mass of dry air 88
4.3 Conservation of atmospheric moisture 93
4.4 Conservation of momentum on a rotating sphere 96
4.4.1 The equation of motion 96
4.4.2 "Apparent" gravity and hydrostatic balance 99
4.4.3 The equations of horizontal and vertical motion 101
4.5 Conservation of thermodynamic energy 103
4.6 Summary of the basic equations in height coordinates 106
4.7 The hydrostatic and quasi-static approximations 107
4.8 Conservation of mechanical energy 108
4.9 Where does mechanical energy come from? 111
4.10 Conservation of total energy 112
4.11 Static energies 113
4.12 The equations of motion in spherical coordinates 115
4.13 Conservation of angular momentum 117
4.14 Potential temperature as a vertical coordinate 119
4.14.1 Introduction 119
4.14.2 The isentropic system 119
4.14.3 The isentropic system with the quasi-static approximation 122
4.14.4 The mechanical energy equation in isentropic coordinates 123
4.15 Vorticity and potential vorticity 126
4.15.1 Turn, Turn, Turn 128
4.15.2 Stretching and twisting 130
4.15.3 The solenoidal term 131
4.15.4 Potential vorticity 132
4.15.5 The isentropic potential vorticity 135
4.15.6 Fellow travelers 137
4.15.7 The vertical component of the vorticity, and the horizontal vorticity vector 138
4.16 The primitive equations 139
4.17 The quasi-geostrophic system 141
4.18 The shallow water equations 147
4.19 Segue 148
4.20 Problems 149

5 Surface fluxes and the turbulent boundary layer **152**
5.1 Small-scale but important 152
5.2 What is reanalysis? 152

5.3 The surface fluxes . . . 153
5.3.1 The surface moisture flux . . . 153
5.3.2 The surface sensible heat flux . . . 153
5.3.3 The surface momentum flux . . . 156
5.4 Energy transports and dissipation due to small-scale motions . . . 156
5.4.1 Work done on the ocean and land surface . . . 156
5.4.2 Dissipation . . . 156
5.5 Segue . . . 159

6 Go with the flow **160**
6.1 Overview . . . 160
6.2 An isentropic view of the mass circulation . . . 162
6.3 Moisture transports . . . 168
6.4 Moist static energy transports . . . 177
6.5 Angular momentum transports . . . 189
6.6 Segue . . . 202
6.7 Problems . . . 203

7 The energy cycle **207**
7.1 The flow of energy across the atmosphere . . . 207
7.2 The Earth's entropy budget . . . 209
7.3 The vertically integrated enthalpy . . . 213
7.4 Available potential energy . . . 214
7.5 The gross static stability . . . 221
7.6 How available potential energy is converted into kinetic energy . . . 225
7.7 Examples: The available potential energies of three simple systems . . . 226
7.7.1 Example #1: The APE associated with static instability . . . 226
7.7.2 Example #2: The APE associated with meridional temperature gradients . . . 227
7.7.3 Example #3: The APE associated with surface pressure variations . 232
7.8 Where do variances come from? . . . 233
7.9 Generation of available potential energy, and its conversion into kinetic energy . . . 239
7.10 The governing equations for the eddy kinetic energy, zonal kinetic energy, and total kinetic energy . . . 245
7.11 Observations of the energy cycle . . . 247
7.12 The role of heating . . . 251
7.13 Moist available energy . . . 252
7.14 Segue . . . 253
7.15 Problems . . . 253

8 The roles of moist convection **259**
8.1 What convection is, and what it does . . . 259

8.2 The basics of dry and moist convection . . . 260
8.2.1 What is buoyancy? . . . 260
8.2.2 Dry convection . . . 262
8.2.3 Convection in uniformly saturated air . . . 265
8.2.4 Conditional (or cumulus) instability . . . 267
8.3 Upward energy transport by deep convection . . . 270
8.4 The gross moist stability . . . 271
8.5 The tight coupling of heating and vertical motion in the tropics . . . 275
8.6 Radiative-convective equilibrium . . . 276
8.7 Some comments on radiative-convective equilibrium . . . 284
8.8 Mass fluxes . . . 284
8.8.1 Origin story . . . 284
8.8.2 The simplest mass flux model . . . 285
8.8.3 Generalization . . . 288
8.8.4 EDMF . . . 289
8.9 Large-scale budgets . . . 289
8.10 Updrafts and environment . . . 291
8.11 Why the updraft fractional area is small . . . 292
8.12 How small a small updraft area simplifies things . . . 295
8.13 Large-scale budgets again . . . 295
8.14 A simple cumulus cloud model . . . 297
8.15 Compensating subsidence . . . 298
8.16 The vertical distribution of the mass flux . . . 300
8.17 What determines the intensity of the convection? . . . 302
8.18 Conditional symmetric instability . . . 306
8.19 Segue . . . 310
8.20 Problems . . . 311

9 Circulation regimes and cloud regimes 313
9.1 How the surface winds influence the tropical and subtropical SST . . . 313
9.2 Tropical and subtropical cloud regimes . . . 314
9.3 Representative soundings . . . 316
9.4 Entrainment across the cloud-top inversion . . . 318
9.5 The Walker Circulation . . . 321
9.6 Segue . . . 329

10 A taxonomy of eddies 330
10.1 Not all eddies are waves . . . 330
10.2 Free and forced oscillations of a thin atmosphere on a rotating sphere . . . 332
10.2.1 The basic state and the linearized governing equations . . . 332
10.2.2 Boundary conditions . . . 338
10.2.3 Solution procedure . . . 339

10.2.4 Free oscillations . . . 340
10.2.5 Forced oscillations . . . 342
10.2.6 The scope of Laplace's analysis . . . 342
10.2.7 Solving the LTE for the case of no rotation . . . 343
10.2.8 With rotation: Free Oscillations of the First and Second Kinds . . . 344
10.3 Observations of stationary and transient eddies in middle latitudes . . . 348
10.4 Annular modes . . . 353
10.5 Propagation of planetary waves . . . 356
10.6 Theory of orographically forced stationary waves . . . 361
10.7 Tropical waves . . . 370
10.8 The response of the tropical atmosphere to stationary heat sources and sinks 387
10.9 Monsoons . . . 391
10.10 The Madden-Julian Oscillation . . . 400
10.11 Tropical Easterly Waves . . . 407
10.12 Tropical cyclones . . . 407
10.13 Segue . . . 410
10.14 Problems . . . 411

11 Interactions and non-interactions of waves with the mean flow 414
11.1 Interactions and non-interactions of gravity waves with the mean flow . . . 414
11.2 Rossby waves . . . 418
11.2.1 The effects of a mean zonal flow . . . 418
11.2.2 Angular momentum transports by Rossby waves . . . 421
11.2.3 Energy transports by Rossby waves . . . 422
11.3 More about the interaction of planetary waves with the mean flow . . . 424
11.3.1 How the eddies affect the zonally averaged quasigeostrophic PV . . . 424
11.3.2 Wave fluxes of potential temperature, wave energy, and potential vorticity . . . 425
11.3.2.1 Baroclinically amplifying waves . . . 425
11.3.2.2 Upward propagating neutral waves . . . 426
11.3.2.3 A non-interaction theorem for the quasigeostrophic system 427
11.3.2.4 Acceleration of the mean flow by upward propagating neutral waves . . . 427
11.3.2.5 Wave momentum exchange with the Earth's surface . . . 428
11.3.3 Momentum transport by upward-propagating neutral waves . . . 429
11.4 The Transformed Eulerian Mean (TEM) system . . . 430
11.5 The Eliassen-Palm theorem in isentropic coordinates . . . 437
11.6 Segue . . . 444
11.7 Problems . . . 444

12 Examples of wave-mean flow interactions 446
12.1 Blocking . . . 446

12.2 Sudden stratospheric warmings . . . 456
12.3 The quasi-biennial oscillation . . . 457
12.4 The semi-annual oscillation . . . 461
12.5 The Brewer-Dobson Circulation . . . 461
12.6 Segue . . . 463
12.7 Problems . . . 463

13 The global circulation as turbulence 464
13.1 The nature of turbulence . . . 464
13.2 The kinetic energy cascade . . . 468
13.3 Dimensional analysis of the kinetic energy spectrum . . . 468
13.4 Nonlinearity and scale interactions . . . 471
13.5 Two-dimensional turbulence . . . 472
13.6 Quasi-two-dimensional turbulence . . . 476
13.7 The anti-cascade of two-dimensional turbulence . . . 480
13.8 Observations of the kinetic energy spectrum . . . 482
13.9 Dissipating enstrophy but not kinetic energy . . . 484
13.10 What does the blender blend? . . . 486
13.11 Sensitive dependence on initial conditions . . . 487
13.12 Quantifying the limits of predictability . . . 494
13.12.1 The dynamical approach . . . 495
13.12.2 The empirical approach . . . 496
13.12.3 The dynamical-empirical approach . . . 499
13.13 Ensemble forecasting . . . 504
13.14 The response of the atmosphere to changes in sea surface temperature . . . 505
13.15 Segue . . . 507
13.16 Problems . . . 508

14 The effects of changes in external forcing 510
14.1 Forced chaotic systems . . . 510
14.1.1 Pushing the attractors around . . . 510
14.2 Transitive and intransitive systems . . . 511
14.3 Anthropogenic climate change . . . 514
14.4 Problems . . . 520

Appendices 522

A Vectors, Coordinates, and Coordinate Transformations 522
A.1 Physical laws and coordinate systems . . . 522
A.2 Scalars, vectors, and tensors . . . 522
A.3 Differential operators . . . 525
A.4 Vector identities . . . 527
A.5 Spherical coordinates . . . 529

A.5.1 Vector operators in spherical coordinates . . . 529
A.5.2 Horizontal and vertical vectors in spherical coordinates . . . 530
A.5.3 Derivation of the gradient operator in spherical coordinates . . . 532
A.5.4 Applying vector operators to the unit vectors in spherical coordinates 533
A.6 Solid body rotation . . . 534
A.7 Formulas that are useful for two-dimensional flow . . . 535
A.8 Vertical coordinate transformations . . . 536
A.8.1 Basics . . . 536
A.8.2 Some useful operators . . . 537
A.9 Concluding summary . . . 539

B Dimensional Analysis, Scale Analysis, and Similarity Theories 540
B.1 Dimensions and units . . . 540
B.2 Consistent use of dimensional and nondimensional quantities . . . 542
B.3 The Buckingham Pi Theorem . . . 543
B.4 An example of dimensional analysis . . . 544
B.5 Scale analysis . . . 547
B.6 Similarity theories . . . 550
B.7 Summary . . . 552

C Derivation of the Isentropic System 554
C.1 The horizontal pressure-gradient force . . . 554
C.2 The vertical pressure-gradient force . . . 555
C.3 The continuity equation . . . 556
C.4 The vertical mass flux . . . 558
C.5 The equation for an intensive scalar . . . 558
C.6 Final forms of the prognostic equations . . . 559

D Means, eddies, and transients 560

E Why is the Dissipation Rate Positive? 564

F The Moist Adiabatic Lapse Rate 568

G Eddy Kinetic Energy and Zonal Kinetic Energy 572

H Spherical Harmonics 583

I Hermite Polynomials 594

Bibliography 597

Preface

This is a graduate-level introductory overview of the global circulation of the atmosphere, a subject that is closely tied to atmospheric dynamics. It is an updated, extended, and corrected version of my book with the same title, which was published in 2015 by Princeton University Press. Color figures were not permitted in the published book, but I am using them extensively in this updated version, which may (or may not) eventually become the second edition of the book.

Graduate-level studies in atmospheric dynamics are essential as preparation for reading this book. I have assumed that the reader is familiar with basic concepts of atmospheric dynamics such as the equation of motion, the approximate hydrostatic and geostrophic balances, potential temperature, vorticity, and planetary waves. More advanced dynamical concepts are introduced as needed. Chapter 4, in particular, gives a brief but fairly detailed overview of the concepts used throughout the book. Some instructors may choose to refer to Chapter 4 only as needed when concepts are used in the later chapters. A course on dynamics tends to focus on basic physical concepts and methods for their analysis, however, while a course on the global circulation must focus on what the atmosphere actually does, and why.

I have also included quite a bit of discussion of the mechanisms that give rise to the thermal driving of the global circulation, and the feedbacks of the circulation on those diabatic processes. We can't claim to understand the circulation unless we understand the heating that drives it, as well as how the circulation influences the heating. This part of the story is particularly interesting, and also particularly challenging.

Tropical meteorology has been at least partially subsumed into the study of the global circulation. This is natural, in part because the tropics covers about half the planet, in terms of surface area. In addition, the tropics is home to major heating and cooling processes that invigorate the global circulation.

It is difficult to draw a line between the global circulation and climate. The two subjects are growing closer together, as the roles of heating and dissipation in the global circulation emerge as key issues. Such topics as monsoons, the hydrologic cycle, and the planetary energy budget can be included under either "climate" or "global circulation," although perhaps with different slants. Climate is the bigger subject. This book skirts the edges of

physical climatology.

I have chosen to discuss many of the topics in terms of their original sources, rather than the latest papers. This quasi-historical approach gives credit to the pioneers of our field, and highlights some of the human aspects of the research.

The global circulation involves many named "emergent" phenomena. I think it's important to tie the discussion of these phenomena back to first principles, so that the students will learn more than just the high-level lingo.

Some of the problems at the ends of the chapters involve working with observations. The Internet has made it unnecessary to include the data with the book. Many of the figures in this book were made using the Interim Reanalysis of the European Centre for Medium Range Weather Forecasts, which would be a suitable source of global atmospheric data for use with the problems, (although of course there are other possibilities).

Mark Branson and Don Dazlich ably produced many of the figures. I could not have finished without their help.

Chapter 1

Perpetual Motion

The atmosphere *circulates*. The circulation is global in extent (Fig. 1.1). The circulating mass consists of "dry air" and three phases of water. Energy and momentum are carried with the air, but evolve in response to various processes along the way. Many of those same processes add or remove moisture.

The circulation is sustained by thermal forcing, which ultimately comes from the Sun. On the average, the Earth absorbs about 240 W m^{-2} of incoming or "incident" solar energy, of which roughly 2% is converted to maintain the kinetic energy of the global circulation against frictional dissipation. Additional, "primordial" energy leaks out of the Earth's interior, but at the relatively tiny rate of 0.1 W m^{-2} (Sclater et al., 1980; Bukowinski, 1999). The thermal forcing of the global circulation is strongly influenced by the circulation itself, e.g., as clouds form and disappear. The interactions between the circulation and the heating are fascinating but complicated.

When averaged over time, the global circulation has to satisfy various balance requirements. For example, in a time average, the infrared radiation emitted at the top of the atmosphere must balance the solar radiation absorbed, precipitation must balance evaporation, and angular momentum exchanges between the atmosphere and the ocean-solid Earth system must sum to zero. We will discuss the global circulation from this classical perspective. We will also supplement this discussion, however, with descriptions and analyses of the many and varied but inter-related phenomena of the circulation, including such things as the Hadley and Walker circulations, monsoons, stratospheric sudden warmings, the Southern Oscillation, subtropical highs, and extratropical storm tracks.

In addition, we will discuss the all-important diabatic and frictional processes that maintain the circulation, and the ways in which these processes are affected by the circulation itself.

The circulations of energy and water are closely linked. It takes about 2.5 million Joules of energy to evaporate one kilogram of water from the oceans, and the same amount of energy is released when the water vapor condenses to form a cloud. This "latent heat"

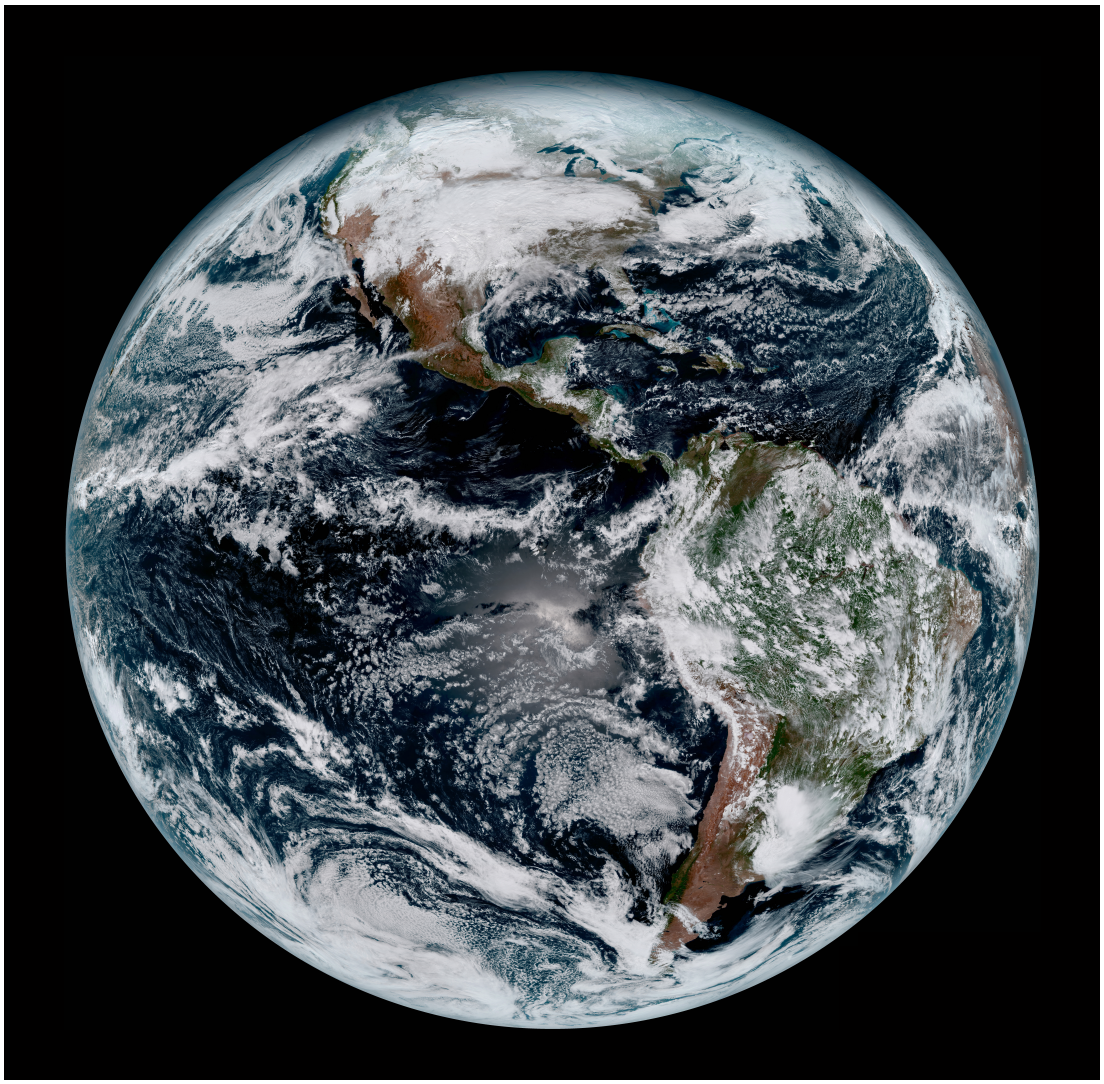


Figure 1.1: A full disk image of the Earth taken by the GOES-16 satellite at 1:07 p.m. EDT on January 15, 2017 , looking down on the Equator, with North and South American in view. Many elements of the global circulation can be seen in this picture, including the "inter-tropical" rain band in the eastern North Pacific, swirling midlatitude storms, and the low clouds associated with the high-pressure systems over the eastern subtropical oceans. From `https://www.nesdis.noaa.gov/content/goes-16-color-composite-images`.

release through condensation drives thunderstorm updrafts that in one hour or less can penetrate to heights of ten or even twenty kilometers. The cloudy outflows from such storms reflect sunlight back to space, and block infrared radiation from the warm surface below. Shallower clouds cast shadows over vast expanses of the oceans. One of the aims of this book is to give proper emphasis to the role of moisture in the global circulation of the atmosphere.

It is conventional and useful, although somewhat arbitrary, to divide the atmosphere

into parts. For purposes of this quick sketch, we will divide the atmosphere vertically and meridionally, only briefly mentioning the longitudinal variations. Let's start at the bottom, and work our way into the sky.

Most of the solar radiation that the Earth absorbs is captured by the surface, rather than within the relatively transparent atmosphere. Several processes act to transfer the absorbed energy upward, from the ocean and land surface into the lower portion of the atmosphere. The layer of air that is closely coupled with the Earth's surface is, by definition, the "planetary boundary layer," or PBL. The top of the PBL is often very sharp and well defined (Fig. 1.2). The depth of the PBL varies dramatically in space and time, but a ball-park value to remember is 1 km. The air in the PBL is turbulent, and the turbulence is associated with rapid exchanges of "sensible heat" (essentially temperature), moisture, and momentum between the atmosphere and the surface. These "fluxes" have dimensions of energy per unit area per unit time. They are produced by the turbulence, and also promote the turbulence, through mechanisms that will be discussed later. The most important exchanges are of moisture and momentum. The surface latent heat flux is a key source of energy for the global circulation, and turbulent mixing in the PBL is a key mechanism for dissipation of the kinetic energy of the global circulation. Friction is also a major driver of the circulation of the oceans.

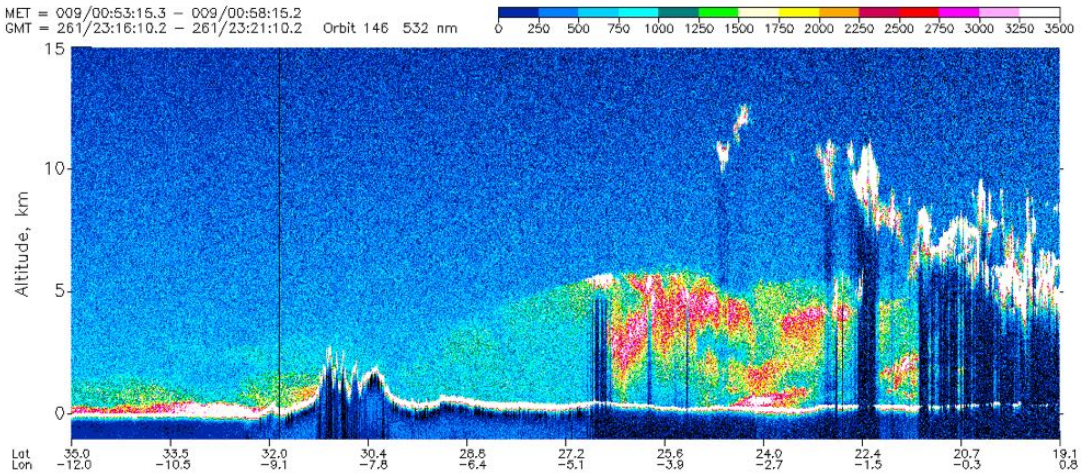


Figure 1.2: This figure shows lidar backscatter from aerosols and clouds. The figure was created using data from LITE, a lidar that flew on the space shuttle in 1994 (Winker et al., 1996). The lidar beam cannot penetrate through thick clouds; this accounts for the dark vertical stripes in the figure. The PBL can be seen in the data because the aerosol concentration decreases sharply upward there. The track shown here starts over the North Atlantic Ocean on the left side, and moves towards the southeast, over Africa. Longitudes and latitudes are given along the bottom of the figure. The Atlas mountains are visible near latitude 31 ° N latitude. From `http://www-lite.larc.nasa.gov/n_saharan_dust.html`.

The troposphere includes the PBL, but it is conventional to add the prefix "free" to

specify the part of the troposphere that lies above the PBL. The PBL is often well mixed, so that air can easily move up or down. The free troposphere has a positive static stability, which means that buoyancy forces resist vertical motion. The depth of the troposphere varies strongly with latitude and season.

A turbulent process called "entrainment" gradually incorporates free-tropospheric air into the PBL. Over the oceans, entrainment is, with a few exceptions, relatively slow but steady. Over land, entrainment is promoted by strong daytime heating of the surface, which helps to generate turbulence. As a result, the turbulent PBL rapidly deepens during the day. When the sun goes down, the processes that promote turbulence and entrainment are abruptly weakened, and the PBL reorganizes itself into a much shallower nocturnal configuration, leaving behind a nno-longer-turbulent layer of air that was part of the PBL during the afternoon. This diurnal deepening and shallowing of the PBL acts as a kind of "pump," which captures air from the free troposphere and adds it to the PBL starting shortly after sunrise, modifies the properties of that air during the day through strong turbulent exchanges with the surface, and then releases the modified air back into the free troposphere at sunset. Diurnal pumping is one way that the PBL exerts an influence on the free troposphere (Denning et al., 1995, 1999).

In addition, however, moisture and energy are carried upward from the PBL into the free troposphere by several mechanisms. Throughout the tropics and the summer-hemisphere middle latitudes, the most important of these mechanisms is cumulus convection. Cumulus clouds typically grow upward from the PBL. The updrafts inside the clouds carry PBL air into the free troposphere, where it is left behind when the clouds decay (Fig. 1.3). One of the effects of this process is to remove air from the PBL, and add it to the free troposphere.

Frontal circulations can also carry air from the PBL into the free atmosphere, essentially by "peeling" the PBL off from the Earth's surface, like the rind from an orange, and lofting the detached air towards the tropopause. This process is especially active in the middle latitudes in winter.

Fig. 1.4 shows somewhat idealized observed midlatitude vertical distributions of temperature, pressure, density, and ozone mixing ratio, from the surface to the 70-km level. For reference, the "top of the atmosphere" is often considered to be 100 km above the surface, but of course this definition is somewhat arbitrary and should be adjusted depending on the context (McDowell, 2018). In the lowest 12 km, the temperature usually decreases monotonically upward; this is the troposphere. The troposphere is cooled radiatively, because it emits infrared radiation much more efficiently than it absorbs solar radiation. The net radiative cooling is balanced mainly by the release of the latent heat of water vapor, as clouds form and precipitate.

The upper boundary of the troposphere is called the tropopause, The height of the tropopause varies from 17 km or so in some regions of the tropics, to about half that near the poles. Above the tropopause, the temperature becomes uniform with height and then begins to increase upward. This region is the stratosphere. The upward increase of tem-

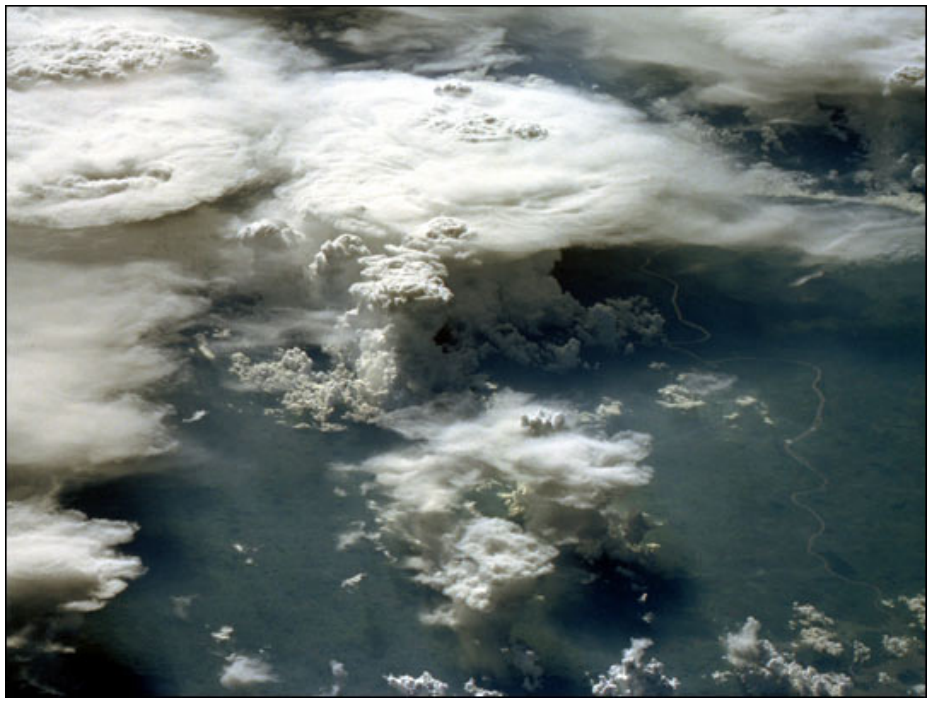


Figure 1.3: A space shuttle photograph of tropical thunderstorms over Brazil. The storms are topped by thick anvil clouds. Much shallower convective clouds can be seen in the foreground. From `https://eoimages.gsfc.nasa.gov/images/imagerecords/2000/2443/STS41B-41-2347_lrg.JPG`.

perature in the stratosphere is due to the absorption of solar radiation by ozone, which is created in the stratosphere by photochemical processes. Without ozone there would be no stratosphere. The summer-hemisphere stratosphere is almost devoid of active weather, and has warm air over the pole. The summer stratosphere is warmed by the perpetual sunlight, and the winds of the summer stratosphere are predominantly "easterly," which means that they blow from east to west. In contrast, the winter-hemisphere stratosphere has very cold air over the pole, with strong westerly winds, and experiences much more active weather, mainly due to waves propagating upward from the troposphere below. During winter, the polar stratosphere is occasionally disturbed by "Sudden Stratospheric Warmings," which are dramatic changes in temperature (and wind) that occur in many years in the Northern Hemisphere, and much less frequently in the Southern Hemisphere (Baldwin et al., 2021). The tropical stratosphere features an amazing periodic reversal of the zonal (i.e., west-to-east) wind, called the Quasi-Biennial Oscillation, or QBO (Baldwin et al., 2001). Its period is slightly longer than two years.

Even though stratosphere is very dry, its moisture budget is quite interesting. It receives small amounts of moisture from the troposphere, and it also gains some moisture through the oxidation of methane.

The upper boundary of the stratosphere, which is called the stratopause, occurs near the

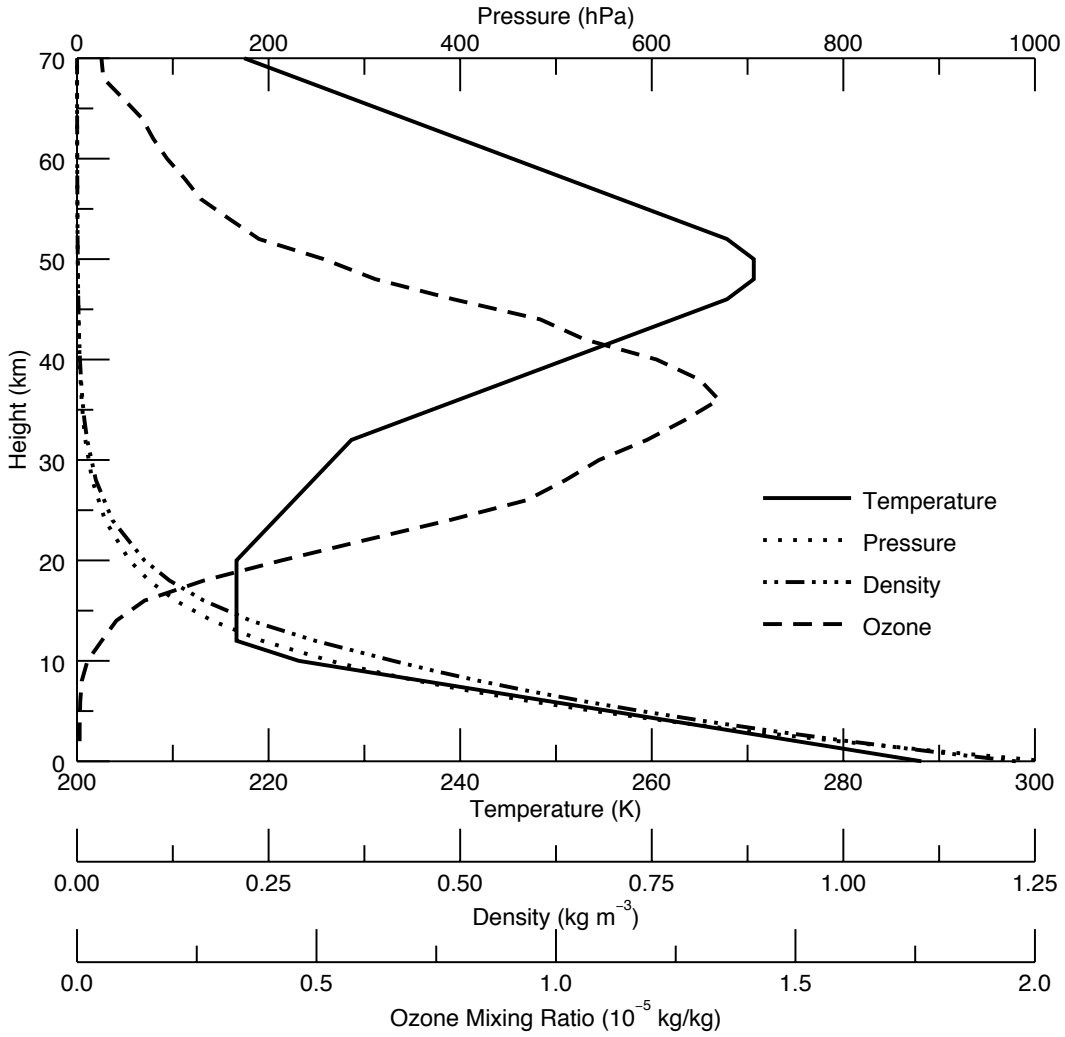


Figure 1.4: Idealized midlatitude temperature, pressure, density and ozone profiles, for the lowest 70 km of the atmosphere. The temperature, pressure, and density profiles are based on the U.S. Standard Atmosphere (National Oceanic and Atmospheric Administration, 1976). The ozone profile is from Krueger and Minzner (1976).

1 hPa ($\sim$ 50 km) level. In this book, we consider the troposphere and the stratosphere, but we will not discuss the layers above the stratopause.

For meteorological purposes, the tropics can be defined as the region from about 20° S to 20° N. Although the tropical temperature and surface pressure are remarkably uniform, the winds and rainfall are quite variable. In many parts of the tropics deep cumulus and cumulonimbus clouds, i.e., thunderstorms, produce lots of rain and transport energy, moisture, and momentum vertically, essentially continuing the job begun closer to the surface by the turbulence of the PBL. The convective clouds often produce strong exchanges of air between the PBL and the free troposphere, in both directions: positively buoyant PBL air

"breaks off" and drifts upward to form the cumuli, while negatively buoyant downdrafts associated with the evaporation of falling rain can inject free-tropospheric air into the PBL. In the convectively active parts of the tropics, the air is slowly rising in an area-averaged sense.

The mean flow in the tropical PBL is easterly; these are the "tradewinds." The tropical temperature and surface pressure distributions are generally very flat and monotonous, for simple dynamical reasons, discussed in Chapter 3, that are connected to the smallness of the tropical Coriolis parameter. The tropical moisture and wind fields are more variable than the temperature, however. The tropics is home to a variety of distinctive traveling waves and vortices, which organize the convective clouds on scales of hundreds to thousands of kilometers. Finally, the tropics is dominated by powerful and very large-scale monsoon systems, which are associated with continental-scale land-sea contrasts, and which actually extend into the subtropics and even middle latitudes.

The tropical atmosphere acquires the angular momentum of the Earth's rotation from the continents and oceans. The global atmospheric circulation carries the angular momentum to higher latitudes, where it is "put back" into the continents and oceans.

The tropical troposphere hosts some circulation phenomena that do not occur in higher latitudes. Most famously, tropical cyclones produce tremendous amounts of rainfall and strong winds. They are relatively small in scale, and highly seasonal. Monsoons are driven by land-sea contrasts. They also very seasonal, and much larger in spatial scale. The Madden-Julian Oscillation, or MJO, is a powerful tropical weather system that influences rainfall across about half of the tropics. El Niño, La Niña, and the Southern Oscillation, which are are collectively known as "ENSO," comprise a strong, quasi-regular oscillation of the ocean-atmosphere system, with a period of a few years (Philander, 1990). In an El Niño, the sea-surface temperatures warm in the eastern (or sometimes central) tropical Pacific, while in a La Niña they cool. The Southern Oscillation is a shift in the pressure and wind fields of the tropical Pacific region, which occurs in conjunction with El Niño and La Niña.

The subtropical portion of each hemisphere is roughly the region between 20° and 30° from the Equator. In many parts of the subtropical troposphere, the air is sinking in large anticyclonic circulation systems called, appropriately enough, "subtropical highs." The subsidence suppresses precipitation, which is why the major deserts of the world are found in the subtropics. Surface evaporation is very strong over the subtropical oceans, which have extensive systems of weakly precipitating shallow clouds. The subtropical upper troposphere is home to powerful "subtropical jets," which are westerly currents that are particularly strong in the winter hemisphere (Fig. 1.5).

The tropical rising motion and subtropical sinking motion can be seen as the vertical branches of a "cellular" circulation in the latitude-height plane. This "Hadley circulation" transports energy and momentum poleward, and it transports moisture toward the Equator. The Hadley circulation interacts strongly with the monsoons.



Figure 1.5: This photograph taken from a space shuttle shows cirrus clouds being carried along by the subtropical jet stream. From `http://earth.jsc.nasa.gov/sseop/efs/lores.pl?PHOTO=STS039-601-49`.

The region that we call the middle latitudes extends, in each hemisphere, from about 30° to 70° from the Equator. The midlatitude surface winds are primarily westerly. The midlatitude free troposphere is filled with vigorous weather systems called baroclinic eddies, which have scales of a few thousand km and grow as instabilities through a process in which warm air shifts upward and poleward, and is replaced by colder air that descends as it slides toward the Equator (Fig. 1.6). These storms transport energy and moisture poleward and upward, primarily in the winter hemisphere, but also to some extent during summer. They transport westerly momentum poleward and downward, driving currents in the oceans and rustling the leaves on the continents. The energy of the baroclinic eddies is derived from horizontal temperature differences that can only be sustained outside the tropics. They can produce massive cloud systems and heavy precipitation.

On the average, the polar troposphere is characterized by sinking motion and radiative cooling to space. The North Pole is in the Arctic Ocean, which is covered with sea ice (Fig. 1.7) and often blanketed by extensive cloudiness, while the South Pole is in the middle of a dry, mountainous continent (Fig. 1.8). Near the surface, the polar winds tend be easterly, but weak.

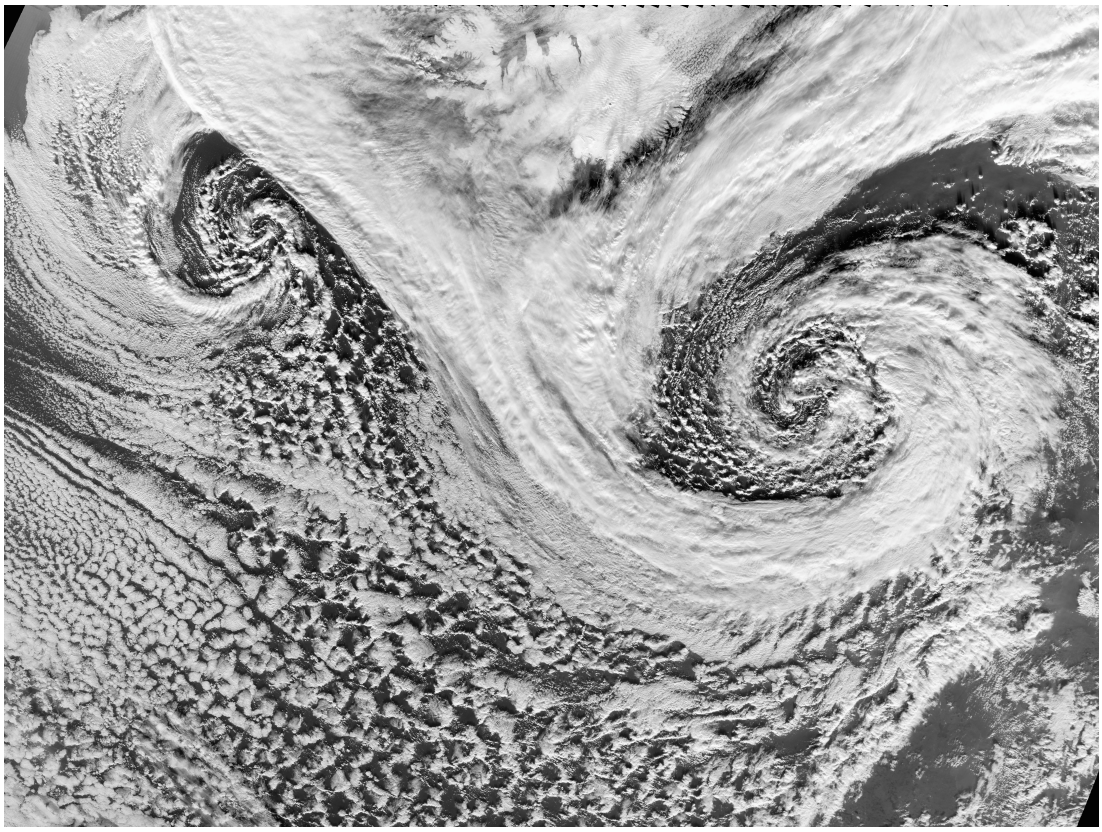


Figure 1.6: A pair of beautiful winter storms over the North Atlantic Ocean in winter. From `http://earthobservatory.nasa.gov/IOTD/view.php?id=7264`.

The polar regions and middle latitudes are home to prominent "annular modes," which by definition are visible even when the data are averaged over all longitudes. The annular modes fluctuate on a variety of time scales, almost uniformly in longitude. They are seen in both the stratosphere and troposphere, and make major contributions to the overall variability of the global circulation.

Analysis of energy sources, sinks, and transformations can aid in understanding the global circulation. The atmosphere cools radiatively, in an overall sense, and this cooling is balanced primarily by the release of latent heat, which in turn is made possible by surface evaporation. The net flow of energy is upward and towards the poles, carried by thunderstorms and the Hadley Circulation in the tropics, and by baroclinic eddies in middle latitudes. The energy escapes to space via infrared radiation at all latitudes, but especially in the subtropics.

The organization of this book is as follows. Chapter 2 provides an overview of the upper and lower boundary conditions on the global circulation. At the "top of the atmosphere," the observed pattern of radiation implies net poleward energy transports by the atmosphere and ocean together. Lower boundary conditions include the distributions of oceans and



Figure 1.7: An ice station used to collect data in and above the Arctic ocean during 1997-8. From `https://en.wikipedia.org/wiki/Surface_Heat_Budget_of_the_Arctic_Ocean#/media/File:Ice_Station_SHEBA-CRREL.jpg`.

continents and mountains, the pattern of sea surface temperature and the directly related pattern of sea-surface saturation vapor pressure, the heat capacity of the surface, the distribution of vegetation on the land surface, and the distributions of sea ice and continental glaciers and ice sheets. These lower boundary conditions strongly affect the flows of energy and moisture across the Earth's surface. Chapter 2 closes with a brief overview of the vertically integrated energy and moisture budgets of the surface and atmosphere, and the connections between them.

With this background information out of the way, Chapter 3 presents an observational overview of some basic aspects of the global circulation. We start with the global distribution of atmospheric mass, then move to winds, temperature, and moisture. The amount of interpretation is deliberately kept to a minimum in Chapter 3. By the end of Chapter 3, many important elements of the global circulation have been introduced, giving a basis for further discussion in later chapters.

While Chapter 3 focuses on the observations, Chapter 4 gives a brief but intensive review of concepts that will be used later in the book, starting with the dynamics of fluid motion on a rotating sphere. Angular momentum conservation is derived. We then turn to a detailed discussion of energy transports and transformations. The key subject of potential vorticity is introduced near the end of the chapter. The quasi-geostrophic approximation and the shallow water equations are also introduced.

The term "zonal average" refers to an average over all longitudes. Chapter 6 discusses



Figure 1.8: Cold-looking clouds over the dry mountains of Antarctica. From `https://wp.natsci.colostate.edu/walllab/research/antarctica/wall-valley-antarctica-information/`.

the zonally averaged circulation, and how it is influenced by sources and sinks of moisture, energy, and momentum. The chapter also presents first look at the effects of "eddies" that transport dry air, moisture, energy, and momentum along "isentropic" surfaces on which the entropy is uniform.

Chapter 8 gives an introduction to the effects of convective energy sources and sinks, starting with a brief review of the nature of buoyant convection in both dry and moist atmospheres. We then outline the famous observational study of convective energy transports by Riehl and Malkus (1958). This is followed by a discussion of the idealized but important concept of radiative-convective equilibrium, and an overview of convective cloud regimes, ranging from deep convection in the tropics to shallow stratocumulus convection in higher latitudes. The concept of convective mass flux is introduced, along with a theory that explains why cumulus updrafts tend to be separated by broad expanses of relatively dry, sinking air. We then show how the convective mass flux can be used to understand the heating and drying associated with cumulus clouds, and what determines the intensity of the convection as a function of the large-scale weather regime. Finally, we introduce the concept of conditional symmetric instability, which is particularly relevant in middle latitudes.

Chapter 7 presents the energetics of the global circulation, beginning with an in-depth discussion of available potential energy and the related concept of the gross static stability. Following Lorenz, these ideas are developed through the use of isentropic coordinates, although we also show how they can be expressed using pressure coordinates. We then discuss the mechanisms by which vertical and meridional gradients of the zonally averaged circulation can be converted into zonally varying features called eddy variances. This leads to a discussion of how available potential energy is generated, and how it is converted into eddy kinetic energy. The chapter closes with a discussion of the observed energy cycle of the global atmosphere.

Chapter 10 introduces various types of eddies, beginning with a brief discussion of the Laplace Tidal Equations, omitting the details of the mathematics. The chapter presents observations of the distribution of energy with scale for midlatitude eddies. We summarize the theory of Rossby waves forced by flow over topography, and Matsuno's theory of Equatorial waves (Matsuno, 1966). This is followed by a discussion of the monsoons, the east-west Walker circulation of the tropical Pacific, and the energy balance of the tropics and subtropics. The chapter closes with discussions of tropical cylcones and the Madden-Julian Oscillation.

Chapter 11 deals with interactions of eddies with the mean flow. We begin with the original "non-interaction theorem" of Eliassen and Palm (1961), as applied to gravity waves excited by flow over mountains. This is followed by a brief discussion of the importance of gravity-wave drag for the global circulation. We then turn to the quasi-geostrophic wave equation, and its use by Charney, Drazin, Matsuno and others to interpret the interactions of Rossby Waves with the zonally averaged circulation. This leads to a discussion of sudden

stratospheric warmings.

Next we discuss the poleward and upward fluxes of sensible heat and momentum associated with winter storms. This leads to a discussion of the Eliassen-Palm Theorem as it relates to balanced flows. The theorem is developed first in pressure coordinates. We then show the additional insights and greater simplicity that come from an analysis in isentropic coordinates, including the relationship between the divergence of the Eliassen-Palm flux and the flux of potential vorticity. The chapter closes with discussions of the Quasi-Biennial Oscillation, blocking, and the Brewer-Dobson circulation (Butchart, 2014) as additional examples of the interactions between eddies and the zonally averaged circulation.

Chapter 13 discusses the global circulation as a kind of large-scale turbulence, starting with a discussion of what turbulence is, framed in terms of vorticity dynamics. We discuss the very different flows of energy between scales in two-dimensional and three-dimensional turbulence, and connect it with a description of the distribution of atmospheric kinetic energy with spatial scale. We then discuss mechanisms that allow dissipation of enstrophy (squared vorticity) without dissipation of kinetic energy. This is followed by a discussion of mixing along isentropic surfaces.

Next, we discuss the predictability of the weather in a chaotic circulation regime. Finally, we explain how climate prediction differs from weather prediction.

The closing chapter briefly summarizes current trends and projected future changes in the global circulation due to increasing concentrations of atmospheric greenhouse gases.

As outlined above, the study of the thermally driven global circulation of the atmosphere brings together concepts from all areas of atmospheric science. We will touch on large-scale dynamics, convection, turbulence, cloud processes, and radiative transfer, with an emphasis on their interactions. Often these various topics are presented as if they were somehow neatly separated from each other. This book will help you to see how they fit together.

Because the circulation is global and spans all seasons, it brings together the concepts of atmospheric science across a very wide range of conditions and contexts. For example, surface friction occurs *everywhere*: over the convectively disturbed tropical oceans, in the chaotic storm track north of Antarctica, over the Himalayan mountains, and in the leafy canopies of the tropical jungles. Unfortunately, however, our understanding of boundary-layer meteorology is mostly applicable only to relatively simple, horizontally uniform conditions, such as might be encountered on a quiet summer morning in Kansas or in a stratocumulus-capped boundary layer over the subtropical oceans. When we try to understand the global circulation, we are forced to confront these and many more limitations of our understanding. This makes the study of the global circulation a particularly challenging and exciting field – as you are about to see for yourselves.

Chapter 2

What Makes It Go?

2.1 The Earth's radiation budget: An "upper boundary condition" on the global circulation

Radiation is (almost) the only mechanism by which the Earth can exchange energy with the rest of the Universe[1]. The most important upper boundary condition on the global circulation of the atmosphere is the incident solar radiation, also called the insolation. The insolation varies with geographical location, and with time. It is determined by the energy output of the Sun, the spherical geometry of the Earth itself, and the geometry of the Earth's orbit (see Fig. 2.1), which can be described in terms of the obliquity (the angle that the Earth's axis of rotation makes with the plane of the Earth's orbit around the sun), the eccentricity (a measure of the degree to with the shape of the Earth's orbit differs from a perfect circle), and the dates of the equinoxes. These all vary over geologic time (e.g., Crowley and North, 1991).

The solar energy flux at the mean radius of the Earth's orbit is about 1365 W m^{-2}. One way to get an intuitive grasp of this number is to imagine fourteen 100-Watt light bulbs per square meter. Another is to consider that 1365 W m^{-2} is equivalent to 1.365 GW km^{-2}, which is the energy output of a large power plant for each square kilometer of area normal to the solar beam.

The globally averaged top-of-the-atmosphere radiation budget is summarized in Table 2.1. The numbers given in the table are now known to three significant digits. The Earth's albedo is near 0.30, independent of season; this number has been known to better than 10 per cent accuracy only since the 1970s. The energy absorbed by the Earth is given by

[1]The other mechanisms are gravitational tides, and the exchanges of mass across the top of the atmosphere, including dust and larger meteors falling in, and the (fortunately) very slow escape of atmospheric mass to space.

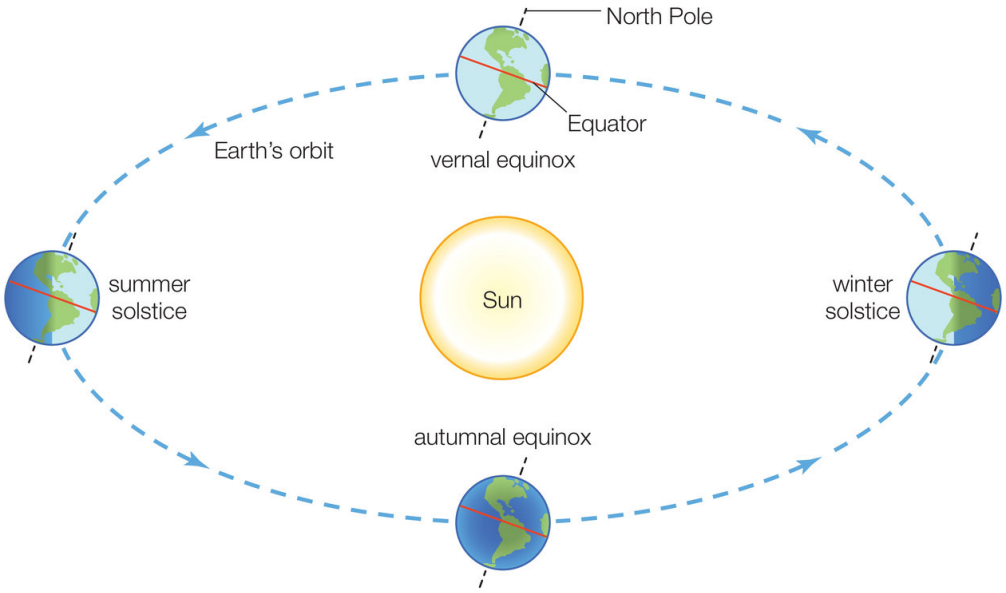


Figure 2.1: March of the seasons. As the tilted Earth revolves around the sun, changes in the distribution of sunlight cause the succession of seasons. From `https://dept.astro.lsa.umich.edu/ugactivities/Labs/seasons/SeasonsIntrotroppics.html`.

$$\begin{aligned} S_{abs} &= S\left(\frac{\pi a^2}{4\pi a^2}\right)(1-\alpha) \\ &= \frac{1}{4}S(1-\alpha)\ . \end{aligned} \tag{2.1}$$

Here S_{abs} is the globally averaged absorbed solar energy per unit area (given in Table 2.1), S is the globally averaged insolation, a is the radius of the Earth, and α is the planetary albedo. The global average used to compute S includes the zeroes on the night-side of the Earth. That is why S is multiplied by πa^2, the area of the absorbing disk, and divided by $4\pi a^2$, the area of the sphere.

As a matter of common experience, the insolation varies both diurnally and seasonally. At a given moment, it also varies strongly with longitude. Because a year is much longer than a day, the *daily-mean* insolation is (almost) independent of longitude, but it varies strongly with latitude in a way that depends on the season, as summarized in Fig. 2.2. As we move from the solar Equator (i.e., the latitude immediately "under" the Sun) to the summer pole, the insolation initially decreases, because at a given local time (e.g., local noon) the sun appears to be lower in the sky. On the other hand, the length of day increases at high latitudes in summer, and this tends to make the daily-mean insolation increase. Near the poles, the length-of-day effect dominates, so that at high latitudes in summer the insolation actually increases towards the pole. That is why there is a minimum of the

Table 2.1: Summary of the annually averaged top-of-the-atmosphere radiation budget, after Trenberth et al. (2009).

Incident solar radiation	341 W m^{-2}
Absorbed solar radiation	239 W m^{-2}
Planetary albedo	0.30
Outgoing longwave radation	239 W m^{-2}

insolation about 23° away from the pole in the summer hemisphere, as shown in the figure.

Notice that the region of insolation stronger than 500 W m^{-2} (in high latitudes of the summer hemisphere) is larger in January than in July. The reason is that, in the present era, the Earth is closer to the Sun in January than in July. The globally averaged insolation is actually about 7% larger in January than in July. The orbit of Mars is more eccentric; as a result, the globally averaged Martian insolation varies seasonally by about 31%.

Seasonal and, to a lesser extent, diurnal cycles are clearly evident in the circulation patterns. Around the time of the solstices, no insolation at all occurs near the winter pole (the polar night), but at the same time, near the summer pole, the daily mean insolation is very strong despite low sun angles, simply because the sun never sets (the polar day). As is well known, these effects arise from the sun-Earth geometry shown in Fig. 2.1. In addition, the distance from the sun to the Earth varies slightly with time of year, resulting in a few percent more globally averaged insolation in January than in July, in the current epoch. The month of maximum insolation varies over geologic time. According to the widely accepted astronomical theory of the ice ages, extensive glaciation is favored when the minimum insolation occurs during the Northern Hemisphere summer, because the Northern Hemisphere contains about twice as much land as the Southern Hemisphere (e.g., Crowley and North, 1991).

The Earth emits about as much infrared radiation as required to balance the absorbed solar radiation; both energy flows are at the rate of about 240 W m^{-2}. This near balance has been directly confirmed by analysis of satellite data. The balance is observed to hold within one Watt per square meter, which is comparable to the uncertainty of the measurements. The actual annual-mean imbalance is thought to be about 0.5 W m^{-2} (Loeb et al., 2012; Trenberth et al., 2014).

Fig. 2.3 shows aspects of the Earth's radiation budget as observed from satellites (Wielicki et al., 1998). As discussed above, the zonally averaged incident (i.e., incoming) solar radiation at the top of the atmosphere varies seasonally as the Earth orbits the sun. The zonally averaged albedo, which is the fraction of the zonally averaged incident radiation that is reflected back to space, is highest near the poles, due to cloudiness as

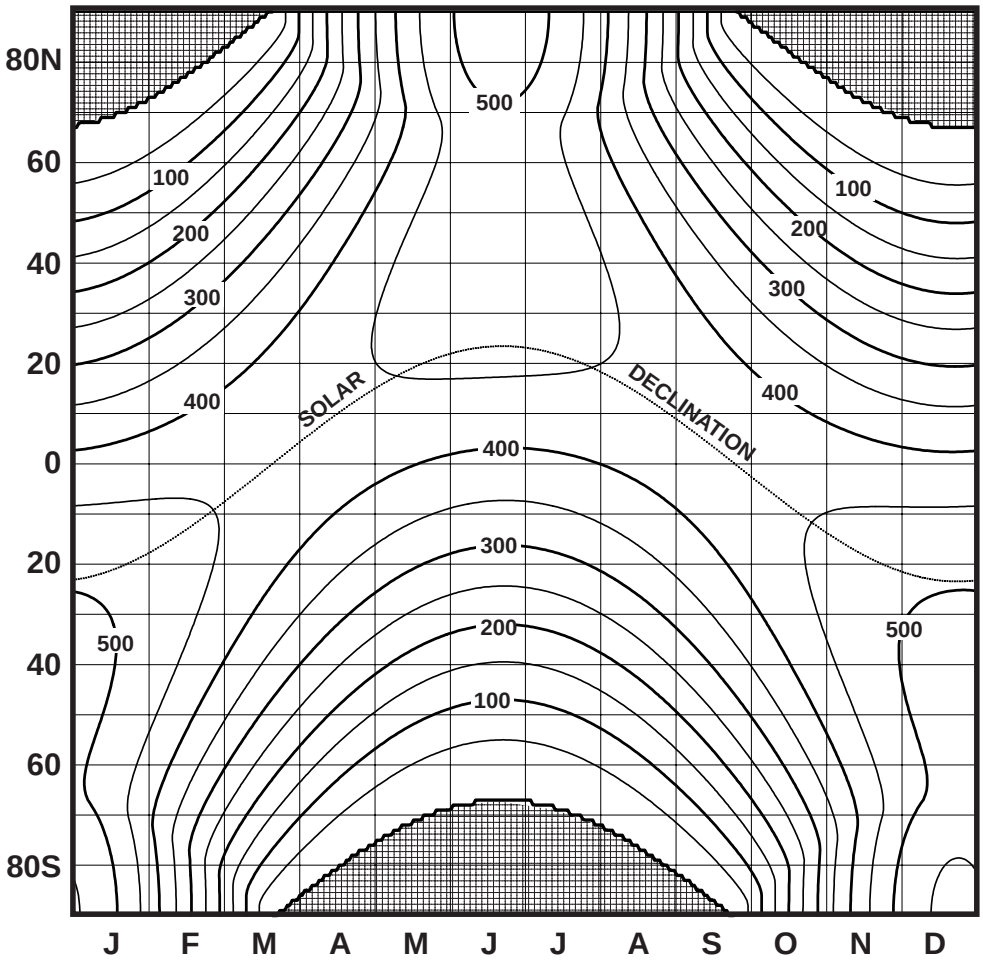


Figure 2.2: The seasonal variation of the zonally (or diurnally) averaged insolation at the top of the atmosphere. The units are W m^{-2}.

well as snow and ice. It tends to have a weak secondary maximum in the tropics, associated with high cloudiness there. The zonally averaged terrestrial radiation at the top of the atmosphere, also called the outgoing longwave radiation or OLR, has its maxima in the subtropics. It is relatively small over the cold poles, but it also has a minimum in the warm tropics, due to the trapping of terrestrial radiation by the cold, high tropical clouds, and by water vapor.

The net radiation at the top of the atmosphere, which is the difference between the absorbed solar radiation and the OLR, is positive in the tropics and negative in higher latitudes. This implies that energy is transported poleward somehow, inside the system. A portion of the energy is transported by the atmosphere, and the rest is transported by the oceans.

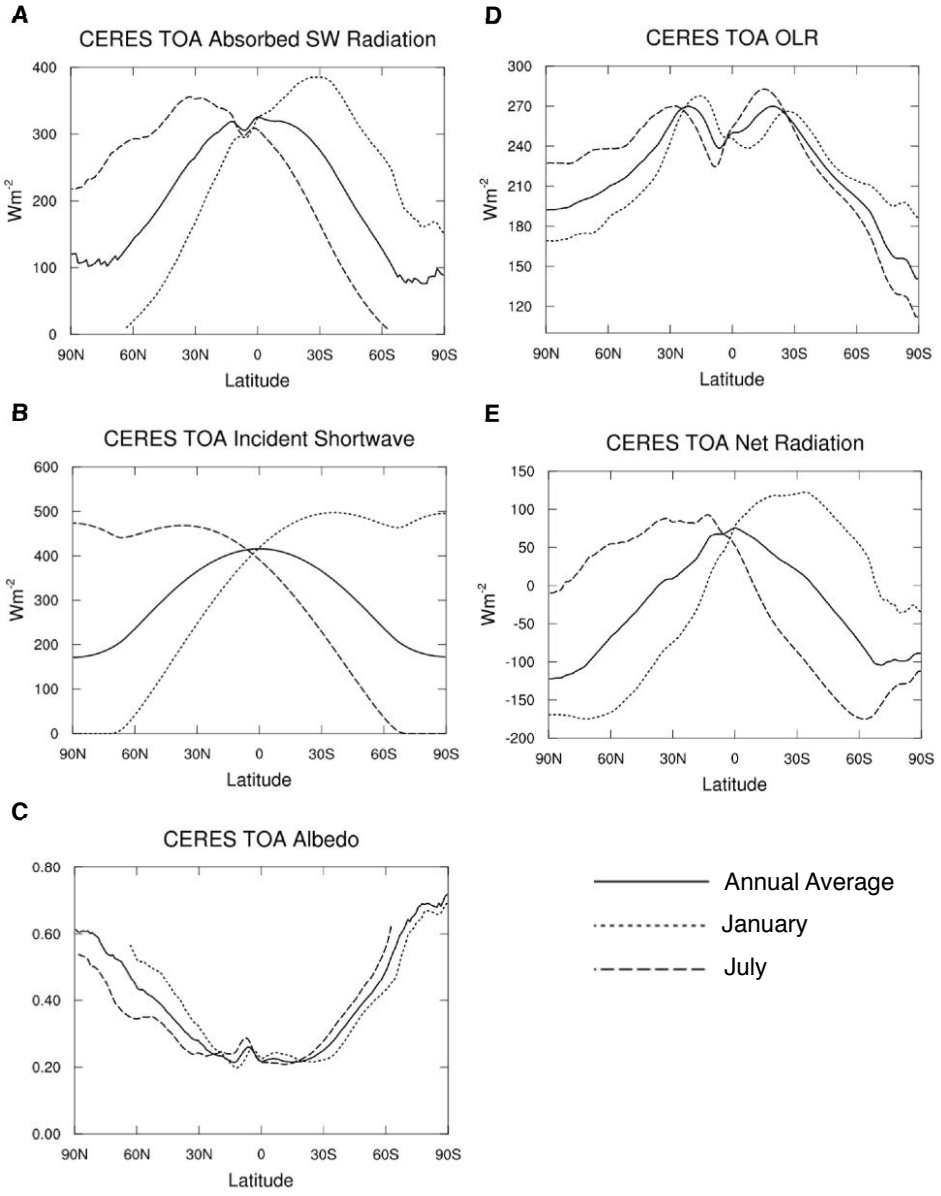


Figure 2.3: The zonally averaged A) absorbed solar radiation at the top of the atmosphere, B) incident solar radiation, C) albedo, D) outgoing longwave radiation (OLR), and E) net radiation at the top of the atmosphere, as observed using satellites in a project called “Clouds and the Earth’s Radiant Energy System” (CERES; Wielicki et al., 1998).

2.2 Meridional energy transports by the atmosphere-ocean system

Considering the energy balance of the atmosphere-ocean system, the variation with latitude of the long-term average net radiation at the top of the atmosphere implies energy transports

inside the system. These transports are produced by the circulations of both the atmosphere and the oceans, and it is tempting to regard the global circulations of the atmosphere and oceans as a "response" to this pattern of net radiation. An important point, however, is that the distributions of the albedo and the outgoing longwave radiation are determined in part by the circulation itself. It is an egregious oversimplification to regard the net radiation at the top of the atmosphere as an "external" forcing function.

Consider the energy budget of a column which extends from the center of the Earth to the "top of the atmosphere:"

$$\frac{\partial E}{\partial t} = N_\infty - \nabla \cdot \mathbf{G}_\infty \, . \tag{2.2}$$

Here E is the energy per unit area stored in the column; t is time; and N_∞ is the net downward flux of energy at the top of the atmosphere, which is entirely due to radiation. The dimensions of N_∞ are energy per unit time per unit area (e.g., W m^{-2}). The energy transport, $\mathbf{G}_\infty$, represents the movement of energy, in the zonal and meridional directions, due to both the winds and the ocean currents. It is a vector with both zonal and meridional components, and it has dimensions of energy per unit length per unit time (e.g., W m^{-1}). The subscript ∞ on $\mathbf{G}_\infty$ signifies that it includes energy transports in all parts of the Earth system, from the center of the Earth out to space.

Suppose that we average over a time interval Δt:

$$\frac{E(t+\Delta t) - E(t)}{\Delta t} = \overline{N_\infty}^t - \nabla \cdot \overline{\mathbf{G}_\infty}^t \, . \tag{2.3}$$

Here $\overline{(.)}^t$ represents a time average; this is a notation that we will use throughout the book. Because the Earth is close to energy balance, $E(t+\Delta t)$ and $E(t)$ cannot be wildly different from each other, so the numerator on the left-hand side of (2.3) is bounded within a finite range, regardless of how large Δt is. Therefore, as Δt increases, the left-hand side of (2.3) decreases in absolute value, and eventually becomes negligible compared to the individual terms on the right-hand side. The physical meaning is that energy storage inside the Earth system can be neglected if the time-averaging interval is long enough; here the minimum interval would be one year, but ideally the average should be taken over many years. When we apply such a time average, the net radiation across the top of the column must be balanced by transports inside; this can be written as

$$\nabla \cdot \overline{\mathbf{G}_\infty}^t = \overline{N_\infty}^t \, . \tag{2.4}$$

The global mean of $\nabla \cdot \mathbf{G}_\infty$ must be exactly zero, not just in a time average, but at each instant. The reason is purely mathematical, rather than physical: the global mean of the divergence of any vector is zero. You are asked to prove that in Problem 1 at the end of this chapter. The only way that (2.4) can be true is if the global mean of $\overline{N_\infty}^t$ is equal to zero, i.e., if the Earth is in energy balance. To the extent that this is not true, the left-hand side of (2.3) is not negligible.

We now break $\mathbf{G}_\infty$ into its zonal and meridional components, i.e.,

$$\mathbf{G}_\infty = (G_\infty)_\lambda \mathbf{e}_\lambda + (G_\infty)_\varphi \mathbf{e}_\varphi \ . \tag{2.5}$$

Here λ and φ represent longitude and latitude, respectively, and $\mathbf{e}_\lambda$ and $\mathbf{e}_\varphi$ are unit vectors pointing towards the east and north, respectively. We expand the divergence operator in spherical coordinates (see Appendix A) as follows:

$$\nabla \cdot \mathbf{G}_\infty = \frac{1}{a\cos\varphi}\frac{\partial (G_\infty)_\lambda}{\partial \lambda} + \frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left[(G_\infty)_\varphi \cos\varphi\right] \ . \tag{2.6}$$

Here a is the radius of the Earth.[2] Multiply both sides of (2.6) by $a\cos\varphi$ and integrate over all longitudes to obtain

$$\begin{aligned} \int_0^{2\pi} (\nabla \cdot \overline{\mathbf{G}_\infty}^t)\, a\cos\varphi d\lambda &= 2\pi \frac{\partial}{\partial \varphi}\left[\overline{(G_\infty)_\varphi}^{\lambda,t}\cos\varphi\right] \\ &= 2\pi \overline{N_\infty}^{\lambda,t} a\cos\varphi \quad . \end{aligned} \tag{2.7}$$

Here we use the notation $\overline{(.)}^\lambda \equiv \frac{1}{2\pi}\int_0^{2\pi}()\,d\lambda$ to denote a zonal mean; the combination of a time average and a zonal mean can be denoted by either $\overline{(.)}^{\lambda,t}$ or $\overline{(.)}^{t,\lambda}$. In (2.7), the zonal derivative has dropped out as a result of the integration with respect to longitude. The second line of (2.7) comes by comparison with (2.4).

Eq. (2.7) gives us a way to compute the meridional derivative of the poleward energy transport, i.e., $\dfrac{\partial}{\partial \varphi}\left[\overline{(G_\infty)_\varphi}^{\lambda,t}\cos\varphi\right]$, in terms of $\overline{N_\infty}^{\lambda,t}$. What we really want, however, is a

[2] Strictly speaking, a should be the radial distance from the center of the Earth to the "top of the atmosphere," but that is approximately the same as the radius of the Earth because the Earth's atmosphere is sufficiently shallow. This is an example of what is called the "thin atmosphere approximation," which will be discussed further in Chapter 4.

formula for the poleward energy transport itself, rather than its divergence. To get this, we multiply (2.7) by a, and integrate with respect to latitude, from the South Pole ($\varphi = -\pi/2$) to an arbitrary latitude φ, to obtain

$$\overline{\Theta}^t(\varphi) = 2\pi a^2 \int_{-\pi/2}^{\varphi} \overline{N_\infty}^{\lambda,t} \cos\varphi' \, d\varphi' \,, \tag{2.8}$$

where we define

$$\Theta(\varphi) \equiv 2\pi a \cos\varphi \overline{(G_\infty)}_\varphi^\lambda \,, \tag{2.9}$$

and we have used the boundary condition

$$\Theta(-\pi/2) = 0 \,. \tag{2.10}$$

The dimensions of $\Theta(\varphi)$ are energy per unit time, e.g., Watts. The boundary condition (2.10) is exact; if it were not true, a finite amount of energy per unit time would be flowing into or out of the South Pole, which is a "point" of zero mass. A similar condition must apply at the North Pole. The right-hand side of (2.8) is the area integral of $\overline{N_\infty}^t$ over the "south polar cap" that extends from the South Pole up to latitude φ. When the upper limit of meridional integration in (2.8) is set to $\pi/2$, the right-hand side of (2.8) reduces to the global mean of $\overline{N_\infty}^t$. But since $\Theta(\pi/2)$ must be equal to zero, it follows that the global mean of $\overline{N_\infty}^t$ is zero.

Fig. 2.4 gives a plot of $\Theta(\varphi)$, computed by using (2.9), with values of based on satellite data from a project called CERES (Wielicki et al., 1996, 1998). A poleward energy transport is clearly apparent in both hemispheres. The curve of the transport has a pleasingly simple shape, roughly like $\sin(2\varphi)$. VonderHaar and Oort (1973) were the first to diagnose $\Theta(\varphi)$ from satellite measurements of the Earth's radiation budget. The maximum absolute values in middle latitudes of both hemispheres are on the order of 6 PW (a PetaWatt is 10^{15} Joules per second). Because this total energy transport by the climate system can be inferred directly from satellite measurements of the Earth's radiation budget, it is known with relatively good accuracy now.

Fig. 2.4 shows that that $\overline{\Theta}^t(\varphi)$ is exactly zero at both poles. We built that in for the South Pole, by using (2.10), but what makes $\overline{\Theta}^t(\varphi)$ miraculously return to zero at the North

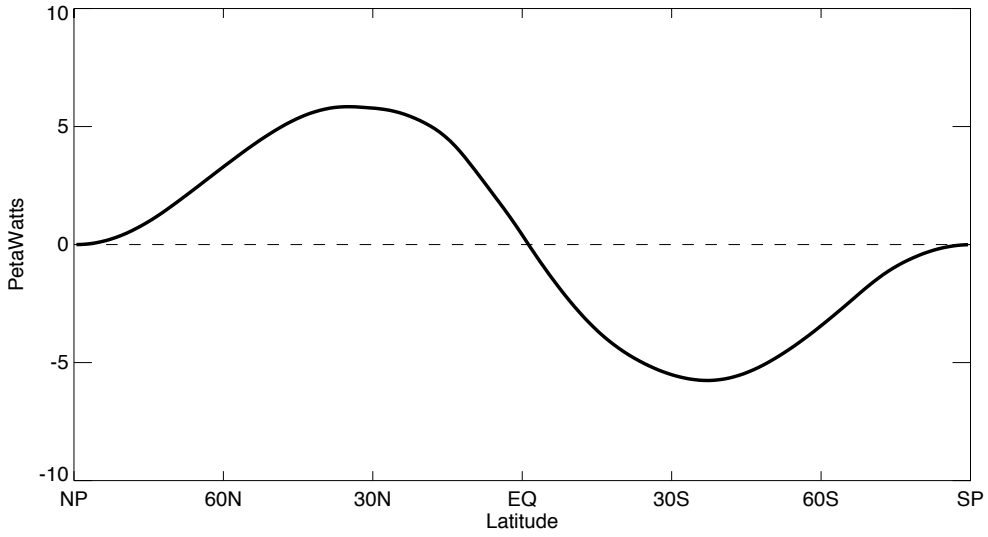


Figure 2.4: The annually averaged northward energy transport by the atmosphere and ocean combined, $\overline{\Theta}^t(\varphi)$, as inferred from the observed annually averaged net radiation at the top of the atmosphere using (2.9). A Petawatt is 10^{15} W. The plot was created using CERES data.

Pole? The answer is that we have forced it to be zero at the North Pole, by "correcting" the data. The observations show that the global mean of N_∞ is small compared to the local values. It is zero within the uncertainty of the measurements, but of course it is not exactly zero. Before computing $\overline{\Theta}^t(\varphi)$ from (2.9), we apply a small, globally uniform correction to N_∞, such that after the correction the global mean of N_∞ is exactly zero. As can be seen from (2.9), this is sufficient to ensure that $\overline{\Theta}^t(\varphi)$ is zero at the North Pole.

The value of N_∞ is is not exactly zero for several reasons:

- The Earth is not in perfect energy balance. As mentioned earlier, observations suggest that the global mean of N_∞ is about 0.5 W m^{-2} (Loeb et al., 2012; Trenberth et al., 2014).
- The data has been averaged over just a few decades, so even if the long-term global mean of N_∞ was exactly zero, the time derivative term of (2.2) might not be not completely negligible.
- The data are not perfectly accurate.

We can say that the "job" of the global circulations of the atmosphere and oceans is to carry out the meridional energy transport shown in Fig. 2.4. If the transport of energy from place to place by the atmosphere and oceans could somehow be prevented, then each part of the Earth would have to come into local energy balance, by adjusting its temperature, water vapor, and cloudiness so that the outgoing longwave radiation locally balanced the absorbed solar radiation. Such a hypothetical state is referred to as "radiative-convective

equilibrium;" modeling studies of radiative-convective equilibrium are discussed in Chapter 8. Radiative-convective equilibrium would presumably entail much warmer temperatures in the tropics, and much colder temperatures at the poles. The global circulation of the atmosphere and oceans has a moderating effect on the global distribution of temperature, tending to warm the higher latitudes and cool the tropics. As discussed in Chapter 7, these same thermal contrasts represent a source of energy (called "available potential energy") that makes the global circulations of the atmosphere and oceans possible.

If we consider that the global circulation of the atmosphere exists in order to produce the energy transports shown in Fig. 2.4, then we can imagine that the "strength of the circulation," as measured for example by its total kinetic energy, is determined by the magnitude of the required energy transports.

Much further discussion of the observations and theory of energy transports by the atmosphere and oceans is given later in this book, especially in Chapter 6.

2.3 Surface boundary conditions

The the global atmospheric circulation is strongly affected by the properties of the Earth's surface, and their geographical variations. The most important properties of the Earth's surface are as follows:

2.3.1 Temperature

The temperature of the Earth's surface varies strongly and rapidly over land, and considerably less over the oceans. The reason for this difference between land and sea will be discussed below, in the subsection on the surface heat capacity.

The oceans cover about two thirds of the Earth's surface. Their average depth is about 4 km. Water is heavy stuff; the mass of 1 m^3 of water is 10^3 kg. The mass of the oceans is about 1.3 x 10^{21} kg. For comparison, the mass of the atmosphere is about 250 times less, roughly 5 x 10^{18} kg.

Not only is water dense, it has a very high specific heat: about 4200 J kg^{-1} K^{-1}. In contrast, the specific heat of air (at constant pressure) is a little less than a quarter of that, i.e., 1000 J kg^{-1} K^{-1}. The total heat capacity of the oceans is thus about 1000 times larger (250 x 4) than the total heat capacity of the atmosphere. When the oceans say "Jump," the atmosphere says "How high?"

The density of the atmosphere decreases exponentially upward with height, and can change by 10 per cent or so in the course of a year, at a given location, due to changes in temperature and pressure. Liquid water is *much* less compressible that air. For example, the density of the water at the bottom of the Mariana Trench, almost 11,000 m below sea level, is only about 5% larger than the density of water at the sea surface, despite the fact that the pressure there is about 1,000 times greater than the pressure at the surface. The

density of sea water is a complex but fairly weak function of temperature, salinity, and pressure. The effects of pressure on the density of sea water are relatively unimportant, so variations are mainly due to changes in temperature and salinity. Warmer and fresher water is less dense and tends to float on top; colder and saltier water is more dense and tends to sink. Surface cooling and evaporation create dense water; surface heating and precipitation create light water. Note that the properties of sea water are altered mainly near the surface; below the top hundred meters or so, the properties of water parcels remain nearly invariant, over decades or even centuries.

In the study of the atmosphere, we sometimes treat the sea-surface temperature (SST) as a seasonally varying lower boundary condition. Fig. 2.5 shows the observed distributions of the SST for January and July. Note the warm currents off the east coasts of North America and Asia, and the cold currents off all west coasts. The warm SSTs, at a particular latitude, are generally associated with poleward currents; the two best-known of these are the Gulf Stream and the Kuroshio. The colder SSTs are generally associated with either equatorward flow (as for example in the case of the California current) or with upwelling (again, in the region of the California current, and also along the Equator in the eastern Pacific).

As discussed in Chapter 10, the pattern of upwelling is very closely related to the low-level winds, and at the same time the low-level winds are strongly tied to the spatial distribution of the SST. The seasonal change of the SST is largest in the Northern Hemisphere, particularly on the western sides of the ocean basins. Note that the seasonal forcing is capable of changing the SSTs by tens of degrees in some middle and high latitude locations. The depth to which this seasonal change penetrates is of course variable, but is typically on the order of 100 m. Of course, the temperature of the water at great depth undergoes virtually no seasonal change. In the study of the global atmospheric circulation we often consider the spatial and seasonal distribution of the SST to be "given," but of course in reality it is determined in part by what the atmosphere is doing, or rather what the atmosphere has been doing over time. For example, the distribution of cloudiness strongly affects the flow of solar radiation into the upper ocean, and over time this can tend to reduce the SST where clouds are prevalent and the solar insolation at the top of the atmosphere is strong, relative to what the SST would be if the cloudiness were somehow prevented from occurring. The role of clouds in determining the distribution of the SST is a major complication hindering our understanding of the atmosphere and ocean as a coupled system.

2.3.2 Wetness

One of the most important properties of the Earth's surface is that roughly 70 per cent of it is permanently wet, and so represents a huge source of moisture. The vapor pressure immediately above a wet surface, called the "saturation vapor pressure," is a strong function of temperature only. It is approximately given by

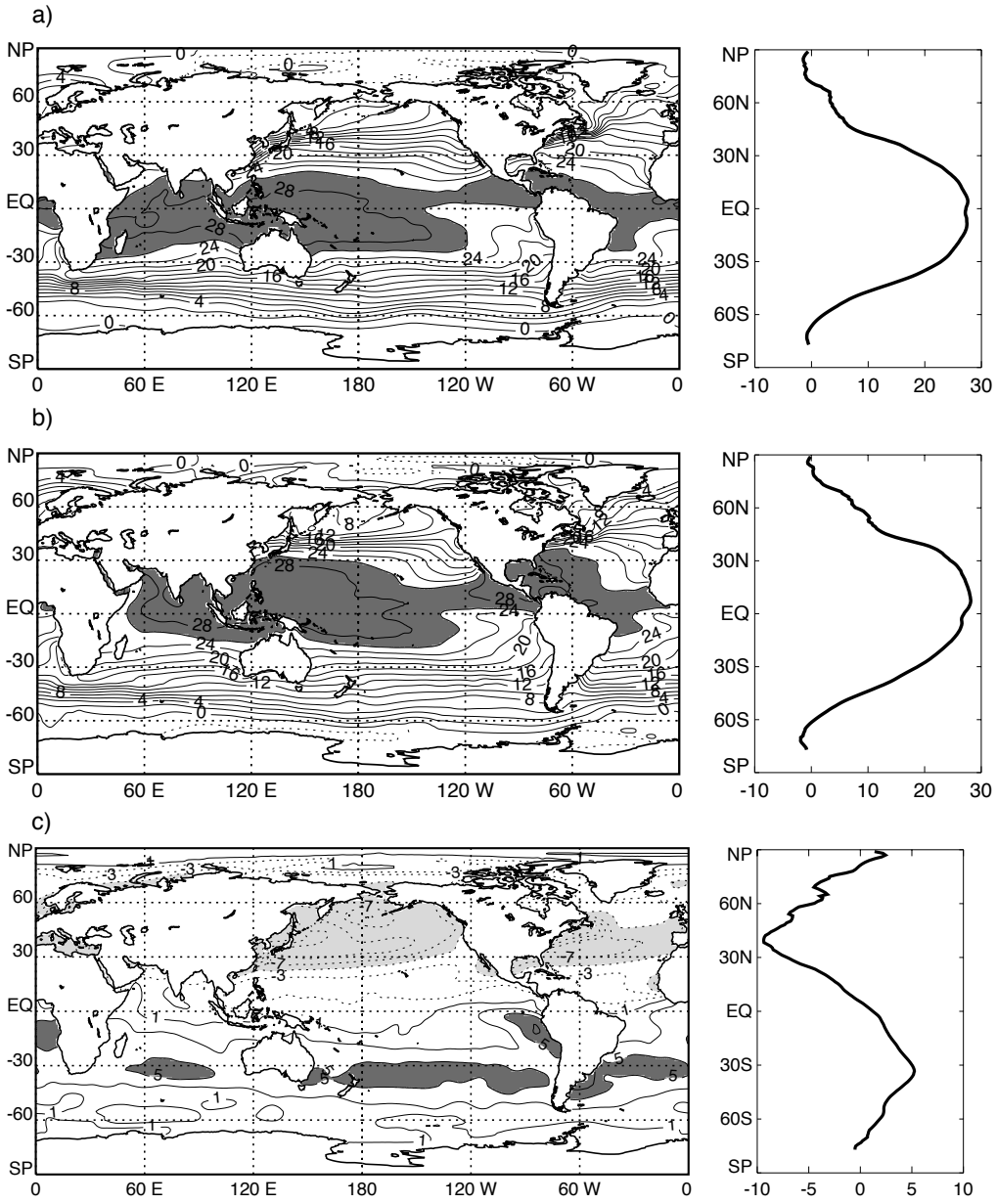


Figure 2.5: a) SST distribution for January. The contour interval is 2°C, with shading for values higher than 28° C. b) Same for July. c) SST difference between March and September, with light shading for values below -5° C and dark shading for values above 5°C. In each panel, zonal means are shown on the right.

$$e_{\text{sat}}(T) \cong 6.11 \exp\left[\frac{L}{R_v}\left(\frac{1}{273} - \frac{1}{T}\right)\right] \quad \text{hPa}\,, \tag{2.11}$$

where the temperature is given in K, $L = 2.52$ x 10^6 J kg^{-1} is the latent heat of water vapor, and $R_v = 461$ J K^{-1} kg^{-1} is the gas constant for water vapor. The enormous value of L is one of the reasons why moisture strongly influences the Earth's climate and the global circulation of the atmosphere. The exponential temperature-dependence of $e_{\rm sat}$ is plotted in Fig. 2.6. Held and Soden (2006) pointed out that for typical surface temperatures, the rate of increase is a spectacular 7% per Kelvin. Fig. 2.7 shows the geographical variation of based on the SSTs plotted in Fig. 2.5. As discussed later, near the surface over the oceans the actual vapor pressure of the air, i.e., the partial pressure of water vapor, "tries" to be $e_{\rm sat}(T)$, but usually falls short by 20% or so. At any rate, the "effective wetness" of the ocean increases rapidly with the SST. The largest values of $e_{\rm sat}$ occur in the tropics, of course, and are close to 40 hPa. In those very humid regions, about 4% of the near-surface air is water vapor.

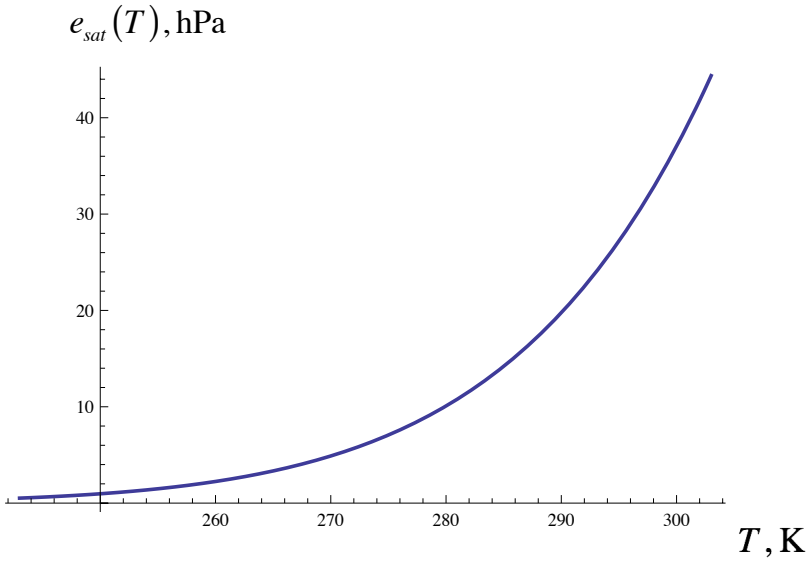


Figure 2.6: The saturation vapor pressure, $e_{\rm sat}$, as a function of temperature.

The availability of land-surface moisture to the atmosphere is much more complicated, and is discussed separately below.

2.3.3 Topography

You will not be surprised to hear that mountains have a strong effect on the global atmospheric circulation. Fig. 2.8 shows the locations of the Earth's mountain ranges. Mountains block the wind; this can be described as a mechanical forcing. The air can flow around a mountainous obstacle, or it can flow over. Which actually happens depends in part on the scale of motion. The distribution of surface pressure across a mountain range can exert a net force on the solid Earth, and an equal and opposite net force on the atmosphere. Chapter 6 explains how this works.

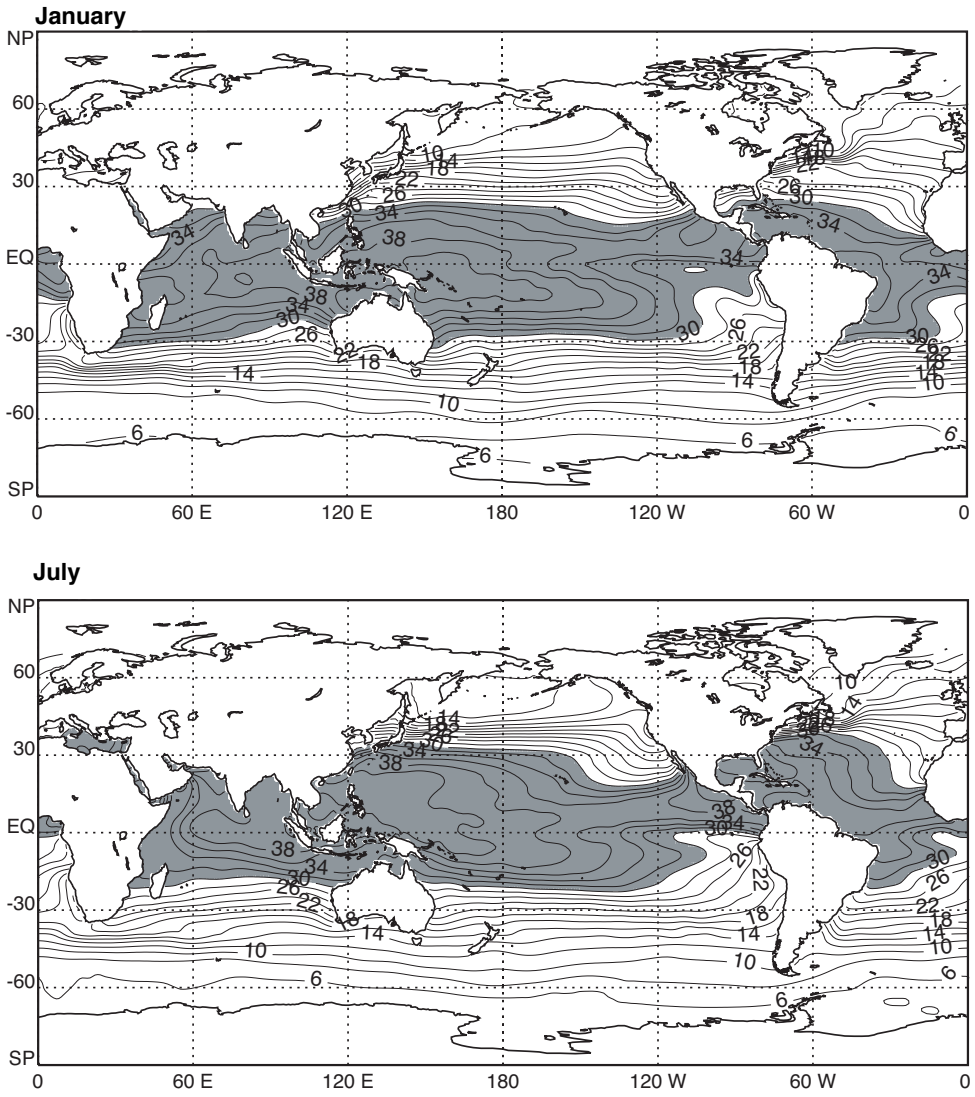


Figure 2.7: The geographical pattern of the saturation vapor pressure based on the SSTs shown in Fig. 2.5. The contour interval is 2 hPa. Values larger than 30 hPa are shaded.

Mountains can also exert a thermal forcing on the atmosphere, because the surface of a mountain can have a temperature quite different from that of the surrounding air at the same height. For example, during the northern summer the Tibetan plateau produces a "warm spot" in the middle troposphere, and this contributes to the thermal forcing that produces the Indian summer monsoon. Further discussion is given in Chapter 10.

Finally, mountains strongly influence the geographical distribution of precipitation.

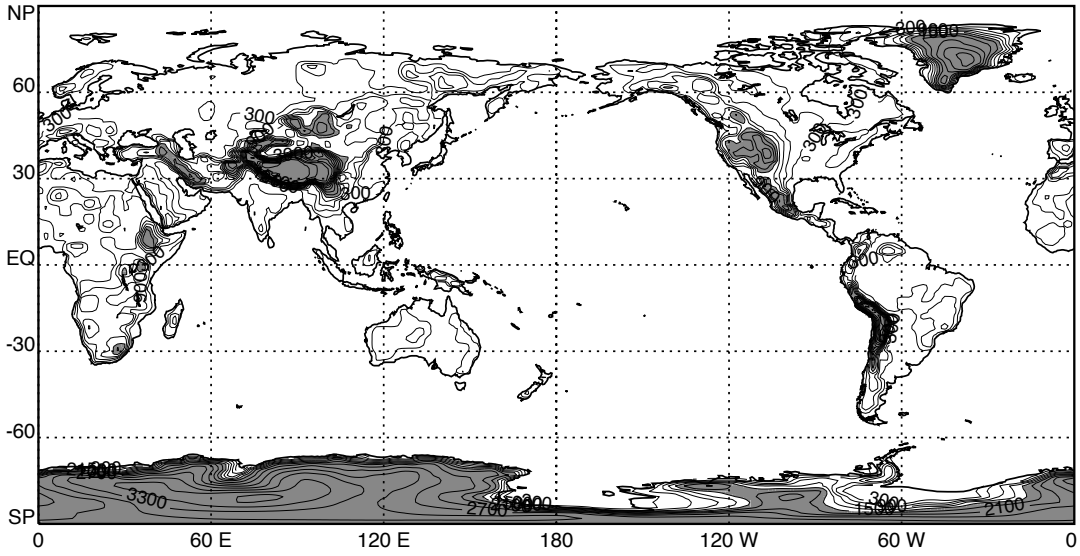


Figure 2.8: The Earth's orography, averaged onto a mesh with a grid spacing of 1° of longitude by 1° of latitude. The contour interval is 300 m. Values higher than 1500 m are shaded.

Rain and snow are enhanced where topography forces the air to flow uphill, and also where surface heating on sunlit slopes promotes moist convection. On the other hand, precipitation is often reduced on the downstream side of mountainous regions, because the air often descends there, and also because the moisture content of the air can be depleted by upstream precipitation maxima.

2.3.4 Heat capacity

The heat capacity of the Earth's surface is the amount of energy needed to change the "skin temperature" of the surface by a given amount. Here the skin temperature, T_S, is defined as the temperature of an equivalent black body that would emit infrared radiation at a rate equal to the actual infrared emission by the Earth's surface. The heat capacity is highly variable in space, and also varies somewhat less dramatically with time. The concept of heat capacity sounds simple, but it is actually somewhat subtle. The time change of the skin temperature satisfies an equation very similar to (2.2), i.e.,

$$C\frac{\partial T_S}{\partial t} = N_S - \nabla \cdot \mathbf{G}_S \tag{2.12}$$

Here C is the heat capacity of the surface, N_S is the net downward energy flux at the Earth's surface (due to radiation and other processes discussed later), and $\mathbf{G}_S$ is the horizontal energy transport "inside" the Earth's surface. For the land, we can assume that $\mathbf{G}_S = 0$, but for the oceans we expect that energy transport by currents will lead to $\mathbf{G}_S \neq 0$. The

heat capacity, C, depends on the composition of the material both at and below the surface, because energy flowing into the surface can be stored through a finite depth. It also depends on the speed with which energy is transported down into the material below the surface.

In general, the oceans have very high heat capacity. The geographical distribution of SST does fluctuate seasonally, however, and varies considerably with both longitude and latitude, as shown in Fig. 2.5.

The land-surface has a much smaller heat capacity. This implies that the net surface energy flux averages to nearly zero over land, even for a single day. To understand why, note that with $C \to 0$ and $\mathbf{G}_S = 0$ (appropriate for land), Eq. (2.12) implies that $N_S = 0$. For the ocean, with large values of C, daily mean values of N_S can be much larger. The large heat capacity of the ocean means that the SST changes slowly, because

$$\frac{\partial T_S}{\partial t} = \frac{N_S - \nabla \cdot \mathbf{G}_S}{C} \tag{2.13}$$

i.e., a large value of C in the denominator on the right-hand side of (2.13) reduces $\partial T_S/\partial t$ for a given value of $N_S - \nabla \cdot \mathbf{G}_S$.

2.3.5 Albedo

The degree to which the surface reflects solar radiation obviously affects its response to the sun. The surface albedo depends on surface composition and sun angle, among other things. The ocean has an albedo close to 0.06 when the sun is high in the sky, i.e., it is quite dark. At low sun angles, however, the ocean can reflect considerably more of the incident solar radiation. The albedo of the land surface varies widely, due to differing compositions of the soil or rock at the surface, differing types and amounts of vegetation cover (discussed further below), and of course the presence or absence of snow.

2.3.6 Roughness

"Rough" surfaces exert a drag on the wind more readily than smooth ones. The surface roughness is another example of a lower boundary condition that is at least partially mechanical in nature. The ocean is relatively smooth, depending on the wind speed, and presents little "roughness" to stimulate momentum exchange with the atmosphere. The land surface is much rougher than the ocean.

2.3.7 Vegetation

The vegetation on the land-surface regulates the flow of moisture from the soil, as discussed further below. It also affects both the roughness and albedo of the surface. The pattern of vegetation on the land surface affects the atmosphere in very complicated ways. Fig. 2.9

shows a highly simplified distribution of vegetation types on the land surface, and Fig. 2.10 shows the "leaf area index," which is a measure of the greenness of the vegetation. Obviously, the type, density, and even the health of the land-surface vegetation can affect the surface albedo and surface roughness. These characteristics of the vegetation vary with season, especially in middle latitudes. They can also vary interannually. The degree to which the plants allow moisture to transpire from leaves into the atmosphere strongly regulates the surface fluxes of sensible and latent heat; strong transpiration cools the surface and reduces the sensible heat flux. Sellers et al. (1997) provide an introductory overview.

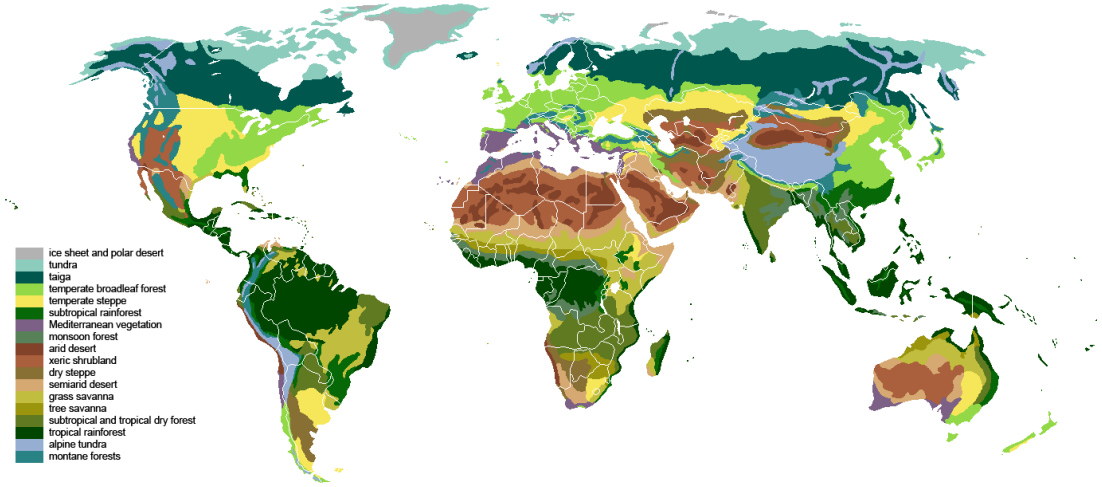


Figure 2.9: A simplified map of vegetation types on the land surface. From `https://en.wikipedia.org/wiki/Biome`.

2.3.8 Sea ice

The distribution of sea ice (Fig. 2.11) also acts as a thermal lower boundary condition. There are strong seasonal changes in ice cover in the Southern Hemisphere, but not in the Northern Hemisphere. In addition to the obvious strong effect of sea ice on the surface albedo, the ice also acts as an insulator that separates the relatively warm ocean water from the air. Because sea ice is a good insulator, its upper boundary can be much colder than the water beneath. Sea ice is also very smooth, so that little surface drag occurs for a given wind speed. Until recently, the Arctic ocean has been ice-covered all year, while the North Atlantic and the Southern Oceans have long experienced seasonal melting. Of course, the thickness of the ice also varies both geographically and seasonally, and the thickness strongly determines the insulating power of the ice. In addition, several percent of open water typically occurs, especially when the ice is thin. Some of the open water is found in cracks in the ice, called "leads." The water in the leads can be much warmer than the ice nearby, especially in winter. Under such conditions, the large-scale average sensible and latent heat fluxes can be dominated by the contributions from the leads, even though leads

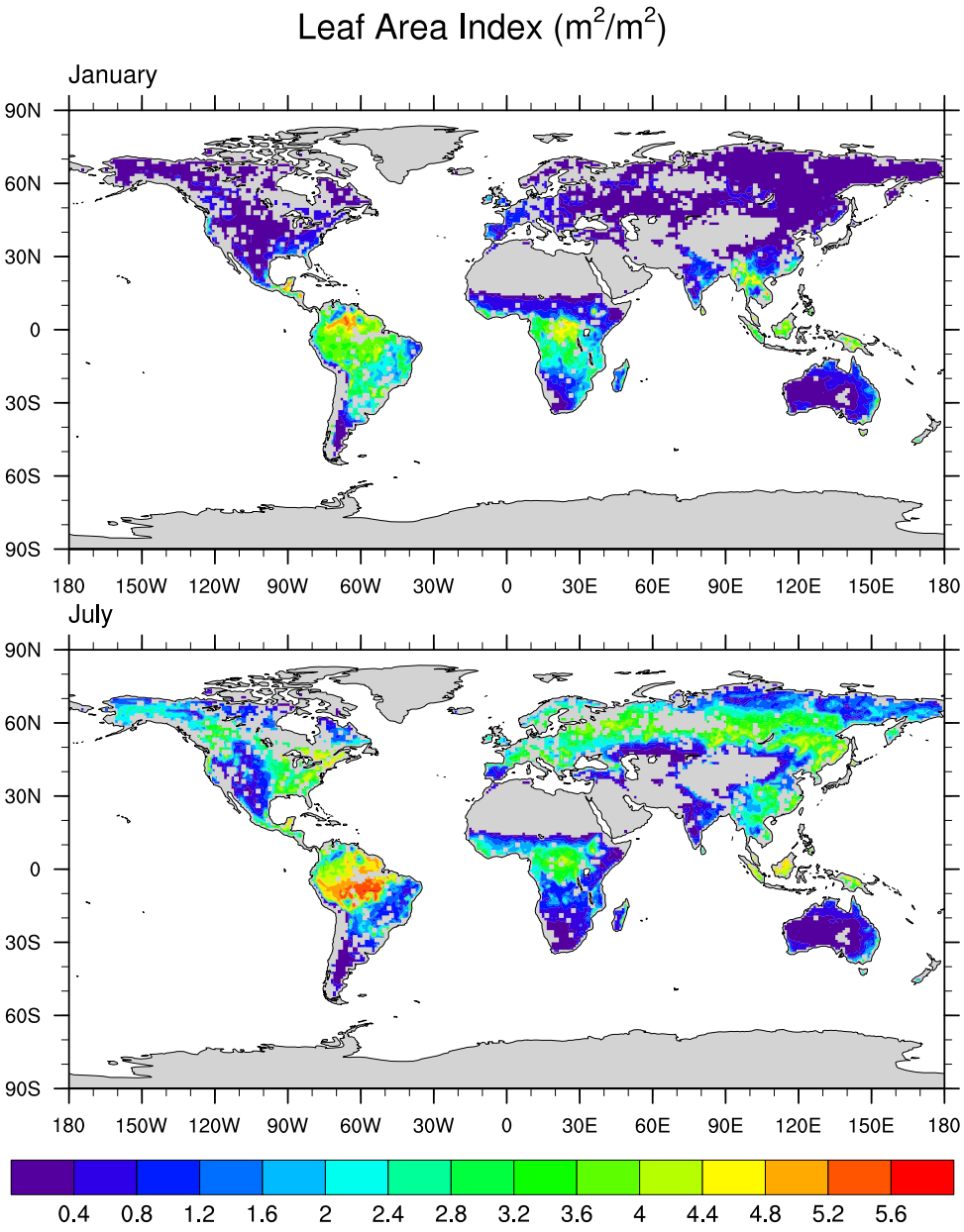


Figure 2.10: The leaf area index, a measure of greenness, for January and July. The data were obtained from `https://daac.ornl.gov/VEGETATION/guides/Mean_Seasonal_LAI.html`, and have been plotted with one-degree resolution.

may cover only a few percent of the area. Snow that falls on the sea ice insulates it and protects it from the effects of the sun, helping to prevent the ice from melting.

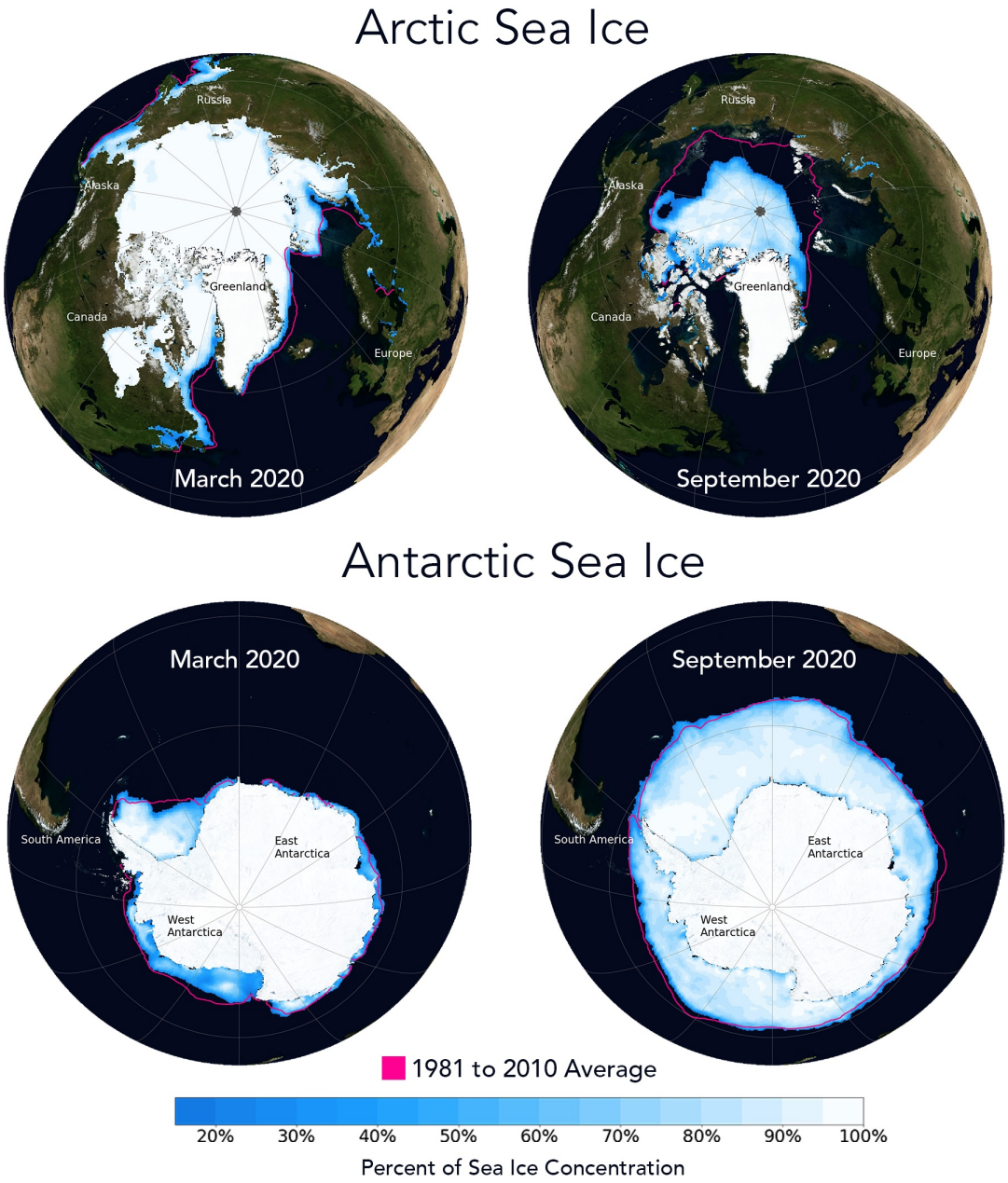


Figure 2.11: The Arctic and Antarctic distributions of sea ice for March and September of 2020. The red curves outlines the average ice distributions for the years 1981 to 2020. From `https://nsidc.org/learn/parts-cryosphere/sea-ice`.

2.3.9 Land ice

The land-sea distribution and the locations of "permanent" (or, more accurately, non-seasonal) land ice (e.g., the ice sheets that cover Antarctica and Greenland) strongly affect the surface albedo. Over land, the geographical and seasonal variations of the surface

albedo are largely determined by the distribution of vegetation, but of course they also depend on snow cover. Permanent land ice is mainly confined to Antarctica and Greenland, in the present climate, although there are many smaller glaciers throughout the world. The Greenland and Antarctic ice sheets are thousands of meters thick in places, and so increase the effective local topographic height of the Earth's surface by substantial amounts. The distribution of land ice can vary dramatically on time scales of thousands of years and longer (e.g., Imbrie and Imbrie, 1986; Crowley and North, 1991).

2.4 Energy and moisture budgets of the surface and atmosphere

Some aspects of the global atmospheric circulation can be regarded as more or less direct responses to the various boundary conditions mentioned above. Examples include the equator-to-pole energy flux by the atmosphere, planetary waves produced by flow over mountains, and monsoons that are strongly tied to the land-sea distribution and the seasonally varying insolation. Of course, there are many additional time-dependent features of the circulation that are less directly tied to the boundary conditions, but instead arise from the internal dynamics of the atmosphere, which include winter storms, tropical cyclones, and many other things.

The planetary radiation budget has already been discussed. We now consider the energy and moisture budgets of the Earth's surface and the atmosphere, as shown in Table 2.2. All of the numbers in this table have uncertainties of several per cent. Of the 239 W m^{-2} that is absorbed by the Earth-atmosphere system, 161 W m^{-2} is absorbed by the Earth's surface. Thus only about 239 - 161 = 78 W m^{-2} of solar radiation is absorbed by the atmosphere. That is only about 1/3 of the total solar radiation absorbed by the Earth-atmosphere system. The surface receives a total (LW $\downarrow$ + SW; see the notation defined in Table 2.2) of 494 W m^{-2} of "incoming radiation. Note that LW $\downarrow$ is about twice as large as SW! This is the "greenhouse" effect on the surface energy budget. The incoming energy due to longwave and solar radiation absorbed by the surface is given back to the atmosphere in the form of $(\mathrm{LW}\uparrow)_S$, $L\left(F_q\right)_S$ and $(F_s)_S$. By far the largest of these is $(\mathrm{LW}\uparrow)_S$. The oceans can transport energy from one place to another, so the energy absorbed by the oceans is not necessarily given back in the same place. Also, the large heat capacity of the upper ocean allows energy storage on seasonal time scales. In contrast, the continents cannot transport energy laterally at any significant rate, and their limited heat capacity forces near energy balance, everywhere, on time scales of a few days at most. Table 2.2. shows that the net radiative heating of the surface, which amounts to 98 W m^{-2}, is balanced primarily by evaporative cooling of the surface at the rate of 80 W m^{-2}. In other words, the surface cools itself by evaporating water.

The globally averaged energy budget of the atmosphere is shown in 2.3. The atmosphere sheds energy through infrared radiation at the rate required to balance the various forms of energy input, and the temperature of the atmosphere adjusts to allow the necessary

Table 2.2: Components of the globally and annually averaged surface energy budget, after Trenberth et al. (2009). A positive sign means that the surface is warmed.

Absorbed solar radiation, S_S	161 W m^{-2}
Downward terrestrial radiation, $(\text{LW}\downarrow)_S$	333 W m^{-2}
Upward terrestrial radiation, $(\text{LW}\uparrow)_S$	-396 W m^{-2}
Net terrestrial radiation, $(\text{LW})_S$	-63 W m^{-2}
Net radiation, R_S	98 W m^{-2}
Latent heat flux, $L\left(F_q\right)_S$	-80 W m^{-2}
Sensible heat flux, $(F_s)_S$	-17 W m^{-2}

infrared emission. The net radiative cooling of the atmosphere, at the rate of -98 W m^{-2}, is primarily balanced by the latent energy source due to surface evaporation. Of course, the latent energy is converted into sensible heat when water vapor condenses. A fraction of the condensed water re-evaporates inside the atmosphere. The net condensation rate within the atmosphere is closely balanced by the rate of precipitation at the Earth's surface; this means that the amount of condensed water in the atmosphere is neither increasing nor decreasing with time. The rate at which evaporation introduces moisture into the atmosphere has to be balanced by the rate at which precipitation removes it. Keep in mind that these various balances apply in a globally averaged sense, rather than locally in space, and in a time-averaged sense, rather than instantaneously.

Table 2.3: The globally and annually averaged energy budget of the atmosphere. obtained by combining the numbers in Tables 2.1 and 2.2. A positive sign means that the atmosphere is warmed.

Absorbed solar radiation	78 W m^{-2}
Net loss of terrestrial radiation	-176 W m^{-2}
Net radiative heating	-98 W m^{-2}
Latent heat input	80 W m^{-2}
Sensible heat input	17 W m^{-2}

The globally averaged rate of precipitation, and the globally averaged rate of evaporation, are measures of the "speed" or intensity of the hydrologic cycle. The preceding discussion suggests a second interpretation of the atmospheric energy budget: To a first

approximation, the speed of the hydrologic cycle is "determined by" the rate at which the atmosphere is cooling radiatively. Of course, this does not mean that the geographical and temporal distributions of precipitation are determined by the corresponding distribution of radiative cooling; in fact, the local rate of precipitation tends to be negatively correlated with the local atmospheric radiative cooling, because precipitation systems produce high, cold clouds (see below) that reduce the infrared emission to space.

The local rate of precipitation is controlled mainly by dynamical processes, and the rate of evaporation from the Earth's surface is influenced by the surface wind speed. The overall strength of the global circulation of the atmosphere must be consistent with the speed of the hydrologic cycle that is required to balance the globally averaged rate of atmospheric radiative cooling.

The net radiative cooling of the atmosphere is strongly affected by the high, cold cirrus clouds, many of which are formed within precipitating cloud systems. The cirrus clouds absorb the infrared radiation emitted by the warm atmosphere and surface below; the cirrus themselves emit much more weakly because they are very cold. This means that the cirrus effectively trap infrared radiation inside the atmosphere. For this reason, as the cirrus cloud amount increases, the radiative cooling of the atmosphere decreases.

Consider together the following points, which have been made in the last few paragraphs:

- The radiative cooling of the atmosphere is primarily balanced by latent heat release in precipitating cloud systems.
- Precipitating weather systems produce cirrus clouds.
- Cirrus clouds tend to reduce the radiative cooling of the atmosphere.

In combination, these ideas suggest a negative feedback loop that tends to regulate the strength of the hydrologic cycle. Consider an equilibrium in which atmospheric radiative cooling and latent heat release are in balance. Suppose that we perturb the equilibrium by increasing the speed of the hydrologic cycle, including the rate of latent heat release. The same perturbation will (presumably) increase the rate of cirrus cloud production, which will reduce the rate at which the atmosphere is radiatively cooled. The radiative cooling acts to promote cloud formation through moist convection, so when the radiative cooling rate decreases, the rate of cirrus production decreases. In this way, the initial perturbation is damped. Further discussion is given in Chapter 6.

The "effective altitude" for infrared emission by the Earth-atmosphere system is near 5 km above sea-level. This simply means that the outgoing longwave radiation at the top of the atmosphere is equivalent to that from a black body whose temperature is that of the atmosphere near the 5 km level. Roughly speaking, then, atmospheric motions must carry energy upward from the surface through the first 5 km of the atmosphere, and infrared emission carries the energy the rest of the way out to space. This upward energy transport

by moving air occurs on both small scales, notably in boundary-layer turbulence and cumulus convection, and also on large scales, notably through midlatitude baroclinic eddies and the tropical Hadley circulation., which are discussed in later chapters. In short, the atmospheric circulation carries energy upward as well as poleward.

We now examine in more detail the fluxes of various quantities at the Earth's surface. In addition to the surface solar and terrestrial radiation, we must also consider the turbulent fluxes of momentum, sensible heat, and latent heat. In principle, we should also consider the fluxes of various chemical species, but this important aspect of the climate system is neglected here.

An example of the seasonal variations of the surface shortwave and longwave radiation at a particular station is given in Fig. 2.12, which shows the variations of the upward and downward shortwave (SW) and longwave (LW) near-surface radiation at a field site in Oklahoma. The data cover the three years 2006 - 2008. The seasonal cycle is clearly evident. High frequency fluctuations are primarily due to cloudiness. Note that the upward solar radiation has ocasional maxima during the winter. These are associated with the increased albedo of the ground following snow storms.

Data like those shown in Fig. 2.12 are available for only a few stations around the world. Most of our ideas about the global pattern of surface radiation are based on various estimates, which have been carefully worked out but are subject to significant errors (e.g., Wielicki et al., 1996). Panel A) of Fig. 2.13 shows the meridional distribution of the zonally averaged solar radiation absorbed by the Earth's surface, as a function of latitude, and for January, July, and the annual mean. Seasonal changes are clearly visible and easy to interpret. Near 50° N in July there is a slight dip or "shoulder" in the meridional profile of the surface absorbed solar radiation. This is associated with cloudiness, and indicates that the clouds are having a major impact on the energy budget of the ocean in those latitudes. Cloudiness also leads to a weak tropical minimum, just north of the Equator. The annual mean curve is fairly symmetrical about the Equator, but shows a minimum near 10° N associated with tropical rain systems. Note also that in the annual mean the southern high latitudes absorb less than the northern high latitudes.

The zonally averaged net surface longwave energy flux is shown in panel B) of Fig. 2.13. In January, the strongest cooling occurs over Antarctica, and in the subtropics of the winter hemisphere. The weakest cooling occurs in cloudy regions, e.g., over the Southern Oceans and in the storm tracks of both hemispheres. Although the surface temperature is warmer in summer than in winter, at some latitudes the net longwave cooling of the surface is actually stronger in winter than in summer! The explanation is that the downward radiation from the atmosphere to the surface increases from winter to summer due to both the warming of the air and the increase in the atmospheric emissivity due to seasonally increased water vapor content and also seasonal changes in cloudiness. This increase in the downward component is so strong that it sometimes overwhelms the increase in the upward component, giving a net decrease in surface infrared cooling from winter to summer. Panel

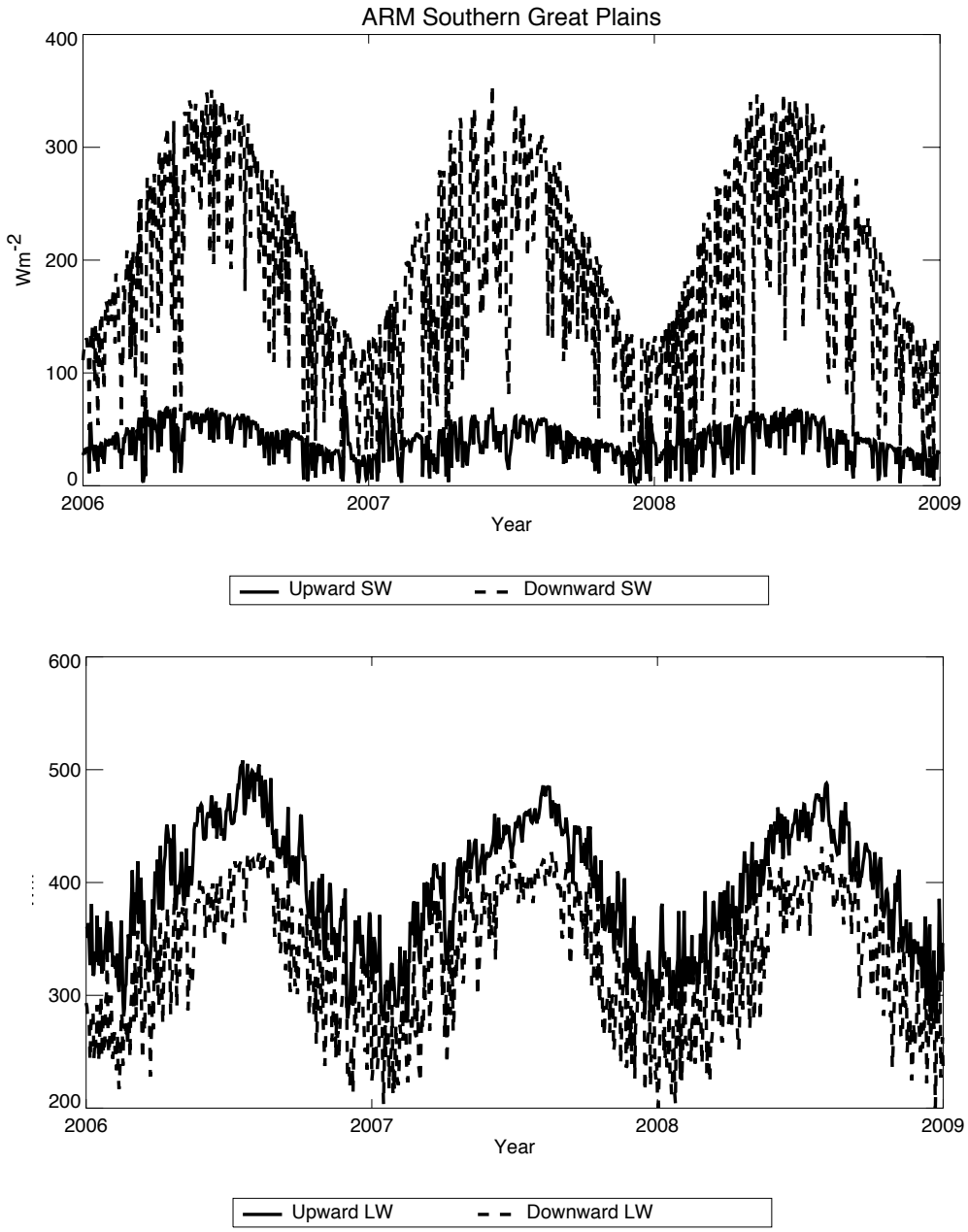


Figure 2.12: Observations of the upward and downward solar (SW) and terrestrial (LW) radiation at the Southern Great Plains observing facility of the Atmospheric Radiation Measurement Program, in Oklahoma, USA. The data are for the years 2006 - 2008.

C) shows the zonally averaged net surface radiation (solar and terrestrial combined). High latitudes experience net radiative cooling of the surface in winter, as would be expected. The annual-mean net radiation into the surface is positive at all latitudes. It follows that the

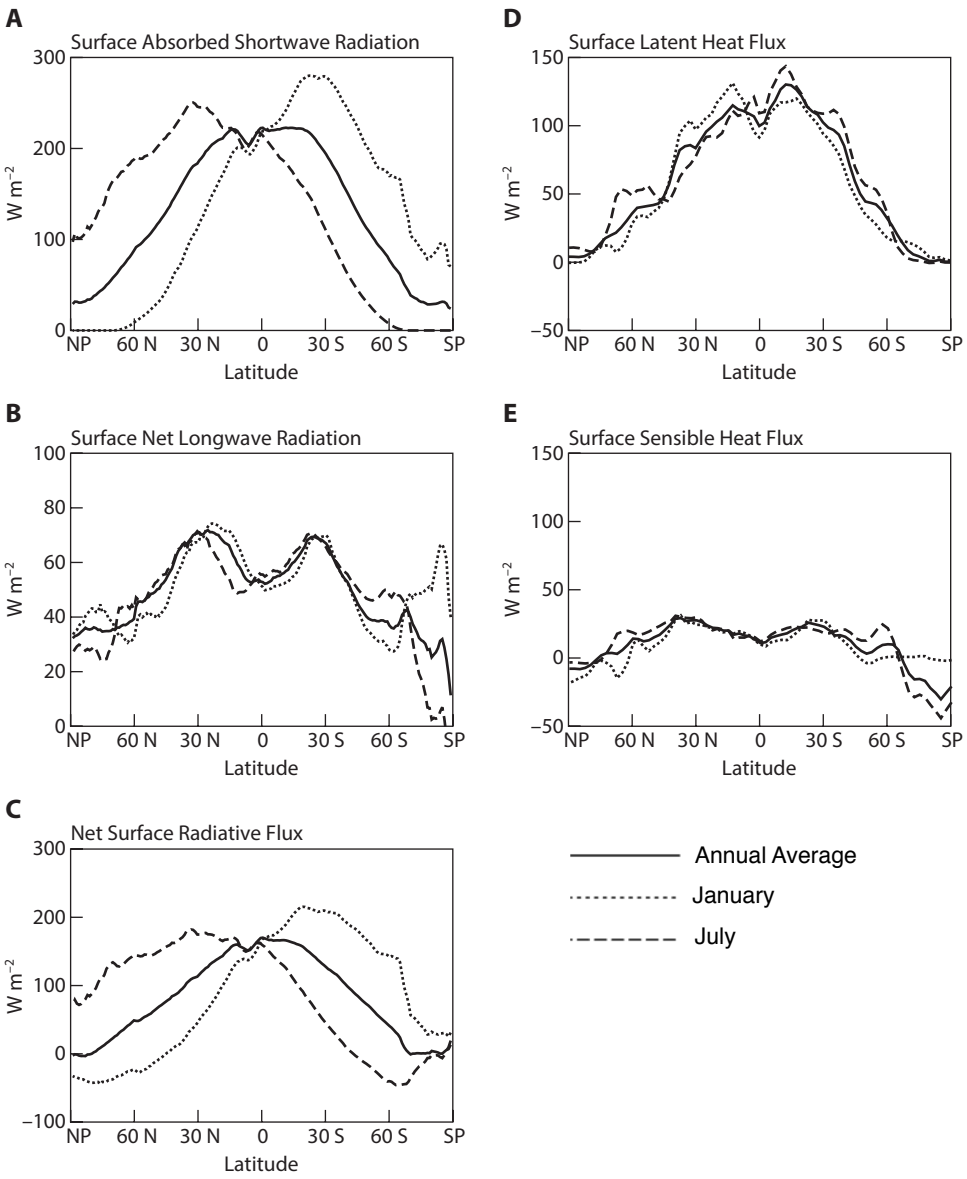


Figure 2.13: A summary of the zonally averaged energy budget of the Earth's surface. A) The zonally averaged (land and ocean) net solar radiation absorbed by the Earth's surface. B) The zonally averaged (land and ocean) net infrared cooling of the Earth's surface. C) The zonally averaged net surface radiation, obtained by combining the data from panels A) and B). D) The zonally averaged surface latent heat flux, positive upward. E) The zonally averaged surface sensible heat flux, positive upward. Panels A) - C) are based on Wielicki et al. (1996). Panels E) - F) are based on the ECMWF reanalysis. Note: These data are not true observations, although they are based on observations.

surface must cool by non-radiative means.

Panel D) shows the zonally averaged net latent heat flux. Positive values represent a moistening of the atmosphere and a cooling of the surface. The latent heat flux compensates, to a large extent, for the net radiative heating of the surface shown in the previous figure. Note that the maxima of the latent heat flux occur in the subtropics. Recall that the precipitation maxima occur in the tropics and middle latitudes. This implies that moisture is transported from the subtropics into the tropics, and from the subtropics into middle latitudes. Panel E) shows the corresponding curves for the surface sensible heat flux. Note that the surface sensible heat flux is generally smaller than the surface latent heat flux. Maxima occur in the winter hemisphere, especially in the Northern winter in association with cold-air outbreaks over warm ocean currents at the east coasts of North America and Asia. Local heat flux maxima associated with such cold outbreaks can be on the order of 1000 W m^{-2}, on individual days.

We close this chapter with a figure that illustrates the globally and annually averaged vertical flows of energy by various processes. We use CERES data for the net vertical flux of energy due to radiation, as a function of height. The net radiation curve is constrained by measurements of the radiation at the top of the atmosphere, but the vertical profile is computed using a radiative transfer model with the observed state of the atmosphere as input. The net radiation is zero at the top of the atmosphere, and downward at all other levels. This downward radiative energy flux must be balanced by upward fluxes due to other processes. The energy flux due to large-scale circulations was computed from the ECMWF Interim reanalysis, and is snown by the red curve. It reaches a maximum in the lower middle troposphere, around 700 hPa, and decreases above. This means that the large-scale ciruclation removes energy from the lower troposphere and deposits it in the upper troposphere. The energy flux due to convection and turbulence, plotted as a green curve, was computed as a residual, using the assumption that the total globally and annually averaged flux of energy must be zero at all heights. It is largest at the surface, and decreaes upward at all levels. All three curves hug zero at and above the 100 hPa level.

The surface value of the net radiation shown in the figure differs from the value shown in Table 2.2, and surface flux due to convection and turbulence in the figure also differs from the sum of the sensible and latent heat fluxes shown in Table 2.2. These differences reflect various uncertainties in the calculations.

A key conclusion from the figure is that *the globally and annually averaged upward flux of energy due to convection and turbulence exceeds that due to large-scale motions throughout the lower and middle troposphere*. The two fluxes are approximately equal in the upper troposphere. Remember, however, that the convective activity is determined in part by the large-scale circulation, and vice versa. These two contributions to the upward energy flux are jointly produced by what amounts to a single, strongly coupled system of processes, operating on a wide range of spatial and temporal scales. That is a major theme of this book.

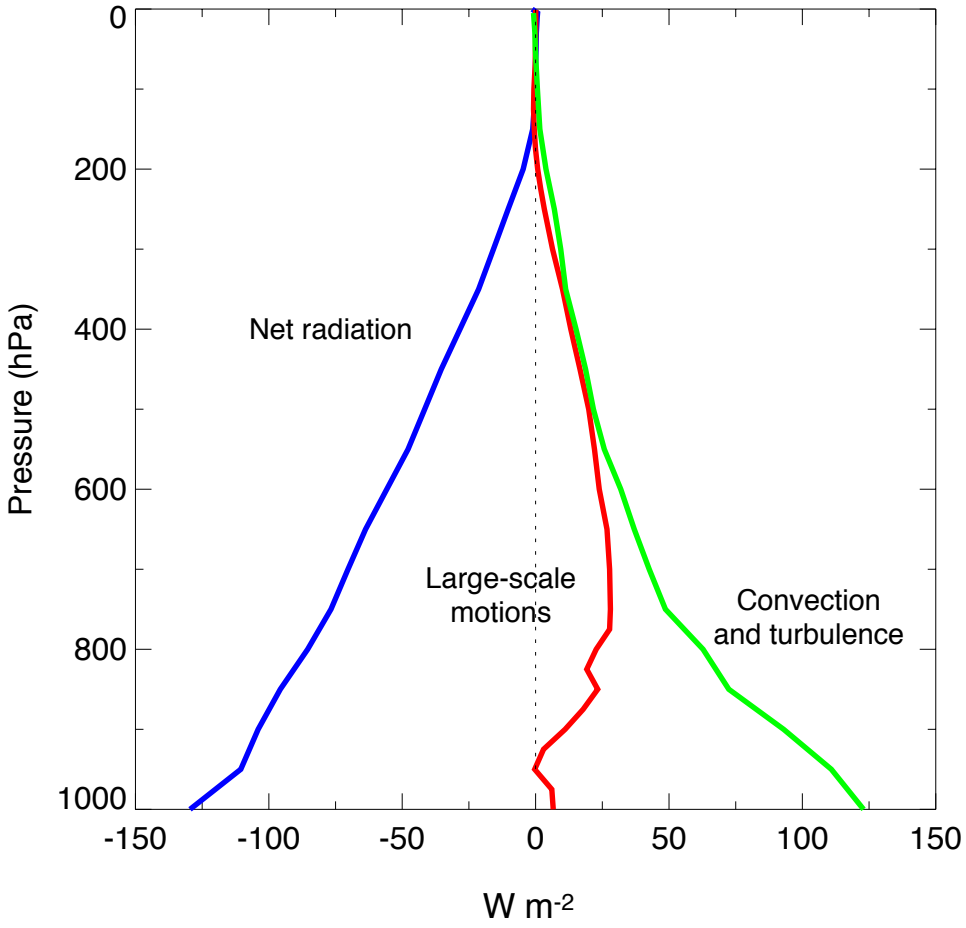


Figure 2.14: Estimates of the globally and annually averaged vertical flows of energy due to radiation, large-scale motions, and convection and turbulence. The radiation fluxes are based on CERES data. The fluxes due to large-scale motions are computed from the ECMWF reanalysis. The fluxes due to convection and turbulence are computed as a residual.

2.5 Segue

This chapter provides a brief overview of the energy fluxes at the top of the atmosphere, at the Earth's surface, and across the atmosphere. The meridional structure of the net radiation at the top of the atmosphere implies transports by the ocean-atmosphere system. The meridional structure of the net surface energy flux implies energy transports by the ocean. The meridional structure of the net flow of energy into the atmosphere, across its upper and lower boundaries, implies a net transport of energy by the atmosphere. In later chapters we will discuss the nature of these energy transports in more detail, and as well as the meridional transports of angular momentum and moisture.

Among the most important points in this chapter is that the net radiative heating of the

Earth's surface is balanced mainly by evaporative cooling, and the net radiative cooling of the atmosphere is balanced mainly by latent heat release. We also mentioned the important effects of water vapor and clouds on the Earth's radiation budget. Finally, we discussed the lower boundary conditions on the global atmospheric circulation associated with the distributions of continents and oceans, sea surface temperature, topography, vegetation, and ice and snow.

With this preparation, we are now ready to take a look at some of the observed features of the global circulation.

2.6 Problems

1. (a) Prove that, for any vector $\mathbf{Q}$,

$$\int_S \nabla \cdot \mathbf{Q}\, dS = 0, \tag{2.14}$$

where the integral is taken over a closed surface, e.g., the surface of a sphere. We assume that $\mathbf{Q}$ is everywhere tangent to the surface, i.e., it "lies in" the surface and so can be described as a "horizontal" vector. Eq. (2.14) shows that the globally averaged divergence of any horizontal vector is zero.

(b) Also prove that

$$\int_S \mathbf{k} \cdot (\nabla \times \mathbf{Q})\, dS = 0, \tag{2.15}$$

where the integral is taken over a closed surface. Here $\mathbf{k}$ is a unit vector everywhere perpendicular to the surface. Eq. (2.15) implies that the global mean of the vertical component of the vorticity is zero.

2. (a) Suppose that 1 W m^{-2} is supplied to a column of water 100 m deep. Assume that the temperature changes uniformly with depth throughout the entire column. How much time is needed to increase the temperature of the water by 1 K?

(b) Estimate the heat capacity of the entire global ocean in J K^{-1}. If all of the solar radiation incident at the top of the atmosphere were used to warm the ocean uniformly, how long would it take for the temperature of the entire ocean to increase by 1 K?

Chapter 3

A quick overview of the circulation

3.1 Introduction

This chapter is intended to provide a brief summary of some of the important phenomena of the observed seasonally varying global circulation of the atmosphere, with a few additional comments about non-seasonal variability. Selected fields will be shown and described, with an emphasis on mass, the winds, temperature, and moisture. Among the most important phenomena introduced are the tropical Hadley and Walker circulations, the monsoons, planetary waves, and some aspects of the hydrologic cycle. All of these things will be discussed in more detail in later chapters. Lots of questions will be raised, but for the most part the answers will be deferred until later.

Many of the plots shown in this and later chapters are based on analyses (actually, reanalyses) created at the European Centre for Medium Range Weather Forecasts (ECMWF; Uppala et al., 2005). The existence of such analyses, and their ready availability online, makes a book like this one enormously easier to create now than in decades past. The nature of a reanalysis is briefly explained in Chapter 6.

3.2 The global distribution of atmospheric mass

Mass is arguably the most fundamental quantity in any physical description of the atmosphere. The density, ρ, is defined as the mass of air per unit volume. With high accuracy, the surface pressure is equal to the weight of the air above, per unit area, which is given by

$$p_S = \int_{z_S}^{\infty} \rho g dz. \tag{3.1}$$

.

Here p is pressure, the subscript S denotes a surface value, g is the acceleration of gravity,

and z is height. The surface pressure varies from place to place, in part due to the effects of topography and in part due to the circulation of the atmosphere. An example is shown in Fig. 3.1. The figure shows surface pressure scatter-plotted against surface elevation, for 00Z (i.e., midnight in Greenwich, England) on January 1 and July 1, 2000. The main point of the figure is that a large fraction of the spatial variability of surface pressure is due to topography; you already knew this, but perhaps you have never seen the data plotted this way before. For each hemisphere, the data fall roughly onto two curves. For January, the Northern Hemisphere points on the upper curve come mainly from Greenland, and the Southern Hemisphere points on the upper curve come mainly from Antarctica. The scatter of the data about their mean for each surface elevation indicates the dynamical variability of the surface pressure. The range of variability is particularly large when the surface height is zero, which means that it is large for the oceans. An interpretation is that weather systems are particularly vigorous over the oceans.

Maps of surface pressure are dominated by minima associated with mountain ranges. Because of this, maps of surface pressure maps do not clearly show how the horizontal pressure-gradient force varies with changes in the weather. To get around this problem, it is conventional to define "sea-level pressure," which is computed by adding to the surface pressure a correction that is designed to represent the additional pressure that would occur if the mountains were not present. Fig. 3.2 shows monthly mean maps of sea-level pressure for January and July. These maps are much smoother than weather maps plotted for particular observation times, because the moving highs and lows associated with individual weather systems have been smoothed out by the time-averaging. The plots on the right side of Fig. 3.2 show the corresponding zonally averaged distributions of the sea-level pressure for January and July.

Throughout this book, we will distinguish between the zonally averaged circulation, one aspect of which is the zonally averaged sea-level pressure shown in the right-side panels of Fig. 3.2, and departures from the zonal average, which we will call *"eddies"*. The highs and lows of the sea-level pressure that can be seen in Fig. 3.2 are associated with eddies.

Especially in the Northern Hemisphere, there is a very pronounced tendency for low pressure over the oceans and high pressure over the continents in winter, and vice versa in summer. This seasonal shift of mass between the oceans and continents is associated with a seasonal shift in surface temperature. Fig. 3.3 shows the departure of the surface temperature from its zonal mean at each latitude, for January and July. We use the term "eddy" to refer to departures from the zonal mean, so we can say that these are plots of the "eddy surface temperature." The figures show that in middle latitudes, especially in the Northern Hemisphere, the surface air temperature is generally colder over the continents than over the oceans in winter, and vice versa in summer. Particularly cold temperatures occur on the eastern sides of the Northern Hemisphere continents in January. This is partly because the flow is generally from west to east, so that the air on the east sides of the continents has traveled all the way across the cold continent in order to reach the eastern side, gradually cooling along the way. The strong temperature contrast on the east coasts

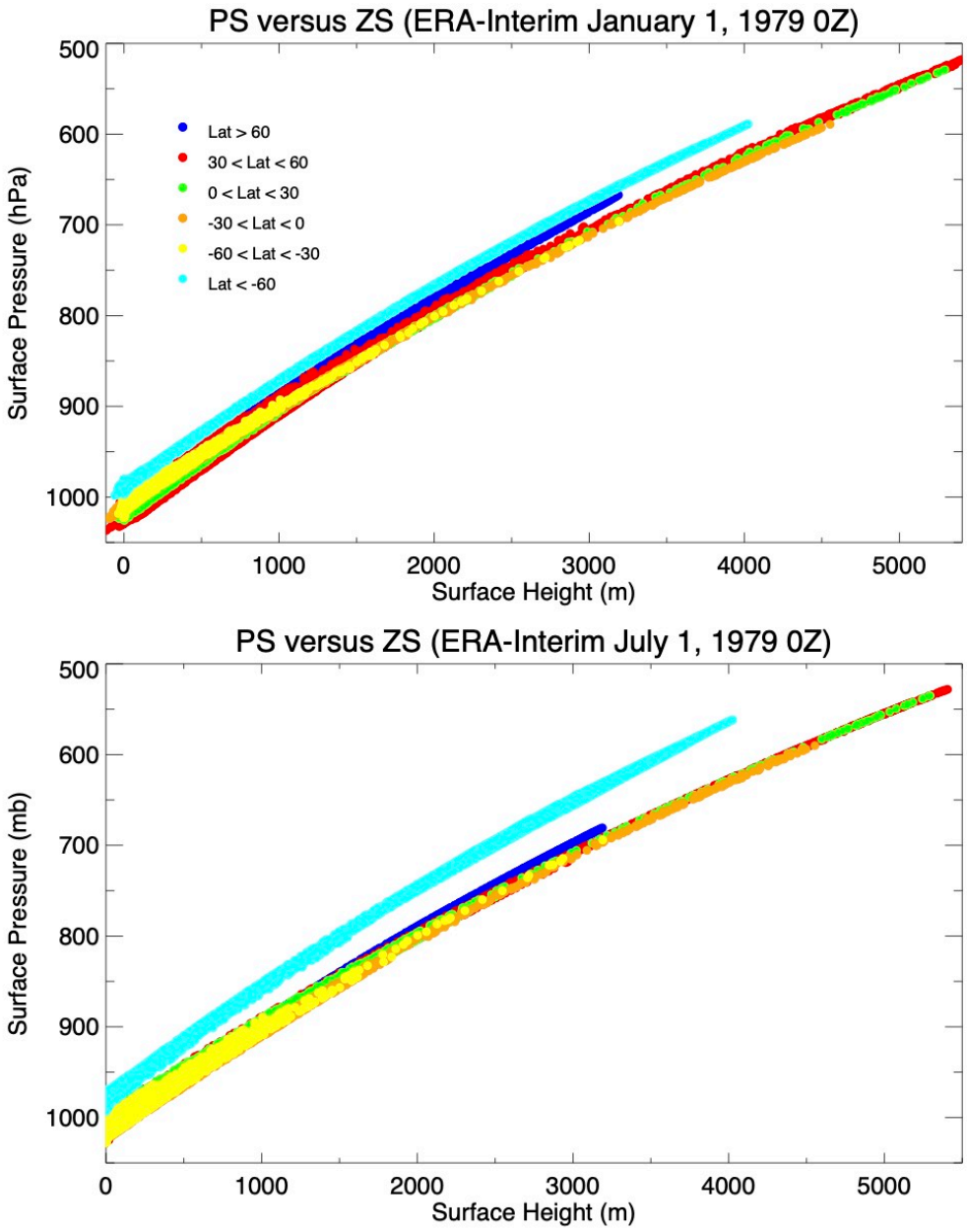


Figure 3.1: Surface pressure plotted against surface height, for particular observation times. Note that the vertical scale decreases upward. Color is used to indicate roughly which latitudes the data points come from. The dots along the dark blue curve come from Greenland, and the dots along the light blue curve come from Antarctica. The data were taken from a latitude-longitude grid, so the individual points do not represent equal areas.

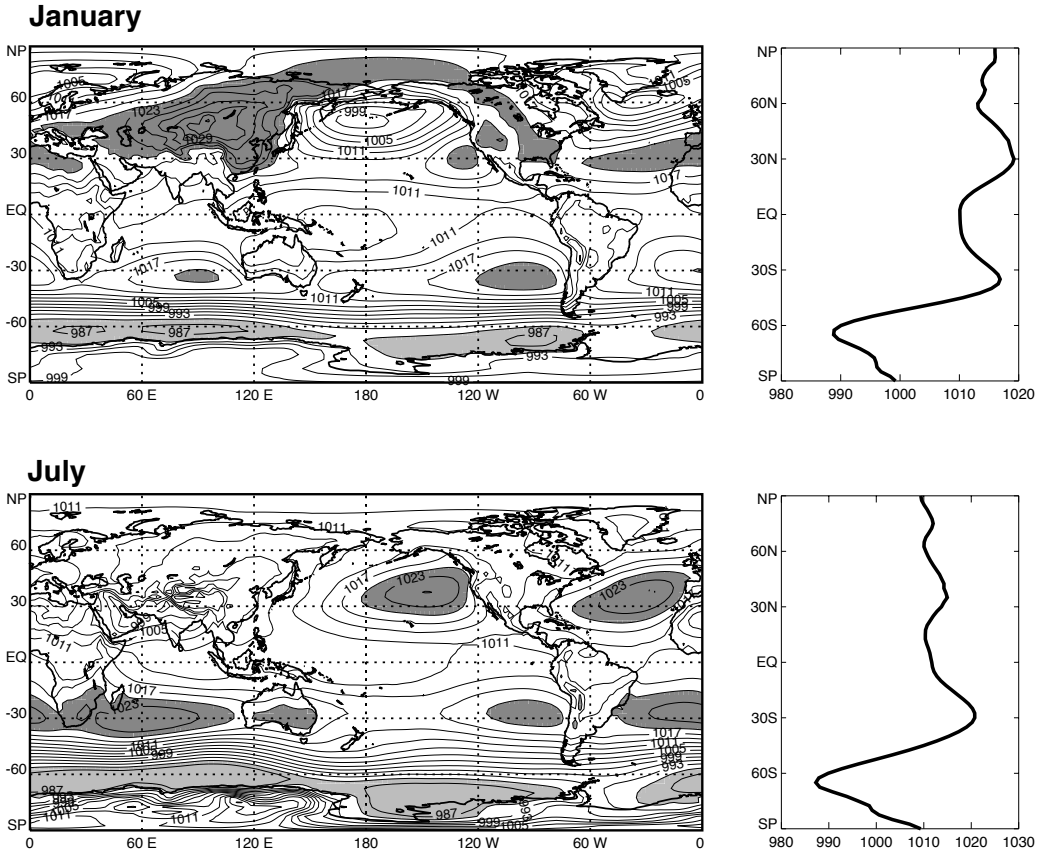


Figure 3.2: Sea-level pressure maps. The contour interval is 3 hPa. Values higher than 1020 hPa have dark shading, and those lower than 990 hPa have light shading. Averages with respect to longitude are shown on the right.

of the Northern Hemisphere continents in winter leads to the frequent formation of winter storms, which are often called "baroclinic eddies." Further discussion is given later.

Comparison of Figs. 3.2 and 3.3 shows that there is a tendency for high sea-level pressures to be associated with cold surface temperatures, and vice versa. To understand this, recall that cold air has a higher density than warm air (for a given pressure), so that a "pile" of cold air of a given geometrical thickness will contain more mass, and therefore weigh more, than a pile of warm air of the same geometrical thickness. This can be seen mathematically by combining the ideal gas law with (3.1) to obtain

$$p_S = \int_{z_S}^{\infty} \left(\frac{pg}{RT}\right) dz. \tag{3.2}$$

.

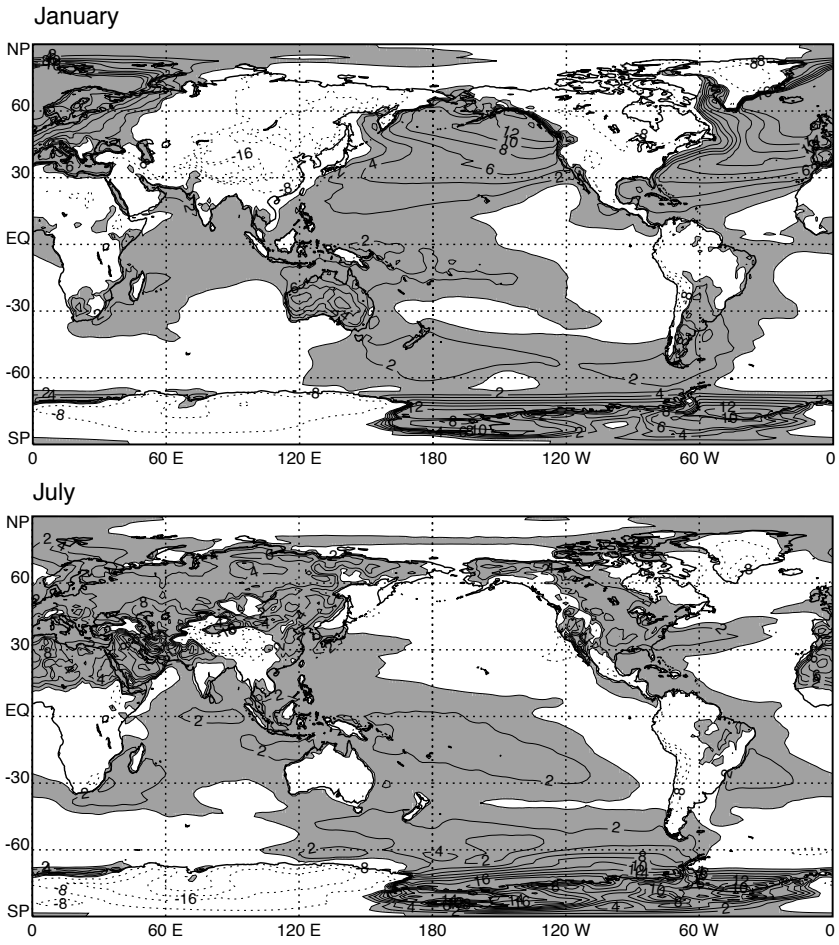


Figure 3.3: The departure of the 2 m temperature from its zonal mean at each latitude, for January and July. Positive values are shaded. The contour interval is 2 K.

Here R is the gas constant for dry air, and T is temperature.

Fig. 3.2 shows that for both January and July there is a tendency for high sea-level pressure to occur in the subtropics, near $\pm 30^\circ$. These subtropical highs typically appear as "cells," e.g., in the North Atlantic and North Pacific in July, or off the west coast of South America in January. In many cases, the subtropical highs are found over the eastern parts

of the oceans. In the Northern Hemisphere, they are particularly strong in the northern summer (Hoskins, 1996). Strong highs are also apparent in middle latitudes during winter, e.g., in Siberia and western North America. Both regions are quite mountainous.

In spherical coordinates, and using pressure as the vertical coordinate, geostrophic balance is expressed by

$$0 = f v_g - \frac{1}{a \cos\varphi} \left(\frac{\partial \phi}{\partial \lambda} \right)_p , \tag{3.3}$$

and

$$0 = -f u_g - \frac{1}{a} \left(\frac{\partial \phi}{\partial \varphi} \right)_p , \tag{3.4}$$

where u_g and v_g are the zonal and meridional components of the geostrophic wind, respectively; f is the Coriolis parameter, and ϕ is the geopotential. With appropriate caution, we can use (3.3) and (3.4) to relate the pressure gradients to the winds. For example, throughout the year, there is a zonally oriented belt of low pressure across the tropics. Geostrophy tells us that the decrease of sea-level pressure from the subtropics towards the Equator implies tropical easterlies near the surface, although we have to be careful about invoking geostrophy near the Equator, where the Coriolis parameter passes through zero. Because the sea-level pressure generally decreases from the subtropics to middle latitudes, the geostrophic relation leads us to expect surface westerlies on the poleward side of the subtropical highs. The relatively high pressure over the poles suggests surface easterlies near the south pole, although the "sea-level" pressures plotted for Antarctica must be taken with a grain of salt because the surface of Antarctica is far above sea level.

In the Northern Hemisphere during northern winter, prominent low-pressure cells appear, most notably near the Aleutian Islands and Iceland. These are regions where storm systems are often found on individual days. There is a tendency for a minimum of the sea-level pressure near 60°N, especially in January but also to some extent in July. A very pronounced belt of low pressure is found over the "southern ocean" north of Antarctica, throughout the year, although it is more intense in July (winter) than January (summer).

Generally there is less seasonal change in the Southern Hemisphere than in the Northern Hemisphere. Also, the departures from the zonal means are much stronger in the Northern Hemisphere than the Southern Hemisphere.

The tropical sea level pressure distribution is generally very smooth and featureless, compared to that of middle latitudes. A simple explanation for this was given by Jule Charney (1963), in terms of the differences in dynamical balance between the tropics. Charney



Figure 3.4: Jule G. Charney, whose many achievements include scale analyses of both extratropical and tropical motions, development of the quasi-geostrophic theory, development (in his Ph.D. thesis at the University of California at Los Angeles) of a classical theory of baroclinic instability, pioneering work on numerical weather prediction, analysis of the interactions of cumulus convection with large-scale motions in tropical cyclones, development of a theory of planetary waves propagating through shear, analysis of blocking, and a theory of desertification. Figure used with permission of the MIT Museum.

(Fig. 3.4) was one of the giants of twentieth-century meteorology. His work is discussed for the first time in this chapter, but his name will come up again and again throughout the remainder of this book. Charney's explanation of the flatness of tropical sea-level pressures is based on the fact that in middle latitudes the effects of the Earth's rotation are much more important than particle accelerations, while the opposite is true in the tropics. He began his scale analysis (see Appendix B) with the equation of horizontal motion, which can be written in simplified form as

$$\frac{D\mathbf{v}}{Dt} + f\mathbf{k} \times \mathbf{v} = -\nabla_p \phi. \tag{3.5}$$

Here $\mathbf{v}$ is the horizontal wind vector, and $\mathbf{k}$ is a unit vector pointing upward. The three terms shown in (3.5) represent most of the "action" throughout most of the atmosphere. For middle latitudes, their orders of magnitude can be estimated as follows:

$$\left|\frac{D\mathbf{v}}{Dt}\right| \sim \frac{V^2}{L}, \tag{3.6}$$

$$|f\mathbf{k} \times \mathbf{v}| \sim f_{midlat} V, \tag{3.7}$$

$$\left|\nabla_p \phi\right| \sim \frac{|\delta\phi|}{L}. \tag{3.8}$$

Here V is a "velocity scale," which might be on the order of 10 m s^{-1}, L is a length scale, which might be on the order of 10^6 m, and $|\delta\phi|$ is the magnitude of a typical fluctuation of the geopotential height. Note that $|\delta V| \sim V$, but $|\delta\phi|$ is generally much less than ϕ. The numerical values of these scales have been chosen to be representative of "large-scale" motions on the Earth; if we wanted to analyze small-scale motions, we would choose different numerical values. The same numerical values of V and L can be used for both the tropics and middle latitudes because the term "large-scale" is used in the same way for both regions.

In middle latitudes, the acceleration following a particle, $D\mathbf{v}/Dt$, is usually negligible in (3.5), compared to the rotation term. A typical value of $D\mathbf{v}/Dt$ can be estimated as $V^2/L \sim 10^2/10^6 = 10^{-4}$ m s^{-2}. A representative midlatitude value of the Coriolis parameter is 10^{-4} s^{-1}, so that $f_{midlat} V \sim 10^{-3}$ m s^{-2}, about one order of magnitude larger than $D\mathbf{v}/Dt$. Geostrophic balance is, therefore, approximately satisfied in mid-latitudes, i.e.,

$$f_{midlat} V \sim \frac{(\delta\phi)_{midlat}}{L}, \tag{3.9}$$

or

$$(\delta\phi)_{midlat} \sim f_{midlat} V L. \tag{3.10}$$

According to (3.10), rotation can balance pressure gradients in midlatitudes.

The Coriolis parameter vanishes on the Equator, so it is reasonable to expect that sufficiently close to the Equator geostrophic balance breaks down (a point discussed further in Chapter 10), and particle accelerations tend to balance the pressure gradient force, much as they do on small scales almost everywhere in the atmosphere, and in many engineering contexts (e.g., the flow of water in a pipe):

$$\frac{V^2}{L} \sim \frac{(\delta\phi)_{tropics}}{L}, \tag{3.11}$$

or

$$(\delta\phi)_{tropics} \sim V^2. \tag{3.12}$$

Comparing (3.11) and (3.12), we see that

$$\frac{(\delta\phi)_{tropics}}{(\delta\phi)_{midlat}} \sim \frac{V}{f_{midlat} L} \equiv \mathrm{Ro}_{midlat}, \tag{3.13}$$

where Ro_{midlat} is a midlatitude Rossby number. By substituting the numerical values given above, we find that $\mathrm{Ro}_{midlat} \cong 0.1$. Eq. (3.13) therefore tells us that geopotential height fluctuations on pressure surfaces are much smaller in the tropics than in middle latitudes. It follows that pressure fluctuations on height surfaces are an order of magnitude smaller in the tropics than in middle latitudes. The real message is that the horizontal pressure-gradient force is much smaller in the tropics than in middle latitudes. This conclusion holds at any level, so it holds at all levels.

At this point, we can bring in the hydrostatic equation to show that large-scale fluctuations of temperature and surface pressure are also much smaller in the tropics than in middle latitudes. Suppose that the pressure-gradient force is small at some particular height. If the temperature changes rapidly in the horizontal, this will imply large horizontal pressure gradients at other heights. The implication is that if the horizontal pressure gradient is small *at*

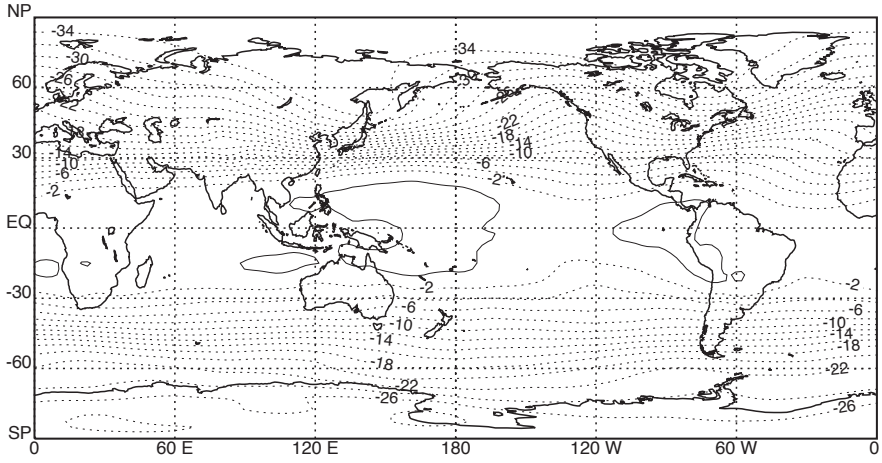


Figure 3.5: The departure of the 500 hPa temperature from its value at Darwin, Australia, for January. The contour interval is 2 K, and the zero contour is solid while negative contours are dashed. Note the amazing uniformity of the tropical temperatures. This is a consequence of the smallness of the Coriolis parameter in the tropics, as explained by Charney (1963).

all levels, then the horizontal temperature gradients must also be small. Fig. 3.5 illustrates this.

The smallness of horizontal temperature gradients in the tropics is now widely invoked to justify what is called the "weak temperature gradient approximation" (e.g., Sobel et al., 2001). It is important to remember, however, that, as explained above, Charney's primary conclusion is that horizontal *pressure* gradients are weak in the tropics; the weakness of the tropical horizontal temperature gradients is a secondary conclusion (Romps, 2012).

3.3 Zonal wind

Fig. 3.6 shows the latitude-height distribution of the zonally averaged zonal wind for January and July, respectively. The plot extends from the surface to the middle stratosphere. We plot the wind components and other variables against height, rather than pressure, because the pressure coordinate tends to "squash" the stratosphere into a thin region at the top of the plot, obscuring its structure. Although the stratosphere contains only a small fraction of the mass of the whole atmosphere, it is home to some spectacular phenomena, and its dynamical influence extends downward into the troposphere. The stratosphere should be of interest even to people who are mainly interested on the global circulation of the troposphere.

In both January and July, easterlies extend through the entire depth of the tropical troposphere. They are somewhat stronger in July. They are concentrated in the Northern Hemisphere in July and the Southern Hemisphere in January. The near-surface easterlies are stronger in the winter hemisphere.

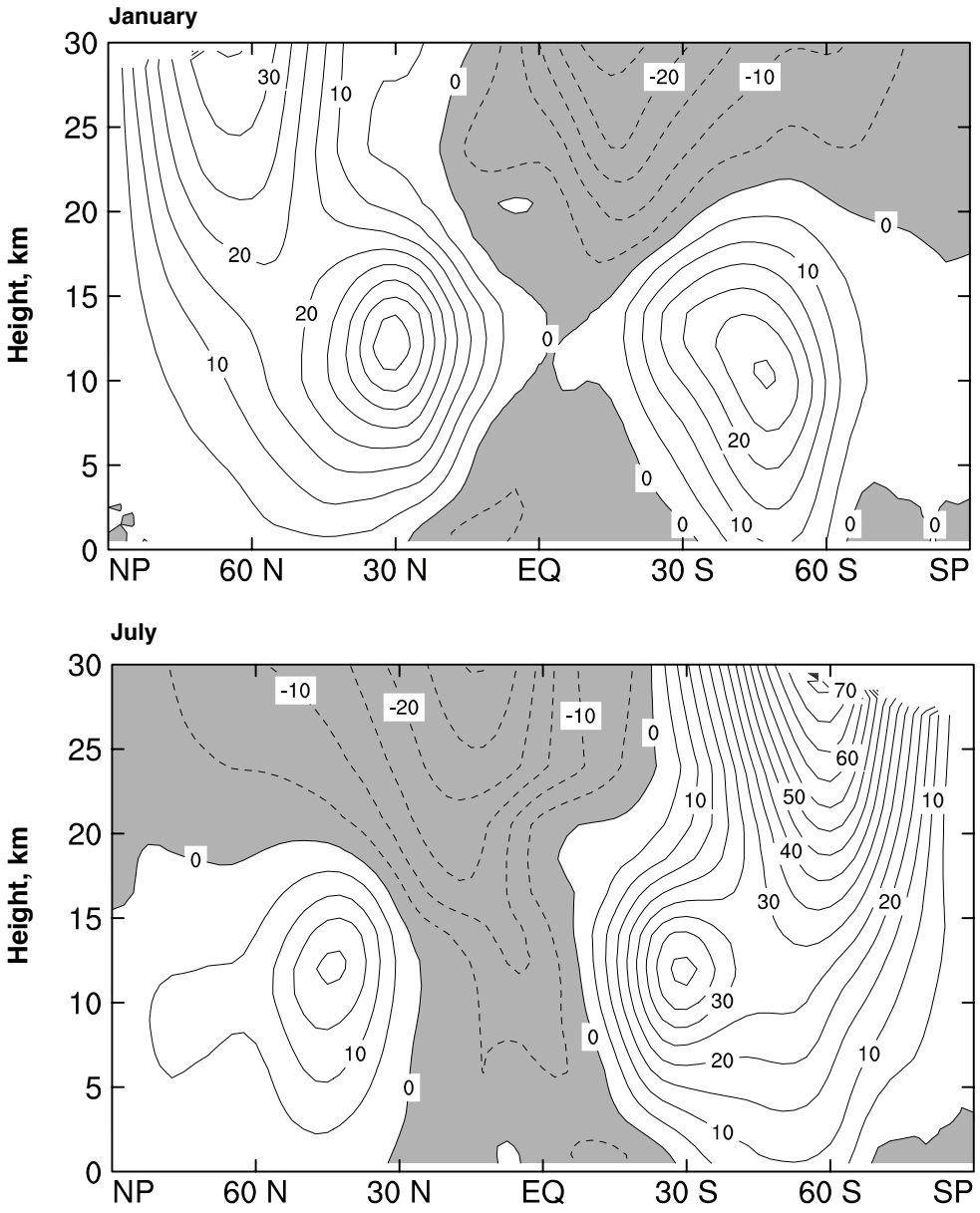


Figure 3.6: Latitude-height section of the zonal wind. The contour interval is 5 m s^{-1}. Easterlies are shaded.

Westerly jet streams are quite prominent in the upper troposphere, especially in the winter hemisphere. These are called the subtropical jets, and they reach their maximum strength during winter at about 30° latitude. Directly below the winter subtropical jet, the surface zonal wind is very weak. The jet extends poleward, however, with a hint of a second maximum in the troposphere near 50 or 60° from the Equator, where the zonally averaged

surface wind is definitely westerly. The subtropical jets are weaker in summer, and are shifted to about 45° latitude. The subtropical jet maxima are consistently found near the 200 hPa level, which is the cruising altitude of commercial airliners.

Suppose that the zonal surface wind is nearly geostrophic. Then, in the absence of mountains (e.g., over the oceans) the surface pressure must have a meridional maximum at a latitude where u_S passes through zero, i.e., where surface easterlies meet surface westerlies. Comparison of Fig. 3.2 with Fig. 3.6 shows that in fact the subtropical surface-pressure maxima occur at about the same the latitudes where the zonal component of the surface wind passes through zero. The subtropical highs sit under the subtropical jets.

Strong westerly jets also occur in the winter stratosphere. In the Southern Hemisphere in July, there is a clear minimum in the westerlies near 150 hPa, at about 40°S. Above and poleward of this minimum is a very powerful stratospheric westerly jet, called the polar night jet. As shown below, these strong stratospheric westerlies are associated with very cold stratospheric temperatures over the poles. A similar but weaker polar night jet occurs in the Northern Hemisphere winter. The stratospheric jets are separated from the troposphere jets by a (weak) minimum of the westerlies.

Easterlies fill the summer-hemisphere stratosphere. As discussed later, the summer stratosphere is radiatively controlled, while the winter stratosphere is strongly influenced by dynamics, including upward wave propagation from the troposphere.

The zonally averaged surface winds are quite weak near the poles.

Fig. 3.7 shows maps of the 850 hPa zonal wind for January and July, respectively. Again, keep in mind that many intense small-scale ($\sim$1000 km) features would appear in daily maps, but have been smoothed out here by time averaging. The monthly mean maps show very obvious alternating bands of easterlies and westerlies, which are qualitatively reminiscent of Jupiter (see Fig. 3.8), although Jupiter has more bands; generally speaking, the Earth's atmosphere features easterlies in the tropics, westerlies in middle latitudes, and easterlies again near the poles. We can also see features associated with the strong cells in the sea level pressure maps, e.g., the easterlies in the extreme North Pacific in January, associated with the Aleutian Low. Again, variations with longitude are much more apparent in the Northern Hemisphere than the Southern Hemisphere. Generally speaking, however, the features seen in the maps have a very zonal orientation, with strong north-south gradients and relatively weak east-west gradients. Note the intensification of both the zonal-mean flow and the eddies of the midlatitude westerlies in winter, in each hemisphere. In winter, the Northern Hemisphere westerlies are particularly strong over the oceans.

The strong positive maximum in the Arabian Sea in July is associated with the Indian monsoon; this is discussed in Chapter 10. The westerlies north of Australia (but south of the Equator) in January are indicative of the Winter Monsoon. In both regions, the sign of the zonal wind reverses seasonally.

Fig. 3.9 shows the corresponding maps for 200 hPa. The winds are generally stronger

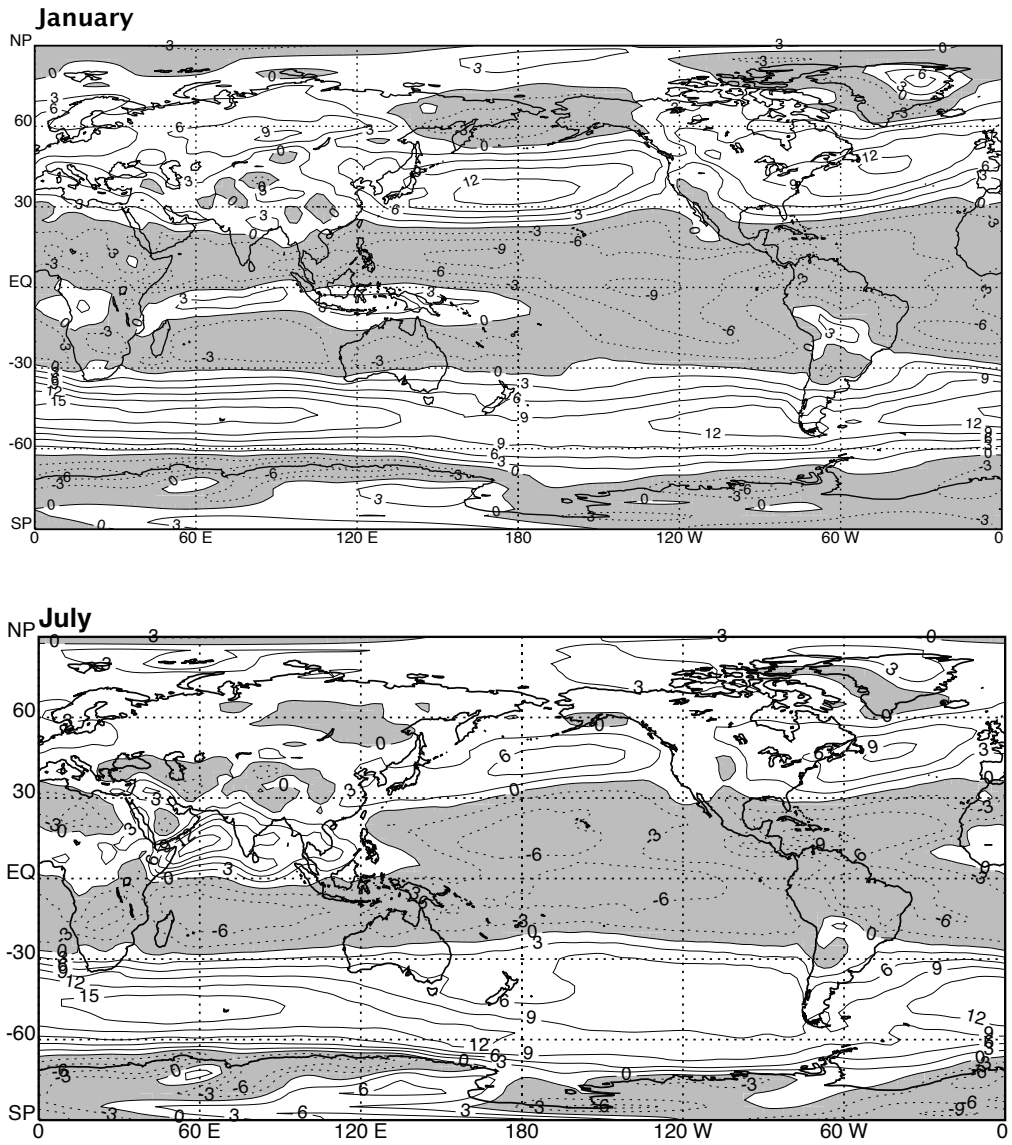


Figure 3.7: Maps of the 850 hPa zonal wind, for January and July. The contour interval used is 3 m s^{-1}. Easterlies are shaded.

aloft than near the surface; this is true of both the zonal-mean flow and the eddies. Note the very prominent January westerly jet maxima off the east coasts of North America and, especially, Asia. There is also a westerly jet maximum at about 30° S, near the Date Line. In the Northern winter, there are "dipoles" consisting of easterly-westerly pairs, straddling the equator, near the Americas and also near the longitude of Australia. At some longitudes, the westerlies extend to the Equator in January. In July, there are equatorial easterlies at all longitudes, and the westerlies have intensified in the midlatitudes of the

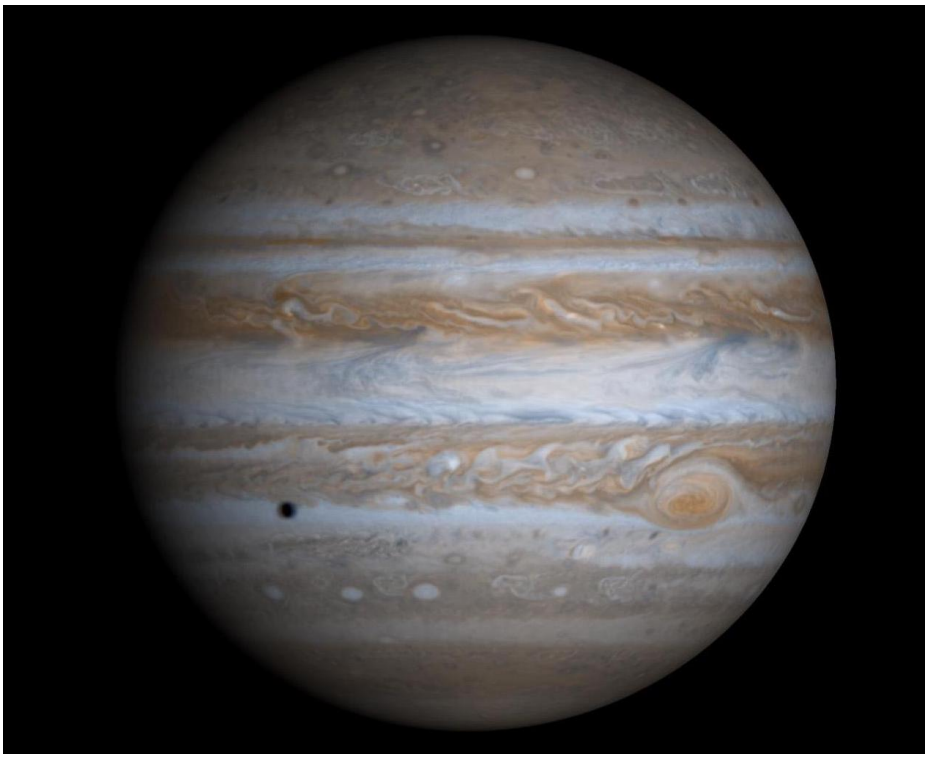


Figure 3.8: An image of Jupiter obtained from the Cassini probe on December 7, 2000. Note the zonal bands, and the Great Red Spot in the southern hemisphere. The black dot is the shadow of Jupiter's moon Europa. From `https://www.jpl.nasa.gov/spaceimages/details.php?id=PIA02873`.

Southern Hemisphere, and weakened in the midlatitudes of the Northern Hemisphere. Two intrusions of westerlies are seen in the Northern Hemisphere tropics: one just east of the Date Line, and another over the Atlantic Ocean. As discussed later, these regions of mean westerlies allow waves to propagate from middle latitudes into the tropics.

3.4 Meridional wind

Fig. 3.10 shows the latitude-height distribution of the zonally averaged meridional wind for January and July, respectively. The zonal means reach about 2 m s^{-1} in absolute value; the strongest values occur in the tropics. In the winter-hemisphere tropics, in both months, there is a conspicuous dipole structure, with poleward flow aloft and equatorward flow near the surface. Evidently there is convergence near the equator at low levels, and divergence aloft. The convergence zone near the surface shifts from the Southern Hemisphere in January to the Northern Hemisphere in July. These features are associated with the Hadley circulation, which is discussed later in this chapter and again in Chapter 6. Poleward flow is also found near the surface in middle latitudes, with weak equatorward flow above. We can also see weak regions of low-level convergence near 60°S and 60°N.

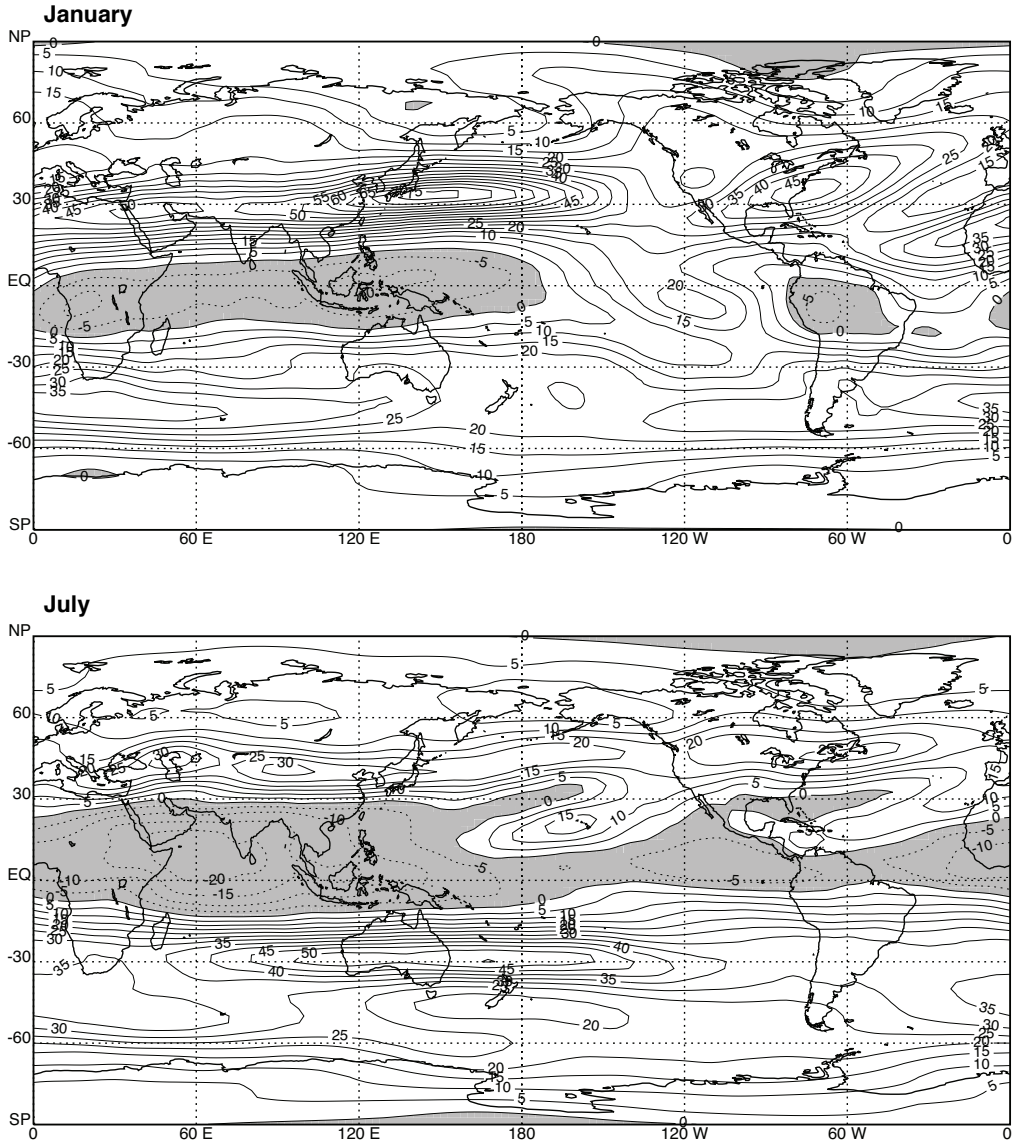


Figure 3.9: Maps of the 200 hPa zonal wind for January and July. The contour interval is 5 m s^{-1}. Easterlies are shaded.

As with most other fields, the seasonal changes of the meridional wind in the mid-latitudes of the Southern Hemisphere are quite weak, relative to those in the Northern Hemisphere.

The mass-weighted vertical mean of the time- and zonally averaged meridional wind must be very close to zero at all latitudes. This can be understood from the surface pressure-tendency equation,

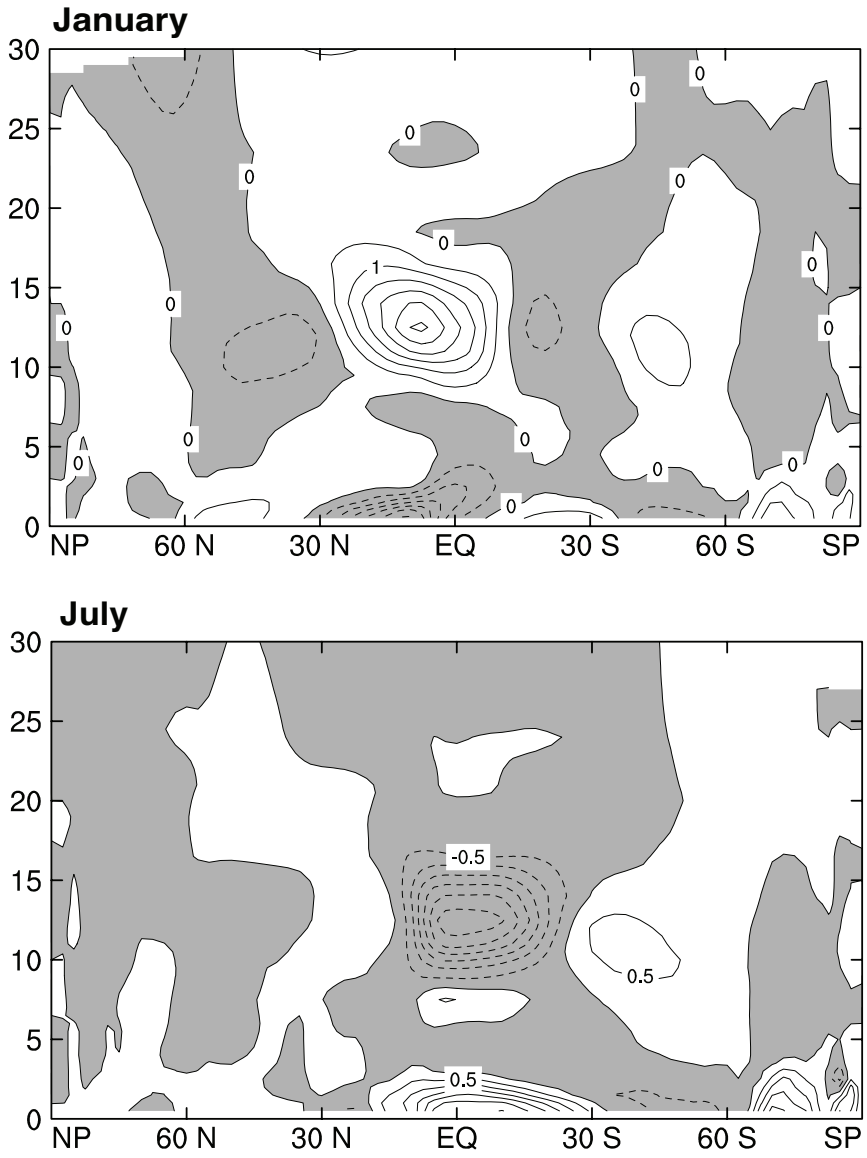


Figure 3.10: Latitude-height plots of the zonally averaged meridional wind, for January and July. The contour interval is 0.5 m s^{-1}.

$$\frac{\partial p_S}{\partial t} + \nabla \cdot \left(\int_0^{p_S} \mathbf{v} dp \right) = 0, \tag{3.14}$$

which is derived in Chapter 4. In (3.14), $\mathbf{v}$ is the horizontal wind vector. The physical meaning of (3.14) is that the surface pressure changes in time due to the convergence or

divergence of mass in the column above. Averaging (3.14) around a latitude circle gives

$$\frac{\partial \overline{p_S}^{\lambda}}{\partial t} = \frac{-1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\int_0^{p_S} v\,dp}^{\lambda}\cos\varphi\right). \tag{3.15}$$

In an average over a sufficiently long time, $\dfrac{\partial \overline{p_S}^{\lambda}}{\partial t}$ must become negligible at each latitude, because p_S is bounded within a fairly narrow range, so that (3.15) reduces to

$$\frac{\partial}{\partial\varphi}\left(\overline{\int_0^{p_S} v\,dp}^{\lambda,t}\cos\varphi\right) = 0. \tag{3.16}$$

Eq. (3.16) means that $\overline{\int_0^{p_S} v\,dp}^{\lambda,t}\cos\varphi$ is independent of latitude. Since $\cos\varphi = 0$ at both poles, we conclude that

$$\boxed{\overline{\int_0^{p_S} v\,dp}^{\lambda,t} = 0 \text{ at all latitudes.}} \tag{3.17}$$

The physical interpretation of (3.17) is very simple. Suppose, for example, that $\overline{\int_0^{p_S} v\,dp}^{\lambda,t}$ was positive at the Equator, so that air was systematically flowing out of the Southern Hemisphere and into the Northern Hemisphere. That can really happen at a given instant, but if it continued over time the surface pressure in the Southern Hemisphere would eventually decrease to zero and the surface pressure in the Northern Hemisphere would increase to roughly double its normally observed average value. The pressure-gradient force would of course resist such a scenario. The implications of (3.17) for angular momentum transport are discussed in Chapter 6.

It follows from (3.3) that the zonally averaged meridional component of the geostrophic wind is *exactly* zero:

$$\boxed{\overline{v_g}^{\lambda} = 0} \tag{3.18}$$

on each pressure surface. Eq. (3.18) implies that all zonally averaged meridional circulations are completely ageostrophic. An implication is that the important large-scale circulations are not necessarily close to geostrophic balance. We note, however, that the strongest features in Fig. 3.10 are found in the tropics, where geostrophy is expected to lose its grip.

Eq. (3.15) shows that the mass-weighted vertical integral of the meridional wind (i.e., the vertically integrated meridional mass flux) can lead to systematic changes with time in the meridional distribution of atmospheric mass. Fig. 3.11 shows the variation with season of the vertically and zonally averaged meridional velocity, as inferred from the seasonal changes of the distribution of mass. The top panel shows variations with latitude, and the bottom panel shows the seasonal cycle at the Equator. The values are tiny: on the order of 1 millimeter per second.

The globally averaged surface pressure is very nearly invariant with time, apart from small changes associated with the seasonal cycle of atmospheric water vapor. Trenberth et al. (1987) discuss observations of the seasonal changes of the hemispherically averaged and globally averaged surface pressures associated with dry air and with water vapor. These are shown in Fig. 3.12. The fluctuations are on the order of 0.1% of the atmosphere's mass. Further discussion of the observed distribution of water vapor is given later in this chapter.

Fig. 3.13 shows maps of the 850 hPa meridional wind for January and July, respectively. Fig. 3.14 shows the corresponding maps for the 200 hPa surface. Unlike the zonal wind, the time-averaged meridional wind does not show a banded, east-west structure; the east-west gradients are at least as strong as the north-south gradients. In the Northern Hemisphere, there is a tendency for alternating southerly and northerly flows, with a structure that resembles zonal wave number 3 or 4, i.e., 3 or 4 maxima and minima around a latitude circle. The time-averaged meridional currents in the Southern Hemisphere are generally weaker than those in the Northern Hemisphere. The intensities of the meridional currents are stronger at 200 hPa than at 850 hPa. In the Northern Hemisphere especially, there is a tendency for stronger features in winter than in summer.

As can be seen in Fig. 3.10, the zonally averaged meridional flow in the tropics of the winter hemisphere is generally toward the summer pole at low levels, and toward the winter pole aloft. This is not very apparent in Fig. 3.13 or Fig. 3.14, however. Fig 3.13 shows a strong southerly flow at 850 hPa just north of the Equator in July, associated with the Indian summer monsoon. The northerly flow near 120°E in January is associated with the winter monsoon, but is relatively inconspicuous.

In many parts of the world, the mean meridional wind reverses seasonally. Examples include the Arabian Sea, most of the North Pacific, and the southern Great Plains of North America.

Fig. 3.15 shows the streamlines at 850 hPa, for January and July, and Fig. 3.16 shows

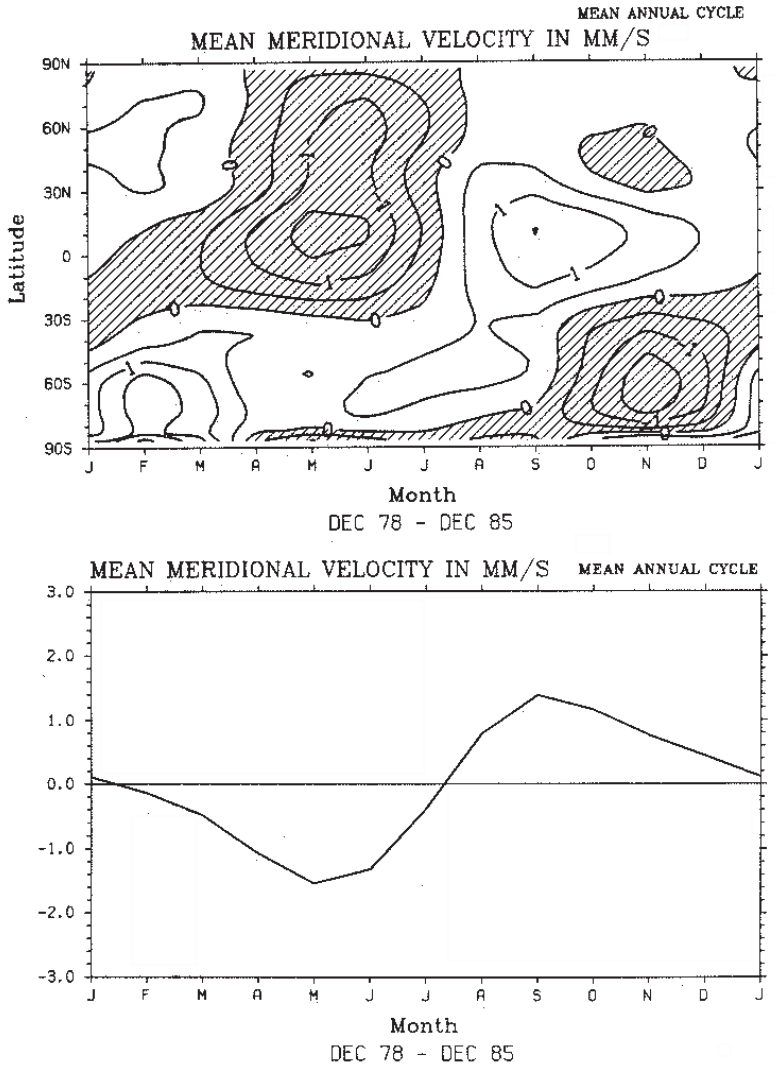


Figure 3.11: Seasonal variations of the mass-weighted vertically averaged meridional velocity, as inferred from the seasonally changing meridional distribution of surface pressure. The top panel shows variations with latitude, and the bottom panel shows the climatological seasonal cycle at the Equator. From Trenberth et al. (1987).

the corresponding streamlines at 200 hPa. Streamlines indicate direction but not magnitude. Such features as the subtropical highs and midlatitude lows of sea-level pressure are clearly evident in the 850 hPa winds. There are strong cross-equatorial flows at 850 hPa in both the Pacific and Indian Oceans, in July. The 200 hPa streamlines show the wavy patterns in the midlatitude winter, and also tropical phenomena including a strong monsoon-induced anticyclone over the Indian subcontinent in July.

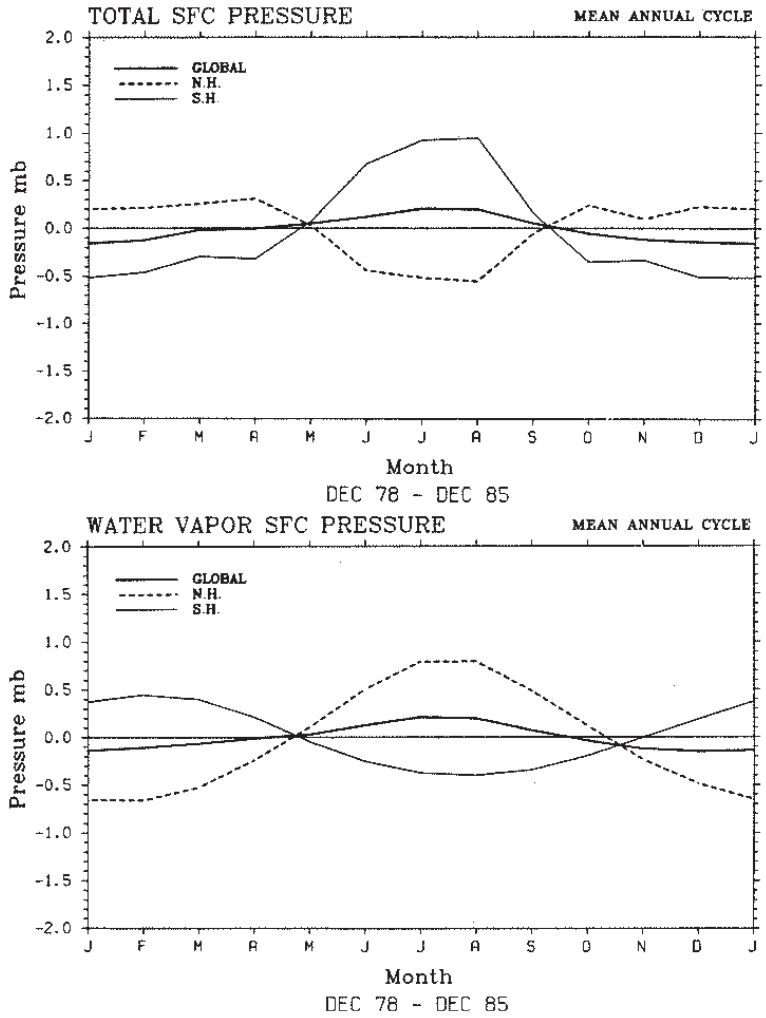


Figure 3.12: The variations with season of the hemispherically averaged and globally averaged surface pressures associated with dry air and with water vapor. From Trenberth et al. (1987).

3.5 Vertical velocity and the mean meridional circulation

The "mean meridional circulation" (MMC) is the name given to the circulation of mass in the the latitude-height plane. It can be analyzed by starting from the continuity equation in pressure coordinates, which is

$$\nabla_p \cdot \mathbf{v} + \frac{\partial \omega}{\partial p} = 0, \tag{3.19}$$

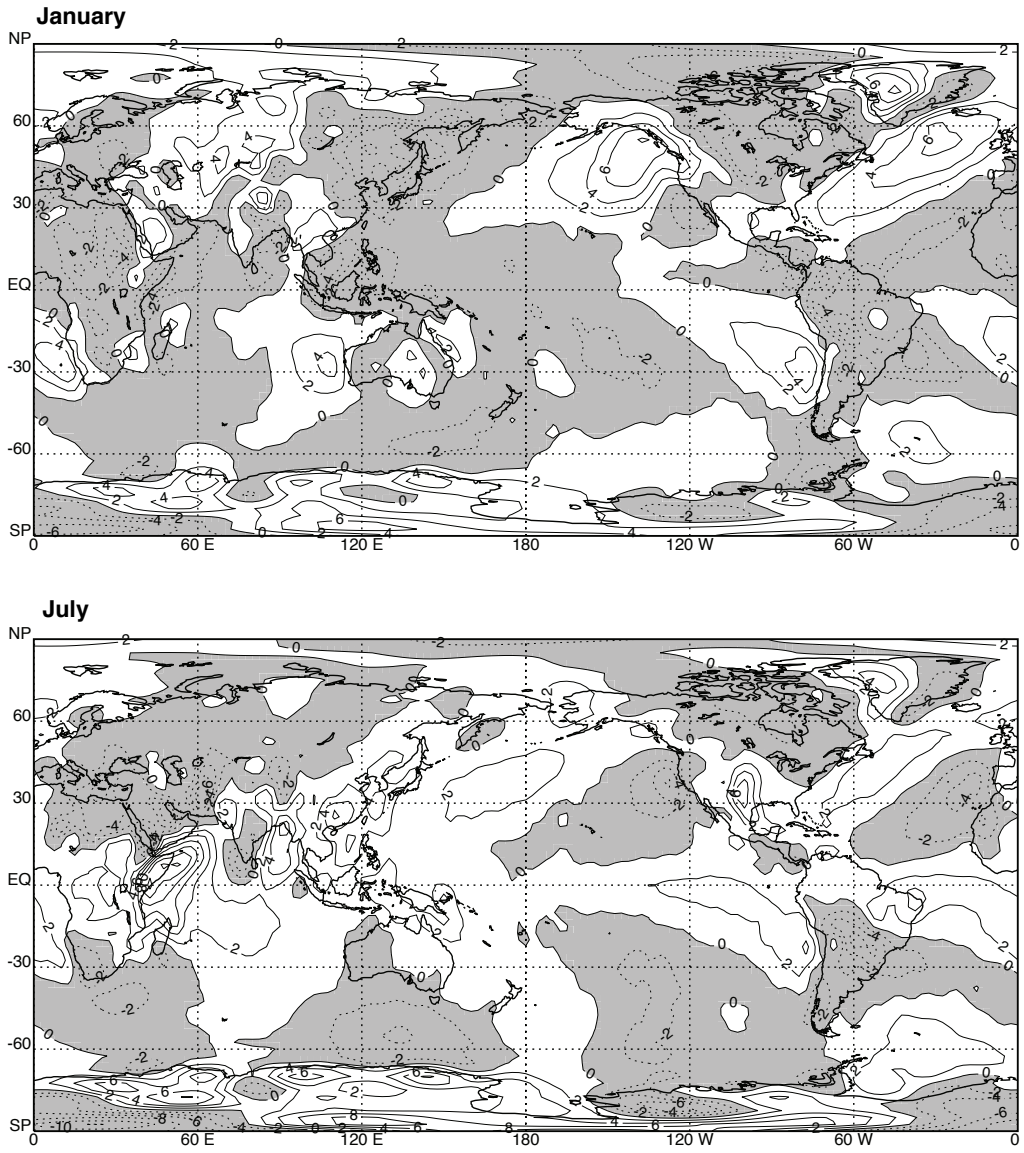


Figure 3.13: Maps of the 850 hPa meridional wind, for January and July. The contour interval is 2 m s^{-1}. Northerlies are shaded.

where ω is the vertical velocity as seen in pressure coordinates, i.e., the time rate of change of pressure following a particle of air. Taking the zonal average of (3.19), and using the form of the divergence operator in spherical coordinates that was discussed in Chapter 2, we find that

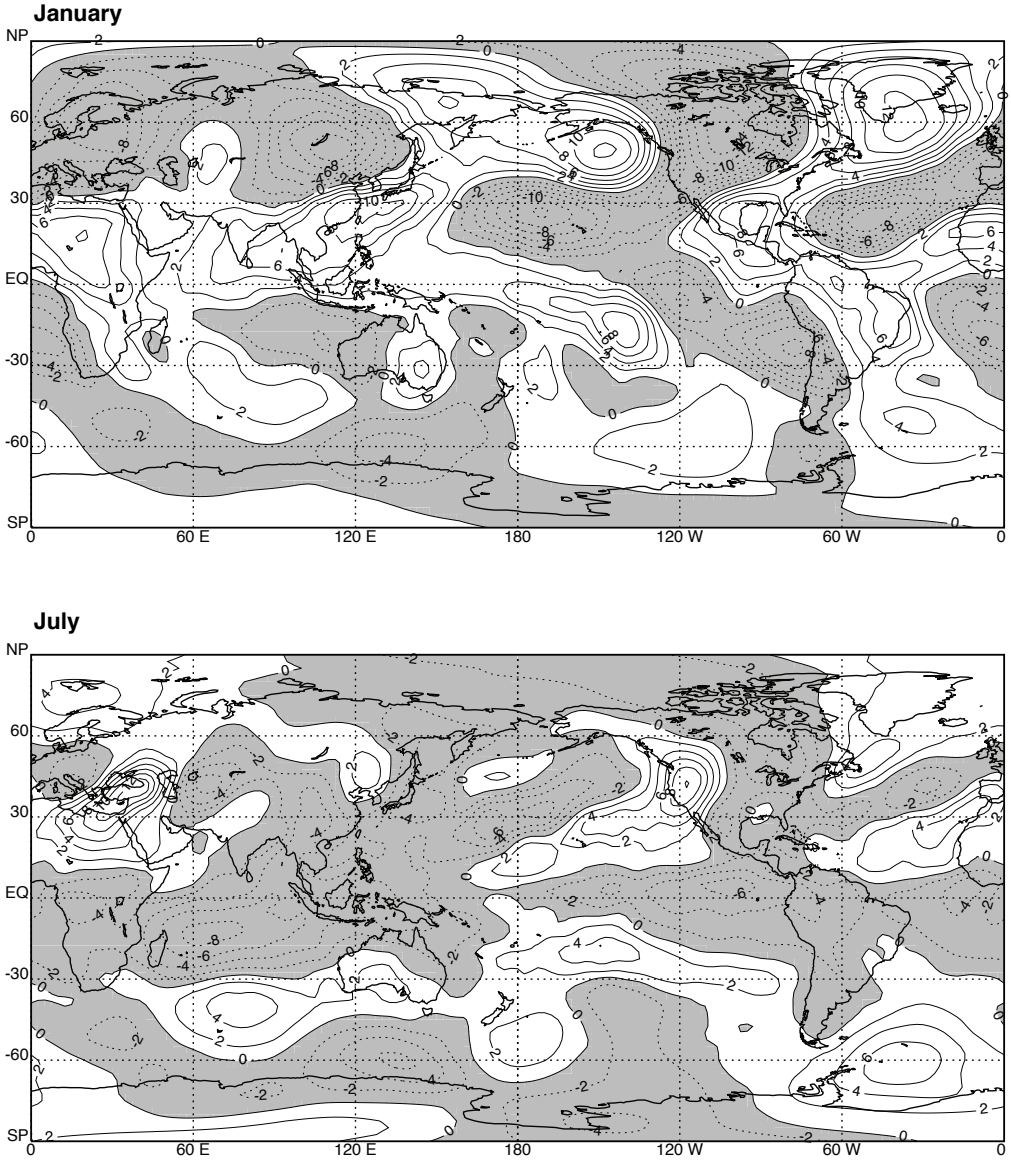


Figure 3.14: Maps of the 200 hPa meridional wind, for January and July. The contour interval is 2 m s^{-1}. Northerlies are shaded.

$$\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v}^{\lambda}\cos\varphi\right)+\frac{\partial\overline{\omega}^{\lambda}}{\partial p}=0. \tag{3.20}$$

It is useful to discuss the MMC in terms of a "stream function," ψ. Plots of ψ conveniently depict the zonally averaged vertical velocity and the zonally averaged meridional velocity,

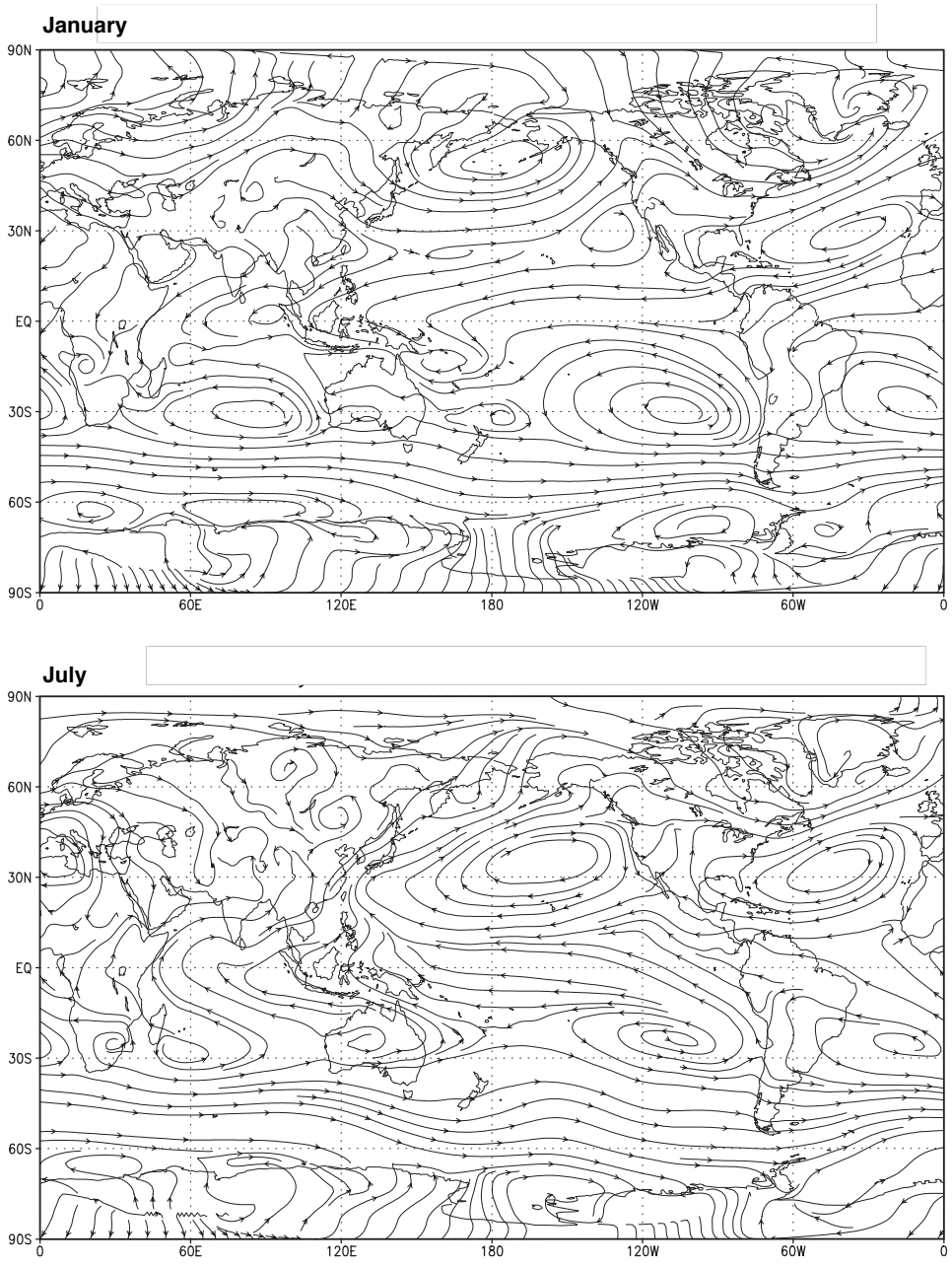


Figure 3.15: Streamlines of the 850 hPa wind, for January and July.

together in one diagram. The definition of ψ is embodied in the two equations

$$\bar{v}^{\lambda} 2\pi a \cos\varphi \equiv g\frac{\partial \psi}{\partial p}, \tag{3.21}$$

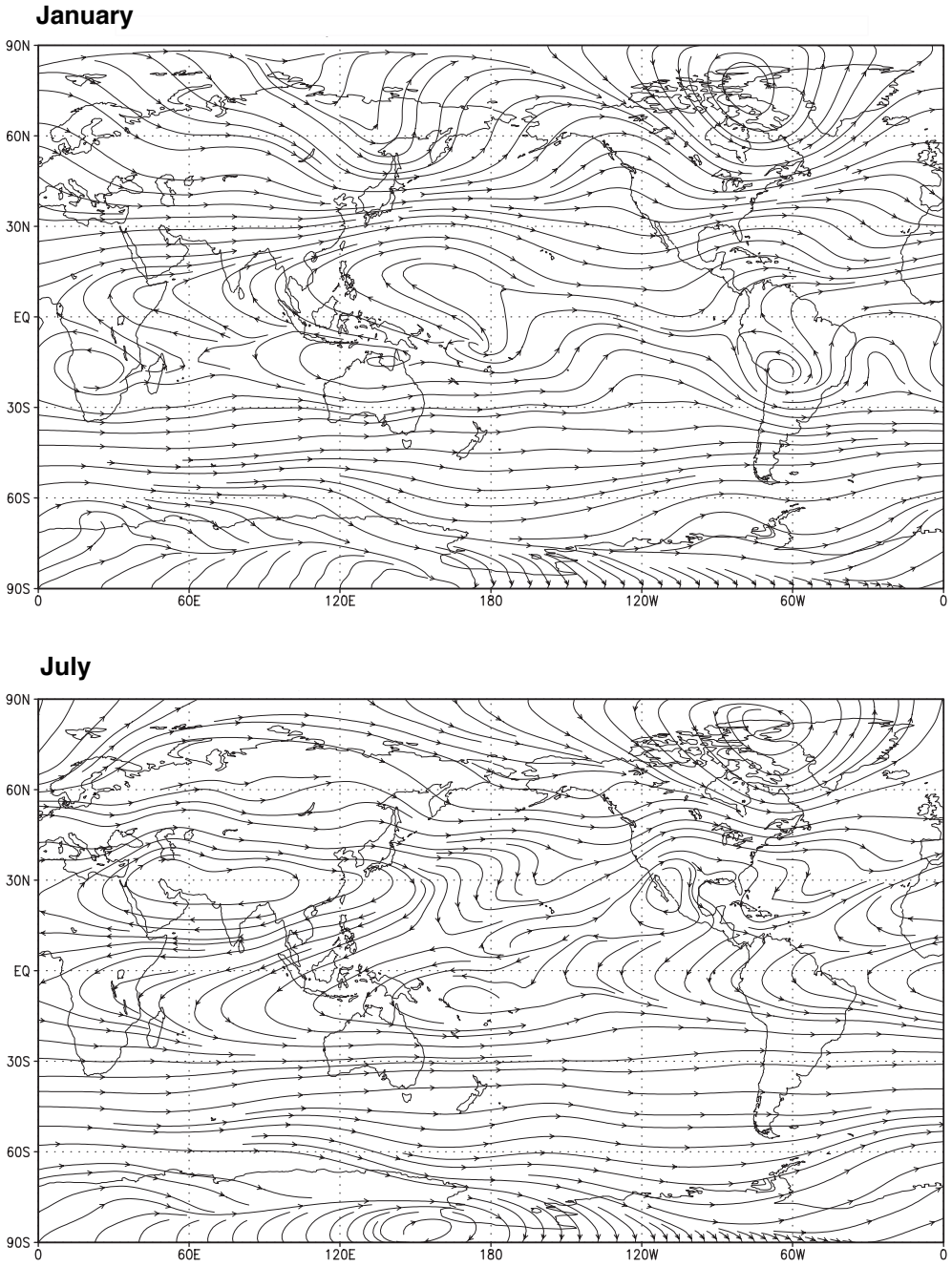


Figure 3.16: Streamlines of the 200 hPa wind, for January and July.

$$\overline{\omega}^{\lambda} 2\pi a^2 \cos\varphi \equiv -g\frac{\partial \psi}{\partial \varphi}. \tag{3.22}$$

The motivation for this definition is that, for any given distribution of ψ, the zonally averaged continuity equation, (3.20), is automatically satisfied; this is easy to verify by substitution. Note that ψ is independent of longitude, because it is defined in terms of the zonal averages of v and ω. Because ψ is defined in terms of its derivatives, it can only be determined within an arbitrary constant. In other words, an arbitrary constant can be added to ψ without changing $\overline{v}^{\lambda}$ or $\overline{\omega}^{\lambda}$.

A minor technical difficulty is that the zonal averages used to define ψ cannot be carried out for pressure surfaces that intersect the ground. We are going to ignore that issue. It can be skirted, for example by using a terrain-following vertical coordinate system, but the details are too technical to consider here.

In view of (3.21) and (3.22), we can compute ψ from either $\overline{v}^{\lambda}$ or $\overline{\omega}^{\lambda}$, but observations of $\overline{v}^{\lambda}$ are generally considered more reliable, so $\overline{v}^{\lambda}$ is the better choice. In order to do the vertical integration, we need a boundary condition. At $p = 0$, we have $\omega = 0$; this means that the atmosphere is neither gaining nor losing mass by exchanges with space. It follows that

$$\frac{\partial \psi}{\partial \varphi} = 0 \text{ at } p = 0, \tag{3.23}$$

i.e., the top of the atmosphere is a line of constant stream function. Similarly, lines of constant ψ cannot intersect the Earth's surface; if they did, that would imply a flow of air through the Earth's surface. We can therefore set

$$\frac{\partial \psi}{\partial \varphi} = 0 \text{ at } p = p_S, \tag{3.24}$$

In fact, since the zonal mean of the vertically integrated meridional wind is approximately zero in a time average, ψ must take *the same* latitude-independent value at $p = 0$ and $p = p_S$. It is conventional to choose this value to be zero, i.e., we use the upper boundary condition

$$\psi = 0 \text{ at } p = 0; \tag{3.25}$$

this choice determines the arbitrary constant mentioned above.

Note from (3.21) and (3.22) that

$$\psi \sim \left(\frac{\delta p}{g}\right) \cdot \frac{L}{t} \cdot L \sim \frac{M}{L^2} \cdot \frac{L^2}{T} = \frac{M}{T} \tag{3.26}$$

where M denotes dimensions of mass, L denotes dimensions of length, and T denotes dimensions of time. According to (3.26), ψ has dimensions of mass per unit time. It is usually expressed in units of 10^{12} g s^{-1}, which is the same as 10^9 kg s^{-1}. This unit is sometimes called a "Sverdrup" (abbreviated Sv), especially in the oceanographic literature.

Fig. 3.17 shows the latitude-height distribution of the stream function of the MMC, for January and July, respectively. The observed meridional wind has been used to create these plots, by vertical integration of (3.21), and use of (3.25). The figure shows that deep rising motion occurs in the summer-hemisphere tropics, with sinking motion on either side. The strongest tropical rising motion is near 300 hPa, but notice that weak rising motion continues into the tropical stratosphere. The strongest subsidence is in the winter hemisphere subtropics, again near 300 hPa. Rising motion occurs in middle latitudes, and is strongest in the winter. Maximum values tend to occur near 500 hPa. Sinking motion is found near the poles, mainly in the lower troposphere. The cover of this book shows a portion of a Hadley cell over Africa in August, when the "large" Hadley cell has its rising branch north of the Equator, and its sinking branch in the southern subtropics. The rising branch of the is marked by the bright thunderstorms across Africa and extending out over the Atlantic Ocean, slightly north of the Equator. The sinking branch is associated with clear air over southern Africa, and shallow stratocumulus clouds over the South Atlantic.

The dominant cellular structures in the tropics are called "Hadley cells." They are very important branches of the global circulation. There is a "large" Hadley cell at each solstice, with its rising branch in the summer-hemisphere tropics and its body extending into the winter-hemisphere subtropics. Its peak magnitude is about 160 x 10^{12} g s^{-1}. A weaker Hadley cell is found in the summer hemisphere. Both Hadley cells are "direct" circulations, which means that their rising branches are warmer than their sinking branches. As discussed later, direct circulations convert potential energy into kinetic energy.

Because of the seasonal growth and decay of the Hadley cells in the two hemispheres, the zonally averaged meridional wind at the Equator reverses seasonally. Near the solstices it is from the winter hemisphere into the summer hemisphere at low levels, and from the summer hemisphere into the winter hemisphere in the upper troposphere. Bowman and Cohen (1997) discuss the inter-hemispheric transports associated with the seasonally changing Hadley circulations.

There are also indirect circulations in the middle latitudes, most clearly in the Southern

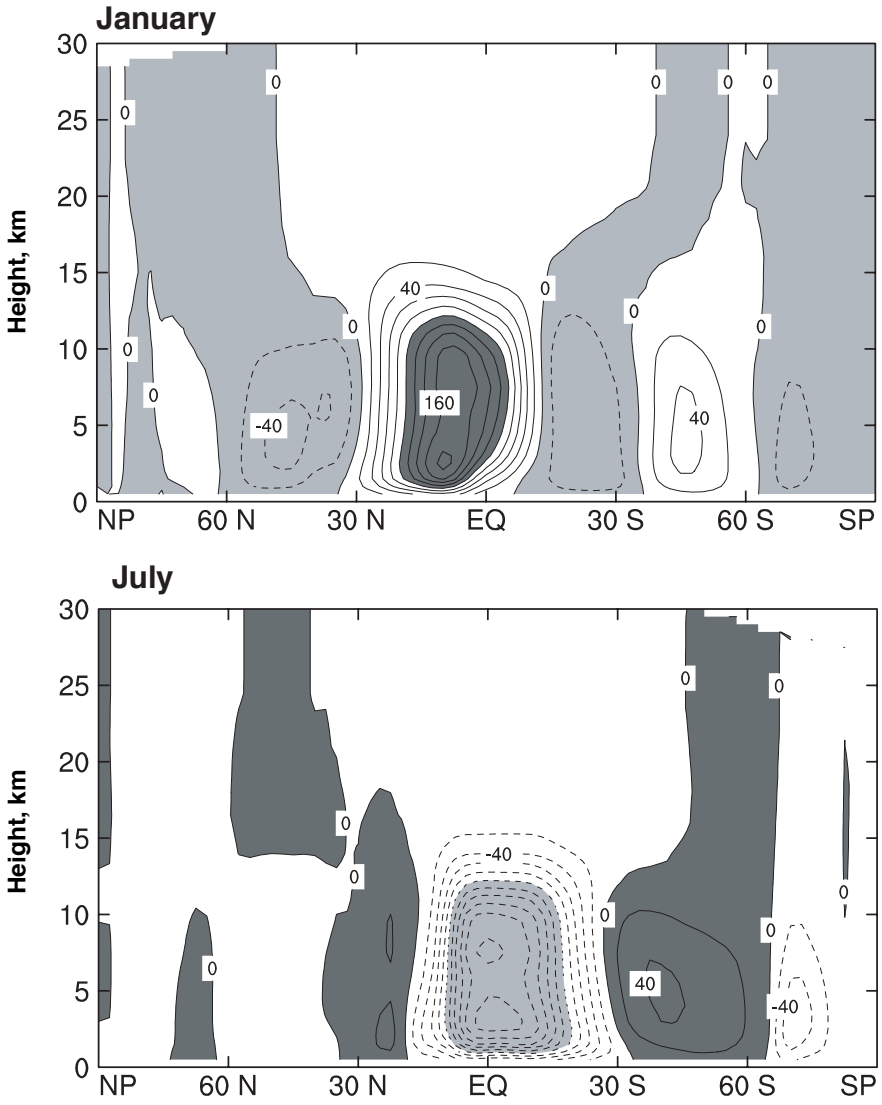


Figure 3.17: The stream function of the mean meridional circulation. Positive values represent counterclockwise circulations, while negative values represent clockwise circulations. The units are 10^{12} g s^{-1}.

Hemisphere in both seasons and both hemispheres. These are called "Ferrel cells." The sinking branches of the Ferrel cells are adjacent to the sinking branches of the Hadley cells; both are found in the subtropics, near 30° of latitude both north and south of the Equator. In these latitude belts, the sinking air diverges away horizontally, both toward the pole and toward the Equator. The poleward branch is the inflow to the rising motion of the Ferrel cell, and the Equatorward branch is the inflow to the rising branch of the Hadley cell. Recall that the zonally averaged sea-level pressure has its maximum in the subtropics; we can think of the diverging subtropical meridional flows as being pushed by the meridional

pressure-gradient force, from the sea-level pressure maximum towards lower pressure on both sides. Further physical interpretation of the Ferrel cells is given later.

Finally, the polar regions play host to weak direct circulations.

There is an important but very slow meridional circulation in the stratosphere, called the Brewer-Dobson circulation (Butchart, 2014). It is too weak to show up in Fig. 3.17, but it carries mass upward across the tropopause in the tropics, and then towards the winter pole. It produces important transports of ozone and angular momentum. It is discussed further in Chapter 11.

It is interesting to examine the seasonal change of the MMC, as shown in Fig. 3.18. Rough symmetry between the hemispheres is seen approximately one month after the equinoxes. Near the solstices, the Hadley cells are best developed on the winter side, and least developed on the summer side. Explanations for this for this have been proposed by Lindzen and Hou (1988); Hack et al. (1989), and Dima and Wallace (2003). The rising motion on the summer side of the Equator, near the solstice, is associated with the summer monsoons (Newell et al., 1972; Schulman, 1973; Dima and Wallace, 2003), which are discussed in Chapter 10.

The correspondence between the zonally averaged vertical motion and the zonally averaged meridional motion is fairly evident. The meridional currents can be interpreted as outflows from or inflows to the vertical currents. Fig. 3.19 shows maps of the 500 hPa vertical velocity for January and July, respectively. The units are nanobars per second. The strongest maxima and minima have absolute values of roughly 1000 nanobars per second, which is about 100 hPa per day. Fig. 3.17 shows that there are regular bands of rising and sinking motion, arranged along latitude circles. The bands are apparent in Fig. 3.19, especially in the July data. There is some tendency for rising motion in the tropics; sinking motion in the subtropics, especially in the winter hemisphere; rising motion in middle latitudes; and sinking motion over the poles. Sinking motion tends to be associated with surface pressure maxima, and rising motion with surface pressure minima. For example, the subtropical highs are clearly identifiable as regions of large-scale sinking motion in the middle troposphere. The seasonal change of the large-scale vertical motion is very spectacular in the region of the Tibetan plateau. Rising motion occurs in summer, and sinking motion in winter. These changes are associated with the Indian monsoon, as discussed in Chapter 10. There are some cases of rising motion upstream of mountain ranges, and sinking motion downstream; examples are the Rocky Mountains and the Himalayas, in winter. Looking at the other major mountain ranges of the world, however, it is not easy to see a clear pattern of orographically forced vertical motions. They do exist, but a more refined analysis is needed to detect them.

Note the rising motion over southern Africa and tropical South America in January, in the same regions where we will see later that there are water vapor maxima at 850 hPa in January. There is a tendency for large water vapor mixing ratios in regions of rising motion, and small water vapor mixing ratios in regions of sinking motion. In particular,

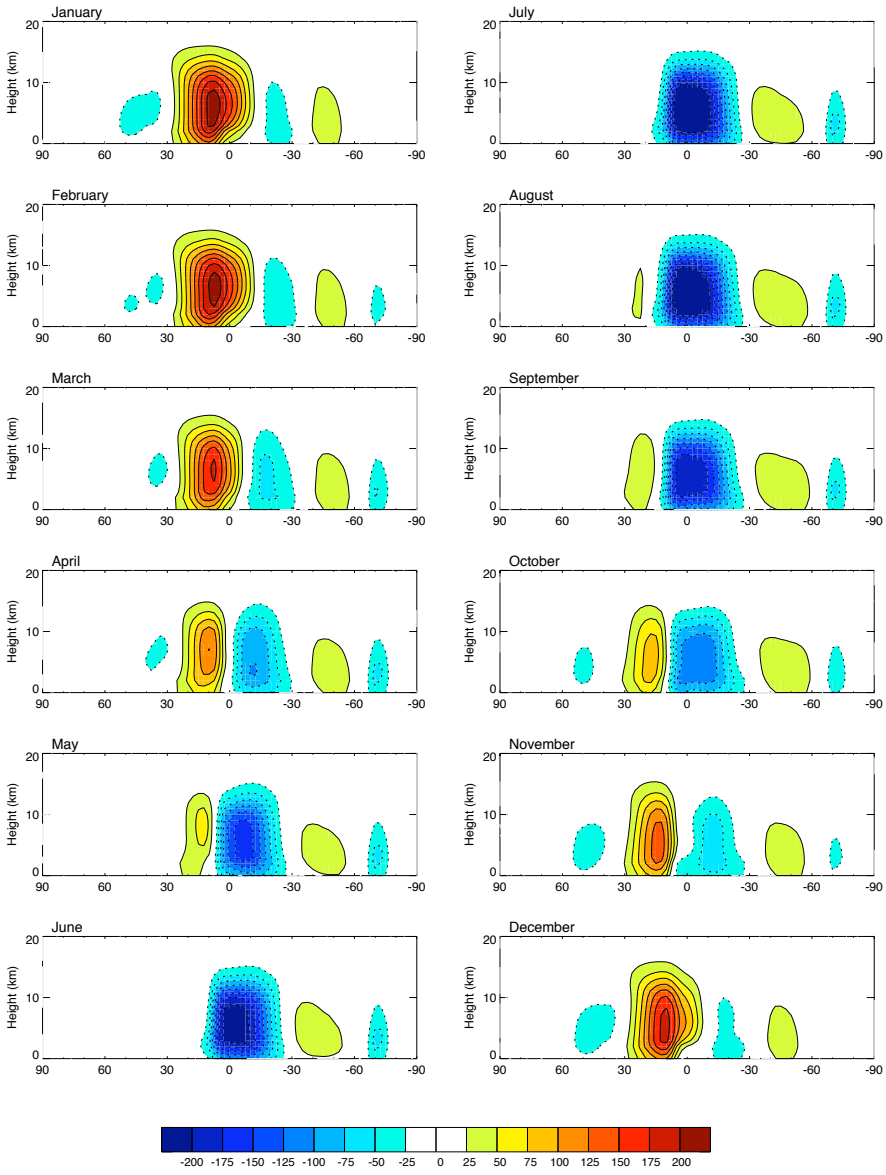


Figure 3.18: Seasonal change of the stream function of the mean meridional circulation. A similar figure appeared in Dima and Wallace (2003).

deserts, such as the Sahara, are regions of tropospheric sinking motion

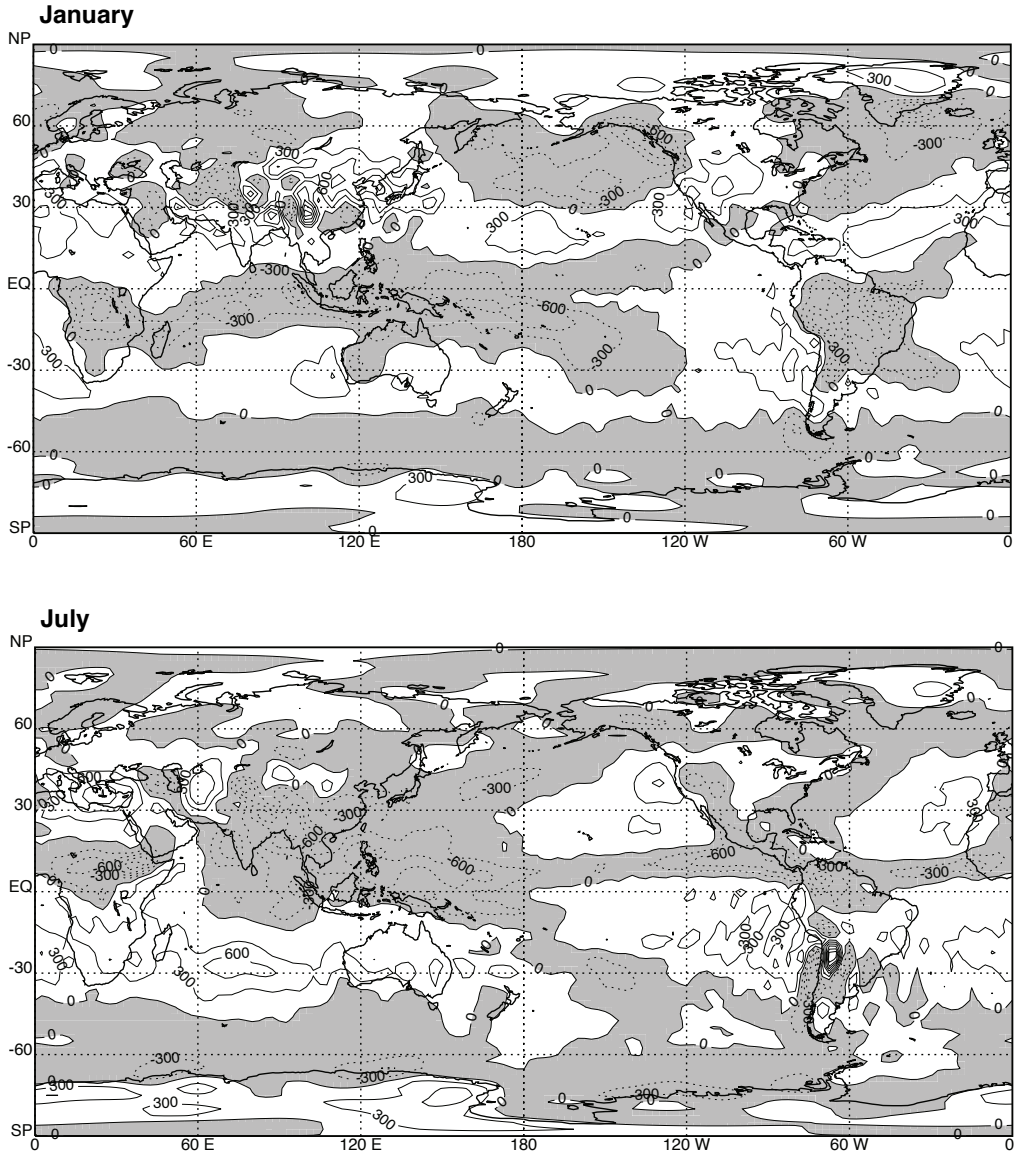


Figure 3.19: Maps of ω, the vertical "pressure velocity," for the 500 hPa surface. The units are nb s^{-1}. Negative values (corresponding to rising motion) are shaded.

3.6 Temperature

Fig. 3.20 shows the latitude-height distributions of the zonally averaged temperature for January and July, respectively. At low levels, the warmest air is near the Equator, but near 100 hPa the *coldest* air is over the Equator. In fact, some of the lowest temperatures in the atmosphere are found near the tropical tropopause. It is also true that the tropopause is

highest in the tropics and lowest near the poles. The tropopause height is nearly discontinuous in the subtropics, particularly in the winter hemisphere. In the stratosphere, extremely cold temperatures are found above the winter pole, especially in the Southern Hemisphere. The summer pole is much warmer due to the absorption of solar radiation by ozone.

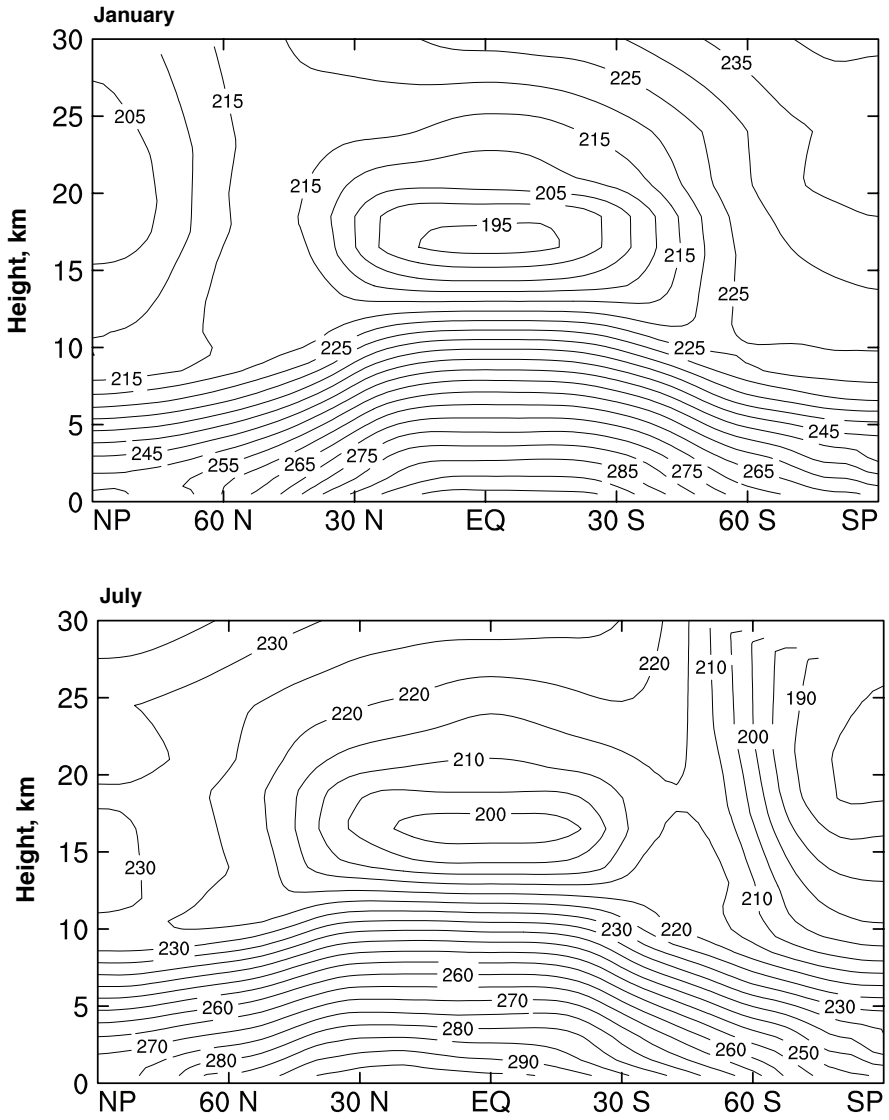


Figure 3.20: Latitude-height cross sections of temperature, for January and July. The contour interval is 5 K.

It has been suggested that, for a wide class of sufficiently massive atmospheres, the tropopause must occur near the 100 hPa level, because at higher pressures the atmosphere is relatively opaque to the thermal infrared radiation, while at lower pressures it is relatively transparent (Robinson and Catling, 2014).

The midlatitude low-level temperature gradients are quite strong in the winter hemisphere. Above 200 hPa, the summer pole is considerably warmer than the winter pole; in fact, the zonally averaged temperature increases monotonically from the Equator to the summer pole near 100 hPa. This suggests easterlies in the summer stratosphere, which are observed. In the winter hemisphere, the warmest 100 hPa temperatures occur in middle latitudes. The strong decrease of temperature between midlatitudes and the poles is consistent with the polar night jets mentioned earlier.

Generally speaking, the lapse rate, $-\partial T/\partial z$, is largest in the tropics, and smaller (or even negative) near the poles. In January, a temperature "inversion" (i.e., temperature increasing upward) appears at low levels near the North Pole.

The thermal wind equations relate the vertical shear of the geostrophic wind to the horizontal temperature gradient. They are given by

$$\frac{\partial u_g}{\partial p} = \frac{R}{fpa}\left(\frac{\partial T}{\partial \varphi}\right)_p \tag{3.27}$$

and

$$\frac{\partial v_g}{\partial p} = \frac{-R}{fpa\cos\varphi}\left(\frac{\partial T}{\partial \lambda}\right)_p. \tag{3.28}$$

Thermal wind balance between the meridional temperature gradient and the vertical shear of the zonal wind is well satisfied. This can be qualitatively confirmed by comparison of Fig. 3.6 and Fig. 3.20.

Fig. 3.21 shows maps of the 850 hPa temperature for January and July, and Fig. 3.22 shows the corresponding results for 200 hPa. The expected winter-to-summer warming at 850 hPa is clearly evident in the Northern Hemisphere, but not in the Southern Hemisphere, except over land. Monthly mean temperatures over the high Antarctic terrain reach about -50°C in July, while those over the Arctic ocean in January do not fall below -35°C. In the tropics there is very little seasonal change, and the temperature distribution is very uniform, for reasons that we have already discussed.

Naturally, the temperature gradient points mainly from the poles toward the tropics at 850 hPa, but wavy patterns are plainly visible in the Northern Hemisphere in January. There are some regions in which the mean temperature actually increases poleward, at 850 hPa, e.g., from Northern Africa across to India in July, and north of Australia in January. From thermal wind considerations, we might expect easterlies aloft in these regions. This expectation is borne out in Fig. 3.9. In winter, there is a tendency for the eastern sides of the Northern Hemisphere continents to be colder than the western sides. This leads to

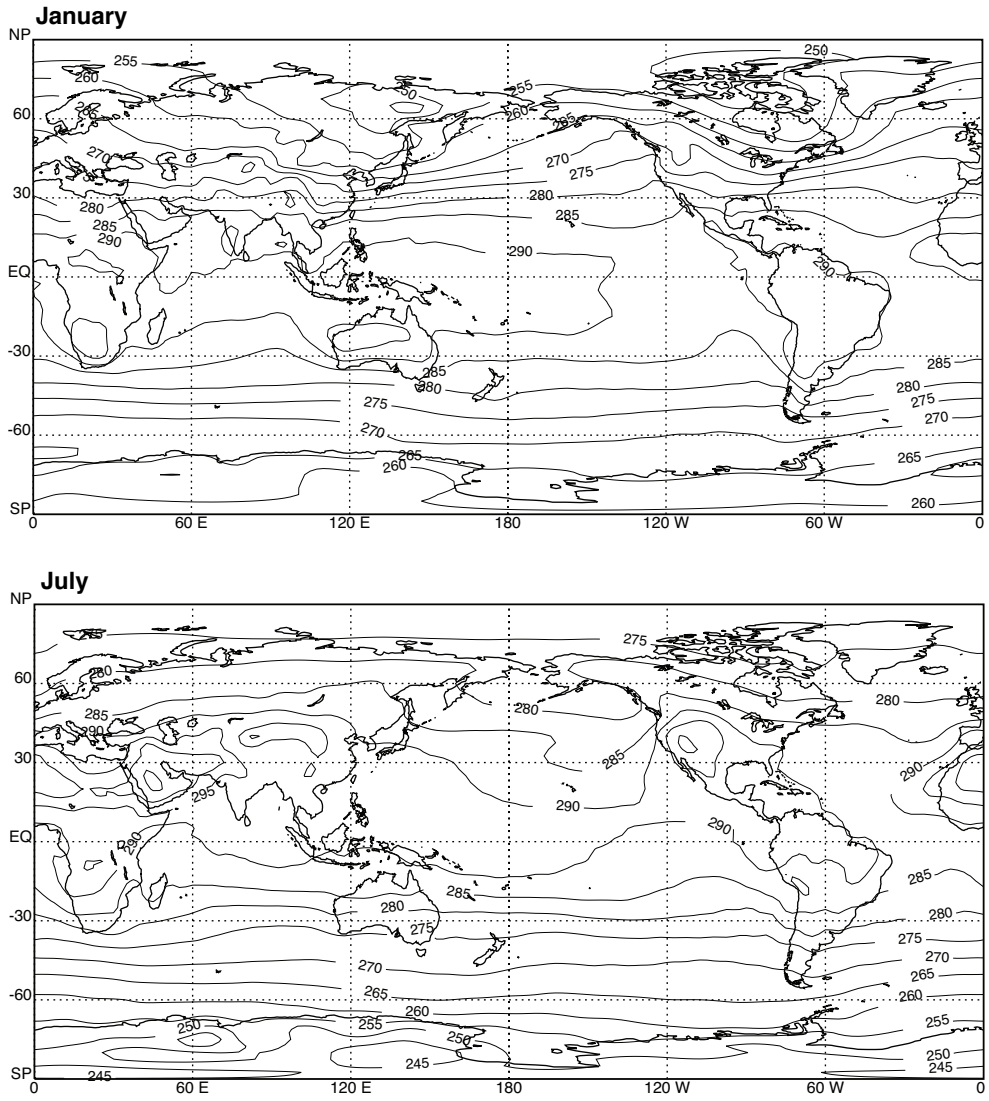


Figure 3.21: Maps of the 850 hPa temperature for January and July. The contour interval is 5 K.

particularly strong meridional temperature gradients on the east coasts. We know that such strong temperature gradients favor a rapid upward increase of the westerlies; also, such highly baroclinic regions are preferred centers of cyclogenesis.

The strongest temperature gradients at 200 hPa are found in high latitudes, especially in winter. The wave-like eddy pattern is much stronger at 200 hPa than at 850 hPa. Particularly noticeable are the maxima over the North Pacific in January, over eastern North America in January, and over southern Asia in July. Note that in each of these regions the 200 hPa temperature increases as we move from the tropics towards middle latitudes. This implies a tendency for the westerlies to weaken above this level.

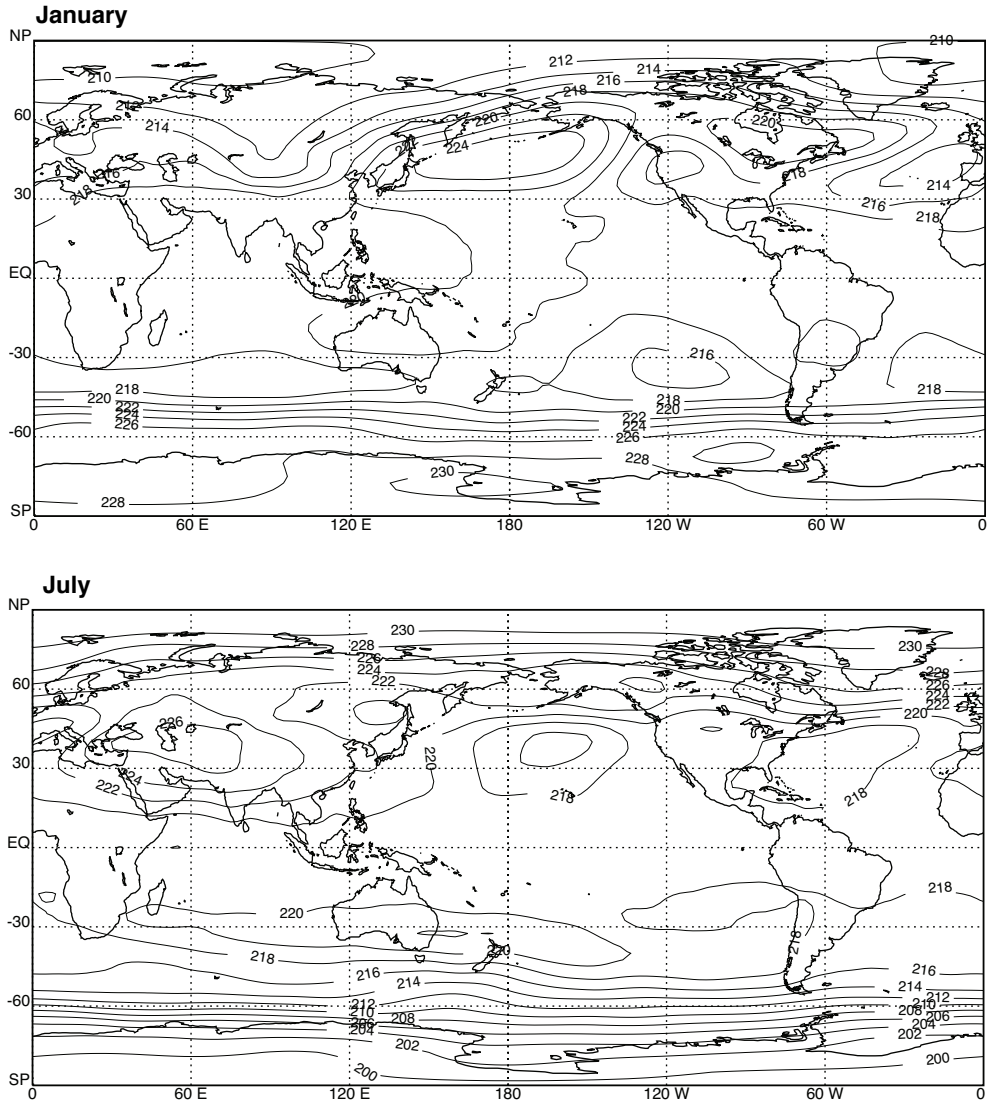


Figure 3.22: Maps of the 200 hPa temperature for January and July. The contour interval is 5 K.

Recall that the potential temperature is defined by

$$\theta \equiv T\left(\frac{p_0}{p}\right)^{\kappa}, \tag{3.29}$$

where p_0 is a constant reference pressure, usually chosen to be 1000 hPa, and $\kappa = R/c_p$, where R is the gas constant and c_p is the specific heat of air at constant pressure. As

discussed in Chapter 4, important facts about the potential temperature are that it is conserved under dry adiabatic processes, and that (with minor exceptions) it increases upward throughout the atmosphere. This upward increase is evident in Fig. 3.23. In the troposphere, potential temperature surfaces slope downwards from the polar regions to the tropics, while in the stratosphere the opposite occurs. The stratosphere is easily identified because the potential temperature increases upward very sharply there, showing that the static stability is very strong. On the other hand, the static stability is particularly weak in the tropical upper troposphere.

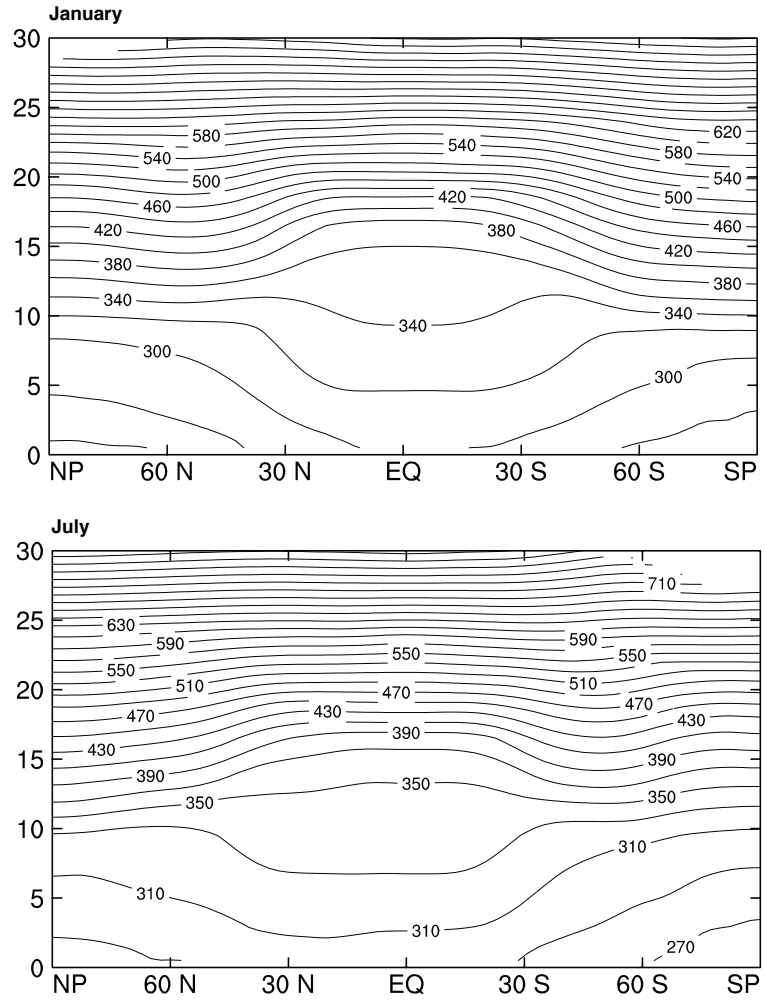


Figure 3.23: Latitude-height cross sections of potential temperature, for January and July. The contour interval is 10 K.

Hoskins (1991) distinguishes the following three regimes, which can be identified in Fig. 3.23: There is an "Overworld," in which (by definition) the potential temperature surfaces are everywhere above the tropopause. From the data, we see that such air has potential

temperatures of about 390 K or larger, i.e., the tropical tropopause roughly coincides with the surface on which the potential temperature is 390 K. The "Middleworld" is defined to have potential temperature surfaces that cross the tropopause, which means that air moving along isentropic surfaces in the Middleworld can move between the troposphere and the stratosphere. The Middleworld is found outside the tropics, in middle latitudes. Finally, the "Underworld" has potential temperature surfaces that intersect the Earth's surface, so that air moving isentropically in the Underworld can "sample" the properties of the Earth's surface and communicate these to the atmosphere. Much of the Underword is found in the high-latitude troposphere. From Fig. 3.23, it appears that the largest potential temperatures in the Underworld are on the order of 300 K, although in reality much larger values of the potential temperature do occur locally near the Earth's surface, e.g., on a summer afternoon in the Sahara desert.

As you may know, potential temperature can be used as a vertical coordinate, and such an approach has many advantages for interpreting the observations, for theoretical analyses, and for numerical modeling (e.g., Hoskins et al., 1985). We will frequently use potential temperature as a vertical coordinate in this book. Chapter 4 introduces the relevant concepts.

3.7 Storm tracks

We need a notation that allows us to distinguish between eddies and zonal means. We will occasionally (but not systematically) distinguish between "stationary" eddies, which are anchored to features (such as mountain ranges) on the Earth's surface, and so appear in time-averaged (e.g. monthly mean) maps, and "transient" eddies that move and so are smeared out to invisibility in sufficiently long time averages. We adopt the notations shown in Table 3.1, some of which have already been used. In the following discussion, we consider various statistics derived from the spatial and temporal distributions of fields called v and T. We use these symbols only for convenience. In principle, v and T can be any variables. They could be the same variable.

As discussed in Appendix D, the temporal and zonal mean of the product of two arbitrary variables v and T can be expanded as follows:

$$\overline{vT}^{t,\lambda} = \overline{v}^{t,\lambda}\overline{T}^{t,\lambda} + \overline{\overline{v^*T^*}^t}^{\lambda} + \overline{\overline{v'}^{\lambda}\overline{T'}^{\lambda}}^{t} + \overline{v'^*T'^*}^{\lambda,t} \,. \tag{3.30}$$

Here a prime denotes a departure from a time average, and a star denotes a departure from the zonal mean. The terms of (3.30) have the following meanings:

Table 3.1: Terminology used in the discussion of the zonally averaged flow and eddy covariances.

Notation	Meaning
$\overline{(\ )}^t$	Time mean
$(\ )'$	Departure from the time mean, or *transient* component
$\overline{(\ )}^\lambda$	Zonal mean
$(\ )^*$	Departure from the zonal mean, or *eddy* component

- the term $\overline{v}^{t,\lambda}\overline{T}^{t,\lambda}$ represents the contribution to $\overline{vT}^{t,\lambda}$ from the product of the temporally and zonally averaged values of v and T;
- the term $\overline{\overline{v^*}^t\overline{T^*}^t}^\lambda$ represents the contribution from "stationary eddies," i.e., eddies that appear in time averages;
- the term $\overline{\overline{v'}^\lambda\overline{T'}^\lambda}^t$ represents the contribution from temporal fluctuations of the zonal means, sometimes called the transient symmetric contribution; and
- the term $\overline{v'^*T'^*}^{\lambda,t}$ represents the contribution from "transient eddies," including baroclinic waves in the extratropics.

Finally, the "total eddy" contribution is the sum of the stationary and transient eddy contributions.

Fig. 3.24 shows maps of the transient eddy meridional flux of temperature at 850 hPa, for January (upper panel) and July (lower panel). The plots on the right show the zonal averages. The maxima over the midlatitude oceans in winter are called the "storm tracks." They exist in summer too, but are much weaker (especially in the Nortern Hemisphere) and shifted poleward.

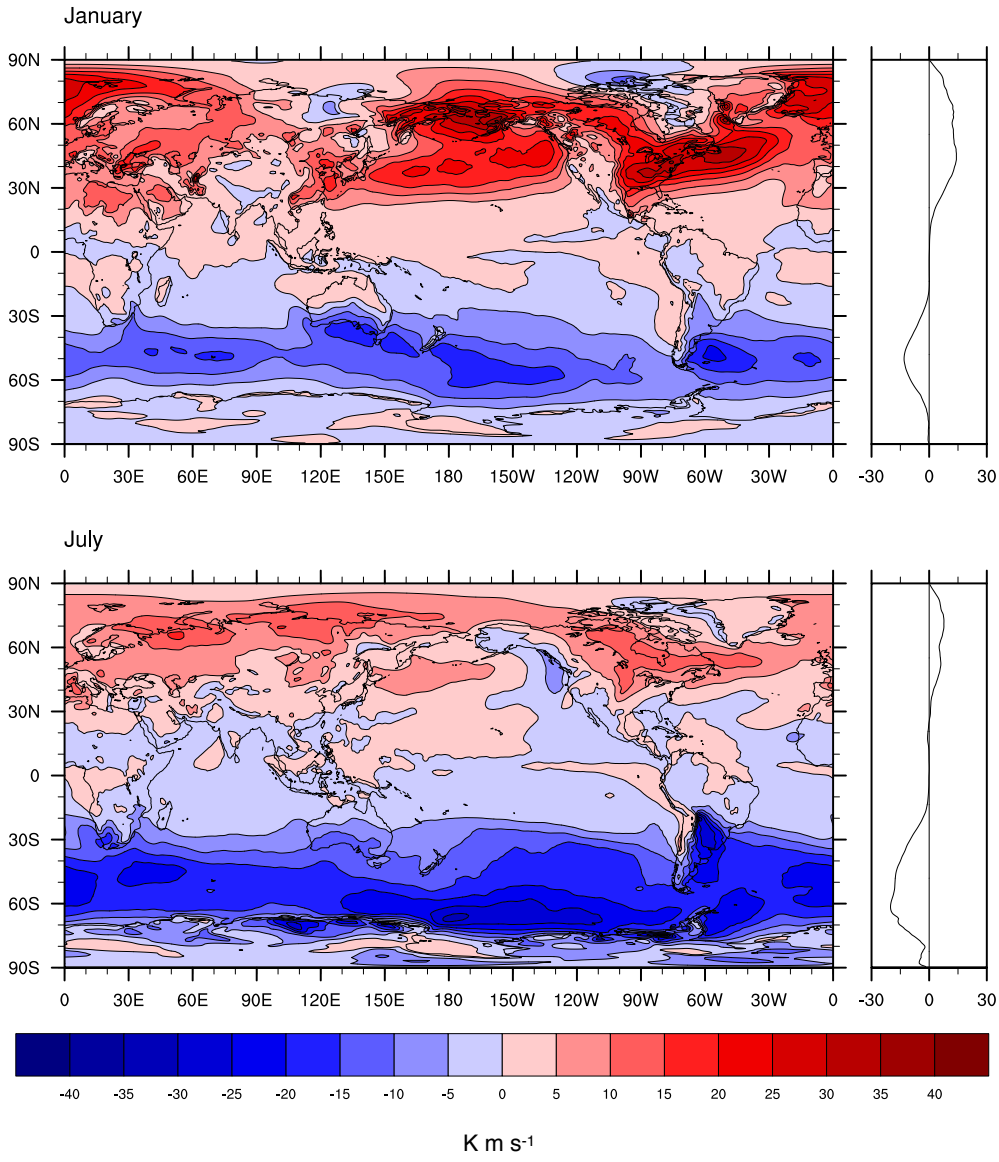


Figure 3.24: Maps of the transient eddy meridional flux of temperature at 850 hPa, for January (upper panel) and July (lower panel). The plots on the right show the zonal averages.

3.8 Moisture

The globally averaged evaporation and precipitation rates, which must balance in a time average, are not accurately known, but are a little less than 3 mm day^{-1}. It follows that an average water molecule spends a "residence time" of about 8 days in the atmosphere, between its introduction by surface evaporation and its removal by precipitation.

The water vapor mixing ratio is the density of water vapor divided by the density of

dry air. Fig. 3.25 shows the latitude-height distributions of the zonally averaged water vapor mixing ratio for January and July, respectively. Caution is needed here, because the data presented in these figures have been fed through the analysis/forecast system of a numerical weather prediction center, which can easily distort the distribution of water vapor. The figures show the most humid air near the equator, and the driest air near the winter pole. The seasonal change in the Northern Hemisphere is quite dramatic. There is an extremely rapid upward decrease of the mixing ratio at all latitudes. The largest average values, near the surface in the tropics, are close to 18 g kg^{-1}, which means that about 2% of the air is water vapor. Even minute quantities of water vapor in the upper troposphere and stratosphere can be very important radiatively.

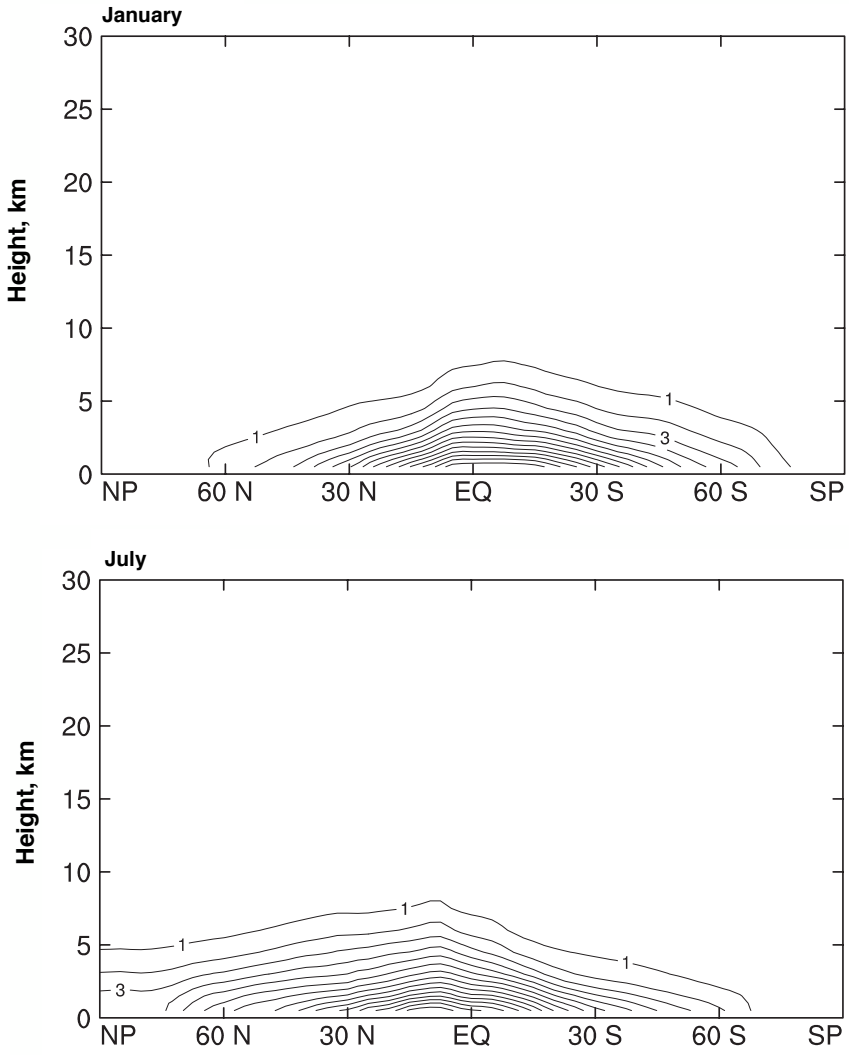


Figure 3.25: Latitude-height section of water vapor mixing ratio. The contour interval is 1 g kg^{-1}.

Because the mixing ratio is greatest near the surface, regions of low-level mass convergence, such as those apparent in the zonally averaged meridional wind, tend to be regions of vertically integrated moisture convergence as well. Although as much mass diverges at upper levels as converges at lower levels, the diverging air aloft is dry, while the converging air near the surface is humid. A net convergence of water vapor occurs in such regions. The situation is reversed in desert regions, where the dry upper tropospheric air converges and descends.

The vertically integrated water vapor content of the air, with dimensions of mass per unit area, is called the "column water vapor" or "the precipitable water." Fig. 3.26 shows the seasonal cycle of the zonally averaged precipitable water, as a function of latitude. The tropical maxmimum shifts meridionally, favoring the summer hemisphere. The seasonal variations are especially pronouced in the Northern Hemisphere.

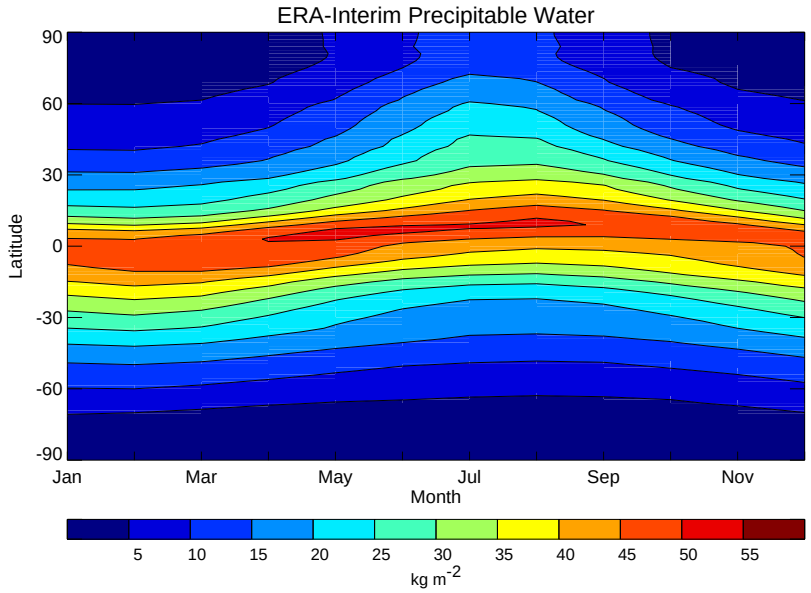


Figure 3.26: the seasonal cycle of the zonally averaged precipitable water, as a function of latitude.

Fig. 3.27 shows maps of the 850 hPa water vapor mixing ratio for January and July, respectively. As would be expected, the largest values occur in the tropics and the summer hemisphere. A very clear maximum extends around the circumference of the Earth in the tropics, mainly somewhat north of the Equator. Meridional moisture gradients are often quite sharp. There are also strong east-west variations, however. For example, in January there are strong maxima over southern Africa and in the Amazon basin. Minima are found in the subtropical highs. There are very dramatic seasonal changes over the midlatitude continents, with larger values in summer. Major desert regions like the Sahara and western North America are clearly associated with water vapor minima.

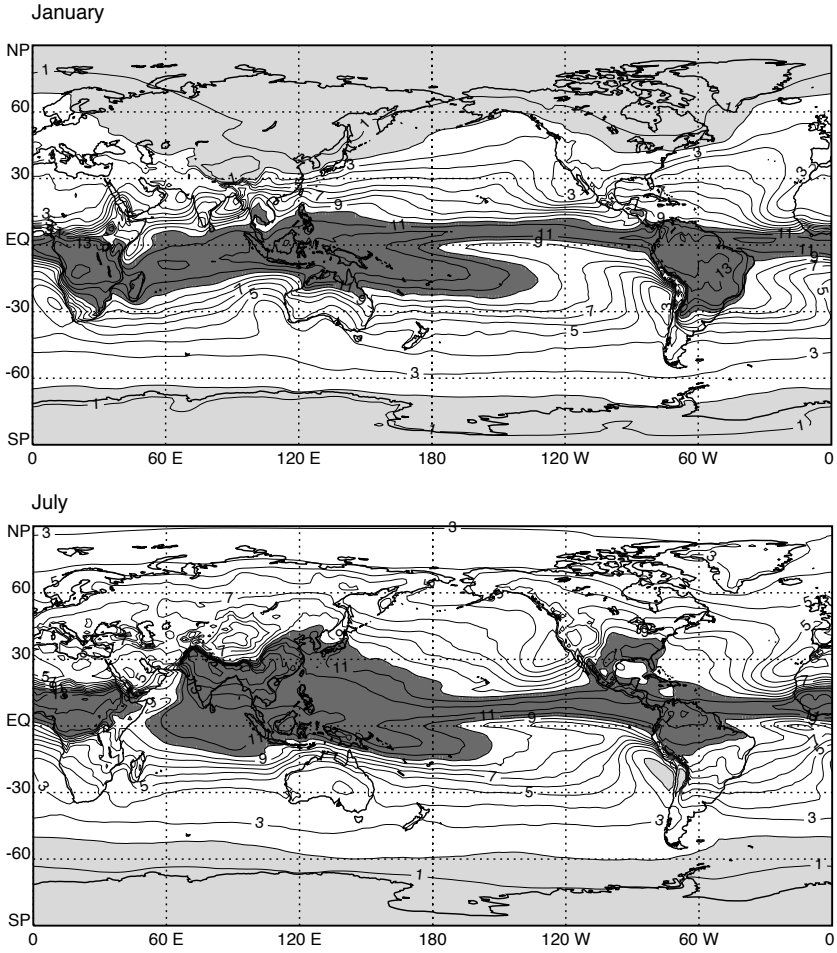


Figure 3.27: Maps of the 850 hPa water vapor mixing ratio. The contour interval is 1 g kg^{-1} . Values larger than 10 g kg^{-1} are darkly shaded.

The relative humidity can be defined as the ratio of the actual water vapor pressure to the saturation vapor pressure; the latter is an increasing function of temperature only, and was discussed already in Chapter 2. Locally, a relative humidity of 1 or more, corresponding to water vapor "saturation," indicates the presence of cloud. The zonally averaged relative humidity is shown in Fig. 3.28. The tropical troposphere is very humid in rainy regions

of the tropics, and high relative humidity is also found in middle latitudes, especially in winter, where strong low-pressure systems are frequent.

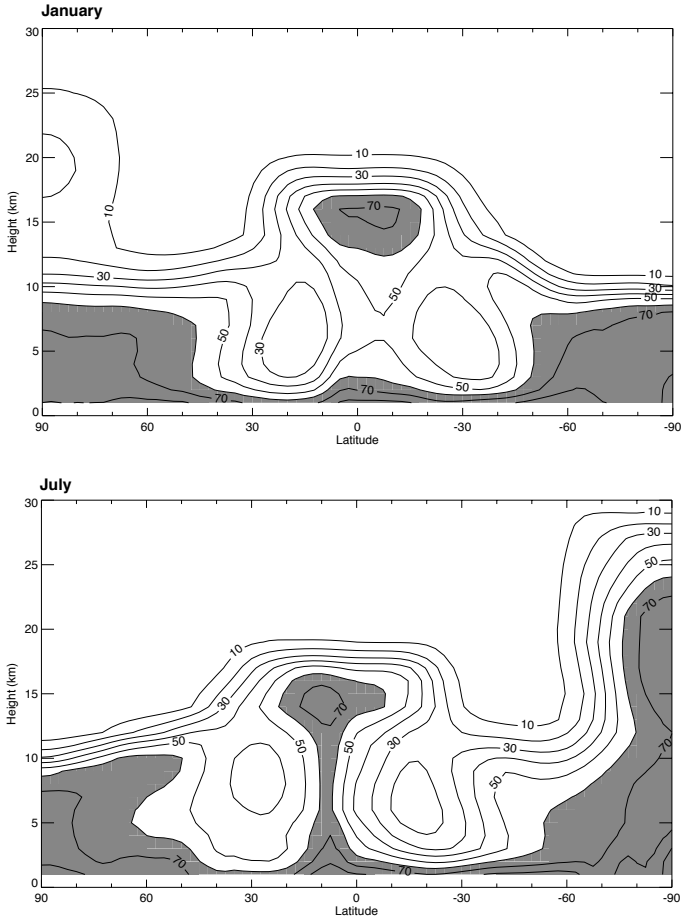


Figure 3.28: The observed latitude-height distribution of the zonally averaged relative humidity, in %, as analyzed by ECMWF. Values larger than 60% are shaded.

3.9 Lots of questions

This chapter is designed to impart some familiarity with basic features of the observed global circulation, but without much attempt to explain why the circulation appears as it does. The observations presented raise many questions. For example:

- What determines the magnitude of the pole-to-equator gradient of the surface temperature?
- Why are seasonal changes generally weaker in the Southern Hemisphere than in the Northern Hemisphere?

- What determines the observed lapse rate of temperature? Why does the lapse rate change as we move from the tropics to the middle latitudes to the poles?
- What determines the intensities and latitudes of the jet streams? Why do the winter jet maxima occur at particular longitudes?
- Why are the subtropical highs typically found on the eastern sides of the ocean basins?
- Why is the Intertropical Convergence Zone mainly north of the Equator?
- What accounts for the observed patterns of large-scale rising and sinking motion?
- Why is there such a strong belt of low pressure around Antarctica? What determines the locations and intensities of the Aleutian and Icelandic Lows?
- What determines the vertical distribution of water vapor?
- Why is the upper-level circulation "wavy"? Why is the lower-level circulation "lumpy?"
- What determines the height of the tropopause, as a function of latitude? Why is the tropical tropopause so cold? Why does the tropopause height have a discontinuity near the subtropical jets?
- What mechanisms generate the observed variations with longitude of the monthly-mean fields?
- What accounts for the observed large-scale pattern of winds in a monsoon?
- How much do individual Januaries and Julys, for particular years, differ from the "average" January and July conditions shown in this chapter? What causes such year-to-year variations?
- What are the geographical patterns of the day-to-day weather fluctuations that accompany the monthly mean maps shown here, and how do these fluctuations affect the time means?
- Why does the global circulation appear "smooth," rather than "noisy?"

These and many other questions will be discussed in the remainder of this book. To address them, we will need the ideas presented in Chapter 4.

3.10 Problems

1. Estimate the total water vapor content of the atmosphere, in kg. Explain how you arrive at your answer.
2. Make a rough estimate of the total kinetic energy of the atmosphere, in joules. If all of the solar radiation absorbed by the Earth were used to supply this kinetic energy,

how long would it take to accumulate the observed amount? Note: In reality the rate of kinetic energy generation in the atmosphere is (fortunately!) much less than the rate at which the Earth absorbs solar radiation.

Chapter 4

The rules of the game

4.1 Introduction

This chapter gives a quick review of concepts to be used later. Topics covered including the conservation principles for momentum, moisture, energy in its various forms, and potential vorticity. We show the general vector forms of the equations, and also the particular forms that apply in spherical coordinates. We also work out the forms of the equations that apply with potential temperature as the vertical coordinate. We also briefly introduce some important approximations, including the quasi-hydrostatic and quasi-geostrophic approximations.

4.2 Conservation of the mass of dry air

Mass is arguably the most fundamental quantity in any physical description of the atmosphere. The circulation is the movement of mass from place to place. Let $\mathbf{V}$ denote the three-dimensional wind vector. The vector $\rho\mathbf{V}$ is the flux of mass across a unit area. Consider a "control volume," CV, through which fluid can flow (Fig. 4.1). Conservation of mass is expressed by

$$\frac{d}{dt}\int_{\mathrm{CV}} \rho \, dV = \int_{\mathrm{CV}} \frac{\partial \rho}{\partial t} \, dV = -\int_{\sigma} \rho \mathbf{V} \cdot \mathbf{d}\sigma, \tag{4.1}$$

where the first two integrals are taken over the control volume, σ is the bounding surface of the control volume, and $\mathbf{d}\sigma$ is an outward-normal vector whose magnitude is that of a (differential) element of the bounding surface. Eq. (4.1) is a statement of mass conservation. It says that the mass in the volume changes *only* due to mass flowing in or out across the boundaries of the volume.

Gauss's Theorem tells us that

$$\int_{\mathrm{CV}} (\nabla \cdot \mathbf{A})\, dV = \int_{\sigma} \mathbf{A} \cdot \mathbf{d}\sigma \,, \tag{4.2}$$

where $\mathbf{A}$ is an arbitrary three-dimensional vector. Using Gauss's Theorem, we can rewrite (4.1) as

$$\int_{\mathrm{CV}} \left[\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) \right] dV = 0 \,. \tag{4.3}$$

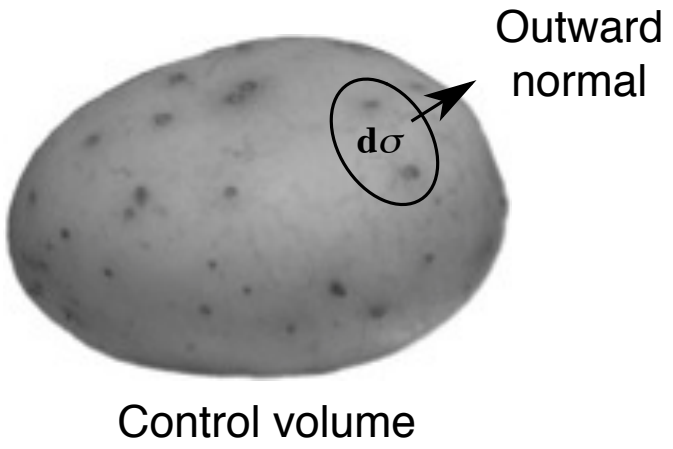


Figure 4.1: Sketch illustrating the control volume used to derive the continuity equation. Here $\mathbf{d}\sigma$ is a small element of area on the boundary of a control volume. Control volumes are always shaped like potatoes; hence the expression "Great Geophysical Potato."

Since the control volume is arbitrary, the integrand must be zero everywhere, i.e.,

$$\boxed{\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0} \,. \tag{4.4}$$

This expresses conservation of mass in what is called a "flux form;" it says that the density of the air at a fixed point changes only due to the convergence or divergence of the mass flux.

Mass conservation can also be expressed using a Lagrangian framework, in which we consider the time rate of change of the density following a particle. For this purpose, we use the Lagrangian (particle-following) time derivative, D/Dt, which can be expanded as

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla \tag{4.5}$$

To demonstrate (4.5) s, let $\mathbf{r}$ be the position vector that corresponds to the particle's location, and write the Lagrangian time derivative of an arbitrary variable A as

$$\frac{DA}{Dt} = \lim_{\Delta t \to 0} \left[\frac{A(\mathbf{r} + \Delta\mathbf{r}, t + \Delta t) - A(\mathbf{r}, t)}{\Delta t} \right] . \tag{4.6}$$

Think of the two values of A on the right-hand side of (4.6) as "measured values" for a particular particle of air. The measurements are made at times t and $t + \Delta t$, when the particle's positions are $\mathbf{r}$ and $\mathbf{r} + \Delta\mathbf{r}$, respectively. Writing a series expansion about $(t, \mathbf{r})$, we find that

$$A(\mathbf{r} + \Delta\mathbf{r}, t + \Delta t) - A(\mathbf{r}, t) = \left(\frac{\partial A}{\partial t} \right)_{\mathbf{r}} \Delta t + (\nabla A)_t \cdot \Delta\mathbf{r} + \text{higher order terms.} \tag{4.7}$$

Dividing both sides of (4.7) by Δt, using $\Delta\mathbf{r}/\Delta t \equiv \mathbf{V}$, and letting Δt and $\Delta\mathbf{r}$ go to zero, we obtain

$$\frac{DA}{Dt} = \left(\frac{\partial A}{\partial t} \right)_{\mathbf{r}} + \mathbf{V} \cdot (\nabla A)_t \ . \tag{4.8}$$

The left-hand side of (4.8) is a Lagrangian description of the change of A experienced by a moving particle, and the right-hand side is the sum of the "Eulerian" time-rate-of-change of A as seen at the point $\mathbf{r}$ and a term representing the effects of advection as seen in the same Eulerian coordinate system. Both the left-hand side of (4.8) and the *sum* of the terms on the right-hand side are independent of coordinate system. The individual terms on the right-hand side of (4.8) *do* depend on the coordinate system, however. Starting here, we simplify the notation by dropping the subscripts $\mathbf{r}$ and t that appear in (4.7) and (4.8).

By using (4.8), we can rewrite (4.4) in the alternative form

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{V} = 0, \tag{4.9}$$

which can be rearranged to

$$\nabla \cdot \mathbf{V} = -\frac{1}{\rho}\frac{D\rho}{Dt} = \frac{1}{\alpha}\frac{D\alpha}{Dt}, \tag{4.10}$$

where $\alpha \equiv 1/\rho$ is the specific volume. An (idealized) incompressible fluid obeys $D\rho/Dt = 0$ and $D\alpha/Dt = 0$, i.e., the density (and specific volume) of an incompressible particle cannot change. Eq. (4.10) therefore implies that

$$\nabla \cdot \mathbf{V} = 0 \tag{4.11}$$

for an incompressible fluid, i.e., the three-dimensional wind field must be nondivergent. Eq. (4.11) is called the Boussinesq approximation to the continuity equation; it is often used in oceanography, and it can also be applied to sufficiently shallow layers of the atmosphere such as the PBL.

The mass budget of a column of air can be deduced from (4.4) as follows. Integrate with respect to height, through the entire column, to obtain

$$\int_{z_S}^{\infty} \frac{\partial \rho}{\partial t} dz + \int_{z_S}^{\infty} \nabla \cdot (\rho \mathbf{V})\, dz = 0. \tag{4.12}$$

Here z_S is the surface height, which varies in space and could conceivably vary with time (e.g., during an earthquake). Moving the integrals inside the derivatives, and taking into account the possible spatial and temporal variability of z_S, we obtain

$$\frac{\partial}{\partial t}\left(\int_{z_S}^{\infty} \rho dz\right) + \nabla \cdot \left(\int_{z_S}^{\infty} \rho \mathbf{v} dz\right) + \rho_S \left(\frac{\partial z_S}{\partial t} + \mathbf{v}_S \cdot \nabla z_S - w_S\right) = 0 \tag{4.13}$$

In deriving (4.13), we have used the fact that $\rho \to 0$ as $z \to \infty$. The subscript S denotes the Earth's surface, $\mathbf{v}$ is the horizontal velocity vector, and w is the vertical velocity, so that

$$\mathbf{V} = \mathbf{v} + w\mathbf{e}_r\,, \tag{4.14}$$

where $\mathbf{e}_r$ is a unit vector that points radially outward (i.e., upward), away from the center of the Earth. You should be able to see that, in (4.13), the expression $\rho_S \left(\dfrac{\partial z_S}{\partial t} + \mathbf{v}_S \cdot \nabla z_S - w_S \right)$ represents the flux of mass across the Earth's surface. As a physical lower boundary condition, we declare that this mass flux is zero:

$$\boxed{\frac{\partial z_S}{\partial t} + \mathbf{v}_S \cdot \nabla z_S - w_S = 0 \, .} \tag{4.15}$$

Eq. (4.15) will be used repeatedly later in this book. The *form* of (4.15) carries over to vertical coordinate systems other than height. For example, with the pressure coordinate (4.15) is replaced by

$$\frac{\partial p_S}{\partial t} + \mathbf{v}_S \cdot \nabla p_S - \omega_S = 0 \, . \tag{4.16}$$

We can keep the $\partial z_S / \partial t$ term of (4.15), e.g., if we want to study the effects of earthquakes, or waves moving along the sea surface. In such cases, we would specify $\partial z_S / \partial t$ to describe how the lower boundary moves. Normally, of course, the height of the Earth's surface can be assumed to be independent of time,and with a few exceptions we will assume that in this book. In that case (4.15) reduces to

$$w_S = \mathbf{v}_S \cdot \nabla z_S \, . \tag{4.17}$$

Eq. (4.17) says that the three-dimensional wind vector is tangent to the Earth's surface.

The Martian atmosphere consists almost entirely of carbon dioxide (CO_2). There is a non-zero flux of atmospheric mass across the Martian surface, near the poles. In the winter, polar temperatures are low enough so that the atmosphere (partially) condenses onto the surface; during the summer, the atmosphere gains mass near the summer pole as the frozen CO_2 sublimates back to gaseous form (e.g., Lewis and Prinn, 1984; Zent, 1996). This process results in seasonal changes of the Martian surface pressure on the order of 30% (Hess et al., 1977; Tillman et al., 1993; Banfield et al., 2020).

The condensation of water in the Earth's atmosphere is analogous to the condensation of CO_2 in the Martian atmosphere, but of course water makes up only a tiny fraction of the Earth's atmosphere so the resulting changes in the surface pressure are small (see further

discussion in Chapter 6). If the Earth's atmosphere became cold enough to condense nitrogen or oxygen (not gonna happen!), similarly large changes in the surface pressure could result.

With the use of (4.15), (4.13) reduces to

$$\frac{\partial}{\partial t}\left(\int_{z_S}^{\infty}\rho dz\right)+\nabla\cdot\left(\int_{z_S}^{\infty}\rho\mathbf{v}dz\right)=0, \tag{4.18}$$

which expresses conservation of mass for the air column. According to (4.18), the total mass of air in the column, per unit horizontal area, changes in time *only* as a result of fluxes of air across the sides of the column.

Recall that to an excellent degree of approximation a vertical increment of pressure is proportional to the corresponding vertical increment of mass of air per unit area, i.e.,

$$dp\cong-\rho g dz\,. \tag{4.19}$$

This implies that the surface pressure is (approximately) given by

$$p_S\cong\int_{z_S}^{\infty}\rho g dz. \tag{4.20}$$

Use of (4.19) and (4.20) in (4.18) gives

$$\frac{\partial p_S}{\partial t}+\nabla\cdot\left(\int_{0}^{p_S}\mathbf{v}dp\right)\cong 0\,. \tag{4.21}$$

This is the surface-pressure-tendency equation, which was already discussed in Chapter 3.

4.3 Conservation of atmospheric moisture

While we are on the subject of mass, let's consider atmospheric moisture. About 99% of it is water vapor. The remainder is liquid or ice.

The mixing ratio of water vapor, q_v, is the density of water vapor divided by the density of dry air. In the troposphere, ranges from around 20 g kg^{-1} in warm tropical places like Hong Kong in summer, to a minimum of less than 1 g kg^{-1} in cold places like the upper troposphere or near the surface in the Antarctic winter . We can also define mixing ratios for liquid water and ice. The total water mixing ratio, q_T, can be separated into various subspecies by phase and/or particle size, as follows:

$$q_T = q_v + q_c + q_i + q_r + q_s \, . \tag{4.22}$$

Here q_c is the mixing ratio of "cloud water," which consists of droplets small enough to have negligible fall speeds; q_i is the mixing ratio of "cloud ice," which consists of crystals small enough to have negligible fall speeds; q_r is the mixing ratio of rain water; and q_s is the mixing ratio of snow. Much more elaborate break-downs are possible based on fine-grain distinctions among particles of various sizes (and shapes). There is evidence that such complexity does matter in detailed analyses of the global circulation of the atmosphere, but it is beyond the scope of this book, and in fact we will even ignore the ice phase for simplicity. We therefore replace (4.22) by

$$q_T = q_v + l \, , \tag{4.23}$$

where l represents the mixing ratio of liquid water regardless of drop size, including both q_c and q_r.

Conservation of total water is expressed by

$$\frac{\partial}{\partial t}(\rho q_T) + \nabla \cdot \left(\mathbf{V} \rho q_T + \mathbf{F}_{q_T} - \mathbf{P} \right) = 0 \, , \tag{4.24}$$

where $\mathbf{F}_{q_T}$ is the vector flux of total water vapor due to molecular diffusion, and $\mathbf{P}$ is minus the vector flux of liquid water (or ice) due to precipitation, which involves motion relative to the moving air. We include a minus sign in the definition of $\mathbf{P}$ so that a positive value of the vertical component of $\mathbf{P}$ will correspond to downward-moving rain or snow.

We can write separate conservation equations for q_v and l, as follows:

$$\frac{\partial}{\partial t}(\rho q_v) + \nabla \cdot \left(\rho \mathbf{V} q_v + \mathbf{F}_{q_v} \right) = -C, \tag{4.25}$$

$$\frac{\partial}{\partial t}(\rho l) + \nabla \cdot (\rho \mathbf{V} l + \mathbf{F}_l - \mathbf{P}) = C. \tag{4.26}$$

Here C is the net rate of condensation per unit mass; condensation converts vapor into liquid. We have also included in (4.26) a term that can represent a turbulent flux of liquid water. When we add (4.25) and (4.26), the condensation terms cancel, and we recover (4.24).

As discussed in detail later, vertical moisture transport in the atmosphere is often dominated by the fluxes associated with small-scale and mesoscale motions, which include turbulent eddies with scales from centimeters to a few kilometers, convective and wave motions with scales from a few kilometers to a few tens of kilometers, and mesoscale weather system (such as squall lines) with scales from a few tens of kilometers up to a few hundred kilometers. Fluxes due to these small-scale and mesoscale motions can be many orders of magnitude larger than the molecular fluxes, so in the study of the global circulation we can often (but not always) neglect the molecular fluxes.

The small-scale fluxes are vectors that include both horizontal and vertical components, but they mainly affect the large-scale circulation through vertical exchanges, i.e., it is the vertical fluxes that matter most. For example, when considering the effects of small-scale moisture fluxes on large-scale weather systems, we can safely assume that

$$\nabla \cdot \mathbf{F}_{q_T} \cong \frac{\partial}{\partial z}\left(F_{q_T}\right)_z , \tag{4.27}$$

where we now interpret $\mathbf{F}_{q_T}$ as a flux due to small-scale motions rather than molecular diffusion, and $\left(F_{q_T}\right)_z$ is the vertical component of $\mathbf{F}_{q_T}$. Although we have written (4.27) for the total water flux, approximations similar to (4.27) also apply to fluxes of other quantities, including radiative fluxes. The justification for (4.27) is *not* that the horizontal fluxes themselves are smaller than the vertical fluxes. In fact, for small-scale motions, the horizontal and vertical fluxes are often about the same size. The time rate of change is driven by the *flux divergence*, however, rather than the flux itself. The justification for the approximation in (4.27) is that the depth of the atmosphere, over which the vertical flux converges or diverges, is shallow compared to horizontal extent of the large-scale motions, over which the horizontal flux converges or diverges.

The approximation (4.27) is not valid for the fluxes associated with large-scale eddies. The horizontal fluxes due to large-scale eddies, such as winter storms, are considerably stronger than the vertical fluxes. For large-scale eddies, the horizontal and vertical flux divergences are comparable in magnitude; both matter.

Similar reasoning shows that when considering large horizontal scales we can approximate the precipitation term of (4.24) by

$$\nabla \cdot \mathbf{P} \cong \frac{\partial P}{\partial z} \tag{4.28}$$

Substituting (4.27) and (4.28) into (4.24), we obtain

$$\boxed{\frac{\partial}{\partial t}(\rho q_T) + \nabla \cdot (\rho \mathbf{V} q_T) + \frac{\partial}{\partial z}\left[\left(F_{q_T}\right)_z - P\right] = 0} \tag{4.29}$$

which applies on large scales, and will be used later.

4.4 Conservation of momentum on a rotating sphere

4.4.1 The equation of motion

We now consider the "exact" equations of motion in vector form, without using any coordinate system.

The Earth's angular velocity, $\mathbf{\Omega}$, is a vector pointing towards the celestial North Pole, as shown in Fig. 4.2. The magnitude of $\mathbf{\Omega}$ is $2\pi/86,164\text{s} = 7.292 \times 10^{-5}\text{s}^{-1}$, corresponding to a sidereal day of length 86,164 s. This is the actual period of the Earth's rotation (relative to the distant stars), which is slightly shorter that the "nominal" length of day, which is 86,400 s. Can you explain why the two numbers are different? In the discussion below, we assume that $\mathbf{\Omega}$ is independent of time, which is very nearly true.

Newton's laws of motion are assumed to hold in inertial reference frames (hereafter, inertial frames). An inertial frame is one that is not undergoing any acceleration. The principal of "Gallilean invariance" states that the laws of nature are the same in all inertial frames; in this sense, all inertial frames are equivalent. We will use the notation $D_a\mathbf{A}/Dt$ to denote the Lagrangian or particle-following" time-rate-of-change of an arbitrary vector $\mathbf{A}$ (i.e., *any* vector), as seen in as seen in an inertial frame. Newton's second law of motion says that a particle of air with density ρ undergoes an acceleration that satisfies

$$\rho \frac{D_a \mathbf{V}}{Dt} = -\rho \nabla \left(\Phi_{\overset{+}{\bigcirc}} + \Phi_{\text{ET}}\right) - \nabla p - \nabla \cdot \mathbf{S}\,, \tag{4.30}$$

where the terms on the right-hand side are contributions to the net force per unit volume acting on the air.

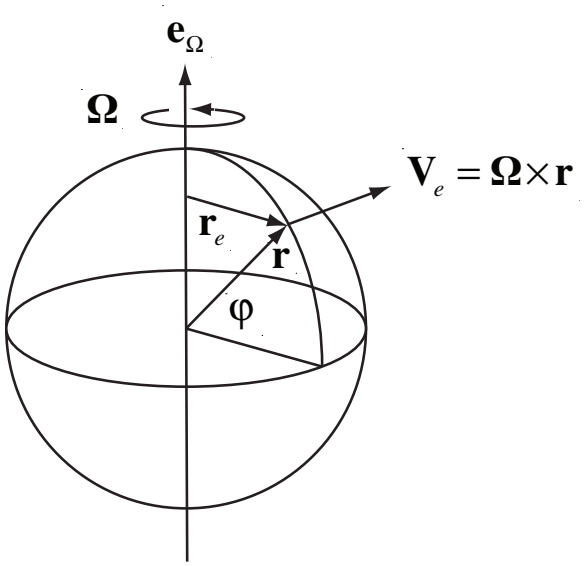


Figure 4.2: Sketch defining notation used in the text.

The first term on the right-hand side of (4.30) represents the effects of gravity. It is expressed here in terms of the gradient of the Earth's gravitational potential, $\Phi_{\text{♁}}$, and also the gradient of Φ_{ET}, a much smaller contribution from the combined gravitational fields of the Moon, the Sun, and other extraterrestrial bodies (see Chapter 10). The subscript ♁ means "Earth", and the subscript "ET" means "extraterrestrial." We include Φ_{ET} because, as discussed in Chapter 10, the locally time-dependent gravitational influences of the Moon, the Sun, and other extraterrestrial bodies can excite tides in both the atmosphere and oceans. Because the atmosphere is very thin compared to the distances between the Earth and gravitationally significant extraterrestrial bodies, the variation of Φ_{ET} with height can be neglected. The variations of Φ_{ET} with longitude, latitude, and time are, of course, key for the tides. Further discussion is given in Chapter 10.

The Earth's gravitational field is not completely uniform for a given distance from the center of the Earth because of small irregularities in the distribution of terrestrial mass. As a result, the vertical component of the force of gravity varies slightly from place to place, and there is a small gravitational force in the horizontal plane, which is measured by some satellite observing systems (e.g., Tapley et al., 2004).

In the second term of (4.30), $-\nabla p$ represents the pressure-gradient force per unit volume. The pressure-gradient vector points almost downward, but its relatively small horizontal components are of course critical for the horizontal winds.

The last term on the right-hand side of (4.30) represents the effects of momentum transports due to small-scale motions, which can be loosely described as "friction." It is written as the divergence of the stress tensor, $\mathbf{S}$ (see Appendix A). The dimensions of $\mathbf{S}$ are force per unit area, e.g., Nt m^{-2} (which is the same as Pa). Note that $\nabla\cdot\mathbf{S}$ is a vector. For

convenience, we will use the notation

$$\rho \mathscr{F}_{\mathbf{V}} \equiv \nabla \cdot \mathbf{S} \,, \tag{4.31}$$

where $\rho \mathscr{F}_{\mathbf{V}}$ is the three-dimensional "friction" vector.

We want to formulate our equations using a reference frame that rotates with the Earth, because that is what we experience as riders on the Earth. A rotating reference frame is not inertial. Let the Lagrangian time-rate-of-change of an arbitrary vector $\mathbf{A}$, as seen in the rotating frame, be denoted by $D\mathbf{A}/Dt$. The two time-rates-of-change are related by

$$\frac{D_a \mathbf{A}}{Dt} = \frac{D\mathbf{A}}{Dt} + \boldsymbol{\Omega} \times \mathbf{A} \,. \tag{4.32}$$

Keep in mind that Eq. (4.32) applies to *any* vector. Let $\mathbf{r}$ be a position vector that points from the Earth's center to a particle of air. See again Fig. 4.2. The "absolute" velocity of the particle, as seen in an inertial reference frame, is denoted by

$$\mathbf{V}_a \equiv \frac{D_a \mathbf{r}}{Dt} \,, \tag{4.33}$$

and the "relative" velocity of the particle as seen in the rotating frame is

$$\mathbf{V} \equiv \frac{D\mathbf{r}}{Dt} \,, \tag{4.34}$$

Applying (4.32) to $\mathbf{r}$, and using (4.33) and (4.34), we can write

$$\mathbf{V}_a = \mathbf{V} + \boldsymbol{\Omega} \times \mathbf{r} \,. \tag{4.35}$$

As an aside, we can define

$$\mathbf{V}_e \equiv \boldsymbol{\Omega} \times \mathbf{r} = (\Omega r \cos\varphi)\, \mathbf{e}_\lambda \tag{4.36}$$

where $\mathbf{e}_\lambda$ is a unit vector pointing east, φ is latitude, and $\mathbf{V}_e$ is the eastward velocity (as seen in the inertial frame) that a particle at radius $\mathbf{r}$ and latitude φ experiences due to the Earth's rotation (Fig. 4.2). At the equator, the magnitude of $\mathbf{V}_e$ is about a thousand miles per hour ($\cong 450$ m s^{-1}).

Next, we apply (4.32) to $\mathbf{V}_a$:

$$\frac{D_a \mathbf{V}_a}{Dt} = \frac{D\mathbf{V}_a}{Dt} + \boldsymbol{\Omega} \times \mathbf{V}_a \,. \tag{4.37}$$

We use (4.35) to eliminate $\mathbf{V}_a$ on the right-hand side of (4.37). Using (4.34) again, the result can be written as

$$\begin{aligned} \frac{D_a \mathbf{V}_a}{Dt} =& \frac{D}{Dt}\left(\mathbf{V} + \boldsymbol{\Omega} \times \mathbf{r}\right) + \boldsymbol{\Omega} \times \left(\mathbf{V} + \boldsymbol{\Omega} \times \mathbf{r}\right) \\ =& \frac{D\mathbf{V}}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{V} + \boldsymbol{\Omega} \times \left(\boldsymbol{\Omega} \times \mathbf{r}\right) \,. \end{aligned} \tag{4.38}$$

Here $2\boldsymbol{\Omega} \times \mathbf{V}$ represents the Coriolis acceleration, whose direction is perpendicular to both $\boldsymbol{\Omega}$ and $\mathbf{V}$, and $\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r})$ represents the centrifugal acceleration, whose direction is perpendicular to $\boldsymbol{\Omega}$.

We can now substitute (4.38) into (4.30) to express Newton's second law of motion in the rotating coordinate system:

$$\boxed{\rho \frac{D\mathbf{V}}{Dt} + 2\Omega \times \rho\mathbf{V} + \rho\boldsymbol{\Omega} \times \left(\boldsymbol{\Omega} \times \mathbf{r}\right) = -\rho\nabla\left(\Phi_{\overset{+}{\bigcirc}} + \Phi_{\text{ET}}\right) - \nabla p - \nabla \cdot \mathbf{S} \,.} \tag{4.39}$$

As shown in (4.38), the terms on the left-hand side of (4.39) sum to give the mass per unit volume times the acceleration as seen in an inertial reference frame.

4.4.2 "Apparent" gravity and hydrostatic balance

It is customary to simplify (4.39) as follows: The centrifugal acceleration can be written as

$$\begin{aligned} \boldsymbol{\Omega} \times \left(\boldsymbol{\Omega} \times \mathbf{r}\right) =& -\Omega^2 \mathbf{r}_e \\ =& -\nabla\left(\frac{1}{2}|\boldsymbol{\Omega} \times \mathbf{r}|^2\right) \,, \end{aligned} \tag{4.40}$$

where

$$\mathbf{r}_e \equiv -\mathbf{e}_\varphi \sin\varphi + \mathbf{e}_r \cos\varphi \tag{4.41}$$

is the vector shown in Fig. 4.2. In (4.41), $\mathbf{e}_\varphi$ and $\mathbf{e}_r$ are unit vectors pointing in the meridional and radial directions, respectively, so that $\mathbf{r}_e$ has both radial and meridional components that depend on latitude. The first line of (4.40) shows that the centrifugal acceleration is perpendicular to the axis of the Earth's rotation. The second line shows that it can be written as minus the gradient of a "centrifugal potential," which is given by $\frac{1}{2}|\boldsymbol{\Omega}\times\mathbf{r}|^2$. Eq. (4.40) allows us to define an "apparent" gravity that combines the effects of the Earth's true gravity and the centrifugal acceleration:

$$\phi \equiv \Phi_{\earth} - \frac{1}{2}|\boldsymbol{\Omega}\times\mathbf{r}|^2 \,, \tag{4.42}$$

Here ϕ is the "geopotential," which was already used in Chapter 3. Because of the centrifugal acceleration, and also because of spatial inhomogeneities of true gravity due to inhomogeneities of the Earth's mass distribution, the isosurfaces of the geopotential are not quite spherical. In particular, the centrifugal acceleration causes the geopotential isosurfaces (and the solid Earth itself) to bulge outward in low latitudes, so that the shape of a geopotential isosurface is (approximately) an "oblate spheroid," whose radius is about 20 km greater at the equator than at the poles; see Staniforth (2022) for further discussion.

We will use the shorthand notation

$$g \equiv \frac{\partial\phi}{\partial z} \,. \tag{4.43}$$

Although g decreases quadratically with distance from the Earth's center, such variations are usually ignored in discussions of the tropospheric and stratospheric circulations.

Using (4.40), (4.42), and (4.31), we can rewrite (4.39) in the relatively simple (and possibly familiar) form

$$\rho\frac{D\mathbf{V}}{Dt} + 2\boldsymbol{\Omega}\times\rho\mathbf{V} = -\rho\nabla\left(\phi+\Phi_{\text{ET}}\right) - \nabla p - \rho\mathscr{F}_{\mathbf{V}}. \tag{4.44}$$

An alternative form of (4.44) is

$$\rho\frac{\partial \mathbf{V}}{\partial t}+(\nabla\times\mathbf{V}+2\boldsymbol{\Omega})\times\rho\mathbf{V}+\rho\nabla K=-\rho\nabla\left(\phi+\Phi_{\rm ET}\right)-\nabla p-\rho\mathscr{F}_{\mathbf{V}}\,. \tag{4.45}$$

Here $\nabla\times\mathbf{V}$ is the three-dimensional vorticity vector, and

$$K\equiv\frac{1}{2}\mathbf{V}\cdot\mathbf{V} \tag{4.46}$$

is the kinetic energy per unit mass. To obtain (4.45) from (4.44), we used the vector identity (A.25). We will have occasion to use both (4.44) and (4.45).

The value of K depends on the choice of coordinate system. For example, the value of the kinetic energy at a given place in the atmosphere will differ a lot between a rotating frame and an inertial frame. Note, however, that (4.77) involves $\partial K/\partial t$ and ∇K, but not K itself. Since $\partial K/\partial t$ and ∇K take the same values in both inertial and rotating coordinate systems, Eq. (4.77) is Galilean invariant.

We can *define* exact hydrostatic balance as a motionless state in which the pressure is uniform along isosurfaces of the geopotential, so that

$$\nabla_\phi p=0\,, \tag{4.47}$$

and (4.44) reduces to

$$0=-\rho\nabla\phi-\nabla p\,. \tag{4.48}$$

4.4.3 The equations of horizontal and vertical motion

We conclude this section by separating the equation of three-dimensional motion, (4.44), into the equations for horizontal and vertical motion. The Coriolis term of (4.44) needs some discussion. As a first step, we separate $\boldsymbol{\Omega}$ into meridional (i.e., horizontal) and vertical vectors:

$$\boldsymbol{\Omega} = \boldsymbol{\Omega}_\varphi + \boldsymbol{\Omega}_z \,, \tag{4.49}$$

where

$$\boldsymbol{\Omega}_\varphi \equiv (\Omega \cos\varphi)\,\mathbf{e}_\varphi \quad \text{and} \quad \boldsymbol{\Omega}_z \equiv (\Omega \sin\varphi)\,\mathbf{e}_r \,. \tag{4.50}$$

Then we can write

$$\begin{aligned} \boldsymbol{\Omega} \times \mathbf{V} &= \left(\boldsymbol{\Omega}_\varphi + \boldsymbol{\Omega}_z\right) \times (\mathbf{v} + w\mathbf{e}_r) \\ &= \boldsymbol{\Omega}_\varphi \times (\mathbf{v} + w\mathbf{e}_r) + \boldsymbol{\Omega}_z \times \mathbf{v} \\ &= \left[\boldsymbol{\Omega}_\varphi \times (w\mathbf{e}_r) + \boldsymbol{\Omega}_z \times \mathbf{v}\right] + \boldsymbol{\Omega}_h \times \mathbf{v} \,. \end{aligned} \tag{4.51}$$

On the third line of (4.51), the quantity in square brackets is a horizontal vector, and the final term is a vertical vector that will go into the equation of vertical motion.

The second step is to separate the three-dimensional friction vector into horizontal and vertical parts. We write

$$\rho \mathscr{F}_\mathbf{V} = \rho \mathscr{F}_\mathbf{v} + \mathbf{e}_r \rho \mathscr{F}_w \,. \tag{4.52}$$

Here $\rho\mathscr{F}_\mathbf{v}$ is the (vector) friction acting on the horizontal wind, and $\rho\mathscr{F}_w$ is the friction acting on the vertical component of the wind vector.

Using these results, and starting from (4.14), we can separate Eq. (4.44) into

$$\rho \frac{D\mathbf{v}}{Dt} + 2\rho\left[\boldsymbol{\Omega}_\varphi \times (w\mathbf{e}_r) + \boldsymbol{\Omega}_z \times \mathbf{v}\right] = -\rho\nabla_z(\phi + \Phi_{\text{ET}}) - \nabla_z p - \rho\mathscr{F}_\mathbf{v} \,, \tag{4.53}$$

which governs the horizontal wind vector, and

$$\rho \frac{Dw}{Dt} + (2\rho\boldsymbol{\Omega}_h \times \mathbf{v}) \cdot \mathbf{e}_r = -\rho g - \frac{\partial p}{\partial z} - \rho\mathscr{F}_w \,, \tag{4.54}$$

which governs the vertical motion. It is conventional to neglect $\nabla_z \phi$ in (4.53), in which case it simplifies to

$$\rho \frac{D\mathbf{v}}{Dt} + 2\rho \left[\mathbf{\Omega}_\varphi \times (w\mathbf{e}_r) + \mathbf{\Omega}_z \times \mathbf{v} \right] = -\rho \nabla_z \Phi_{\text{ET}} - \nabla_z p - \rho \mathscr{F}_\mathbf{v} \,, \tag{4.55}$$

Approximate hydrostatic balance can be obtained by dropping the left-hand side and friction terms of (4.54):

$$0 = -\rho g - \frac{\partial p}{\partial z} \,. \tag{4.56}$$

Compare with (4.48).

Starting here, we will neglect the effects of Φ_{ET} until Chapter 10.

4.5 Conservation of thermodynamic energy

The internal energy of a perfect gas is given by

$$e = c_v T, \tag{4.57}$$

where c_v, the specific heat of air at constant volume, is a constant. For dry air, $c_v = 5R/2 \cong 713 \text{ J kg}^{-1} \text{ K}^{-1}$. More generally, the internal energy also includes the latent heat associated with the potential condensation of water vapor, so that for moist air

$$e \cong c_v T + L q_v \,. \tag{4.58}$$

In atmospheric science it is common to define the internal energy as $c_v T$, and to treat the latent heat as a source or sink of internal energy.

We could also include the latent heats of other atmospheric constituents, e.g., nitrogen, oxygen, and carbon dioxide, to represent the effects of their potential phase changes. We don't bother because (fortunately) those constituents don't change phase in the Earth's atmosphere.

Eq. (4.58) is approximate because we have neglected the specific heat of the water vapor, as well as the specific heat of any liquid (or ice) that might be present. Those minor corrections are now included in many numerical models of the atmosphere.

When thermodynamic energy is added to a system, the energy input equals the sum of the work done and the change in the internal energy:

$$\rho \frac{D}{Dt}(c_v T) + \frac{p}{\alpha}\frac{D\alpha}{Dt} = -\nabla \cdot (\mathbf{R} + \mathbf{F}_s) + \rho LC + \rho \delta \tag{4.59}$$

Here $\mathbf{F}_s$ is the vector flux of internal energy due to molecular diffusion; $\mathbf{R}$ is the vector flux of energy due to radiation (notice the notational conflict with the gas constant, which should not be confusing), L is the latent heat of water vapor, and C is the rate of condensation per unit mass. The dissipation rate appears in (4.59) as a source of internal energy, i.e., "frictional heating." Eq. (4.59) is a statement of the conservation of thermodynamic energy, applied to a moving particle of air. It is sometimes called the "First Law of Thermodynamics," although that terminology seems rather medieval.

The "expansion work" term of (4.59), $(p/\alpha)D\alpha/Dt$, represents the work done by the pressure force to make the specific volume increase. This is the work that you do when you blow up a balloon. The expansion work term will appear again later in the mechanical energy equation, but with the opposite sign. It therefore represents a rate of conversion between mechanical and internal energy. Because α would be a constant for an (idealized) incompressible fluid (e.g., "shallow water," which is discussed near the end of this chapter), the internal energy of an incompressible fluid is nonconvertible, although it is not zero. Because it is nonconvertible, the internal energy of an incompressible fluid plays no role in the energy cycle; it can be ignored.

For convenience, we now define the "heating" as

$$Q \equiv -\nabla \cdot (\mathbf{R} + \mathbf{F}_s) + \rho LC \; . \tag{4.60}$$

With this notation, (4.59) can be written more compactly as

$$\rho \frac{D}{Dt}(c_v T) + \frac{p}{\alpha}\frac{D\alpha}{Dt} = Q + \rho \delta \; . \tag{4.61}$$

An alternative statement of the conservation of thermodynamic energy is

$$\rho \frac{D}{Dt}\left(c_p T\right) = \rho\left(\omega\alpha\right) + Q + \rho\delta \ , \tag{4.62}$$

where

$$c_p = R + c_v \cong \ 1004 \text{ J kg}^{-1} \text{ K}^{-1} \ , \tag{4.63}$$

a nice round number that you can easily remember. Eq. (4.62) can be called the "enthalpy form" of the thermodynamic energy equation, where the enthalpy is defined by

$$c_v T + p\alpha = c_p T \ . \tag{4.64}$$

It turns out that for saturated air containing liquid water the enthalpy is (approximately) given by $c_p T - Ll$, where l is the liquid water mixing ratio (e.g., Iribarne and Godson, 1981).

A third form of the thermodynamic equation is

$$\rho \frac{Dp}{Dt} - \left(\frac{c_p}{c_v}\right) p \frac{D\rho}{Dt} = Q + \rho\delta \ . \tag{4.65}$$

This can be derived from (4.59) or (4.62) by using the equation of state and the continuity equation. We will not use (4.65) in this book, but it is included here for completeness.

The conservation of thermodynamic energy can also be expressed in terms of the potential temperature, which has been defined already in Chapter 3. We can show that

$$\rho c_p \left(\frac{T}{\theta}\right) \frac{D\theta}{Dt} = Q + \rho\delta \ . \tag{4.66}$$

In the absence of heating and dissipation, $D\theta/Dt = 0$, i.e., θ is conserved following a particle. This is one of the reasons that θ is a particularly useful quantity. We can relate T to θ using

$$c_p T = \Pi\theta \ , \tag{4.67}$$

where Π is the Exner function, which is defined by

$$\Pi \equiv c_p \left(\frac{p}{p_0} \right)^{\kappa} . \tag{4.68}$$

In summary, the thermodynamic energy equation can be expressed in the four alternative forms (4.59), (4.62), (4.65), and (4.66).

Finally, the conservation of latent energy is expressed by

$$\rho \frac{D}{Dt} (Lq_v) = -\nabla \cdot \left(L\mathbf{F}_{q_v} \right) - \rho LC . \tag{4.69}$$

4.6 Summary of the basic equations in height coordinates

The equations of horizontal and vertical motion, (4.53) and (4.54), are

$$\rho \frac{D\mathbf{v}}{Dt} + 2\rho \left[\mathbf{\Omega}_\varphi \times (w\mathbf{e}_r) + \mathbf{\Omega}_z \times \mathbf{v} \right] = -\nabla_z p - \rho \mathscr{F}_\mathbf{v} \tag{4.70}$$

and

$$\rho \frac{Dw}{Dt} + (2\rho \mathbf{\Omega}_h \times \mathbf{v}) \cdot \mathbf{e}_r = -\rho g - \frac{\partial p}{\partial z} - \rho \mathscr{F}_w . \tag{4.71}$$

The continuity equation is

$$\left(\frac{\partial \rho}{\partial t} \right)_z + \nabla_z \cdot (\rho \mathbf{v}) + \frac{\partial}{\partial z} (\rho w) = 0 . \tag{4.72}$$

The thermodynamic energy equation can be written as

$$\frac{D\theta}{Dt} \equiv \dot{\theta} = \frac{Q}{\rho \Pi} , \tag{4.73}$$

where Q is the heating rate per unit mass. The latent energy equation is

$$\rho \frac{D}{Dt}\left(Lq_v\right) = -\nabla \cdot \left(L\mathbf{F}_{q_v}\right) - \rho LC. \tag{4.74}$$

Finally, we include the prognostic equation for an arbitrary scalar A, which is

$$\frac{DA}{Dt} = S_A \ , \tag{4.75}$$

where S_A is the source of A per unit mass.

4.7 The hydrostatic and quasi-static approximations

The hydrostatic approximation is a simplified form of (4.71):

$$\frac{\partial p}{\partial z} = -\rho g \ . \tag{4.76}$$

With an appropriate boundary condition, (4.76) allows us to compute $p\left(z\right)$ from $\rho\left(z\right)$. Even when the air is moving, (4.76) gives a good approximation to $p\left(z\right)$, simply because Dw/Dt and the vertical component of the friction force are almost always very small compared to g. Eq. (4.76) as applied to moving air is called the hydrostatic approximation, and it is applicable to virtually all meteorological phenomena, even including violent thunderstorms.

For large-scale circulations, the approximate pressure determined through the use of (4.76) can be used to compute the pressure-gradient force in (4.70), the equation of horizontal motion. This is called the *quasi-static* approximation. It has more limited applicability than the hydrostatic approximation. The quasi-static approximation is accurate for large-scale motions, but it introduces significant errors for small-scale motions, such as thunderstorms and gravity waves. When the quasi-static approximation is used, the effective kinetic energy is due to the horizontal wind only; the contribution of the vertical component, w, is neglected. For large-scale motions, $|w| \ll (|u|, |v|)$, so that this quasi-static kinetic energy is very close to the true kinetic energy. Further discussion is given by Holton (1973).

When the equation of vertical motion is replaced by (4.76), there are two fundamental changes. First, we can no longer use the equation of vertical motion to determine the

vertical motion. The method that is actually used to obtain the vertical motion depends on the vertical coordinate system used, which is expected because the actual *meaning* of the vertical motion also depends on the vertical coordinate system. For example, in the case of pressure coordinates, the vertical motion is $\omega \equiv Dp/Dt$, and we can determine ω using the continuity equation with the upper boundary condition $\omega = 0$ at $p = 0$. In this book, we use several different vertical coordinate systems, and we will determine the vertical motion as needed, on a case-by-case basis.

The second fundamental change that comes with use of (4.76) is that there is only room to predict one (three-dimensional) thermodynamic variable, whereas with the full equation set we can and must predict two. For example, if the density is predicted then (4.76) can be used to determine the pressure by vertical integration from $p = 0$ at the top of the atmosphere, and at that point the ideal gas law can be used to determine the temperature. There is no room to predict both density and temperature.

The quasi-static approximation filters vertically propagating sound waves. That is a good thing, because vertically propagating sound waves have no perceptible effects on weather or climate, and can cause practical difficulties in numerical models; for most purposes, it is better to eliminate them (e.g., Arakawa and Konor, 2009).

4.8 Conservation of mechanical energy

We continue in this section without using the quasi-static approximation.

Mechanical energy is the sum of kinetic and potential energy. The kinetic energy equation can be derived by dotting $\mathbf{V}$ with the momentum equation in the form (4.45), giving

$$\rho \left[\frac{\partial K}{\partial t} + (\mathbf{V} \cdot \nabla) K \right] = -\rho \mathbf{V} \cdot \nabla \phi - \mathbf{V} \cdot \nabla p - \rho \mathbf{V} \cdot \mathscr{F}_{\mathbf{V}} \,, \tag{4.77}$$

or

$$\rho \frac{DK}{Dt} = -\rho \mathbf{V} \cdot \nabla \phi - \mathbf{V} \cdot \nabla p - \mathbf{V} \cdot \rho \mathscr{F}_{\mathbf{V}} \,. \tag{4.78}$$

We now add $\rho D\phi/Dt$ to both sides of (4.78):

$$\rho \frac{D}{Dt} (K + \phi) = -\mathbf{V} \cdot \nabla p - \mathbf{V} \cdot \rho \mathscr{F}_{\mathbf{V}} \,. \tag{4.79}$$

This is a form of the mechanical energy equation. The sum $K+\phi$ is called the mechanical energy per unit mass. In the absence of forces, i.e., if the terms on the right-hand side are set to zero, Eq. (4.79) reduces to the statement that the sum of the kinetic and potential energies per unit mass does not change following a particle. This may remind you of a high-school physics scenario: A roller coaster gains kinetic energy and loses potential energy as it rolls downhill, then gains potential energy and loses kinetic energy as it coasts uphill. The sum of the roller coaster's kinetic and potential energies does not change as it rolls up and down the tracks.

On the right-hand side of (4.78), $-\mathbf{V}\cdot\nabla p$ represents the rate at which work is done by the pressure force, per unit mass. This expression can be manipulated as follows:

$$\begin{aligned} -\mathbf{V}\cdot\nabla p &= -\nabla\cdot(p\mathbf{V}) + p\nabla\cdot\mathbf{V} \\ &= -\nabla\cdot(p\mathbf{V}) + \frac{p}{\alpha}\frac{D\alpha}{Dt}. \end{aligned} \tag{4.80}$$

On the second line of (4.80), we have used the continuity equation in the form (4.10) to eliminate $\nabla\cdot\mathbf{V}$. The $\nabla\cdot(p\mathbf{V})$ term of (4.80) has the form of a flux divergence; it represents a spatial redistribution of energy by the pressure force. The expansion work term, $(p/\alpha)D\alpha/Dt$, was encountered earlier in the thermodynamic energy equation, (4.59). It represents conversion between thermodynamic and mechanical energy, so it has no effect on their sum.

Referring back to (4.81), the friction term of (4.78) can also be expanded to reveal two physically distinct parts, as follows:

$$\mathbf{V}\cdot\rho\mathscr{F}_{\mathbf{V}} = \mathbf{V}\cdot(\nabla\cdot\mathbf{S}) = \nabla\cdot(\rho\boldsymbol{\Lambda}) + \rho\delta\,. \tag{4.81}$$

Here

$$\rho\boldsymbol{\Lambda} \equiv \mathbf{V}\cdot\mathbf{S} \tag{4.82}$$

is the three-dimensional (vector) flux of energy due to frictional work, and

$$\rho\delta \equiv -(\mathbf{S}\cdot\nabla)\cdot\mathbf{V} \tag{4.83}$$

is the (scalar) rate of kinetic energy dissipation per unit volume. The flux divergence $\nabla \cdot (\rho \mathbf{\Lambda})$ in (4.81) represents a spatial redistribution of kinetic energy as neighboring air parcels to do work on each other through shearing stresses. Because this is merely a spatial redistribution of energy, it does not change the total amount of kinetic energy in the atmosphere, except where friction does work on the boundary. In contrast, the dissipation rate, $\rho\delta$, is a true sink of kinetic energy. For reasons discussed in Appendix E,

$$\delta \geq 0 \,. \tag{4.84}$$

This means that dissipation always tends to reduce the kinetic energy.

The final term on the right-hand side of (4.79) is $(\partial\phi/\partial t)_z$. We normally think of the gravitational potential as being independent of time, in which case the term could simply be dropped.

Substitution of (4.80) and (4.81) into (4.79) gives the mechanical energy equation in the form

$$\rho \frac{D}{Dt}(K+\phi) = -\nabla \cdot (p\mathbf{V}) + \frac{p}{\alpha}\frac{D\alpha}{Dt} - \nabla \cdot (\rho \mathbf{\Lambda}) - \rho\delta \,. \tag{4.85}$$

Using the equation of state, we can write

$$\begin{aligned} p\frac{D\alpha}{Dt} &= \frac{D}{Dt}(p\alpha) - \alpha\frac{Dp}{Dt} \\ &= \frac{D}{Dt}(p\alpha) - \omega\alpha \\ &= \frac{D}{Dt}(RT) - \omega\alpha \,, \end{aligned} \tag{4.86}$$

This shows that the expansion-work term is closely related to the product $\omega\alpha$. Substituting (4.86) into (4.85), we find that

$$\rho \frac{D}{Dt}(K+\phi-RT) = -\nabla \cdot (p\mathbf{V} + \rho \mathbf{\Lambda}) - \rho(\omega\alpha) - \rho\delta \,. \tag{4.87}$$

4.9 Where does mechanical energy come from?

Using continuity, the mechanical energy equation (4.85) can be converted to flux form:

$$\frac{\partial}{\partial t}\left[\rho\left(K+\phi\right)\right]+\nabla\cdot\left[\rho\mathbf{V}\left(K+\phi\right)+p\mathbf{V}+\rho\mathbf{\Lambda}\right]=\frac{p}{\alpha}\frac{D\alpha}{Dt}-\rho\delta\ . \tag{4.88}$$

The flux-divergence term on the left-hand side of (4.88) represents transport processes, which merely redistribute energy in space. The expansion work term can be either positive or negative at a given place and time. Recall, however, that the dissipation term is always a sink. It follows that, in an average over the whole atmosphere, and over time, the expansion work term must be positive, i.e., it must act as a source of mechanical energy:

$$\int_V \overline{\frac{p}{\alpha}\frac{D\alpha}{Dt}}^t\, dV=\int_V \overline{\rho\delta}^t dV\geq 0. \tag{4.89}$$

Here we are neglecting effects of the extraterrestrial gravitational forcing. Eq. (4.89) means that, on the average, the pressure force must do positive expansion work to compensate for the dissipation of kinetic energy. Expansion work is a source of mechanical energy that, on the average, balances the sink due to dissipation. It can have either sign locally but must be positive when averaged over the whole atmosphere and over time. To make $\int_V \overline{\frac{p}{\alpha}\frac{D\alpha}{Dt}}^t dV$ positive, expansion ($D\alpha/Dt>0$) must take place, on the average, where p/α is large, and compression ($D\alpha/Dt<0$) must take place, on the average, where p/α is small. For example, the average rate of expansion could be more vigorous in the lower troposphere where p/α is large, and the average rate of compression could be more vigorous in the upper troposphere where p/α is small.

Similarly, in an average over the whole atmosphere, and over time, the $-\omega\alpha$ term of (4.87) must be positive, i.e., it must act as a source of mechanical energy that balances the sink due to dissipation:

$$-\int_V \overline{\rho\left(\omega\alpha\right)}^t dV=\int_V \overline{\rho\delta}^t dV\geq 0\ . \tag{4.90}$$

Comparison of (4.90) and (4.89) shows that

$$-\int_V \overline{\rho\,(\omega\alpha)}^t\,dV = \int_V \overline{\frac{p}{\alpha}\frac{D\alpha}{Dt}}^t\,dV\ . \tag{4.91}$$

We can therefore use $\omega\alpha$ and $(p/\alpha)D\alpha/Dt$ as alternative measures of the rate of conversion between mechanical and thermodynamic energy. Most of the time we will use $\omega\alpha$.

We show below that the mechanical energy is supplied by conversion from the thermodynamic energy of the atmosphere. Given that the dissipation term of (4.88) represents a sink of mechanical energy, we should ask where the energy goes. As discussed below, it appears as a source of thermodynamic energy. Dissipation is, therefore, just another energy conversion process, but it is special because it runs in only one direction.

As mentioned earlier, the globally averaged rates of kinetic energy dissipation and mechanical energy generation have been estimated to be about 5 W m^{-2}. Recall that the solar radiation absorbed by the Earth-atmosphere system is about 240 W m^{-2}. Evidently the climate system is not very efficient at converting the absorbed solar energy into atmospheric mechanical energy.

4.10 Conservation of total energy

When we add (4.59), (4.85), and (4.69) the result is

$$\boxed{\rho\frac{De_T}{Dt} = -\nabla\cdot(p\mathbf{V} + \rho\mathbf{\Lambda} + \mathbf{R} + \mathbf{F}_h)} \tag{4.92}$$

where

$$e_T \equiv K + \phi + c_v T + Lq_v \tag{4.93}$$

is the total energy (per unit mass) and $\mathbf{F}_h$ is the sum of the molecular fluxes of sensible heat (i.e., temperature) and latent heat. $\mathbf{F}_h$ is also called the molecular flux of "moist static energy;" the reason for this terminology will be explained later. The "expansion work", condensation, and dissipation terms of (4.59), (4.85), and (4.69) have all cancelled, because they represent *conversions* between the thermodynamic, mechanical, and latent energies. Those conversions have no effect on the total energy.

The continuity equation can be used to rewrite (4.92) in flux form:

$$\frac{\partial}{\partial t}(\rho e_T) + \nabla \cdot [\rho \mathbf{V}(K + \phi + c_v T + L q_v) + p\mathbf{V} + \rho \mathbf{\Lambda} + \mathbf{R} + \mathbf{F}_h] = 0 \,. \tag{4.94}$$

The only terms in this equation are a time tendency, a flux divergence, and a (usually neglected) term representing gravitational exchanges of energy with other celestial bodies.

4.11 Static energies

Now add $\partial p/\partial t$ to both sides of (4.94), and then use the equation of state and (4.63)

to obtain

$$\frac{\partial}{\partial t}[\rho(K + \phi + c_p T + L q_v)] + \nabla \cdot [\rho \mathbf{V}(K + \phi + c_p T + L q_v) + \mathbf{R} + \mathbf{F}_h + \rho \mathbf{\Lambda}] = \frac{\partial p}{\partial t} \,. \tag{4.95}$$

Take a moment to compare (4.94) and (4.95). The advected quantity in (4.94) is the total energy $K + \phi + c_v T + L q_v$, while the advected quantity in (4.95) is $K + \phi + c_p T + L q_v$. These two quantities are not even approximately the same, because c_p is quite a bit larger than c_v. The transport term associated with $\mathbf{V}$ in (4.94) is $\rho \mathbf{V}(K + \phi + c_v T + L q_v) + p\mathbf{V}$, so it includes both advection of total energy and pressure-work. In (4.95) the transport term is $\rho \mathbf{V}(K + \phi + c_p T + L q_v)$, so it represents only advection. In view of the equation of state, *these two transport terms are actually equal to each other*. On the left-hand side of (4.95) the quantity under the time rate of change is the same as the quantity that is advected by the wind. On the other hand, the $\partial p/\partial t$ term appears on the right-hand side. Recall, however, that $\partial p/\partial t$ becomes negligible in a sufficiently long time average, and can be negligible even instantaneously under some conditions.

Eq. (4.95) is exact. We now make some approximations. An air parcel zipping along at a rather extreme 100 m s^{-1} has a kinetic energy per unit mass of 5 x 10^3 J kg^{-1}, and of course a parcel moving at a more typical 10 m s^{-1} has a kinetic energy 100 times smaller. For comparison, a parcel on the 200 hPa surface has a potential energy per unit mass (relative to sea level) of about 1.2 x 10^5 J kg^{-1}. With a temperature of 200 K a parcel's internal energy per unit mass is about 1.5 x 10^5 J kg^{-1}. These examples illustrate that the contribution of the kinetic energy to the total energy is usually negligible. In addition, the friction and pressure-tendency terms of (4.95) can often be neglected. With these simplifying approximations, (4.95) reduces to

$$\frac{\partial}{\partial t}(\rho h) + \nabla \cdot (\rho \mathbf{V} h + \mathbf{R} + \mathbf{F}_h) \cong 0, \tag{4.96}$$

where

$$\boxed{h \equiv c_p T + \phi + L q_v} \tag{4.97}$$

is called the moist static energy (Chavas and Peters, 2023). According to (4.96), the moist static energy is approximately conserved under both moist adiabatic and dry adiabatic processes. Since precipitation does not affect the water vapor mixing ratio, the moist static energy is approximately conserved even when precipitation is occurring. Our earlier comparison of (4.95) and (4.94) shows that although the moist static energy is not even approximately equal to the total energy, the *transport* of moist static energy is good approximation to the *transport* of total energy. It is sometimes claimed that conservation of moist static energy is based in part on the hydrostatic approximation, but the hydrostatic approximation was not used in the derivation of (4.96).

The discussion above shows that conservation of moist static energy is an approximate form of the total energy equation, rather than the thermodynamic energy equation. That is why there is no dissipation term in (4.96); such a term would of course appear in any version of the thermodynamic energy equation, although we might justify neglecting it under some conditions.

Because the water vapor mixing ratio, q_v, is conserved under dry adiabatic processes, conservation of moist static energy implies that the dry static energy,

$$\begin{aligned} s \equiv & c_p T + \phi \\ = & h - L q_v \end{aligned} \tag{4.98}$$

is also approximately conserved under dry adiabatic processes. The dry static energy normally increases upward in the atmosphere, because, as you will show when you work the problems at the end of this chapter, the rate of change of the dry static energy with height has the same sign as the rate of change of the potential temperature with height.

The moist and dry static energies are used extensively later in this book, and are also very widely used in the research literature. The vertical and geographical distributions of the moist static energy are discussed in Chapters 6 and 8.

4.12 The equations of motion in spherical coordinates

The velocity vector can be expressed in terms of zonal, meridional, and radial components:

$$\mathbf{V} \equiv u\mathbf{e}_\lambda + v\mathbf{e}_\varphi + w\mathbf{e}_r \ . \tag{4.99}$$

Here the velocity components are *defined* by

$$u \equiv r\cos\varphi \frac{D\lambda}{Dt} \ , \tag{4.100}$$

$$v \equiv r\frac{D\varphi}{Dt} \ , \tag{4.101}$$

$$w \equiv \frac{Dr}{Dt} = \frac{Dz}{Dt} \ . \tag{4.102}$$

The Lagrangian time derivative can be written as

$$\begin{aligned} \frac{D}{Dt} \equiv & \frac{\partial}{\partial t} + \frac{D\lambda}{Dt}\frac{\partial}{\partial\lambda} + \frac{D\varphi}{Dt}\frac{\partial}{\partial\varphi} + \frac{Dr}{Dt}\frac{\partial}{\partial r} \\ = & \frac{\partial}{\partial t} + \frac{u}{r\cos\varphi}\frac{\partial}{\partial\lambda} + \frac{v}{r}\frac{\partial}{\partial\varphi} + w\frac{\partial}{\partial r} \ . \end{aligned} \tag{4.103}$$

The directions in which the unit vectors $\mathbf{e}_\lambda$, $\mathbf{e}_\varphi$, and $\mathbf{e}_r$ actually point depend on longitude and latitude. Because of this, the directions of the unit vectors change as an air particle moves from place to place. An expanded version of the acceleration term of (4.44) therefore has a total of six terms:

$$\frac{D\mathbf{V}}{Dt} = \left(\frac{Du}{Dt}\mathbf{e}_\lambda + u\frac{D\mathbf{e}_\lambda}{Dt}\right) + \left(\frac{Dv}{Dt}\mathbf{e}_\varphi + v\frac{D\mathbf{e}_\varphi}{Dt}\right) + \left(\frac{Dw}{Dt}\mathbf{e}_r + w\frac{D\mathbf{e}_r}{Dt}\right). \tag{4.104}$$

Geometrical reasoning leads to the following formulae:

$$\begin{aligned}\frac{D\mathbf{e}_\lambda}{Dt} &= \frac{D\lambda}{Dt}\sin\varphi\mathbf{e}_\varphi - \cos\varphi\frac{D\lambda}{Dt}\mathbf{e}_r \\ &= -\left(\frac{u\tan\varphi}{r}\right)\mathbf{e}_\varphi - \frac{u}{r}\mathbf{e}_r,\end{aligned} \tag{4.105}$$

$$\begin{aligned}\frac{D\mathbf{e}_\varphi}{Dt} &= -\frac{D\lambda}{Dt}\sin\varphi\mathbf{e}_\lambda - \frac{D\varphi}{Dt}\mathbf{e}_r \\ &= \left(\frac{u\tan\varphi}{r}\right)\mathbf{e}_\lambda - \frac{v}{r}\mathbf{e}_r,\end{aligned} \tag{4.106}$$

$$\begin{aligned}\frac{D\mathbf{e}_r}{Dt} &= \cos\varphi\frac{D\lambda}{Dt}\mathbf{e}_\lambda + \frac{D\varphi}{Dt}\mathbf{e}_\varphi \\ &= \frac{u}{r}\mathbf{e}_\lambda + \frac{v}{r}\mathbf{e}_\varphi.\end{aligned} \tag{4.107}$$

You should try to understand where (4.105)-(4.107) come from; one approach is to visualize the movement and rotation of the unit vectors using a pencil and a globe. When (4.105)-(4.107) are taken into account, (4.44) can be expanded into component form as

$$\rho\left(\frac{Du}{Dt} + \frac{uw}{r} - \frac{uv\tan\varphi}{r} - fv + \overline{f}w\right) = -\frac{1}{r\cos\varphi}\frac{\partial p}{\partial\lambda} - (\rho\mathscr{F}_\mathbf{V})_\lambda\ , \tag{4.108}$$

$$\rho\left(\frac{Dv}{Dt} + \frac{vw}{r} + \frac{u^2\tan\varphi}{r} + fu\right) = -\frac{1}{r}\frac{\partial p}{\partial\varphi} - (\rho\mathscr{F}_\mathbf{V})_\varphi\ , \tag{4.109}$$

$$\rho\left[\frac{Dw}{Dt} - \left(\frac{u^2+v^2}{r}\right) - \overline{f}u\right] = -\rho g - \frac{\partial p}{\partial r} - (\rho\mathscr{F}_\mathbf{V})_r\ . \tag{4.110}$$

Here we have used the notation

$$f \equiv 2\Omega \sin\varphi \,, \tag{4.111}$$

and

$$\overline{f} \equiv 2\Omega \cos\varphi \,. \tag{4.112}$$

and we neglect variations of ϕ with longitude and latitude. Eqs. (4.108) - (4.110) are the components of the equation of motion in spherical coordinates. The terms uw/r, $uv\tan\varphi/r$, vw/r, $u^2\tan\varphi/r$, and $\left(u^2+v^2\right)/r$ arise from the spatial variations of the unit vectors in spherical coordinates. They are called "metric terms."

Finally, by using the continuity equation, we can rewrite the three component equations in their flux forms:

$$\frac{\partial}{\partial t}(\rho u) + \nabla\cdot(\rho \mathbf{V} u) + \rho\left(\frac{uw}{r} - \frac{uv\tan\varphi}{r} - fv + \overline{f}w\right) = -\frac{1}{r\cos\varphi}\frac{\partial p}{\partial\lambda} - (\rho\mathscr{F}_{\mathbf{V}})_\lambda \,, \tag{4.113}$$

$$\frac{\partial}{\partial t}(\rho v) + \nabla\cdot(\rho \mathbf{V} v) + \rho\left(\frac{vw}{r} - \frac{u^2\tan\varphi}{r} + fu\right) = -\frac{1}{r}\frac{\partial p}{\partial\varphi} - (\rho\mathscr{F}_{\mathbf{V}})_\varphi \,, \tag{4.114}$$

$$\frac{\partial}{\partial t}(\rho w) + \nabla\cdot(\rho \mathbf{V} w) - \rho\left(\frac{u^2+v^2}{r} - \overline{f}u\right) = -\rho g - \frac{\partial p}{\partial r} - (\rho\mathscr{F}_{\mathbf{V}})_r \,. \tag{4.115}$$

4.13 Conservation of angular momentum

As discussed in introductory physics textbooks, the angular momentum per unit volume of a particle, with respect to some origin where $\mathbf{r} = 0$, is a vector, $\mathbf{L}$, given by the cross

product of the particle's three-dimensional linear momentum vector per unit volume, $\rho\mathbf{V}$, and the displacement vector, $\mathbf{r}$, that separates the particle from the origin:

$$\mathbf{L} \equiv \mathbf{r} \times \rho\mathbf{V} \, . \tag{4.116}$$

The angular momentum of the Earth's atmosphere, with respect to an origin at the center of the Earth, could be computed by applying (4.116) to each air parcel, and integrating over the entire atmosphere. The result would be a vector.

The angular momentum of the air is mostly due to the rotation of the Earth, rather than the motion of the air relative to the Earth. The main reason for this is that (except near the poles) the air is moving slowly relative to the rapidly rotating Earth. In addition, as we have already discussed, the motion of the air relative to the Earth includes strong *mostly zonal* jets. For these reasons, the angular momentum vector for the atmosphere as a whole points roughly at the "North Star," i.e., along the Earth's axis of rotation. In practice, when we discuss the angular momentum of the atmosphere, we are almost always concerned not with the angular momentum vector, but with the *component* of that vector that is parallel to the axis of the Earth's rotation. That component is given by

$$\boxed{M \equiv (\Omega r \cos\varphi + u)\, r \cos\varphi} \, . \tag{4.117}$$

Because $\Omega r \cos\varphi$ is positive and very large compared to u (except near the poles), $M > 0$ almost always and everywhere.

Taking the Lagrangian time derivative of (4.117), and using (4.101), we find that

$$\frac{DM}{Dt} = r\cos\varphi \frac{Du}{Dt} + (2\Omega r \cos\varphi + u)\,(-v\sin\varphi + w\cos\varphi) \, . \tag{4.118}$$

Both the meridional and vertical wind components appear on the right-hand side of (4.118). Eliminating Du/Dt using (4.108), we find that *the Coriolis and metric terms all cancel*, and (4.118) simplifies to

$$\boxed{\rho \frac{DM}{Dt} = -\frac{\partial p}{\partial \lambda} - r\cos\varphi \;\; (\rho \mathscr{F}_{\mathbf{v}})_{\lambda}} \, . \tag{4.119}$$

This shows that M evolves according to a rule that is much, much simpler than (4.108), which is the corresponding rule for u. This simplicity is the reason that angular momentum is a useful concept.

4.14 Potential temperature as a vertical coordinate

4.14.1 Introduction

The most basic requirement for a variable to be used as a vertical coordinate is that it must vary monotonically with height. Even this requirement can be relaxed; e.g., a vertical coordinate can be independent of height over some layer of the atmosphere, provided that the layer is not too deep. We have already used both height coordinates and pressure coordinates in this book. With only minor exceptions, θ increases upward throughout the atmosphere, so it can also be used as a vertical coordinate. Potential temperature coordinates, also called isentropic coordinates, have highly useful properties that have been recognized for many years (e.g., Rossby and Collaborators, 1937; Namias and Stone, 1940; Starr, 1945; Eliassen and Raustein, 1968; Bleck, 1973; Johnson and Uccellini, 1983; Hoskins et al., 1985; Hsu and Arakawa, 1990; Benjamin et al., 2004). Isentropic coordinates bring a conceptual simplicity to both observational and modeling studies.

The most important reason for this simplicity is that, in the absence of heating, θ is conserved following a particle. This means that the vertical motion in θ-coordinates, i.e., $D\theta/Dt$, is proportional to the heating rate. In the absence of heating, there is "no vertical motion," in the sense that a particle that is on a given θ-surface remains on that surface. Partly for this reason, θ-coordinates provide an especially simple pathway for deriving and understanding many important results, including the conservation of Ertel potential vorticity (see section 4.15.5) and the interactions between eddies and the mean flow (Chapter 11).

A complication arises: What about θ-surfaces that intersect the Earth's surface in the given state? Following Lorenz (1955), these are treated as though they continue along the surface, as shown in Fig. 4.3, These "massless layers" between the isentropic surfaces that are following the ground contain have no effect on the physics. Where a θ-surface intersects the Earth's surface, the pressure is $p = p_S$. As discussed below, Lorenz's concept of massless layers makes it relatively easy to deal with the lower boundary condition on the atmospheric circulation.

4.14.2 The isentropic system

The starting point for a derivation of the basic equations as expressed in θ coordinates is the basic equations in height coordinates, which are summarized in Section 4.6. Appendix C shows the details of the derivation. We will start, however, by introducing a couple of key ideas here in the main text.

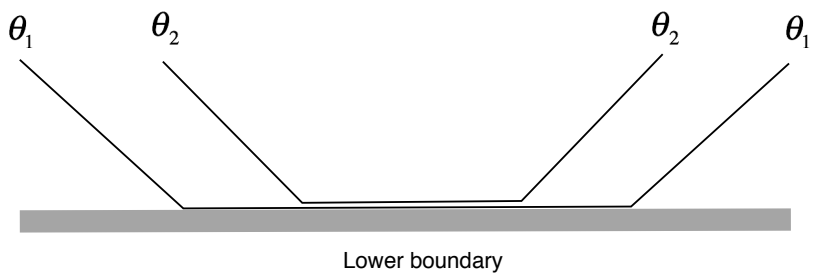


Figure 4.3: A sketch illustrating Lorenz's concept of "massless layers."

The isentropic pseudodensity is defined by

$$\rho_\theta \equiv \rho \frac{\partial z}{\partial \theta} \,. \tag{4.120}$$

It is the amount of mass per unit horizontal area between two θ surfaces that are separated by $d\theta$, just as ρ is the amount of mass per unit horizontal area between two height surfaces that are separated by dz. The spatial distribution of the zonally averaged pseudodensity is shown in Fig. 4.4. It generally decreases upward, but it is zero (shown by white) close to the surface in the tropics, because there is no air with small values of θ in the tropics.

We show in Appendix C that basic equations in height coordinates, which are summarized in Section 4.6, can be transformed into the isentropic system of equations, which are shown below:

$$\frac{D\mathbf{v}}{Dt} + 2\left[\boldsymbol{\Omega}_\varphi \times (w\mathbf{e}_r) + \boldsymbol{\Omega}_z \times \mathbf{v}\right] = -\left(\nabla_\theta \phi - g\nabla_\theta z\right) - \left[\nabla_\theta s - \frac{\partial \theta}{\partial z}\left(\frac{\partial s}{\partial \theta} - \Pi\right)\nabla_\theta z\right] - \mathscr{F}_\mathbf{v} \,, \tag{4.121}$$

$$\frac{Dw}{Dt} + 2\left(\boldsymbol{\Omega}_h \times \mathbf{v}\right) \cdot \mathbf{e}_r = -\frac{\partial \theta}{\partial z}\left(\frac{\partial s}{\partial \theta} - \Pi\right) - \mathscr{F}_w \,. \tag{4.122}$$

$$\rho\left[\left(\frac{\partial z}{\partial t}\right)_\theta + \mathbf{v} \cdot \nabla_\theta z - w\right] = -\rho_\theta \dot{\theta} \,, \tag{4.123}$$

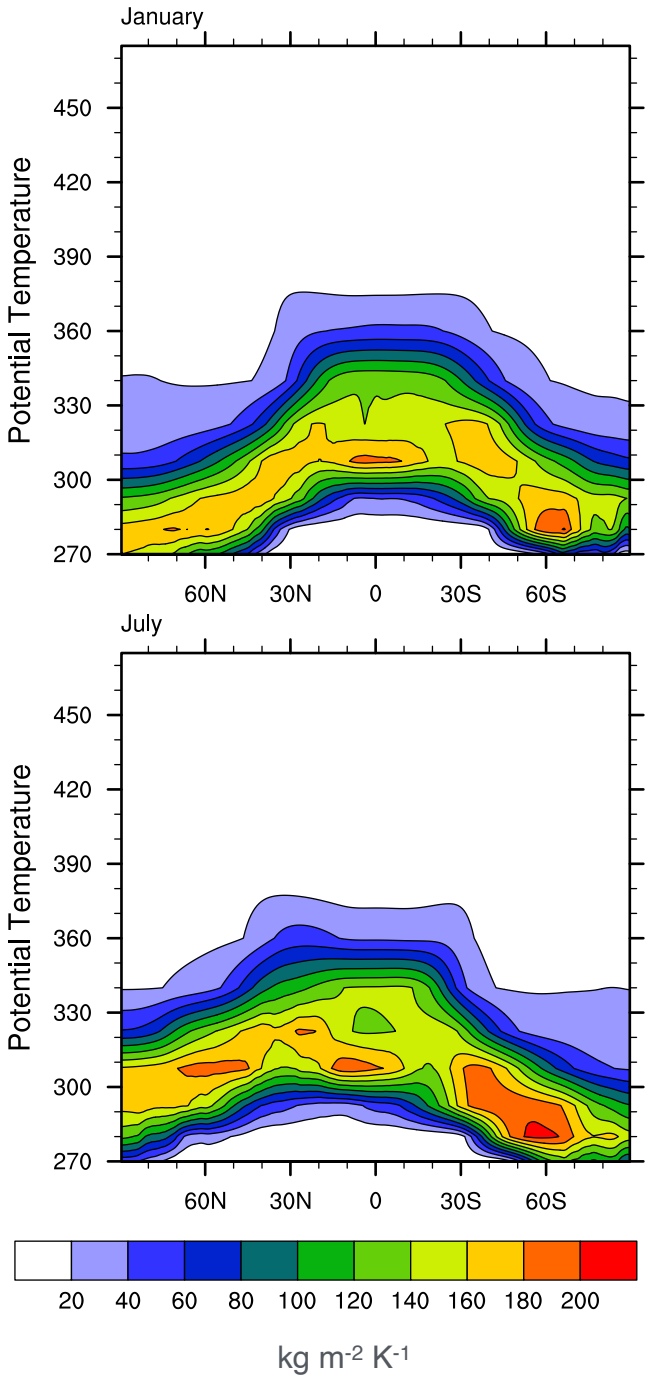


Figure 4.4: The observed zonally averaged isentropic pseudodensity for January (top) and July (bottom). White denotes values close to or equal to zero. White appears near the bottom of the figure because there are no cold values of θ in the tropics.

$$\left(\frac{\partial \rho_\theta}{\partial t}\right)_\theta + \nabla_\theta \cdot (\rho_\theta \mathbf{v}) + \frac{\partial}{\partial \theta}\left(\rho_\theta \dot{\theta}\right) = 0\,, \tag{4.124}$$

and

$$\left[\frac{\partial}{\partial t}\left(\rho_\theta A\right)\right]_\theta + \nabla_\theta \cdot \left(\rho_\theta \mathbf{v} A\right) + \frac{\partial}{\partial \theta}\left(\rho_\theta \dot{\theta} A\right) = \rho_\theta S_A \,, \tag{4.125}$$

Not counting the scalar A, the prognostic variables of the θ-coordinate system are $\mathbf{v}$, w, ρ_θ, and z. The corresponding prognostic variables of the z-coordinate system are $\mathbf{v}$, w, ρ, and θ. In passing from height coordinates to θ coordinates, we trade prediction of θ for prediction of the heights of the θ surfaces.

It remains to determine Π and s. We can use the predicted values of ρ_θ and $z(\theta)$ in (4.120) to find ρ. Knowing ρ, we can determine Π from the equation of state in the form:

$$\Pi = c_p \left(\frac{\rho R \theta}{p_0}\right)^{\frac{\kappa}{1-\kappa}} \,. \tag{4.126}$$

Finally, s can be obtained using

$$s = \Pi\theta + gz \,. \tag{4.127}$$

The pressure and temperature can easily be diagnosed from Π and θ.

4.14.3 The isentropic system with the quasi-static approximation

The equation of vertical motion, (C.14), reduces to the hydrostatic equation in the form

$$\frac{\partial s}{\partial \theta} - \Pi = 0 \,. \tag{4.128}$$

With the hydrostatic approximation, the pseudo-density can be written as

$$\rho_\theta = -\frac{1}{g}\frac{\partial p}{\partial \theta} \,. \tag{4.129}$$

Given a boundary condition, this can be used to compute $p(\theta)$. Once the pressure is known, we can obtain Π from its definition, and the temperature can then be calculated from θ. Then the dry static energy can be obtained by vertically integrating (4.128), starting from the lower boundary condition

$$s = gz_S \text{ where } \theta = 0\,. \tag{4.130}$$

The height of each θ-surface can be obtained from s and the temperature, so that (C.23) is not needed and should not be used. Recall that with height coordinates, use of the quasi-static system means that we can only predict one thermodynamic variable, so that prediction of both $\rho(z)$ and $\theta(z)$ is not allowed. This is a similar situation. We cannot predict both $\rho_\theta(\theta)$ and $z(\theta)$. Predicting $z(\theta)$ is also ruled out because the equation of vertical motion has been replaced by $eq : hydrostaticsthetacoordinate$, so we don't know $w(\theta)$.

When the quasi-static approximation is used, Eqs. (4.121) and (4.122) simplify to

$$\frac{D\mathbf{v}}{Dt} + 2\left[\boldsymbol{\Omega}_\varphi \times (w\mathbf{e}_r) + \boldsymbol{\Omega}_z \times \mathbf{v}\right] = -\left(\nabla_\theta \phi - g\nabla_\theta z\right) - \nabla_\theta s - \mathscr{F}_\mathbf{v} \tag{4.131}$$

and

$$0 = \frac{\partial s}{\partial \theta} - \Pi\,, \tag{4.132}$$

respectively.

As can be seen in (4.131), in the hydrostatic limit the horizontal pressure-gradient force in θ-coordinates is a gradient. This is used later in the derivation of the potential vorticity equation.

4.14.4 The mechanical energy equation in isentropic coordinates

Before closing this section, we derive the mechanical energy equation in isentropic coordinates, which will be used in Chapter 11. We use the quasi-static approximation for simplicity. The result that we want to derive is

$$\boxed{\begin{aligned}\left[\frac{\partial}{\partial t}\left(\rho_\theta K\right)\right]_\theta + \nabla_\theta \cdot \left[\rho_\theta \mathbf{v}\left(K+\phi\right) + \rho_\theta \mathbf{\Lambda}_H\right] + \frac{\partial}{\partial \theta}\left[\rho_\theta \dot{\theta}\left(K+\phi\right) - z\left(\frac{\partial p}{\partial t}\right)_\theta\right] \\ = -\rho_\theta\left(\omega\alpha + \delta\right) .\end{aligned}} \tag{4.133}$$

To derive (4.133), we start with the quasi-static kinetic energy equation in isentropic coordinates, which is

$$\begin{aligned}\left[\frac{\partial}{\partial t}\left(\rho_\theta K\right)\right]_\theta + \nabla_\theta \cdot \left(\rho_\theta \mathbf{v} K + \rho_\theta \mathbf{\Lambda}_H\right) + \frac{\partial}{\partial \theta}\left(\rho_\theta \dot{\theta} K\right) \\ = -\rho_\theta \mathbf{v} \cdot \nabla_\theta s - \rho_\theta \delta .\end{aligned} \tag{4.134}$$

Since we are using the quasi-static approximation, the kinetic energy per unit mass is $K \equiv \mathbf{v}\cdot\mathbf{v}/2$. To make (4.134) look more like (4.133), we add some terms to both sides, to obtain:

$$\begin{aligned}\left[\frac{\partial}{\partial t}\left(\rho_\theta K\right)\right]_\theta + \nabla_\theta \cdot \left[\rho_\theta \mathbf{v}\left(K+\phi\right) + \rho_\theta \mathbf{\Lambda}_H\right] + \frac{\partial}{\partial \theta}\left[\rho_\theta \dot{\theta}\left(K+\phi\right) - z\left(\frac{\partial p}{\partial t}\right)_\theta\right] \\ = \nabla_\theta\left(\rho_\theta \mathbf{v}\phi\right) + \frac{\partial}{\partial \theta}\left[\rho_\theta \dot{\theta}\phi - z\left(\frac{\partial p}{\partial t}\right)_\theta\right] - \rho_\theta \mathbf{v}\cdot\nabla_\theta s - \rho_\theta \delta .\end{aligned} \tag{4.135}$$

The left-hand side of (4.135) is identical to the left-hand side of (4.133), so all we have to do is show that the right-hand side of (4.135) is equivalent to the right-hand side of (4.133), and that will be the case if

$$\nabla_\theta \cdot \left(\rho_\theta \mathbf{v}\phi\right) + \frac{\partial}{\partial \theta}\left[\rho_\theta \dot{\theta}\phi - z\left(\frac{\partial p}{\partial t}\right)_\theta\right] - \rho_\theta \mathbf{v}\cdot\nabla_\theta s = -\rho_\theta \omega\alpha . \tag{4.136}$$

We want to prove that (4.136) is true. Let's take it one term at a time. First we write

$$\begin{aligned}\frac{\partial}{\partial \theta}\left(\rho_\theta \dot{\theta}\phi\right) =& \phi\frac{\partial}{\partial \theta}\left(\rho_\theta \dot{\theta}\right) + \rho_\theta \dot{\theta}\frac{\partial \phi}{\partial \theta} \\ =& \phi\frac{\partial}{\partial \theta}\left(\rho_\theta \dot{\theta}\right) + \rho_\theta \dot{\theta}\frac{\partial \phi}{\partial p}\frac{\partial p}{\partial \theta} \\ =& \phi\frac{\partial}{\partial \theta}\left(\rho_\theta \dot{\theta}\right) - \rho_\theta\left(\dot{\theta}\frac{\partial p}{\partial \theta}\right)\alpha .\end{aligned} \tag{4.137}$$

Here we have used the hydrostatic equation in the form $\partial\phi/\partial p = -\alpha$. Next, we expand the second term in square brackets as

$$\begin{aligned}
\frac{\partial}{\partial\theta}\left[z\left(\frac{\partial p}{\partial t}\right)_\theta\right] =& \frac{\partial z}{\partial\theta}\left(\frac{\partial p}{\partial t}\right)_\theta + z\frac{\partial}{\partial\theta}\left[\left(\frac{\partial p}{\partial t}\right)_\theta\right] \\
=& \frac{\partial z}{\partial p}\frac{\partial p}{\partial\theta}\left(\frac{\partial p}{\partial t}\right)_\theta + z\frac{\partial}{\partial t}\left[\left(\frac{\partial p}{\partial\theta}\right)_\theta\right] \\
=& -\frac{\partial\phi}{\partial p}\rho_\theta\left(\frac{\partial p}{\partial t}\right)_\theta - \phi\left(\frac{\partial\rho_\theta}{\partial t}\right)_\theta \\
=& \alpha\rho_\theta\left(\frac{\partial p}{\partial t}\right)_\theta - \phi\left(\frac{\partial\rho_\theta}{\partial t}\right)_\theta .
\end{aligned} \tag{4.138}$$

Here we have used $\rho_\theta = -(1/g)\,\partial p/\partial\theta$, which can be derived using the hydrostatic approximation. Combining (4.137) and (4.138), we obtain

$$\begin{aligned}
\frac{\partial}{\partial\theta}\left[\rho_\theta\dot\theta\phi - z\left(\frac{\partial p}{\partial t}\right)_\theta\right] =& \phi\frac{\partial}{\partial\theta}\left(\rho_\theta\dot\theta\right) - \rho_\theta\left(\dot\theta\frac{\partial p}{\partial\theta}\right)\alpha - \left[\alpha\rho_\theta\left(\frac{\partial p}{\partial t}\right)_\theta - \phi\left(\frac{\partial\rho_\theta}{\partial t}\right)_\theta\right] \\
=& \phi\left[\frac{\partial\rho_\theta}{\partial t} + \frac{\partial}{\partial\theta}\left(\rho_\theta\dot\theta\right)\right] - \rho_\theta\left[\left(\frac{\partial p}{\partial t}\right)_\theta + \dot\theta\frac{\partial p}{\partial\theta}\right]\alpha \\
=& -\phi\nabla_\theta\cdot\left(\rho_\theta\mathbf{v}\right) - \rho_\theta\left(\omega - \mathbf{v}\cdot p\right)\alpha .
\end{aligned} \tag{4.139}$$

This looks promising. Substituting (4.139) into the left-hand side of (4.136), we find that

$$\begin{aligned}
&\nabla_\theta\cdot\left(\rho_\theta\mathbf{v}\phi\right) + \frac{\partial}{\partial\theta}\left[\rho_\theta\dot\theta\phi - z\left(\frac{\partial p}{\partial t}\right)_\theta\right] - \rho_\theta\mathbf{v}\cdot\nabla_\theta s \\
&= \nabla_\theta\cdot\left(\rho_\theta\mathbf{v}\phi\right) - \phi\nabla_\theta\cdot\left(\rho_\theta\mathbf{v}\right) - \rho_\theta\left(\omega - \mathbf{v}\cdot\nabla_\theta p\right)\alpha - \rho_\theta\mathbf{v}\cdot\nabla_\theta s \\
&= \rho_\theta\mathbf{v}\cdot\nabla_\theta\phi - \rho_\theta\left(\omega - \mathbf{v}\cdot\nabla_\theta p\right)\alpha - \rho_\theta\mathbf{v}\cdot\nabla_\theta\left(c_pT + \phi\right) \\
&= -\rho_\theta\left(\omega - \mathbf{v}\cdot\nabla_\theta p\right)\alpha - \rho_\theta\mathbf{v}\cdot\nabla_\theta\left(c_pT\right) \\
&= -\rho_\theta\left(\omega - \mathbf{v}\cdot\nabla_\theta p\right)\alpha - \rho_\theta\mathbf{v}\cdot\nabla_\theta\left(\Pi\theta\right) \\
&= -\rho_\theta\left(\omega - \mathbf{v}\cdot\nabla_\theta p\right)\alpha - \rho_\theta\theta\mathbf{v}\cdot\nabla_\theta\Pi \\
&= -\rho_\theta\omega\alpha + \rho_\theta\left(\alpha\mathbf{v}\cdot\nabla_\theta p - \theta\mathbf{v}\cdot\nabla_\theta\Pi\right) .
\end{aligned} \tag{4.140}$$

It is easy to show that

$$\alpha \mathbf{v} \cdot \nabla_\theta p - \theta \mathbf{v} \cdot \nabla_\theta \Pi = 0 \, . \tag{4.141}$$

This completes the derivation.

4.15 Vorticity and potential vorticity

Using vector notation, the equation of motion can be written as

$$\frac{\partial \mathbf{V}}{\partial t} + \boldsymbol{\omega}_a \times \mathbf{V} + \nabla (K + \phi) = -\alpha \nabla p - \mathscr{F}_{\mathbf{V}}. \tag{4.142}$$

Here

$$\boldsymbol{\omega}_a \equiv \boldsymbol{\omega} + 2\boldsymbol{\Omega} \tag{4.143}$$

is the absolute vorticity, where $\boldsymbol{\omega}$ is the relative vorticity and $\boldsymbol{\Omega}$ is the angular velocity of the Earth's rotation. Also, K is the kinetic energy per unit mass. We neglect the extraterrestrial gravitation effects in this section.

Helmholtz's theorem states that any sufficiently smooth three-dimensional (or two-dimensional) vector field can be decomposed into a divergent part, denoted here by $\mathbf{V}_\delta$, and a rotational part, denoted by $\mathbf{V}_\omega$, such that

$$\mathbf{V} = \mathbf{V}_\delta + \mathbf{V}_\omega \, . \tag{4.144}$$

The divergent part of the wind vector satisfies

$$\mathbf{V}_\delta = \nabla \chi \, , \tag{4.145}$$

where χ is called the velocity potential. Since $\mathbf{V}_\delta$ is a gradient, its curl is zero:

$$\nabla \times \mathbf{V}_\delta = 0 \,. \tag{4.146}$$

The divergence of the wind is given by

$$\delta \equiv \nabla \cdot \mathbf{V} = \nabla \cdot \mathbf{V}_\delta \,. \tag{4.147}$$

(Here there is a notational conflict with the dissipation rate, but it should not cause confusion.) The rotational part of the wind vector satisfies

$$\mathbf{V}_\omega = \nabla \times \mathbf{\Psi} \,, \tag{4.148}$$

where $\mathbf{\Psi}$ is called the vector potential of the wind. Since $\mathbf{V}_\omega$ is a curl, its divergence is zero:

$$\nabla \cdot \mathbf{V}_\omega = 0 \,. \tag{4.149}$$

The vorticity is a local measure of the angular velocity of a fluid. It is the curl of the velocity:

$$\boldsymbol{\omega} \equiv \nabla \times \mathbf{V} = \nabla \times \mathbf{V}_\omega \,. \tag{4.150}$$

It follows that

$$\nabla \cdot \boldsymbol{\omega} = 0 \,. \tag{4.151}$$

As discussed earlier, the eastward velocity of a particle rotating with the Earth is given by $\mathbf{e}_\lambda \Omega r \cos\varphi$, where $\mathbf{e}_\lambda$ is a unit vector pointing toward the east, r is the distance from the center of the Earth, and φ is latitude. The vorticity associated with this "Earth rotation" is given by

$$\begin{aligned}\boldsymbol{\Omega} \equiv \nabla \times (\mathbf{e}_\lambda \Omega r \cos\varphi) &= \left\{ 0,\ \frac{1}{r}\frac{\partial}{\partial r}\left(r\Omega r\cos\varphi\right),\ \frac{1}{r\cos\varphi}\left[-\frac{\partial}{\partial\varphi}\left(\Omega r\cos\varphi\cos\varphi\right)\right]\right\} \\ &= \left[0,\ \frac{2r\Omega\cos\varphi}{r},\ \frac{\Omega r}{r\cos\varphi}\left(2\cos\varphi\sin\varphi\right)\right] \\ &= 2\Omega\left(\mathbf{e}_\varphi\cos\varphi + \mathbf{e}_r\sin\varphi\right)\ ,\end{aligned} \tag{4.152}$$

where $\mathbf{e}_\varphi$ is a unit vector pointing north, and $\mathbf{e}_r$ is a unit vector pointing outward from the center of the Earth. We note that $\boldsymbol{\Omega}$, like $\boldsymbol{\omega}$, is non-divergent, i.e.,

$$\nabla \cdot \boldsymbol{\Omega} = 0\ , \tag{4.153}$$

and it follows that

$$\nabla \cdot \boldsymbol{\omega}_a = 0\ . \tag{4.154}$$

The divergence is key to acoustic waves and little else. In contrast, the vorticity plays a central role in virtually all aspects of atmospheric dynamics, and in many other applications of fluid dynamics. *Divergence disperses; vorticity coheres.*

4.15.1 Turn, Turn, Turn

The vorticity equation is derived by taking the curl of (4.142). The result is

$$\boxed{\frac{\partial \boldsymbol{\omega}_a}{\partial t} + \nabla \times (\boldsymbol{\omega}_a \times \mathbf{V}) = \nabla p \times \nabla\alpha - \nabla \times \mathscr{F}_\mathbf{V}}\ . \tag{4.155}$$

Here we have assumed that the Earth's angular velocity is independent of time, and we have used

$$\nabla \times (\nabla K) = \nabla \times (\nabla\phi) = 0\ . \tag{4.156}$$

We have also used the identity

$$\nabla \times (\psi \mathbf{A}) = \nabla \psi \times \mathbf{A} + \psi (\nabla \times \mathbf{A}) , \tag{4.157}$$

where $\mathbf{A}$ is an arbitrary vector, to write

$$\begin{aligned} -\nabla \times (\alpha \nabla p) &= -\nabla \alpha \times \nabla p \\ &= \nabla p \times \nabla \alpha . \end{aligned} \tag{4.158}$$

For the special case of constant density, Eq. (4.158) reduces to

$$-\nabla \times (\alpha \nabla p) = 0 , \tag{4.159}$$

and the pressure drops out of the vorticity equation, (4.155). This same conclusion applies when we use the anelastic approximation, which is very useful for the study of turbulence and convection.

It is conventional to manipulate (4.155) into an alternative form. The identity (A.19) allows us to write

$$\begin{aligned} \nabla \times (\boldsymbol{\omega}_a \times \mathbf{V}) = \boldsymbol{\omega}_a (\nabla \cdot \mathbf{V}) - \mathbf{V} (\nabla \cdot \boldsymbol{\omega}_a) \\ - (\boldsymbol{\omega}_a \cdot \nabla) \mathbf{V} + (\mathbf{V} \cdot \nabla) \boldsymbol{\omega}_a . \end{aligned} \tag{4.160}$$

Using (4.151) and (4.153), Eq. (4.160) can be written as

$$\nabla \times (\boldsymbol{\omega}_a \times \mathbf{V}) = \boldsymbol{\omega}_a (\nabla \cdot \mathbf{V}) - (\boldsymbol{\omega}_a \cdot \nabla) \mathbf{V} + (\mathbf{V} \cdot \nabla) \boldsymbol{\omega}_a . \tag{4.161}$$

After substitution from (4.161), we can write (4.155) as

$$\frac{D \boldsymbol{\omega}_a}{Dt} + \boldsymbol{\omega}_a (\nabla \cdot \mathbf{V}) - (\boldsymbol{\omega}_a \cdot \nabla) \mathbf{V} = \nabla p \times \nabla \alpha - \nabla \times \mathscr{F}_{\mathbf{V}} . \tag{4.162}$$

The terms on the left-hand side of (4.162) are purely kinematic, in that they do not involve forces or the mass.

We now use continuity in the form (4.10) to eliminate $\nabla \cdot \mathbf{V}$ in (4.162). The result can be written as

$$\begin{aligned} \alpha \frac{D\boldsymbol{\omega}_a}{Dt} + \boldsymbol{\omega}_a \frac{D\alpha}{Dt} - \alpha \left(\boldsymbol{\omega}_a \cdot \nabla\right) \mathbf{V} \\ = \alpha \left(\nabla p \times \nabla \alpha - \nabla \times \mathbf{F}\right) \ . \end{aligned} \tag{4.163}$$

Combining terms on the left-hand side of (4.163), we find that

$$\boxed{\frac{D}{Dt}\left(\alpha \boldsymbol{\omega}_a\right) - \left(\alpha \boldsymbol{\omega}_a \cdot \nabla\right) \mathbf{V} = \alpha \nabla p \times \nabla \alpha - \nabla \times \left(\alpha \nabla \cdot \mathbf{F}\right).} \tag{4.164}$$

Eq. (4.164) shows that $\alpha \boldsymbol{\omega}_a$ is materially conserved when $-\left(\alpha \boldsymbol{\omega}_a \cdot \nabla\right) \mathbf{V}$ and the forces on the right-hand side of (4.164) vanish. To the extent the product of the specific volume and the vorticity is materially conserved, an increase in the specific volume (i.e., expansion) tends to be associated with a decrease in the magnitude of the absolute vorticity, and vice versa. Note that if the motion is two-dimensional, then $\boldsymbol{\omega}_a \cdot \nabla = 0$.

Some authors (e.g., Johnson, 1989) have discussed the "dynamic vorticity," which is defined as the curl of the product of the velocity vector and the density. Eq. (4.164) shows that the vorticity divided by the density is a more natural variable, in the sense that it obeys a particularly simple evolution equation.

4.15.2 Stretching and twisting

The last term on the left-hand side of (4.164) represents both (i.e., the combination of) stretching and twisting. It can be written as

$$\left(\boldsymbol{\omega}_a \cdot \nabla\right) \mathbf{V} = \left|\boldsymbol{\omega}_a\right| \frac{\partial \mathbf{V}}{\partial s} \ , \tag{4.165}$$

where s is a curvilinear coordinate that points in the direction of $\boldsymbol{\omega}_a$. We can divide $\left|\boldsymbol{\omega}_a\right| \partial \mathbf{V} / \partial s$ into two parts:

- *Stretching*, given by $|\boldsymbol{\omega}_a|\partial V_s/\partial s$, where V_s is the component of $\mathbf{V}$ in the direction of s, and $\mathbf{e_s}$ is a unit vector in the s direction. See Fig. 1. Positive stretching $(\partial V_s/\partial s > 0)$ favors $\dfrac{D}{Dt}|\boldsymbol{\omega}_a| > 0$. The effects of stretching on the vorticity can be understood in terms of conservation of the angular momentum of a cylinder of air with its axis oriented parallel to the vorticity vector. Positive stretching is due to divergence along the axis of the cylinder, which is compensated for by convergence of mass towards the axis. Angular momentum conservation then leads to an increase in the cylinder's angular velocity. Negative stretching does the same thing in reverse.

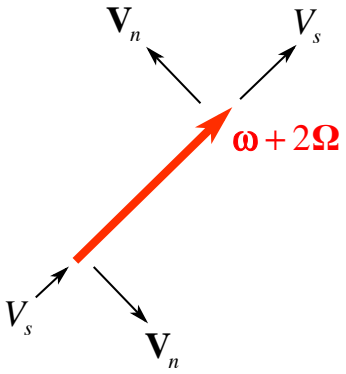


Figure 4.5: The components of the velocity normal and tangent to the vorticity vector, and the stretching and twisting processes associated with these velocity components. The absolute vorticity vector is shown in red.

- *Twisting*, $|\boldsymbol{\omega}_a|\partial \mathbf{V}_n/\partial s$, where $\mathbf{V}_n$ is the two-dimensional vector portion of $\mathbf{V}$ that lies in the plane normal to $\boldsymbol{\omega}_a$. Twisting changes the direction of $\boldsymbol{\omega}_a$, but not its magnitude.

You should be able to see that both stretching and twisting are impossible in two-dimensional motion. This is very important for large-scale turbulence but not very relevant to small-scale boundary layer turbulence and convection. Further discussion is given in Chapter 13.

4.15.3 The solenoidal term

The term $\alpha\nabla p \times \nabla\alpha$ on the right-hand side of (4.162) is often called the solenoidal term, because the vector $\nabla p \times \nabla\alpha$ is nondivergent. The solenoidal term vanishes if the density is spatially uniform, or if the density depends only on the pressure. In either case, the pressure drops out of the vorticity equation altogether.

To facilitate the interpretation of the solenoidal term, it is very useful to distinguish between horizontal and vertical derivatives. We write

$$\boxed{\begin{aligned}\nabla p \times \nabla \alpha &= \left[\left(\nabla_r p + \frac{\partial p}{\partial r}\mathbf{e}_r\right) \times \left(\nabla_r \alpha + \frac{\partial \alpha}{\partial r}\mathbf{e}_r\right)\right] \\ &= (\nabla_r p \times \nabla_r \alpha) + \mathbf{e}_r \times \left(\frac{\partial p}{\partial r}\nabla_r \alpha - \nabla_r p \frac{\partial \alpha}{\partial r}\right) .\end{aligned}} \tag{4.166}$$

The first term on the second line of (4.166) is oriented vertically, so it affects the tendency of the vertical component of the vorticity vector. The second and third terms of (4.166) lie in the horizontal plane, so they affect the tendency of the horizontal vorticity vector. The term $\dfrac{\partial p}{\partial r}\nabla_r \alpha$ represents the effects of buoyancy. To see this, use the hydrostatic approximation in the form

$$\frac{\partial p}{\partial r} \cong -\frac{g}{\alpha}. \tag{4.167}$$

We can then write

$$\frac{\partial p}{\partial r}\nabla_r \alpha \cong -g\left(\frac{\nabla_r \alpha}{\alpha}\right), \tag{4.168}$$

a very accurate approximation, which shows that horizontal vorticity can be generated by horizontal variations of the density in the presence of gravity. This is what happens on the side of a buoyant thermal or plume.

4.15.4 Potential vorticity

Eq. (4.164) is a convenient starting point for a derivation of the potential vorticity equation. Let Λ be an intensive scalar that satisfies

$$\frac{D\Lambda}{Dt} = S_\Lambda \ , \tag{4.169}$$

where S_Λ is a source of Λ. We can write

$$\alpha\left(\boldsymbol{\omega}_a \cdot \nabla\right)\frac{D\Lambda}{Dt} = \alpha\boldsymbol{\omega}_a \cdot \left[\frac{D\left(\nabla\Lambda\right)}{Dt}\right] + \alpha\left\{\left(\boldsymbol{\omega}_a \cdot \nabla\right)\mathbf{V}\right\} \cdot \nabla\Lambda, \tag{4.170}$$

which can be rearranged to

$$\alpha\boldsymbol{\omega}_a\cdot\left[\frac{D(\nabla\Lambda)}{Dt}\right]=-\alpha\left[(\boldsymbol{\omega}_a\cdot\nabla)\,\mathbf{V}\right]\cdot\nabla\Lambda+\alpha\,(\boldsymbol{\omega}_a\cdot\nabla)\frac{D\Lambda}{Dt}. \tag{4.171}$$

We now use (4.169) to eliminate $D\Lambda/Dt$ in (4.171), giving

$$\alpha\boldsymbol{\omega}_a\cdot\left[\frac{D(\nabla\Lambda)}{Dt}\right]=-\alpha\left[(\boldsymbol{\omega}_a\cdot\nabla)\,\mathbf{V}\right]\cdot\nabla\Lambda+\alpha\,(\boldsymbol{\omega}_a\cdot\nabla)\,S_\Lambda. \tag{4.172}$$

When we add (4.172) to $\nabla\Lambda\cdot$ (4.164), the stretching-twisting terms on the right-hand side of (4.164) cancel with the corresponding terms on the right-hand side of (4.172), and we obtain

$$\begin{aligned}\frac{D(\alpha\boldsymbol{\omega}_a\cdot\nabla\Lambda)}{Dt}&=\alpha\,(\nabla\Lambda)\cdot(\nabla p\times\nabla\alpha)\\&\quad-\alpha\,(\nabla\Lambda)\cdot(\nabla\times\mathscr{F}_{\mathbf{V}})+\alpha\,(\boldsymbol{\omega}_a\cdot\nabla)\,S_\Lambda.\end{aligned} \tag{4.173}$$

The pressure-gradient term of (4.164) is still visible in (4.173). However, if we choose Λ to be a thermodynamic function that depends only on α and p, then

$$(\nabla\Lambda)\cdot(\nabla p\times\nabla\alpha)=0, \tag{4.174}$$

and (4.173) reduces to

$$\frac{D(\alpha\boldsymbol{\omega}_a\cdot\nabla\Lambda)}{Dt}=-\alpha\,(\nabla\Lambda)\cdot(\nabla\times\mathscr{F}_{\mathbf{V}})+\alpha\,(\boldsymbol{\omega}_a\cdot\nabla)\,S_\Lambda\ . \tag{4.175}$$

The most common choice for Λ is $\Lambda\equiv\theta$, the potential temperature. With that choice, (4.175) can be written as

$$\rho\frac{D}{Dt}\,\mathrm{PV}=-\,(\nabla\theta)\cdot(\nabla\times\mathscr{F}_{\mathbf{V}})+(\boldsymbol{\omega}_a\cdot\nabla)\,\dot{\theta}\ , \tag{4.176}$$

where

$$\dot{\theta} \equiv \frac{D\theta}{Dt} = S_\theta \tag{4.177}$$

is the Lagrangian time-rate-of-change of θ, which is proportional to the heating rate, and

$$\boxed{\text{PV} \equiv \frac{\boldsymbol{\omega}_a \cdot \nabla\theta}{\rho}} \tag{4.178}$$

is the *potential vorticity*. According to (4.178), the PV is proportional to the component of the absolute vorticity that is perpendicular to an isentropic surface. It is independent of the part of the absolute vorticity that is parallel to the isentropic surface. Combining (4.178) with (4.151), we see that the potential vorticity can be written in terms of a divergence:

$$\rho\ \text{PV} \equiv \nabla \cdot (\theta\, \boldsymbol{\omega}_a)\,. \tag{4.179}$$

Finally, we can use vector identities to write the heating and friction terms of (4.176) as divergences, following Haynes and McIntyre (1987):

$$\begin{aligned}(\nabla\theta)\cdot(\nabla\times\mathscr{F}_{\mathbf{V}}) &= \nabla\cdot(\nabla\theta\times\mathscr{F}_{\mathbf{V}}) + \mathscr{F}_{\mathbf{V}}\cdot(\nabla\times\nabla\theta)\\ &= \nabla\cdot(\nabla\theta\times\mathscr{F}_{\mathbf{V}})\ ,\end{aligned} \tag{4.180}$$

$$\begin{aligned}(\boldsymbol{\omega}_a\cdot\nabla)\,\dot{\theta} &= \nabla\cdot\left(\boldsymbol{\omega}_a\dot{\theta}\right) - \dot{\theta}\,(\nabla\cdot\boldsymbol{\omega}_a)\\ &= \nabla\cdot\left(\boldsymbol{\omega}_a\dot{\theta}\right)\ .\end{aligned} \tag{4.181}$$

Then (4.176) becomes

$$\rho\frac{D}{Dt}\,\text{PV} = \nabla\cdot\left(\boldsymbol{\omega}_a\dot{\theta} - \nabla\theta\times\mathscr{F}_{\mathbf{V}}\right)\ . \tag{4.182}$$

The flux form corresponding to (4.182) is

$$\boxed{\frac{\partial}{\partial t}(\rho\ \mathrm{PV}) + \nabla\cdot\left(\rho\mathbf{V}\ \mathrm{PV} + \nabla\theta\times\mathscr{F}_{\mathbf{V}} - \boldsymbol{\omega}_a\dot{\theta}\right) = 0\,.} \tag{4.183}$$

This amazing result was derived and discussed by Haynes and McIntyre (1987). It says that the time rate of change of the mass-weighted potential vorticity is due only to the divergence of a flux, *even when friction and heating are present*. This very basic and important fact means that potential vorticity can be created or destroyed only at the boundaries of the fluid (also see Bretherton, 1966). In other words, *the atmosphere gains or loses potential vorticity only in the boundary layer.*

Schubert (2004) discusses a generalization of the PV that allows moisture, including phase changes.

4.15.5 The isentropic potential vorticity

We now define

$$\mathrm{IPV} \equiv \frac{\zeta_\theta + f}{\rho_\theta} \tag{4.184}$$

as the isentropic potential vorticity, where

$$\zeta_\theta \equiv \mathbf{e}_r\cdot(\nabla_\theta\times\mathbf{v}) \tag{4.185}$$

can be interpreted as the projection of the vertical component of the vorticity onto the unit vector normal to the isentropic surface. Compare (4.184) with (4.178). Later we will use the IPV to understand the role of planetary waves in the global circulation. To derive the prognostic equation that governs the IPV, we start from the quasi-static equation of horizontal motion in isentropic coordinates:

$$\left(\frac{\partial\mathbf{v}}{\partial t}\right)_\theta + (\zeta_\theta + f)\,\mathbf{e}_r\times\mathbf{v} + \nabla_\theta\,(K+s) + \dot{\theta}\frac{\partial\mathbf{v}}{\partial\theta} = -\mathscr{F}_{\mathbf{V}}\,. \tag{4.186}$$

The vertical component of the vorticity equation in isentropic coordinates can be derived by applying $\mathbf{e}_r\cdot\nabla_\theta\times$ to (4.186), and using the vector identities (A.33) and (A.34). We find that

$$\left(\frac{\partial}{\partial t}\right)_\theta (\zeta_\theta + f) + \nabla_\theta \cdot \left[\mathbf{v}(\zeta_\theta + f) - \mathbf{e}_r \times \left(\dot{\theta}\frac{\partial}{\partial \theta} + \mathscr{F}_\mathbf{v}\right)\right] = 0\,, \tag{4.187}$$

Notice that in (4.187) the contribution from the "vertical advection term" of (4.186) appears inside a horizontal divergence operator! It thus represents a *horizontal* redistribution of PV along θ surfaces. It does not look like vertical advection, and it does not look like a source or sink. This counter-intuitive result, due to Haynes and McIntyre (1987), comes from the use of (A.34).

Eq. (4.187) can be written as

$$\left(\frac{\partial}{\partial t}\right)_\theta (\rho_\theta \text{ IPV}) + \nabla_\theta \cdot \left[\rho_\theta \mathbf{v} \text{ IPV} - \mathbf{e}_r \times \left(\dot{\theta}\frac{\partial \mathbf{v}}{\partial \theta} + \mathscr{F}_\mathbf{v}\right)\right] = 0\,. \tag{4.188}$$

Compare (4.188) with (4.183). Eq. (4.188) is called the "impermeability theorem," because it says that there is no flux of IPV across isentropic surfaces, i.e., the isentropic surfaces are impermeable to the IPV. The impermeability theorem implies that, for isentropic surfaces that do not intersect the Earth's surface, the area-average of the mass-weighted PV cannot change, even in the presence of heating and friction. The local time-rate of change of ρ_θIPV is balanced entirely by a horizontal flux divergence; there is no exchange of IPV with neighboring isentropic surfaces. It's like Las Vegas: What happens on an isentropic surface stays on the isentropic surface. This conclusion does not apply to the "Underworld" of Hoskins (1991), which by definition includes only isentropic surfaces that *do* intersect the Earth's surface. The IPV of the Underworld can change due to boundary effects. For further discussion, see Bretherton and Schär (1993).

The IPV is of interest in part because it obeys a very simple conservation equation, i.e., (4.188). As emphasized by Hoskins et al. (1985), however, a more important reason is that, with appropriate boundary conditions, the isentropic PV essentially determines the wind and temperature distributions for a large class of balanced circulations. Examples are given later.

Fig. 4.6 shows the zonally averaged isentropic IPV, which is (almost) a monotonic function of latitude, and passes through zero very close to the equator. The stratosphere is a region of high IPV because the pseudodensity is small there. The winter polar stratosphere has particularly large values of the IPV. Intrusions of stratospheric air into the troposphere are characterized by anomalously large values of the IPV. The troposphere is a region of nearly uniform IPV, relative to the stratosphere. This suggests that the IPV is mixed or homogenized by the tropospheric weather systems. In middle and high latitudes of each hemisphere the tropopause roughly coincides with a constant-PV surface, namely the PVU surface, where a PVU or "potential vorticity unit" is defined as 10^{-6} m^2 K s^{-1} kg^{-1}.

Kunz et al. (2011) provide some cautionary s on this PV-based definition of the tropopause height. As mentioned earlier, the tropopause height is discontinuous in the subtropics, and one possible definition of the tropics is that it is the region on the equatorward side of these discontinuities.

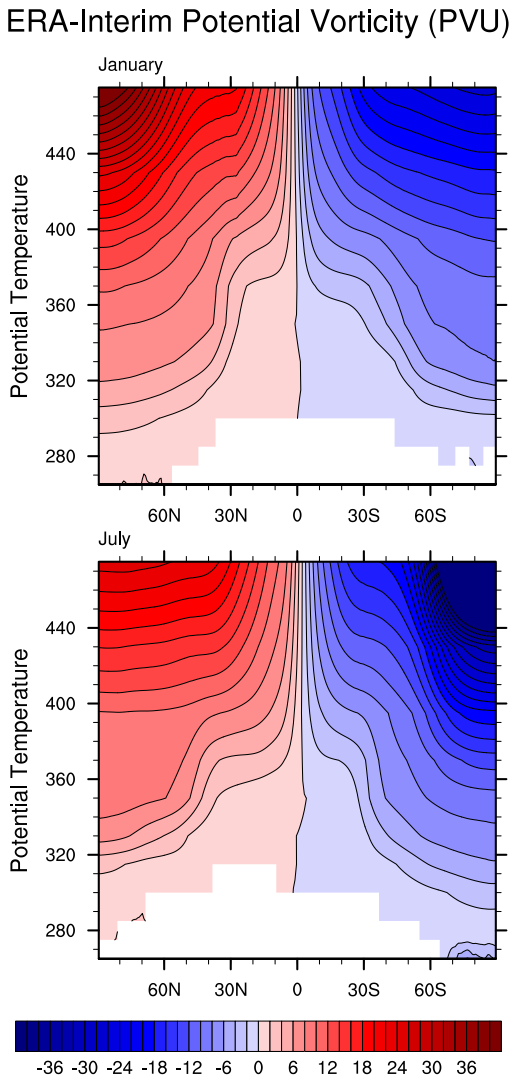


Figure 4.6: The zonally averaged isentropic potential vorticity plotted as a function of latitude and potential temperature. The contour interval is one PVU ("potential vorticity unit"), which is defined as 10^{-6} m^2 K s^{-1} kg^{-1}. The white area at the bottom represents isentropic surfaces that are on the "ground." A similar figure appears in Edouard et al. (1997).

4.15.6 Fellow travelers

Using (4.184), we can rewrite the PV equation, (4.183), as

$$\frac{\partial}{\partial t}(\boldsymbol{\omega}_a \cdot \nabla\theta) + \nabla \cdot \left[(\boldsymbol{\omega}_a \cdot \nabla\theta)\,\mathbf{V} + \nabla\theta \times \mathscr{F}_{\mathbf{V}} - \boldsymbol{\omega}_a \dot{\theta}\right] = 0\,. \tag{4.189}$$

In the absence of friction and heating this reduces to

$$\frac{\partial}{\partial t}(\boldsymbol{\omega}_a \cdot \nabla\theta) + \nabla \cdot \left[(\boldsymbol{\omega}_a \cdot \nabla\theta)\,\mathbf{V}\right] = 0\,. \tag{4.190}$$

Compare this with the exact continuity equation, which is repeated here for convenience:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0\,.$$

The comparison shows that, in the absence of friction and heating, the component of the absolute vorticity that is perpendicular to the isentropic surface moves in the same way as ρ. *The vorticity moves in the same way as the mass.* More to the point, the vorticity moves *with* the mass. This is why vortices seem almost to be alive. Their identities persist as they move. They dance with each other. They are identifiable "discrete things" moving along in and with an otherwise featureless fluid.

4.15.7 The vertical component of the vorticity, and the horizontal vorticity vector

We now separate the three-dimensional vorticity vector into the horizontal vorticity vector, denoted by $\boldsymbol{\eta}$, and the vertical component of the vorticity, denoted by ζ:

$$\boldsymbol{\omega} = \boldsymbol{\eta} + \zeta \mathbf{e}_r. \tag{4.191}$$

It can be shown that

$$\boldsymbol{\eta} = \mathbf{e}_\mathbf{r} \times \left[\frac{1}{r}\frac{\partial}{\partial r}(r\mathbf{v}) - \nabla_r w\right]\,, \tag{4.192}$$

and

$$\zeta \equiv \mathbf{e}_r \cdot \boldsymbol{\omega}\,. \tag{4.193}$$

Note that

$$\frac{1}{r}\frac{\partial}{\partial r}(r\mathbf{v}) \cong \frac{\partial \mathbf{v}}{\partial r}\,, \tag{4.194}$$

because the atmosphere is thin compared to the radius of the Earth.

As an example, in Cartesian coordinates $(x,\, y,\, z)$, with velocity components $(u,\, v,\, w)$ and unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, the horizontal vorticity vector can be written as

$$\begin{aligned}\boldsymbol{\eta} =& \mathbf{k}\times\left(\frac{\partial \mathbf{v}}{\partial z} - \nabla_z w\right)\\ =& \mathbf{i}\left(-\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right) + \mathbf{j}\left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}\right)\,.\end{aligned} \tag{4.195}$$

Eq. (4.195) will be used in Chapter 8.

4.16 The primitive equations

Several important approximations to the equation of motion are commonly used in the analysis of the large-scale circulation systems:

- The quasi-static approximation, which was already discussed earlier in this chapter.
- Treat the effective gravitational force as purely vertical, and neglect its variations with height. This means that

$$\phi \cong \phi\,(z)\,, \tag{4.196}$$

so that

$$\nabla\phi \cong g\mathbf{e}_r\,. \tag{4.197}$$

Here $\mathbf{e}_r$ is a unit vector pointing away from the center of the Earth (see Fig. 4.2). Compare (4.197) with (4.43). The approximation (4.197) can be justified because the non-radial components of $\nabla\phi$, which are due to the centrifugal acceleration, are very small compared to $\partial\phi/\partial z$.

We also neglect variations of $g \equiv \partial\phi/\partial z$ with height through the depth of the troposphere and stratosphere. When we use (4.43) with an approximate spatially constant value of g, we must also approximate the shape of the Earth as a sphere, with superimposed topographical mountains and valleys.

- Replace r by a everywhere, where a is the mean radius of the Earth. An approximation of this form can be justified for an atmosphere that is thin compared to the radius of the planet. It is a good approximation for Earth, but would not be accurate, e.g., for Jupiter.
- Drop the "nontraditional" terms of (4.113) that contain $\overline{f}$. This means that the meridional component of $\mathbf{\Omega}$ disappears from the equations. There is an ongoing discussion as to whether and when this approximation is actually justified (e.g., White and Bromley, 1995; Zeitlin, 2018; Igel and Biello, 2020; Ong and Roundy, 2020).
- Neglect uw/r and vw/r in the equations for u and v, respectively.

Justifications of these approximations can be found in standard textbooks on atmospheric dynamics, Holton (e.g., 1973). The resulting system is often called the "primitive equations."

When we consider large-scale motions, the tendencies due to molecular fluxes are overwhelmed by those associated with turbulence, convection, and gravity waves. In addition, for large-scale motions the flux divergences associated with turbulence, convection, and small-scale waves can be accurately approximated by the vertical derivative of the vertical component of the flux.

With these approximations, (4.108) - (4.110) are replaced by

$$\rho\left(\frac{Du}{Dt} - \frac{uv\tan\varphi}{a} - fv\right) = -\frac{1}{a\cos\varphi}\frac{\partial p}{\partial\lambda} - \frac{\partial F_u}{\partial z}\,, \tag{4.198}$$

$$\left(\frac{Dv}{Dt} + \frac{u^2\tan\varphi}{a} + fu\right) = -\frac{1}{a}\frac{\partial p}{\partial\varphi} - \frac{\partial F_v}{\partial z}\,, \tag{4.199}$$

$$0 = -\rho g - \frac{\partial p}{\partial z}\,. \tag{4.200}$$

Also, the vector equation of horizontal motion can be written as

$$\rho \left[\frac{\partial \mathbf{v}}{\partial t} + (\zeta_z + f)\, \mathbf{e}_r \times \mathbf{v} + \nabla_z K + w\, frac \partial \mathbf{v} \partial z \right] = -\nabla_z p - \frac{\partial \mathbf{F}_\mathbf{v}}{\partial z} \,, \tag{4.201}$$

where

$$\zeta_z \equiv \mathbf{e}_r \cdot (\nabla_z \times \mathbf{v}) \tag{4.202}$$

is the vertical component of the relative vorticity as computed using height coordinates, and

$$K \equiv \frac{1}{2} \mathbf{v} \cdot \mathbf{v} \tag{4.203}$$

is the kinetic energy per unit mass associated with the horizontal wind vector.

4.17 The quasi-geostrophic system

Charney (1990) gave a systematic justification for what is now called the quasi-geostrophic (QG) system of equations. At that time, the QG system was of interest because it provided a way to do numerical weather prediction on the primitive digital computers that were just becoming available (e.g., Harper, 2012). From a modern perspective, the QG system is useful mainly because it makes it possible to give simple (but approximate) explanations for synoptic-scale weather phenomena in middle latitudes. We will use the QG system in Chapter 10.

Holton (1973) and Vallis (2017) provide thorough introductions to QG dynamics. The QG system can be derived via an expansion in powers of the Rossby number, which looks very formal although I've never found it particularly persuasive. Here we briefly review the basic idea, which involves various approximations in the momentum equation and the thermodynamic equation, without trying to justify the approximations in detail.

Using pressure as a vertical coordinate, a simplified version of the equation of horizontal motion can be written as

$$\frac{D\mathbf{v}}{Dt} + f\mathbf{e}_r \times \mathbf{v} = -\nabla\phi \ . \tag{4.204}$$

Here we neglect friction, for simplicity. By definition, the geostrophic wind vector, $\mathbf{v}_{\rm g}$, satisfies

$$f\mathbf{e}_r \times \mathbf{v}_{\rm g} \equiv -\nabla\phi \ , \tag{4.205}$$

where ϕ is the geopotential. Using (4.205), we can rewrite (4.204) as

$$\frac{D\mathbf{v}}{Dt} + f\mathbf{e}_r \times (\mathbf{v} - \mathbf{v}_{\rm g}) = 0 \ . \tag{4.206}$$

or

$$\frac{\partial\mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\,\mathbf{v} + \omega\frac{\partial\mathbf{v}}{\partial p} + f\mathbf{e}_r \times (\mathbf{v} - \mathbf{v}_{\rm g}) = 0 \ . \tag{4.207}$$

As already discussed in Chapter 3, the non-dimensional Rossby number is defined by

$$\mathrm{Ro} \equiv \frac{V}{fL} \ . \tag{4.208}$$

where V and L are, respectively, a wind scale and a horizontal length scale that are characteristic of the types of motion under consideration. Recall from Chapter 3 that $\mathrm{Ro} \ll 1$ for large-scale motions in middle latitudes. Let

$$\beta \equiv \frac{1}{a}\frac{df}{d\varphi} \tag{4.209}$$

denote the rate of change of the Coriolis parameter with latitude. Charney showed that when the conditions

$$Ro \ll 1 \, . \tag{4.210}$$

and

$$\frac{\beta L}{f} \leq Ro \tag{4.211}$$

are satisfied, Eq. (4.207) can be approximated by

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_{\mathrm{g}} \cdot \nabla \right) \mathbf{v}_{\mathrm{g}} + f \mathbf{e}_r \times (\mathbf{v} - \mathbf{v}_{\mathrm{g}}) = 0 \, . \tag{4.212}$$

Eq. (4.212) is the QG form of the horizontal momentum equation. The condition (4.210) means that momentum advection is weak compared to the Coriolis term, and (4.211) means that the Coriolis parameter varies within a fairly small range across the meridional scale of the weather system under consideration, i.e., the weather system is not too "wide" in the meridional direction. In passing from (4.207) to (4.212), the local tendency of the wind has been approximated by the local tendency of the geostrophic wind, horizontal advection of the horizontal wind has been approximated by geostrophic advection of the geostrophic wind. Vertical advection of the horizontal wind has been neglected here, but is included in more complete versions of the QG system.

For the case of no heating, the potential-temperature form of the thermodynamic energy equation, (4.66), can be written using pressure coordinates as

$$\left(\frac{\partial \theta}{\partial t} \right)_p + (\mathbf{v} \cdot \nabla_p) \, \theta + \omega \frac{\partial \theta}{\partial p} = 0 \, . \tag{4.213}$$

In the QG system, Eq. (4.213) is linearized about a "basic state" potential temperature, θ_0, that varies only in the vertical. We write

$$\theta \, (\lambda, \varphi, p, t) = \theta_0 \, (p) + \theta' \, (\lambda, \varphi, p, t) \, . \tag{4.214}$$

where θ' is the departure from the basic state. We approximate (4.213) by

$$\left(\frac{\partial}{\partial t}+\mathbf{v}_{\mathrm{g}}\cdot\nabla_p\right)_p\theta'+\omega\frac{\partial\theta_0}{\partial p}=0\,. \tag{4.215}$$

Here horizontal advection is by the geostrophic wind, and vertical advection acts only on the basic-state potential temperature. Even though vertical advection was neglected in the momentum equation, it has to be retained in the thermodynamic energy equation because the stratification of the Earth's atmosphere is very strong. The use of θ_0 rather than θ in the vertical advection term eliminates any feedback between the vertical velocity and the potential temperature. This filters inertia-gravity waves from the quasi-geostrophic system.

The hydrostatic equation can be written as

$$\begin{aligned}\frac{\partial\phi}{\partial p}&=-\frac{RT}{p}\\&=-\frac{R}{p}\left(\frac{p}{p_0}\right)^{\kappa}\theta\,,\end{aligned} \tag{4.216}$$

from which it follows that

$$\begin{aligned}\frac{\theta'}{\theta_0}&\cong\frac{\partial\phi'}{\partial p}\Big/\frac{\partial\phi_0}{\partial p}\\&=-\frac{1}{\alpha_0}\frac{\partial\phi'}{\partial p}\,,\end{aligned} \tag{4.217}$$

or

$$\theta'=-\frac{\theta_0}{\alpha_0}\frac{\partial\phi'}{\partial p}\,. \tag{4.218}$$

where α_0 is the specific volume of the basic state. With the use of (4.218), we can rewrite (4.215) in terms of the geopotential perturbation:

$$\left(\frac{\partial}{\partial t}+\mathbf{v}_{\mathrm{g}}\cdot\nabla\right)\frac{\partial\phi'}{\partial p}+\omega S_p=0\,. \tag{4.219}$$

Here

$$S_p \equiv -\frac{\alpha}{\theta}\frac{\partial \theta}{\partial p} > 0 \tag{4.220}$$

is called the "static stability."

Next, we form the vorticity equation corresponding to (4.213). It is

$$\frac{\partial \zeta_{\rm g}}{\partial t} + (\mathbf{v}_{\rm g}\cdot\nabla)\,\zeta_{\rm g} + \beta_0 v_{\rm g} + f_0 \nabla\cdot\mathbf{v} = 0\,, \tag{4.221}$$

where

$$\begin{aligned}\zeta_{\rm g} \equiv& \mathbf{e}_r\cdot(\nabla\times\mathbf{v}_{\rm g})\\ \cong& \frac{\nabla^2\phi}{f_0}\end{aligned} \tag{4.222}$$

is the geostrophic vorticity. Compare with (4.187). In the divergence term of (4.221) we have replaced f by f_0 and β by β_0. These are constant parameters appropriate for the "middle" of the latitude band under consideration. To obtain the second line of (4.222), we have neglected a term involving the variation of the Coriolis parameter with latitude. Using (4.222) and the continuity equation,

$$\nabla\cdot\mathbf{v} + \frac{\partial \omega}{\partial p} = 0\,, \tag{4.223}$$

we can rewrite (4.221) as

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_{\rm g}\cdot\nabla\right)\left(\frac{\nabla^2\phi}{f_0} + \beta_0 y\right) - f_0\frac{\partial \omega}{\partial p} = 0\,, \tag{4.224}$$

where we have used

$$\left(\frac{\partial}{\partial t}+\mathbf{v}_{\mathrm{g}}\cdot\nabla\right)y=v_{\mathrm{g}}\,. \tag{4.225}$$

Eq. (4.224) is the QG vorticity equation.

The final step in the derivation of the QG system is to eliminate ω between (4.219) and (4.224). With the use of (4.205), we can show that

$$\frac{\partial}{\partial p}\left[\left(\frac{\partial}{\partial t}+\mathbf{v}_{\mathrm{g}}\cdot\nabla\right)\left(\frac{1}{S_p}\frac{\partial\phi}{\partial p}\right)\right]=\left(\frac{\partial}{\partial t}+\mathbf{v}_{\mathrm{g}}\cdot\nabla\right)\left[\frac{\partial}{\partial p}\left(\frac{1}{S_p}\frac{\partial\phi}{\partial p}\right)\right]\,. \tag{4.226}$$

This allows us to rewrite the thermodynamic equation, (4.219), as

$$\left(\frac{\partial}{\partial t}+\mathbf{v}_{\mathrm{g}}\cdot\nabla\right)\left[\frac{\partial}{\partial p}\left(\frac{1}{S_p}\frac{\partial\phi}{\partial p}\right)\right]+\frac{\partial\omega}{\partial p}=0\,. \tag{4.227}$$

Elimination of $\partial\omega/\partial p$ between (4.224) and (4.227) gives

$$\left(\frac{\partial}{\partial t}+\mathbf{v}_{\mathrm{g}}\cdot\nabla\right)\mathrm{QGPPV}=0\,, \tag{4.228}$$

where

$$\mathrm{QGPPV}=\frac{\nabla^2\phi}{f_0}+\beta_0 y+\frac{\partial}{\partial p}\left(\frac{f_0}{S_p}\frac{\partial\phi}{\partial p}\right)\,. \tag{4.229}$$

is called the QG pseudo-potential vorticity, and (4.228) is called the QGPPV equation. If the QGPPV is known, and with the provision of appropriate boundary conditions, we can solve (4.229) as a boundary-value problem for ϕ. Once ϕ has been determined, we can obtain the geostrophic wind from (4.205), and the potential temperature from (4.218). We say that the QGPPV can be "inverted" to solve for the winds and temperature (e.g., Hoskins et al., 1985). The QGPPV is thus the key to QG dynamics.

We will use the QG system in Chapters 10 and 11.

4.18 The shallow water equations

Finally, we briefly discuss the shallow water equations, which will also be used in Chapters 10 and 11. The shallow water equations are two-dimensional system based on three assumptions or idealizations.

The first assumption is that the fluid is incompressible, and the density is uniform everywhere and for all time. As discussed earlier, water is much less compressible than air; the assumption of incompressibility partially explains the name of the shallow-water system. The assumption of incompressibility has two important consequences. First, the continuity equation simplifies to

$$\nabla \cdot \mathbf{V} = 0 \,. \tag{4.230}$$

See (4.6). Because the three-dimensional divergence vanishes, convergence (divergence) in the horizontal must be accompanied by divergence (convergence) in the vertical. We assume that the fluid is bounded below by an impermeable surface (like the Earth's surface) of possibly variable height h_T, and bounded above by a "free surface" (like the surface of a lake) at height $h_T + h$. See Fig. 4.7. With these assumptions, (4.230) leads to a continuity equation of the form

$$\frac{\partial h}{\partial t} + \nabla \cdot (h\mathbf{v}) = 0 \,, \tag{4.231}$$

where h is the depth of the fluid.

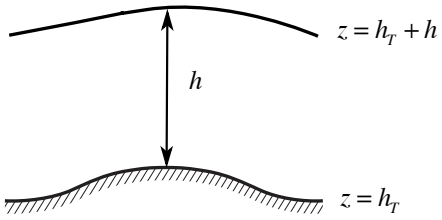


Figure 4.7: Schematic illustrating shallow water flowing over a "mountain."

The second consequence of incompressibility is that the expansion-work term vanishes in the mechanical energy and thermodynamic energy equations [see (4.85) and (4.59)], so that there is no conversion between internal energy and mechanical energy. The internal energy can therefore be ignored, and there is no need for a thermodynamic energy equation,

or a temperature. Any kinetic energy that is dissipated is just "gone," so when dissipative friction is included the shallow water equations do not conserve total energy.

The second assumption is that there is no vertical shear of the horizontal wind. This is where "shallowness" comes in.

The third assumption is that the motion is quasi-hydrostatic, so that the pressure at a given level is proportional to the depth of the water above that level, which varies only due to changes in the height of the free surface.

Using the first and third assumptions, the horizontal pressure-gradient force can be written as

$$\begin{aligned} \mathbf{HPGF} &= -\frac{1}{\rho}\nabla\left[g\rho\left(h_T+h\right)\right] \\ &= -g\nabla\left(h_T+h\right) \ , \end{aligned} \tag{4.232}$$

where h_T is the height of the bottom topography relative to an arbitrary reference level, so that $h_{\rm fs} \equiv h_T + h$ is the height of the free-surface. We can then write the shallow-water form of the horizontal momentum equation as

$$\frac{D\mathbf{v}}{Dt} + f\mathbf{e}_r \times \mathbf{v} = -g\nabla h_{\rm fs} \ . \tag{4.233}$$

A friction term can be added if desired.

4.19 Segue

This chapter has outlined the conservation principles for momentum, mechanical and thermodynamic energy, and potential vorticity. Energy conversion processes have been discussed in detail. We have also discussed a system of equations in which the potential temperature is used as a vertical coordinate. The importance of vorticity has been emphasized. Finally, we have introduced important approximations, including the primitive equations, the quasi-geostrophic system, and the idealized shallow-water system.

These ideas will be used throughout the following chapters. Next up, in Chapter 6, we discuss the balance requirements for moisture, energy, and angular momentum, from the point of view of the zonally averaged circulation.

4.20 Problems

1. Derive the surface pressure tendency equation, (3.14), by starting from (3.19), which is the continuity equation in pressure coordinates. *Do not* use a coordinate transformation.

2. Prove the following about the unit vectors in spherical coordinates:

$$\begin{aligned}
\nabla \cdot \mathbf{e}_\lambda &= 0, \\
\nabla \cdot \mathbf{e}_\varphi &= -\tan\varphi / r, \\
\nabla \cdot \mathbf{e}_r &= 2/r\,, \\
\nabla \times \mathbf{e}_\lambda &= \mathbf{e}_\varphi / r + (\tan\phi / r)\,\mathbf{e}_r, \\
\nabla \times \mathbf{e}_\varphi &= -\mathbf{e}_\lambda / r, \\
\nabla \times \mathbf{e}_r &= 0\,.
\end{aligned} \tag{4.234}$$

3. Starting from the equation of motion, calculate the radius of a geostationary orbit. You will have to take into account the variation of the Earth's gravity with distance from the center of the Earth.

4. Prove that

$$\mathbf{e}_r \cdot (\nabla \times \mathbf{V}_e) = 2\Omega \sin\varphi\,. \tag{4.235}$$

5. (a) Suppose that you can bench press 100 kg at the North Pole. How many kilograms can you bench press at the equator? Assume that the Earth is a perfect sphere and that the acceleration due to the Earth's gravity is horizontally uniform.

   (b) What length of day would be required to make the centrifugal acceleration cancel the gravitational acceleration at the Earth's surface, at 40°N?

6. Prove that for a closed volume

$$\int_V \rho \frac{DA}{Dt} dV = \frac{d}{dt} \int_V \rho A dV\,. \tag{4.236}$$

   Here A is an arbitrary scalar.

7. Show that for an isentropic process

$$\frac{p}{p_0} = \left(\frac{\rho}{\rho_0}\right)^{\gamma} = \left(\frac{T}{T_o}\right)^{\frac{\gamma}{\gamma-1}} , \tag{4.237}$$

where subscript "0" denotes a reference state, and $\gamma \equiv c_p/c_v$.

8. Prove that in a hydrostatic atmosphere the dry static energy, s, satisfies

$$\frac{\partial s}{\partial z} \cong \Pi \frac{\partial \theta}{\partial z} . \tag{4.238}$$

9. Prove that for an isothermal atmosphere (T constant with height), the static stability is given by

$$S_p = \frac{\kappa R T}{p^2} . \tag{4.239}$$

10. Derive an expression for the isentropic pseudodensity $\rho_\theta(\theta)$ for the special case of an isothermal (uniform temperature) and hydrostatic atmosphere. Assuming a surface pressure of 1000 hPa, and a surface temperature of 270 K, plot $\rho_\theta(\theta)$ for θ in the range zero to infinity.

11. Define the axial component of the angular momentum by

$$M \equiv r\cos\varphi\left(\Omega r\cos\varphi + u\right) , \tag{4.240}$$

where r is the radial distance of the parcel from the center of the Earth.

(a) Consider a parcel at rest on the equator and at the surface where $r = a$. Show that, if the parcel is moved to radius r_1, without changing its angular momentum, its zonal velocity is approximately given by

$$u \cong -2\Omega\left(r_1 - a\right) . \tag{4.241}$$

(b) If the parcel rises from the surface to an altitude of 15 km, conserving its angular momentum, what is its zonal velocity?

12. Show that the *three-dimensional* pressure-gradient force can be written as $-\theta\nabla\Pi$. Do not use the hydrostatic approximation.

13. Prove that $\omega\alpha = \theta D\Pi/Dt$.

14. Using the hydrostatic approximation, show that the isentropic pseudodensity is zero if the dry static energy is a linear function of the potential temperature.

15. As discussed in the text, the centrifugal acceleration causes the isosurfaces of the geopotential to bulge at the equator, relative to the poles. *Calculate* the difference, at the Earth's surface, in kilometers. Assume for simplicity that g is uniform.

16. Using

$$f\mathbf{e}_r \times \mathbf{v}_\mathrm{g} = -\nabla\phi \ . \tag{4.242}$$

prove that

$$\frac{\partial}{\partial p}\left[\left(\frac{\partial}{\partial t} + \mathbf{v}_\mathrm{g}\cdot\nabla\right)\left(\frac{1}{S_p}\frac{\partial\phi}{\partial p}\right)\right] = \left(\frac{\partial}{\partial t} + \mathbf{v}_\mathrm{g}\cdot\nabla\right)\left[\frac{\partial}{\partial p}\left(\frac{1}{S_p}\frac{\partial\phi}{\partial p}\right)\right] \ . \tag{4.243}$$

17. Prove that the solenoidal term of (4.162) vanishes if the density depends only on the pressure.

18. Starting from (4.230), derive (4.231).

Chapter 5

Surface fluxes and the turbulent boundary layer

5.1 Small-scale but important

As discussed in Chapter 13, the large-scale circulation is approximately two-dimensional. This includes eddies with horizontal scales larger than about 500 km (about one tenth of the radius of the Earth). Truly three-dimensional eddies occur on scales ranging from about a centimeter to tens of kilometers. Despite their small size, these eddies play a major role in the atmospheric circulation, as shown already in Fig. 2.14. Turbulence can occur anywhere, but it is most active and most important in the boundary layer, inside clouds of all kinds, and where larger-scale eddies including gravity waves are "breaking."

Turbulence is a very challenging subject. In this chapter, we restrict ourselves to a discussion of the turbulent surface fluxes and the dissipation of kinetic energy in the boundary layer. Chapter 8 deals with with the deeper but still three-dimensional eddies associated with cumulus convection. Chapter 13 discusses the basic nature of turbulence in more detail.

5.2 What is reanalysis?

Numerical weather prediction centers have been producing global analyses of the atmospheric circulation since the 1970s (Bauer et al., 2015). These analyses are used to initialize operational weather forecasts. They are produced using data assimilation with the same forecast models. The analysis archives document the circulation and its changes over the past half century.

The data sources used have become much larger and more accurate over time. The models have also evolved drastically over time. In a reanalysis, a current model is used to re-analyze the past data, so that apparent changes in the circulation are not due to changes in the model (Bengtsson and Shukla, 1988; Bengtsson et al., 2007; Uppala et al., 2005; Kalnay et al., 1996; Gelaro et al., 2017; Dee et al., 2011; Hersbach et al., 2020; Bangalath

and Pauluis, 2020). Of course, changes in the data sources can still lead to spurious apparent changes in the reanalyzed circulation.

The models used to create the analyses calculate many quantities that are not well observed. The surface fluxes are an important example of this. The surface fluxes produced by the models are influenced by the assimilated observations, but they area also strongly influenced by the parameterizations used, which are very imperfect. In this chapter, we will present surface flux maps produced by reanalysis. It is important to remember that they are not true observations.

5.3 The surface fluxes

5.3.1 The surface moisture flux

The surface latent heat flux, $L\left(F_q\right)_S$, is the surface evaporation rate multiplied by the latent heat of condensation. It satisfies a so-called "bulk aerodynamic formula" of the form

$$L\left(F_q\right)_S = L\rho_S c_T \left|\mathbf{v}_S\right| \left(q_g - q_a\right) \ . \tag{5.1}$$

Here c_T is a nondimensional transfer coefficient (e.g., Stull, 2012) that has a typical value on the order of 10^{-3}. Over a water surface, q_g is just the saturation mixing ratio evaluated using the surface temperature of the water and the surface air pressure. The "effective" value of q_g over land is much more difficult to determine because it depends on such things as the soil moisture, the amount of vegetation, and the state of the vegetation (e.g., whether or not photosynthesis is occurring). Maps of the surface latent heat flux were shown already in Fig. 5.1. Note the maxima over the subtropical oceans. There are no strong downward surface moisture fluxes, but weak downward fluxes do occur, and lead to the formation of dew or frost.

5.3.2 The surface sensible heat flux

The sensible heat flux is the small-scale flux of dry enthalpy; as discussed in Chapter 2, it is, roughly speaking, a flux of temperature. The surface sensible heat flux satisfies a bulk aerodynamic formula of the form

$$\left(F_s\right)_S = c_p \rho_S c_T \left|\mathbf{v}_S\right| \left(T_g - T_a\right) \tag{5.2}$$

The dimensions of F_s and F_h are energy per unit area per unit time. The subscript g in (5.2) denotes a value representative of the lower boundary (ground or ocean), and the subscript

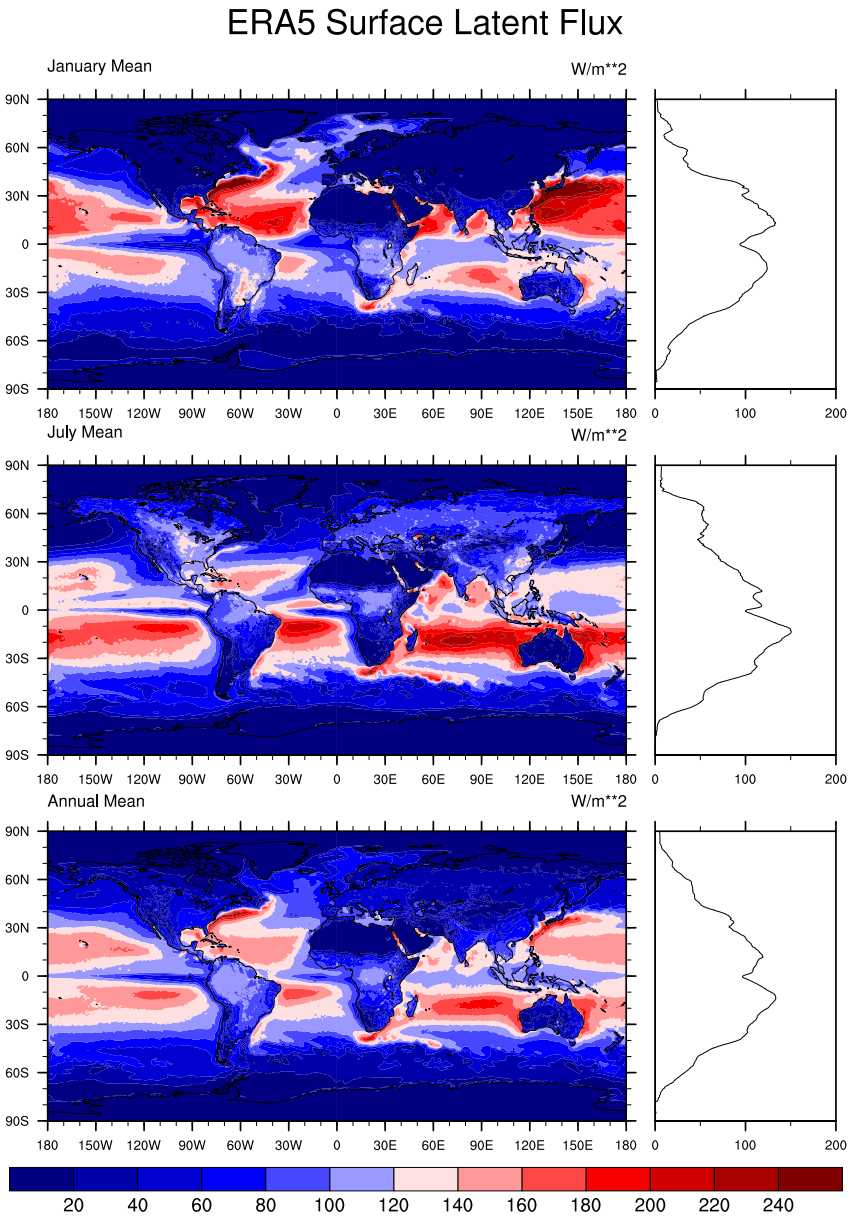


Figure 5.1: Maps of the surface latent heat flux, for January, July, and the annual mean, based on the ECMWF reanalysis. The corresponding zonal averages are shown on the right. These data are not real observations, although they are influenced by observations.

a represents a value representative of a level inside the atmosphere but near the surface. As can be seen from (5.2), the surface sensible heat flux is upward when the ground or ocean is warmer than the air. It tends to cool the ground and warm the air, thus trying to put itself

out of business. The absorption of solar radiation by the ground is an example of a process that can maintain an upward $(F_s)_S$ over time.

Fig. 5.2 shows maps and zonal averages of the surface sensible heat flux for January and July. The largest values occur over the midlatitude oceans in winter, near the eastern coasts of the continents. These strong sensible heat fluxes are associated with fast currents of cold air moving from the cold winter continents out over warm ocean currents such as the Gulf Stream and the Kuroshio. Eq. (5.2) shows that strong sensible heat fluxes are to be expected under such conditions. Large values of $(F_s)_S$ also occur over the summer and tropical continents, especially where the surface is dry and lacking in vegetation, e.g., over the Sahara desert. There are no large negative values of the surface sensible heat flux anywhere, because a downward sensible heat flux tends to damp the turbulence; the buoyancy force does not "want" warm air to move downward or cold air to move upward.

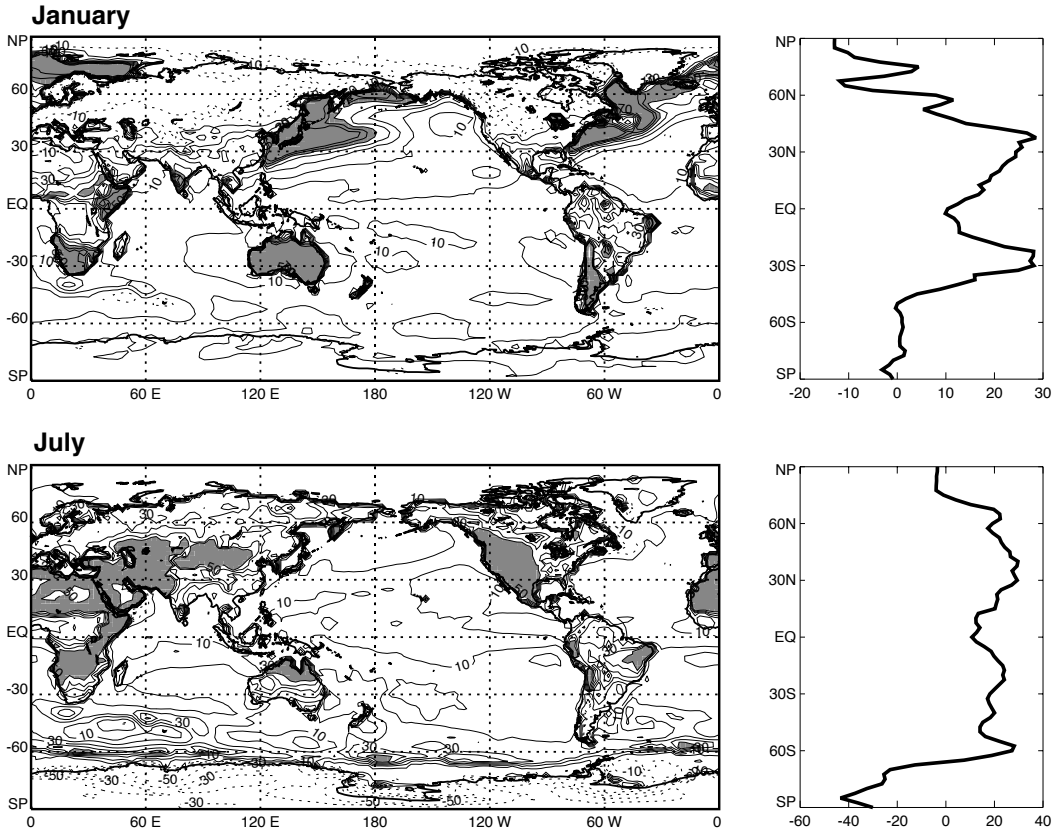


Figure 5.2: Maps of the surface sensible heat flux, for January and July, based on ECMWF reanalyses. The contour interval is 10 W m^{-2}. Values higher than 50 W m^{-2} are shaded. The corresponding zonal averages are shown on the right. These data are not real observations, although they are influenced by observations.

5.3.3 The surface momentum flux

Fig. 5.3 shows maps of the observed magnitude of the surface wind stress over the oceans. These data were obtained using a centimeter-wavelength radars on satellites, which can sense stress-induced centimeter-scale "capillary waves" on the sea surface. The retrieval algorithm works over water surfaces, but not over land. Unfortunately, we have no comparable observations of the wind stress over land. The plots in Fig. 5.3 show the average of the magnitude of the vector, rather than the magnitude of the average of the vector. The surface stress is particularly strong in the storm-track regions. In Chapter 3, we discussed the belt of low pressure and strong low-level westerlies north of Antarctica. Fig. 5.3 shows that the surface wind stress is particularly strong along that entire latitude belt, especially in July, and of course the time average of the zonal component of the wind-stress is particularly strong there. The time-averaged tropical wind stresses are relatively weak, with the exception of a maximum over the Arabian Sea in July, associated with the Somali jet.

5.4 Energy transports and dissipation due to small-scale motions

5.4.1 Work done on the ocean and land surface

Energy exchange occurs when the surface wind stress pushes on the ocean, giving rise to the "wind-driven" ocean circulation. The rate at which the atmosphere does work on the oceans can be roughly estimated as follows: The surface frictional stress is typically less than or on the order of 0.1 Pa. With a few exceptions, the ocean currents have speeds on the order of 0.1 m s^{-1} or slower. The rate at which the ocean gains energy due to the stress applied by the atmosphere is given by the product of the stress with the speed of the current, which, using the values given above, is of order 10^{-2} W m^{-2} – tiny compared to other energy fluxes that we have discussed, such as the net surface radiation. The work that the surface wind stress does on the ocean is obviously quite important for the ocean and for the climate system as a whole. Nevertheless, from the point of view of the atmospheric energy budget, it is utterly negligible.

The surface stress also does a small amount of work as it rustles the leaves of terrestrial vegetation, and raises dust in arid regions.

5.4.2 Dissipation

As discussed in Chapter 4, energy dissipation associated with the molecular viscosity converts macroscopic kinetic energy into microscopic kinetic energy. or "heat." Dissipation is important in the turbulent PBL, and also in regions of precipitation (Pauluis et al., 2000). As discussed in Chapter 4, the dissipation rate appears as a weak but persistent source of internal energy. The global mean of $\rho\delta$ is believed to be about 5 W m^{-2}. It is an amazing fact that even though these molecular processes act on scales of a few millimeters, they play fundamental roles in the energetics of the global-scale circulation of the atmosphere!

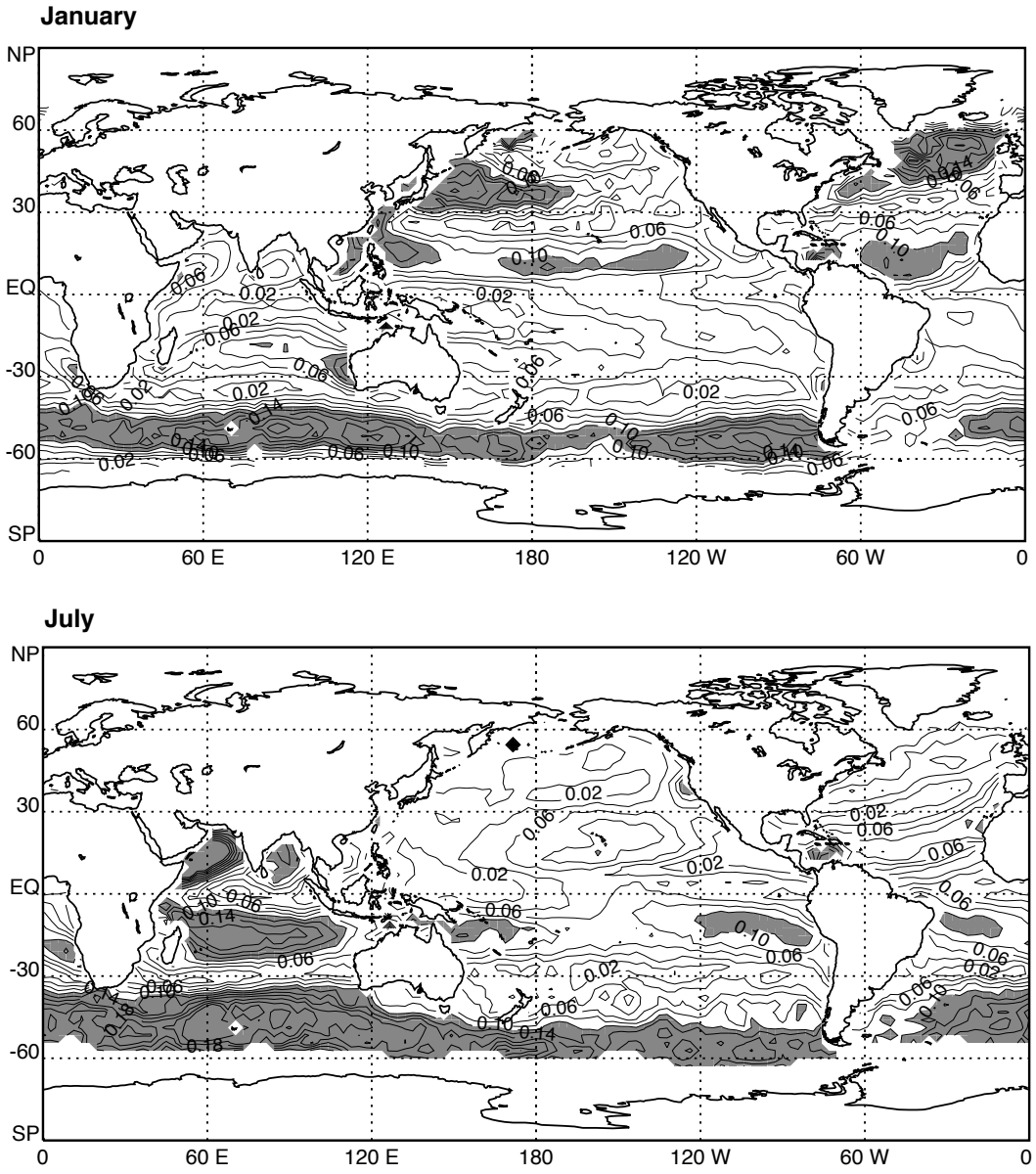


Figure 5.3: Maps of the magnitude of the surface wind stress over the oceans only, based on scatterometer data from ERS-1 (e.g., Liu, 2002). The contour interval is 0.02 Pa. Values higher than 0.1 Pa are shaded.

We can approximate the dissipation rate by

$$\rho\delta \cong -\mathbf{F}_{\mathbf{v}} \cdot \frac{\partial \overline{\mathbf{v}}}{\partial z} \,. \tag{5.3}$$

As discussed in Chapter 7, this is an example of what is called a "gradient production" term. Specifically it represents the rate of production of turbulence kinetic energy (TKE) on small scales by conversion from the kinetic energy of the larger-scale mean flow. The physical picture is that the kinetic energy of the mean flow is *converted* to TKE, i.e., it shifts to smaller scales. This is an example of an energy "cascade;" further discussion is given in Chapter 13. The TKE itself is then dissipated in the true, molecular sense.

We can apply (5.3) to estimate the rate of dissipation in the PBL, as follows: Most of the vertical shear of the horizontal wind typically occurs in the lower part of the PBL, where the momentum flux is fairly close to its surface value. We can therefore approximate the integral of (5.3) through the depth of the PBL by

$$\int_{PBL} \rho \delta dz \cong -(\mathbf{F}_\mathbf{v})_S \cdot \mathbf{v}_M \ , \tag{5.4}$$

where $\mathbf{v}_M$ is the mean horizontal wind in the upper part of the PBL, above the near-surface shear layer. The bulk aerodynamic formula tells us that

$$(\mathbf{F}_\mathbf{v})_S = -\rho_S C_D \left|\mathbf{v}_M\right| \mathbf{v}_M \ , \tag{5.5}$$

where $C_D \cong 10^{-3}$ is a nondimensional "drag coeffcient." Substitution of (5.5) into (5.4) gives

$$\int_{PBL} \rho \delta dz \cong \rho_S C_D |\mathbf{v}_M|^3. \tag{5.6}$$

This shows that the rate of dissipation in the PBL increases very strongly as the wind speed increases. For $|\mathbf{v}_M| = 10$ m s^{-1}, we find that $\int_{PBL} \rho \delta dz \cong 1$ W m^{-2}. The typical rate at which kinetic energy is dissipated in the PBL is thus considerably larger than the typical rate at which the atmosphere does frictional work on the ocean.

Bister and Emanuel (1998) point out that very large dissipation rates occur in hurricanes, because of the large wind speeds in such storms, and the cubic dependence of the dissipation rate on wind speed, as shown in (5.6). They argue that dissipative heating acts to intensify the storm, leading to an increase in the maximum wind speed by as much as 25%. Businger and Businger (2001) elaborate on the importance of the dissipation in regions of strong winds, showing that the vertically integrated dissipation rate in a moderately strong storm can be in excess of 1000 W m^{-2}.

5.5 Segue

The turbulent surface fluxes of moisture, energy, and momentum discussed in this chapter are key sources and sinks for the atmospheric circulation. The next chapter shows how the turbulent surface fluxes and other processes influence the transports of moisture, energy, and momentum across broad stretches of latitude.

Chapter 6

Go with the flow

6.1 Overview

The atmosphere is subject to spatially and temporally varying sources and sinks of energy, moisture, and angular momentum at its upper and lower boundaries. The circulation transports energy, moisture, and angular momentum so as to maintain globally averaged balances over time. At the same time, the sources and sinks themselves are affected by the circulation. In this chapter, we show how the circulation satisfies these balance requirements. We focus on the zonally averaged sources, sinks, and vertical and meridional transports, but with an emphasis on the systematic and powerful effects that the zonally varying component of the circulation has on the zonal averages.

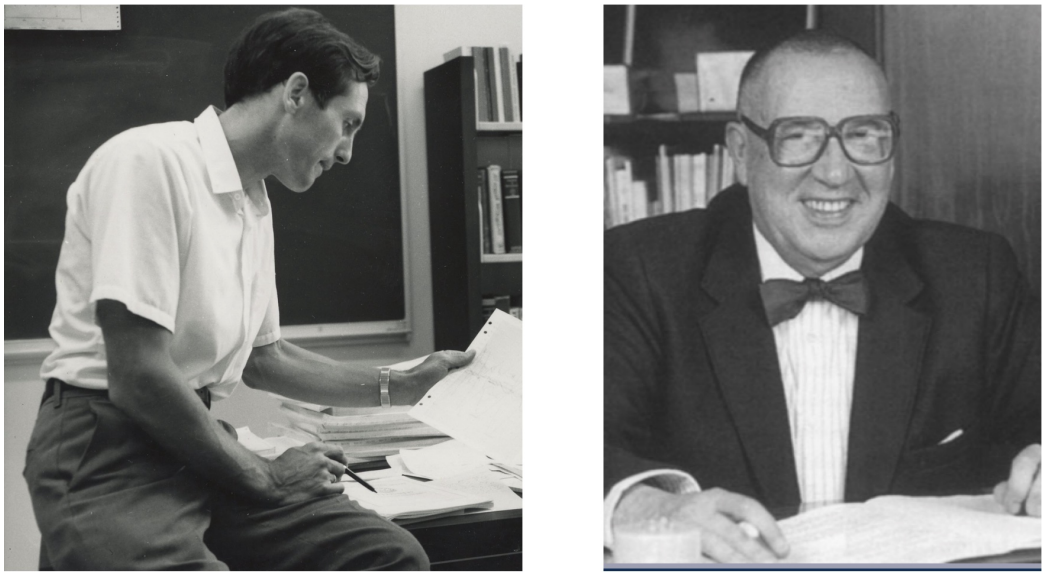
As recounted in the beautiful book by Lorenz (1967), early theories of the global circulation were aimed at explaining the (then very inadequately observed) zonally averaged circulation. Many of the processes that control the zonally averaged circulation depend in an essential way on variations of the winds, temperature, etc., along latitude circles. We cannot understand the zonally averaged circulation without accounting for departures from the zonal mean. As discussed in Chapter 3, we define an "eddy" as a circulation feature that varies with longitude. The important role of eddies in the maintenance of the zonally averaged circulation was only gradually appreciated, by such early scholars as Dove (1837), Defant (1921), Jeffreys (1926), Rossby (1941, 1947), Bjerknes (1948), and Starr (1948), among others. The importance of deep cumulus convection for the global circulation was first understood by Riehl and Malkus (1958), as will be discussed in Chapter 8. Vertical momentum fluxes due to small-scale gravity waves are also important, as discussed in Chapter 11.

During recent decades, the zonally varying structure of the global circulation has been studied intensively for its own sake, rather than just in terms of its effects on the zonally averaged circulation. This more modern perspective will be adopted in Chapter 10 and beyond.

The processes that maintain the observed zonally averaged circulation act on many scales. They involve energy conversions and transports due to radiation and small-scale convection, and transports of angular momentum, energy, moisture, and potential vorticity by a variety of large-scale weather systems. In this chapter, we glimpse those various processes only indirectly, through their effects on the zonally averaged circulation. Later chapters discuss the nature of small-scale convection and large-scale eddies in detail, and explain how they interact with the zonally averaged circulation.

The main goals of this chapter are to present some basic information about the observed global atmospheric circulations of dry air, water, energy, and angular momentum, and to illustrate some useful analysis concepts. Here we emulate the heroic analyses of Peixoto and Oort (1992, see Figure 6.1). Advancing technology and the rise of global forecast centers have made the task much, much easier. The main differences are:

- we have used ECMWF reanalysis products (see 5.2 below) and other modern data sources that did not exist when Peixoto and Oort did their work;
- we have analyzed the data using isentropic coordinates rather than pressure coordinates; and
- we have had the luxury of accessing the data over the Internet and analyzing it using greatly improved computer hardware and software.



Abraham Oort José Peixoto

Figure 6.1: Abraham Oort and José Peixoto, who carried out pioneering observational studies of the global atmospheric circulation.

6.2 An isentropic view of the mass circulation

This chapter makes extensive use of potential temperature as an "isentropic" vertical coordinate; the relevant concepts were introduced in Chapter 4. The many advantages of isentropic coordinates will be pointed out as we go.

The zonally and temporally averaged continuity equation in θ-coordinates is

$$\frac{\partial \overline{\rho_\theta}^\lambda}{\partial t} + \frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{\rho_\theta v}^\lambda \cos\varphi\right) + \frac{\partial}{\partial \theta}\left(\overline{\rho_\theta \dot{\theta}}^\lambda\right) = 0\,. \tag{6.1}$$

In (6.1), it is understood that *the zonal averages, time derivative, and meridional derivative are all taken along isentropic surfaces*. Using the eddy notation introduced in Chapter 3 and Appendix D, the zonally averaged meridional mass flux that appears in (6.1) can be separated into two parts:

$$\boxed{\overline{\rho_\theta v}^\lambda = \overline{\rho_\theta}^\lambda \overline{v}^\lambda + \overline{\rho_\theta^* v^*}^\lambda\,.} \tag{6.2}$$

The first term on the right-hand side of (6.2) is the product of the zonally averaged pseudo-density with the zonally averaged meridional wind. The second term arises from correlated fluctuations of the pseudo-density and the meridional wind. The mass flux $\overline{\rho_\theta^* v^*}^\lambda$ is sometimes called the meridional "bolus mass flux," and $\overline{\rho_\theta^* v^*}^\lambda / \overline{\rho_\theta}^\lambda$ can be called the meridional "bolus velocity."

Here is an analogy: Recall that ρ_θ represents the amount of mass between two isentropic surfaces, per unit horizontal area. Think of ρ_θ as the thickness of a flexible duct that stretches around a latitude circle on a particular θ surface. Think of v as the rate of *meridional* flow through the duct, and assume for simplicity that $\overline{v}^\lambda = 0$, so that the first term on the right-hand side of (6.2) is zero. If the width of the duct and the direction of the flow both fluctuate with longitude in a correlated way, then there can be a net mass transport. For example, suppose that the duct is wider where the flow is toward the north and narrower where the flow is toward the south, in such a way that $\overline{\rho_\theta^* v^*}^\lambda > 0$. Then there will be a net mass flux toward the north, even though $\overline{v}^\lambda = 0$. Eq. (6.2) shows that the mass transport by the zonally averaged meridional wind is supplemented by a kind of "geophysical peristalsis," in which meridional velocity fluctuations are correlated with variations in the pseudo-density, leading to a net eddy mass flux along isentropic surfaces.

Does something similar happen with height coordinates? The zonally averaged horizontal mass flux in height coordinates is $\overline{\rho^* v^*}^\lambda$, where ρ is the (ordinary) density and the

zonal average is taken at a fixed height. We could could decompose ρ so as to define a bolus mass flux in height coordinates, but we don't bother to do so, because variations of ρ on height surfaces are so small that the implied bolus mass flux is negligible. Isentropic coordinates are different. In middle latitudes, especially in winter, large variations of ρ_θ on isentropic surfaces make $\overline{\rho^*_\theta v^*}^\lambda$ large enough to be important, as will be shown below. The variations are small in the tropics, because the horizontal temperature gradients are weak there (as discussed in Chapter 3).

For the reasons explained in Chapter 2, we can neglect the tendency term of (6.1) if we average over a sufficiently long time. This leads to

$$\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\rho_\theta v}^{\lambda,t}\cos\varphi\right)+\frac{\partial}{\partial\theta}\left(\overline{\rho_\theta\dot{\theta}}^{\lambda,t}\right)=0\,. \tag{6.3}$$

Eq. (6.3) shows that heating, represented by $\overline{\rho_\theta\dot{\theta}}^{\lambda,t}$, induces a time-averaged meridional circulation, i.e., $\overline{\rho_\theta\dot{\theta}}^{\lambda,t}\neq 0$ anywhere implies that $\overline{\rho_\theta v}^{\lambda,t}\neq 0$ somewhere. *Without heating there cannot be any time and zonally averaged meridional mass flow on θ-surfaces*. This conclusion follows from mass conservation in θ-space (Dutton, 2002; Townsend and Johnson, 1985; Hsu and Arakawa, 1990; Edouard et al., 1997).

For reasons discussed in Chapter 3, the zonally averaged meridional mass flux is very close to zero when averaged over time and vertically integrated; see Eq. (3.17). The same result can be obtained by integration of (6.3) with respect to θ, over the range $\theta=0$ to $\theta\to\infty$, and using $(\rho_\theta)_{\theta=0}=(\rho_\theta)_{\theta\to\infty}=0$:

$$\overline{\int_0^\infty \rho_\theta v\,d\theta}^{\lambda,t}=0 \text{ at all latitudes.} \tag{6.4}$$

The top two panels of Fig. 6.2 show the distribution of the meridional mass flux, $\overline{\rho_\theta v}^{\lambda,t}$ with latitude and potential temperature, for January (left) and July (right). The middle two panels show the contribution of the "mean" mass flux, $\overline{\rho_\theta}^\lambda\overline{v}^\lambda$, and the bottom two panels show the contributions from the bolus mass flux, $\overline{\rho^*_\theta v^*}^\lambda$, which includes contributions from both stationary and transient eddies. In each panel, there is a "dipole" structure, with a poleward mass flux aloft and an equatorward mass flux below it. In the tropics, the transport by the mean mass flux dominates because ρ^*_θ is small. In mid-latitudes, on the other hand, $\overline{v}^\lambda$ is small due to geostrophy (as discussed in Chapter 3), but ρ^*_θ and the bolus mass flux are large, especially in winter. In mid-latitudes, the bolus mass flux essentially "takes over" from the tropical mean mass flux. The top two panels show that the sum of

$\overline{\rho_\theta}^\lambda \overline{v}^\lambda$ and $\overline{\rho_\theta^* v^*}^\lambda$, i.e., the total zonally averaged mass flux, has about the same magnitude in both the tropics and middle latitudes.

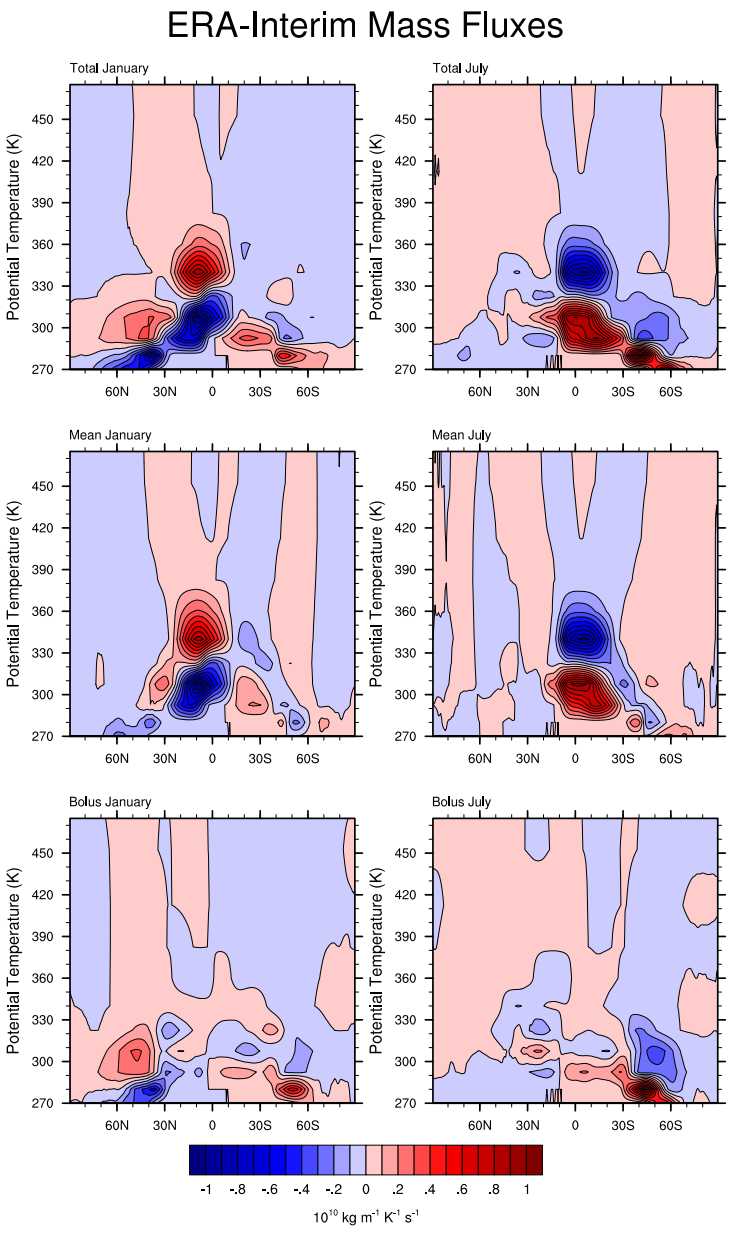


Figure 6.2: Isentropic meridional mass fluxes for January (left) and July (right). The top panels show the total flux, the middle panels show the mean mass fluxes, and the bottom panels show the bolus flux.

Eq. (6.3) shows that changes of $\overline{\rho_\theta v}^{\lambda,t}$ with latitude, along isentropic surfaces, are

indicative of heating or cooling. We can see such changes in Figure 6.2, e.g., between 20°N and 40°N in January. In the upper troposphere, $\overline{\rho_\theta v}^{\lambda,t}$ decreases poleward in January near 20°N, indicating cooling there, and in the lower troposphere it increases poleward near 40°N, indicating heating there. As discussed in Chapters 8 and 10, the subtropical cooling is due to longwave radiation, and the midlatitude heating is due to the combined (and coupled) effects of surface energy fluxes and latent heat release.

A stream function for the mass circulation as seen in θ-coordinates can be defined by

$$\overline{\rho_\theta v}^{\lambda,t} 2\pi a \cos\varphi \equiv -\frac{\partial \psi_\theta}{\partial \theta} \tag{6.5}$$

$$\overline{\rho_\theta \dot{\theta}^{\lambda,t}} 2\pi a^2 \cos\varphi \equiv \frac{\partial \psi_\theta}{\partial \varphi} \, . \tag{6.6}$$

Here the subscript θ on ψ is a reminder that the zonal and temporal averages used to define it have been taken along isentropic surfaces. You should confirm that ψ_θ has the same dimensions (i.e., mass per unit time) as the stream function defined using pressure coordinates in Chapter 3. The sign convention here is consistent with the one used in Chapter 3.

Fig. 6.3 shows plots of the stream function of the seasonally varying stream function of the mean meridional circulation as seen in p-coordinates (right panels) and θ-coordinates (left panels). The actual data used is exactly the same with both coordinate systems. As discussed in Chapter 3, the stream function as seen in p-coordinates features a "large" Hadley cell at each solstice, with its rising branch in the summer-hemisphere tropics and its body extending into the winter-hemisphere subtropics. The peak magnitude of the stream function is about 16 x 10^{10} kg s^{-1}. A weaker Hadley circulation occurs in the summer hemisphere. Both Hadley cells are "direct" circulations. This jargon means that their rising branches are warmer than their sinking branches, so that enthalpy is transported upward. Similar observations were discussed in Chapter 3.

You may be surprised to see, in the left-hand panels of Fig. 6.3, that the stream function plots look very different when θ-coordinates are used. Hadley cells still appear, with somewhat stronger mass circulations, but these direct circulations extend all the way to the poles, especially in the winter hemisphere. Ferrel cells do not appear at all. Here is one part of the explanation: When we use p-coordinates, the zonally averaged potential temperature equation is

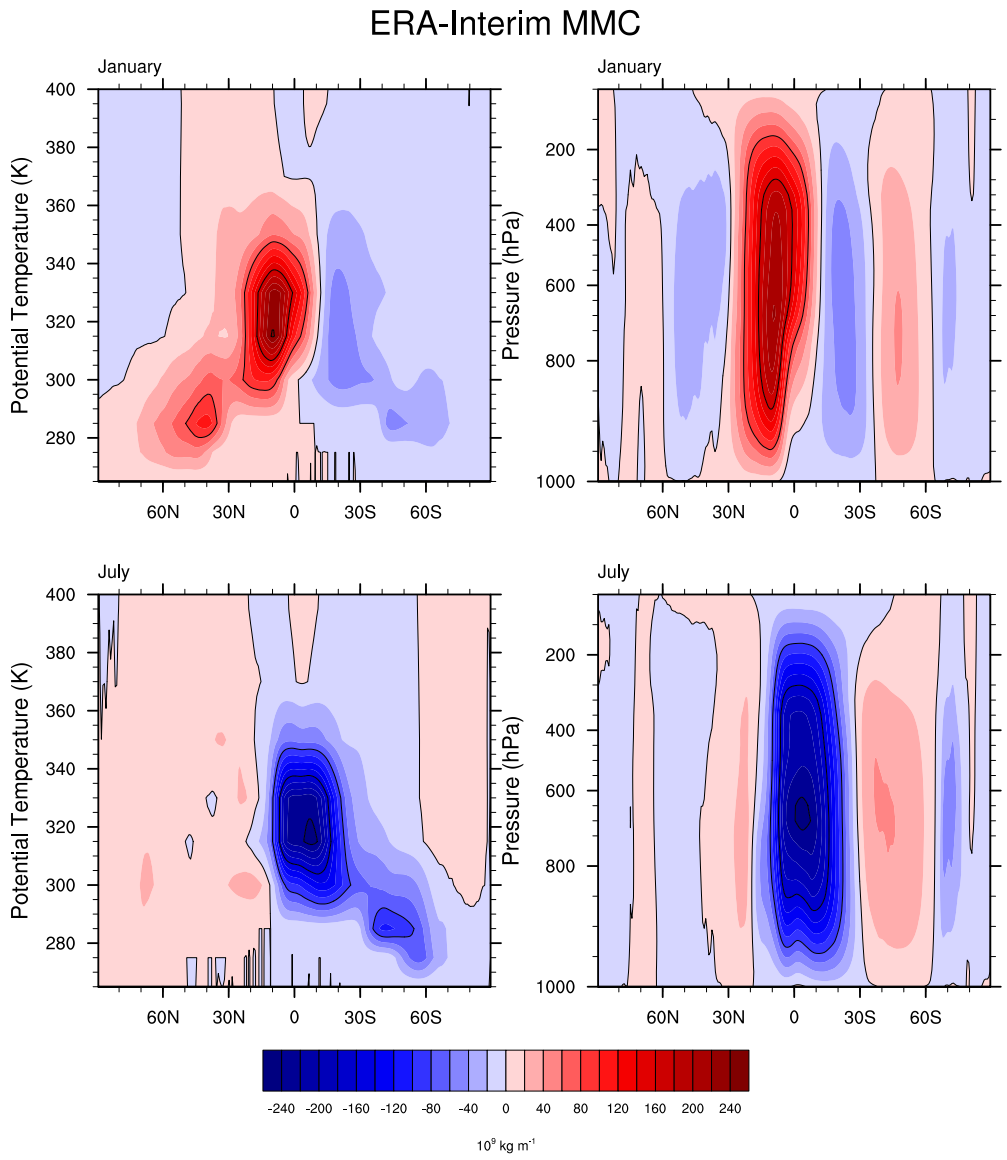


Figure 6.3: The stream function of the seasonally varying mean meridional circulation as seen in p-coordinates (right panels) and θ-coordinates (left panels). Units are 10^9 kg s^{-1}. In the left panels, the curved line near the bottom represents the surface of the Earth, along which the potential temperature is of course a function of latitude. A similar figure was published by Townsend and Johnson (1985).

$$\frac{\partial \overline{\theta}^{\lambda}}{\partial t} + \frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left[\left(\overline{v}^{\lambda}\overline{\theta}^{\lambda} + \overline{v^{*}\theta^{*}}^{\lambda}\right)\cos\varphi\right] + \frac{\partial}{\partial p}\left(\overline{\omega}^{\lambda}\overline{\theta}^{\lambda} + \overline{\omega^{*}\theta^{*}}^{\lambda}\right) = \overline{\dot{\theta}}^{\lambda}. \qquad (6.7)$$

In (6.7), the zonal averages and the temporal and meridional derivatives are taken along isobaric surfaces. We see that meridional fluxes of potential temperature are associated with both the MMC and the eddies. In the tropics and in the subtropics of the winter hemisphere, the MMC is very strong and the fluctuations of potential temperature on isobaric surfaces are quite weak, so the meridional flux of potential temperature along isobaric surfaces is mainly due to the Hadley circulation. In middle latitudes, the Hadley circulation is weak when viewed in pressure coordinates, and the fluctuations of potential temperature on isobaric surfaces are large, especially in the winter hemisphere, due to baroclinic eddies. As a result, the eddy potential temperature flux along isobaric surfaces can and does become large enough to take over the job of transporting potential temperature poleward.

To derive an equation that corresponds to (6.7) in θ-coordinates, simply multiply Eq. (6.1), which is the isentropic-coordinate version of the zonally averaged continuity equation, by θ. Using the fact that the derivatives in (6.1) are taken along isentropic surfaces, on which $\theta^* = 0$, the result can be rearranged to

$$\frac{\partial}{\partial t}\left(\overline{\rho_\theta}^\lambda \theta\right) + \frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{\rho_\theta v}^\lambda \theta \cos\varphi\right) + \frac{\partial}{\partial \theta}\left(\overline{\rho_\theta \dot{\theta}}^\lambda \theta\right) = \overline{\rho_\theta \dot{\theta}}^\lambda, \tag{6.8}$$

which is closely analogous to (6.7). While (6.7) contains "eddy terms," (6.8) does not. Since $\theta^* = 0$ on isentropic surfaces, the eddy flux of potential temperature along isentropic surfaces must be exactly zero. This is one way of understanding why the Hadley circulations have to extend all the way to the poles when depicted in θ-coordinates. The Hadley circulations are the only game in town; there is no eddy flux of θ to hand off to. With isentropic coordinates, we have an eddy mass flux (the bolus flux) but no eddy flux of θ. With pressure coordinates, we have an eddy flux of θ, but no eddy mass flux.

Although the inputs to the calculation of the isentropic stream function are the pseudo-density and the meridional wind, Fig. 6.3 also contains information about atmospheric heating and cooling. Whenever the stream function contours cross isentropic surfaces, heating (for upward motion across isentropic surfaces) or cooling (for downward motion across isentropic surfaces) is implied. For example, Fig. 6.3 shows that there is heating in the tropical rising branches of the Hadley circulations, and cooling in the sinking branches near the poles. As discussed above, there are also signs of cooling in the subtropics, and heating slightly further poleward, in middle latitudes.

The ideas and methods discussed above have been extended by Pauluis et al. (2010) and Pauluis and Mrowiec (2013) to analyze the observed mass circulation in moist entropy space. They find that the mass circulation as seen in moist entropy space is about twice as strong as the mass circulation presented here using θ-coordinates.

6.3 Moisture transports

As discussed in Chapter 4, the total water mixing ratio satisfies

$$\left(\frac{\partial}{\partial t}\right)_z (\rho q_T) + \nabla_z \cdot (\rho \mathbf{v} q_T) + \frac{\partial}{\partial z}\left[\rho w q_T + \left(F_{q_T}\right)_z - P\right] = 0\,, \tag{6.9}$$

where $\left(F_{q_T}\right)_z$ is the upward flux of total water due to small-scale motions, and P is the downward flux of liquid water due to precipitation. We have written the horizontal and vertical advection terms separately and, as indicated by the subscripts on the temporal and horizontal derivatives in (6.9), height is being used as the vertical coordinate. Now we translate (6.9) into θ-coordinates:

$$\left(\frac{\partial}{\partial t}\right)_\theta (\rho_\theta q_T) + \nabla_\theta \cdot (\rho_\theta \mathbf{v} q_T) + \frac{\partial}{\partial \theta}\left[\rho_\theta \dot{\theta} q_T + \left(F_{q_T}\right)_z - P\right] = 0. \tag{6.10}$$

The forms of (6.9) and (6.10) are very similar. From this point on, we will omit the subscripts θ on the temporal and horizontal derivatives, and also the subscript z on F_{q_T}. Averaging (6.10) with respect to longitude and time, we find that

$$\frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{\rho_\theta v q_T}^{\lambda,t}\cos\varphi\right) + \frac{\partial}{\partial \theta}\left(\overline{\rho_\theta \dot{\theta} q_T}^{\lambda,t} + \overline{F_{q_T}}^{\lambda,t} - \overline{P}^{\lambda,t}\right) = 0 \tag{6.11}$$

Finally, vertical integration of (6.11) through the entire depth of the atmosphere gives

$$\frac{\partial}{\partial \varphi}\left(\int_0^\infty \overline{\rho_\theta v q_T}^{\lambda,t}\, d\theta \cos\varphi\right) = \left[\overline{\left(F_{q_T}\right)_S}^{\lambda,t} - \overline{P}_S^{\lambda,t}\right] a\cos\varphi\,. \tag{6.12}$$

Here we have used $(\rho_\theta)_{\theta=0} = (\rho_\theta)_{\theta\to\infty} = 0$, and $\left(F_{q_T}\right)_{\theta\to\infty} = 0$. Because of the vertical integration, Eq. (6.12) would convey exactly the same information regardless of what vertical coordinate system was used in (6.10); the integral on the left-hand side would take slightly different forms with different vertical coordinate systems, but its numerical value would be the same in all cases. Eq. (6.12) shows that meridional changes in the vertically integrated meridional moisture flux balance the surface fluxes of moisture due to turbulence and precipitation.

We can use (6.12) to determine the vertically integrated atmospheric meridional transport of total water as a function of latitude, using the vertical fluxes on the right-hand side as input. In Chapter 2, we used a very similar method to determine the meridional transport of energy by the atmosphere and ocean combined, using the observed net radiation at the top of the atmosphere as input. According to (6.12), moisture must be laterally "imported" into latitude belts where precipitation exceeds evaporation, and it must be exported from latitude belts where evaporation exceeds precipitation. Of course, evaporation and precipitation must balance when averaged over the globe.

Before looking at the meridional transport, we examine the individual terms on the right-hand side of (6.12).

Fig. 6.5 shows maps and zonal means of the surface precipitation rate for January and July. Because precipitation tends to be very "noisy" in both space and time, average values are difficult to determine accurately. The values shown for such remote regions as the South Pacific Ocean cannot be strongly defended, although this situation is gradually improving as more and better data becomes available. It is clear from the figure that the rainiest regions of the world are in the tropics. The tropical maxima are associated with the rising branch of the Hadley circulation and the monsoons. Precipitation minima occur in the subtropics, where many of the Earth's major deserts are found. The seasonal shifts of tropical precipitation are quite spectacular, and are most clearly seen in and near South America, Africa, India, and Southeast Asia. In January, heavy rain falls over the Amazon basin, over southern Africa, the Indian Ocean, the maritime continent north of Australia, in the South Pacific Convergence Zone that extends southeastward from the intersection of the Date Line with the Equator, and across most of the tropical Pacific and Atlantic Oceans north of the Equator. In July, the tropical rains have generally shifted to the north. Heavy rainfall occurs in the extreme northern part of South America, the neighboring Caribbean Sea and tropical North Atlantic Ocean, over northern Equatorial tropical Africa, over India and neighboring regions of Southeast Asia, and to the north of the maritime continent, off the east coast of tropical Asia.

Minima of the zonally averaged precipitation occur in the subtropics. Recall that the descending branch of the Hadley circulation is in the subtropics. The subsidence carries dry air downward and suppresses precipitation.

Secondary precipitation maxima occur in the middle latitudes, and are associated with winter cyclones and summer monsoons. The warm currents off the east coasts of North America and Asia receive heavy precipitation mainly in January. The northwestern portion of the North America receives heavy precipitation in January but not in July.

Regions that receive plentiful precipitation throughout the year include eastern North America, the extreme southern tip of South America, and England.

Fig. 6.6 shows the zonally averaged difference between evaporation and precipitation, for January and July. This is the input needed on the right-hand side of (6.12). Evaporation

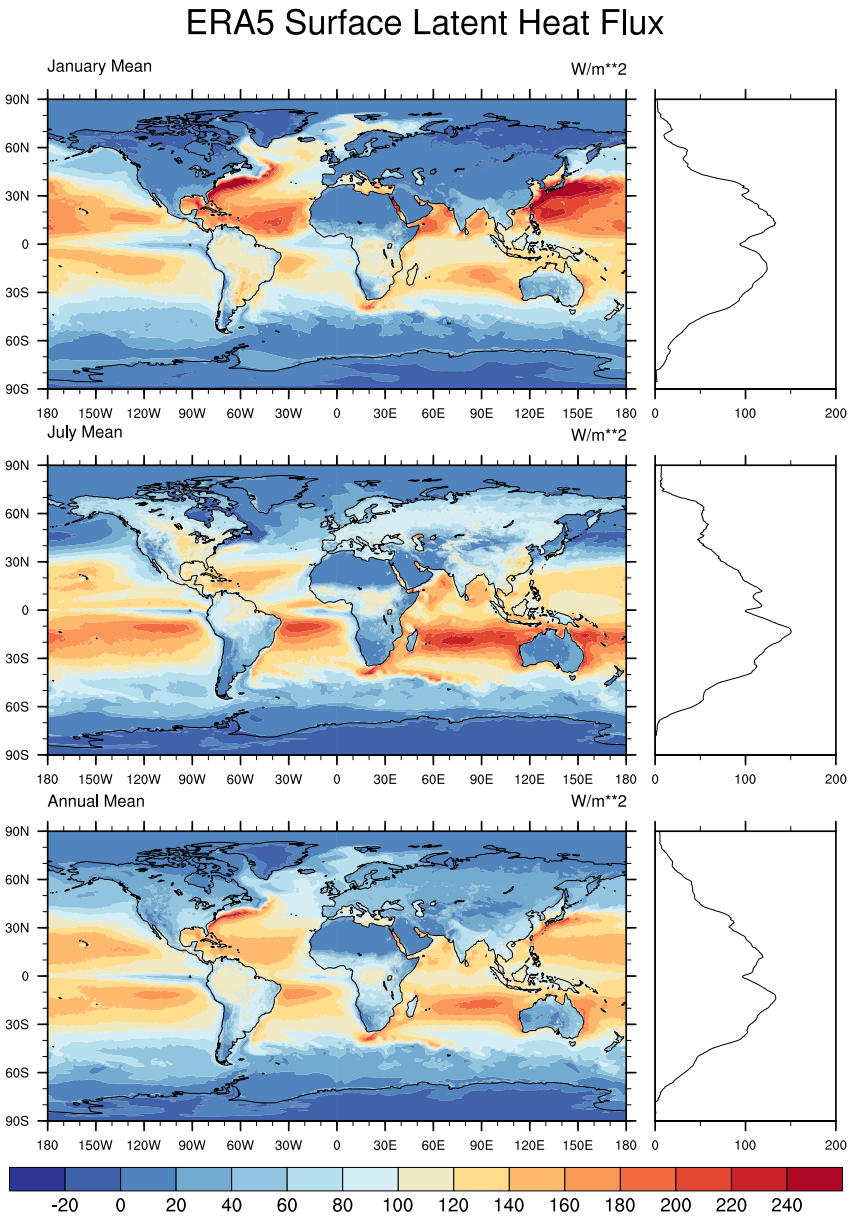


Figure 6.4: Maps of the January, July and annual mean surface latent heat flux, from ERA5. These data are not true observations, although they are influenced by observations.

greatly exceeds precipitation in the subtropics, so in those regions the Earth's surface is a source of total water. In the tropical rain band and the midlatitude storm tracks, precipitation exceeds evaporation, so the atmosphere loses total water.

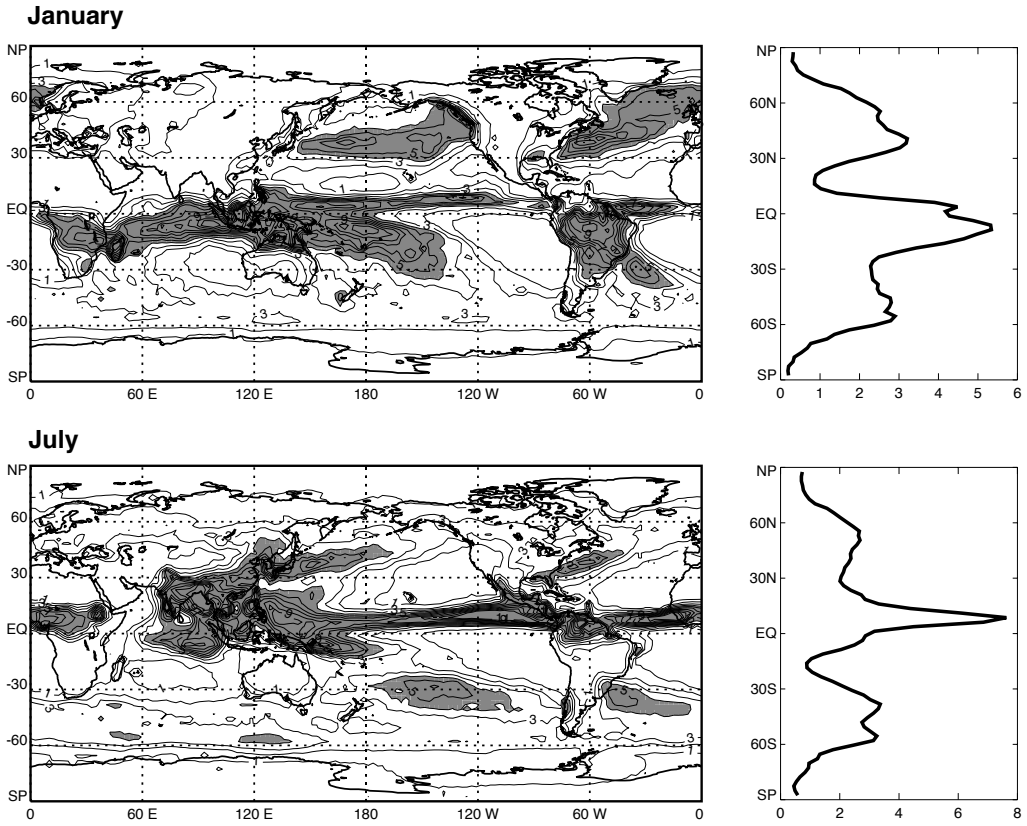


Figure 6.5: Maps of the January and July precipitation rate, from the Global Precipitation Climatology Center (Adler et al., 2003). The contour interval is 1 mm day^{-1}. Values greater than 4 mm day^{-1} are shaded.

We now multiply both sides of (6.12) by $2\pi a$, and integrate with respect to latitude from the South Pole to an arbitrary latitude φ, to obtain

$$\boxed{2\pi a \cos\varphi \int_0^\infty \overline{\rho_\theta v q_T}^{\lambda,t}\, d\theta = 2\pi a^2 \int_{-\frac{\pi}{2}}^{\varphi} \left[\overline{\left(F_{q_T}\right)_S}^{\lambda,t} - \overline{P}^{\lambda,t} \right] \cos\varphi' d\varphi' \,.} \tag{6.13}$$

The quantity on the left-hand side of (6.13) is the total northward transport of moisture across latitude φ. To obtain (6.13), we have used the boundary condition that the northward transport is zero at the South Pole. We used the same method in Chapter 2 to determine the northward transport of energy by the atmosphere and ocean combined, based on observations of the Earth's radiation budget. Compare (6.13) with Eqs. (2.7) and (2.8). For

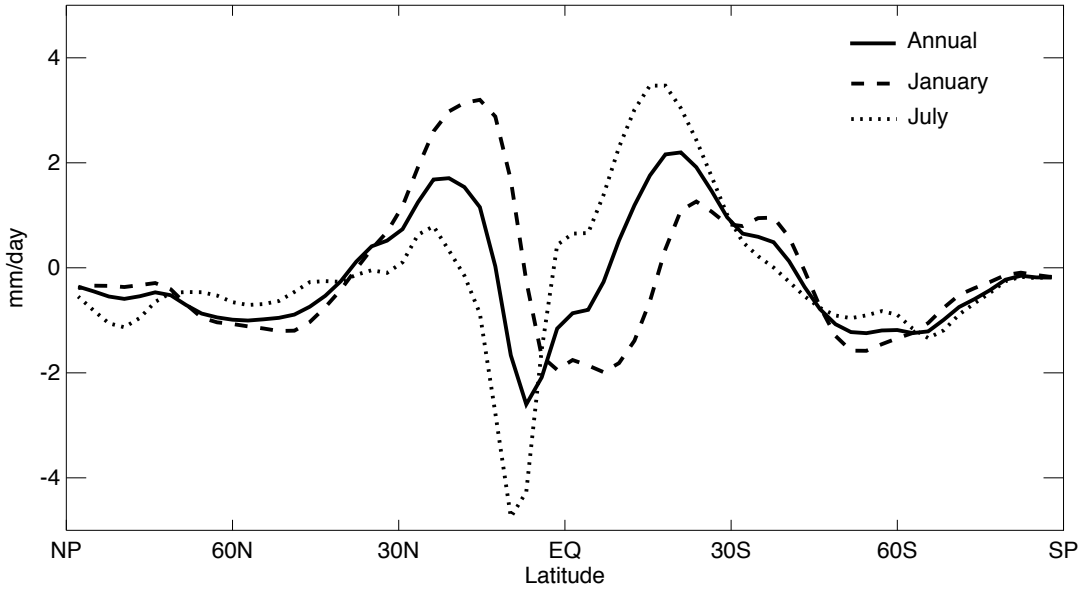


Figure 6.6: The zonally averaged difference between evaporation and precipitation, for January and July, and the annual mean.

reasons discussed in Chapter 2, we have to correct $\overline{\left(F_{q_T}\right)_S}^{\lambda,t} - \overline{P}^{\lambda,t}$, i.e., the time-averaged difference between evaporation and precipitation, to have a global mean of zero before doing the integral in (6.13). The results are shown in Fig. 6.7. The tropical and subtropical latent energy transport is towards the Equator, i.e., it is the "wrong way" to contribute to the poleward energy transport needed to satisfy planetary energy balance. This is expected because most of the water vapor is in the lower troposphere, where the Hadley cells carry it towards the Equator. The midlatitude transport of water vapor (and latent energy) is poleward, but its magnitude is limited by cold temperatures at high latitudes, especially in winter. The ECMWF reanalysis actually includes, as products, the vertically integrated northward transports of both water vapor and liquid water. We could have plotted them in Fig. 6.7, but we chose to show the results of integrating (6.13) in order to illustrate the concept.

With this preparation, we now examine the vertical distribution of the zonally averaged meridional moisture flux, $\overline{\rho_\theta v q_T}^\lambda$, using isentropic coordinates. The flux is the product of three quantities: the pseudodensity, the meridional wind, and the total water mixing ratio. This can be expanded to show four contributions, as follows:

$$\begin{aligned} \rho_\theta v q_T &= \left[\overline{\rho_\theta}^\lambda + (\rho_\theta v)^*\right]\left(\overline{q_T}^\lambda + q_T^*\right) \\ &= \overline{\rho_\theta}^\lambda \overline{q_T}^\lambda + \overline{\rho_\theta}^\lambda q_T^* + (\rho_\theta v)^* \overline{q_T}^\lambda + (\rho_\theta v)^* q_T^* \,. \end{aligned} \tag{6.14}$$

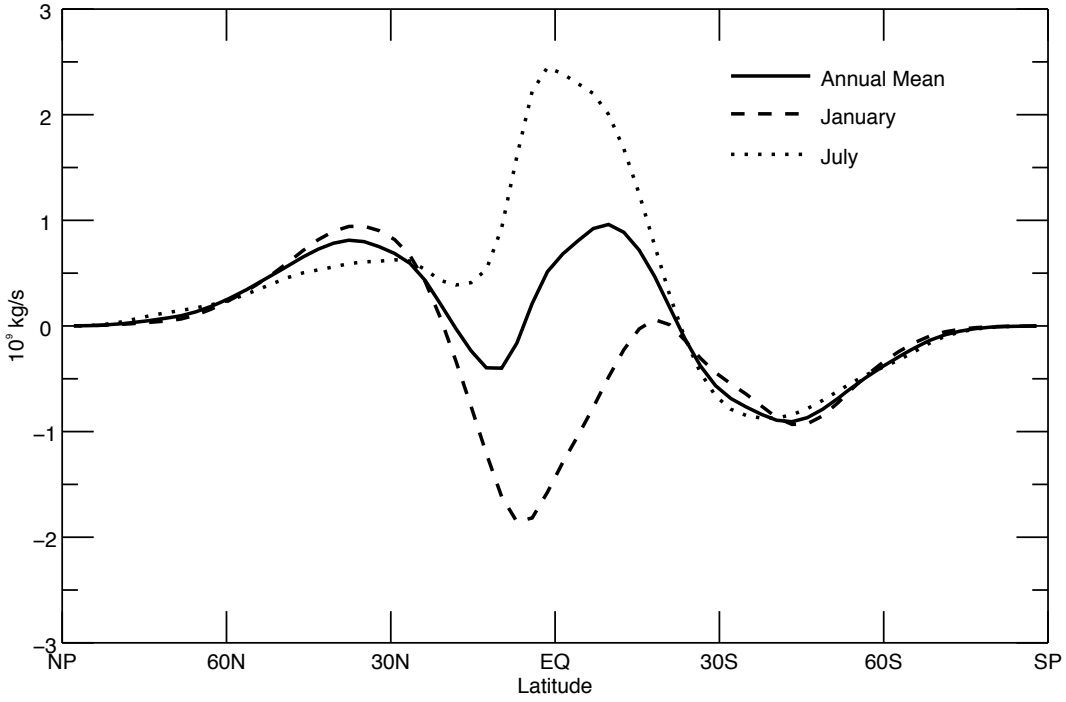


Figure 6.7: The vertically integrated meridional transport of water vapor, for January, July, and the annual mean.

Zonal averaging of (6.14) causes two of the four terms to drop out:

$$\overline{\rho_\theta v q_T}^\lambda = \overline{\rho_\theta v}^\lambda \overline{q_T}^\lambda + \overline{(\rho_\theta v)^* q_T^*}^\lambda \ . \tag{6.15}$$

Substituting for $\overline{\rho_\theta v}^\lambda$ using (6.2), we find that

$$\boxed{\overline{\rho_\theta v q_T}^\lambda = \left(\overline{\rho_\theta}^\lambda \overline{v}^\lambda + \overline{\rho_\theta{}^* v^*}^\lambda\right) \overline{q_T}^\lambda + \overline{(\rho_\theta v)^* q_T^*}^\lambda} \tag{6.16}$$

Eq. (6.16) shows that the zonally averaged total moisture, $\overline{q_T}^\lambda$, is advected by both the mean mass flux and the bolus mass flux (Gent and Mcwilliams, 1990). The third term of (6.16) can be interpreted as the meridional *eddy* flux of total water, because the mass flux involved, i.e., $(\rho_\theta v)^*$, has a mean value of zero, and the term vanishes if q_T is zonally uniform. The bolus flux of moisture does involve eddies, because it is proportional to the bolus mass flux, $\overline{\rho_\theta{}^* v^*}^\lambda$, but it does not involve the eddy part of the total water, q_T^*. We adopt the following shorthand terminology:

- $\overline{\rho_\theta v q_T}^\lambda$ is the total flux,
- $\overline{\rho_\theta}^\lambda \overline{v}^\lambda \overline{q_T}^\lambda$ is the mean flux,
- $\overline{\rho_\theta{}^* v^*}^\lambda \overline{q_T}^\lambda$ is the bolus flux, and
- $\overline{(\rho_\theta v)^* q_T^*}^\lambda$ is the eddy flux.

Substituting (6.16) into (6.11), we find that

$$\begin{aligned} \frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left\{\left[\overline{\left(\overline{\rho_\theta}^\lambda \overline{v}^\lambda + \overline{\rho_\theta{}^* v^*}^\lambda\right)\overline{q_T}^\lambda}^t + \overline{(\rho_\theta v)^* q_T^*}^{\lambda,t}\right]\cos\varphi\right\} \\ + \frac{\partial}{\partial\theta}\left(\overline{\rho_\theta \dot{\theta} q_T}^{\lambda,t} + \overline{F_{q_T}}^{\lambda,t} - \overline{P}^{\lambda,t}\right) = 0\,. \end{aligned} \tag{6.17}$$

Fig. 6.8 shows the total, mean, bolus, and eddy fluxes of q_T, for both January and July. The moisture flux is appreciable only in the lower half of the troposphere, because the upper troposphere is too cold to contain much water vapor. In each hemisphere, the total moisture flux diverges from the subtropics, and converges into both the tropical and midlatitude precipitation maxima, as expected. The mean flux dominates in the tropics, and is equatorward. The bolus flux is weak. The eddy flux is strong in middle latitudes, and carries moisture poleward, i.e., from more humid regions to drier regions.

We are going to define a stream function for global moisture transport, but first we make an adjustment to Eq. (6.17). Let

$$\overline{q_T}^{\mathrm{GM}} \equiv \frac{\int_0^\infty \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} \rho_\theta q_T \, \cos\varphi d\lambda \, d\varphi \, d\theta}{\int_0^\infty \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} \rho_\theta \, \cos\varphi d\lambda \, d\varphi \, d\theta} \tag{6.18}$$

be the mass-weighted average of q_T over the entire atmosphere. Multiply the temporally and zonally averaged continuity equation, (6.3), by $\overline{q_T}^{\mathrm{GM}}$, and subtract the result from (6.17) to obtain

$$\begin{aligned} \frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left\{\left[\overline{\left(\overline{\rho_\theta}^\lambda \overline{v}^\lambda + \overline{\rho_\theta{}^* v^*}^\lambda\right)\left(\overline{q_T}^\lambda - \overline{q_T}^{\mathrm{GM}}\right)}^t + \overline{(\rho_\theta v)^* q_T^*}^{\lambda,t}\right]\cos\varphi\right\} \\ + \frac{\partial}{\partial\theta}\left[\overline{\rho_\theta \dot{\theta}\left(q_T - \overline{q_T}^{\mathrm{GM}}\right)}^{\lambda,t} + \overline{F_{q_T}}^{\lambda,t} - \overline{P}^{\lambda,t}\right] = 0\,. \end{aligned} \tag{6.19}$$

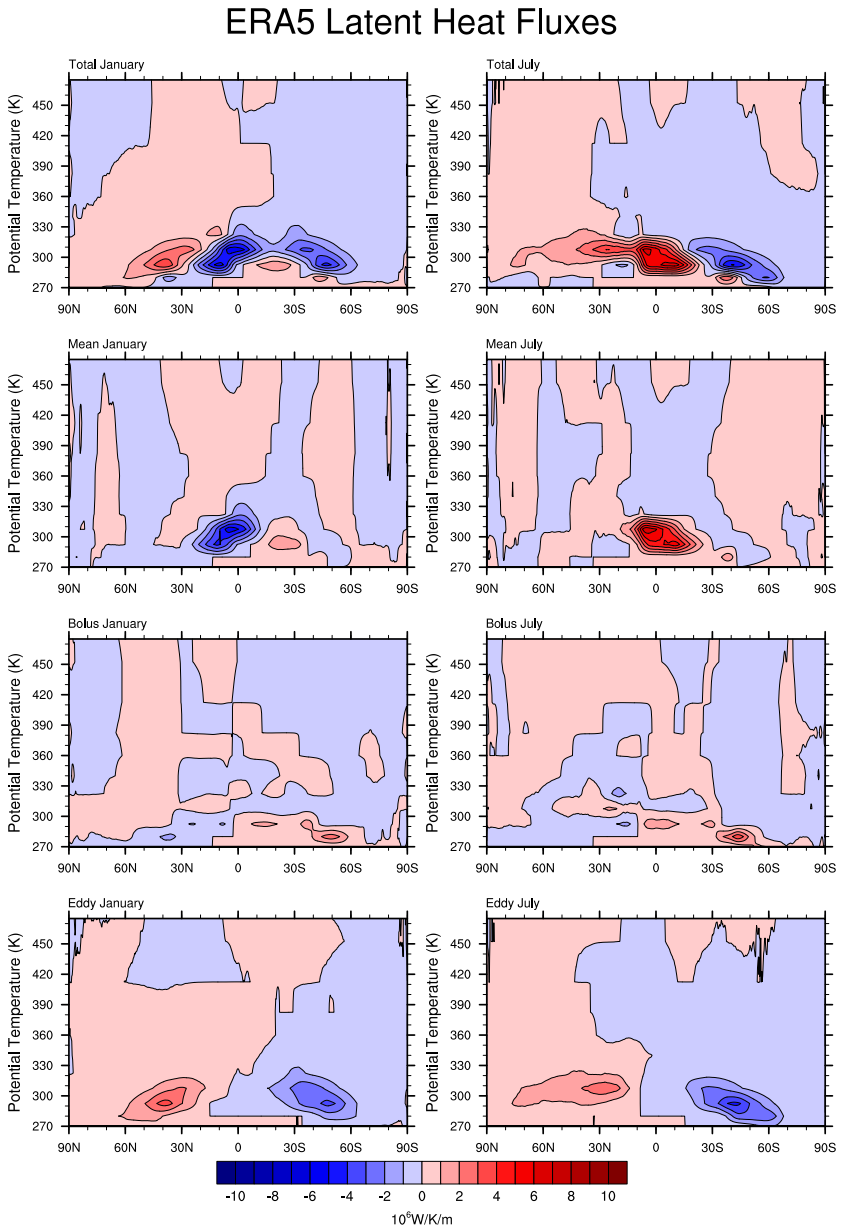


Figure 6.8: Isentropic meridional total water fluxes for January (left) and July (right), multiplied by L and plotted as latent heat fluxees. The top row shows the total flux, the second row shows the flux associated with the zonally averaged meridional wind, the third row shows the bolus flux, and the bottom row shows the eddy flux.

Eq. (6.19) involves both meridional and vertical fluxes of the difference $\overline{q_T}^{\lambda} - \overline{q_T}^{\mathrm{GM}}$, i.e., the difference of $\overline{q_T}^{\lambda}$ from its global mean. Think of $\overline{q_T}^{\lambda}$ as the sum of $\overline{q_T}^{\mathrm{GM}}$ and

$\overline{q_T}^{\lambda} - \overline{q_T}^{\mathrm{GM}}$. The mass circulation carries both $\overline{q_T}^{\mathrm{GM}}$ and $\overline{q_T}^{\lambda} - \overline{q_T}^{\mathrm{GM}}$. The stream function for the transport of $\overline{q_T}^{\mathrm{GM}}$ looks exactly like the stream function for the transport of dry air, multiplied by a constant (namely, $\overline{q_T}^{\mathrm{GM}}$). Any redistribution of $\overline{q_T}^{\mathrm{GM}}$ by the global mass circulation has no effect on the distribution of q_T, simply because $\overline{q_T}^{\mathrm{GM}}$ is spatially constant by definition. For this reason, the flux of $\overline{q_T}^{\lambda} - \overline{q_T}^{\mathrm{GM}}$ is more informative than the flux of $\overline{q_T}^{\lambda}$. Later in this chapter a similar approach will be used with the fluxes of energy and angular momentum.

We are now to define the stream function for the transport of total water:

$$\overline{\rho_\theta v \left(q_T - \overline{q_T}^{\mathrm{GM}}\right)}^{\lambda,t} 2\pi a \cos\varphi \equiv -\frac{\partial \psi_{q_T}}{\partial \theta} \,, \tag{6.20}$$

$$\left[\overline{\rho_\theta \dot{\theta} \left(q_T - \overline{q_T}^{\mathrm{GM}}\right)}^{\lambda,t} + \overline{F_{q_T}}^{\lambda,t} - \overline{P}^{\lambda,t}\right] 2\pi a^2 \cos\varphi \equiv \frac{\partial \psi_{q_T}}{\partial \varphi} \,. \tag{6.21}$$

We calculate ψ_{q_T} by downward integration of (6.20), using the upper boundary condition $\psi_{q_T} = 0$. The integrand is proportional to the zonally averaged meridional flux of total water. This is almost equal to the meridional flux of water vapor, simply because the amount of water vapor in the air is about 100 times larger than the amount of liquid water and ice. Note, however, that the *vertical* transport of total water is very different from the vertical transport of water vapor, because of the strong downward fluxes of precipitating liquid water and ice.

Fig. 6.9 shows contour plots of ψ_{q_T} for January and July. Whereas the stream function contours for the circulation of dry air do not intersect the Earth's surface, the stream function contours for the circulation of total water strongly intersect the Earth's surface because the atmosphere gains moisture there by surface evaporation, and loses it by precipitation. The stream function plots tell us that moisture enters the atmosphere via surface evaporation in the subtropics of the winter hemisphere, flows across the Equator, and then leaves the atmosphere via precipitation in the tropics of the summer hemisphere. It also flows from the subtropics to the midlatitude precipitation zones, mainly in the winter hemisphere. There is very little moisture transport in the arid upper troposphere.

As discussed in Chapter 8, the rising branch of the Hadley circulation contains very strong upward moisture transport by small-scale turbulence and cumulus convection, and very strong downward moisture transport by precipitation. These vertical fluxes extend through almost the entire depth of the troposphere. Fig. 6.9 does not show closely spaced vertical contours of the moisture stream function extending through the troposphere, however. This indicates that through much of the troposphere the upward moisture flux due to

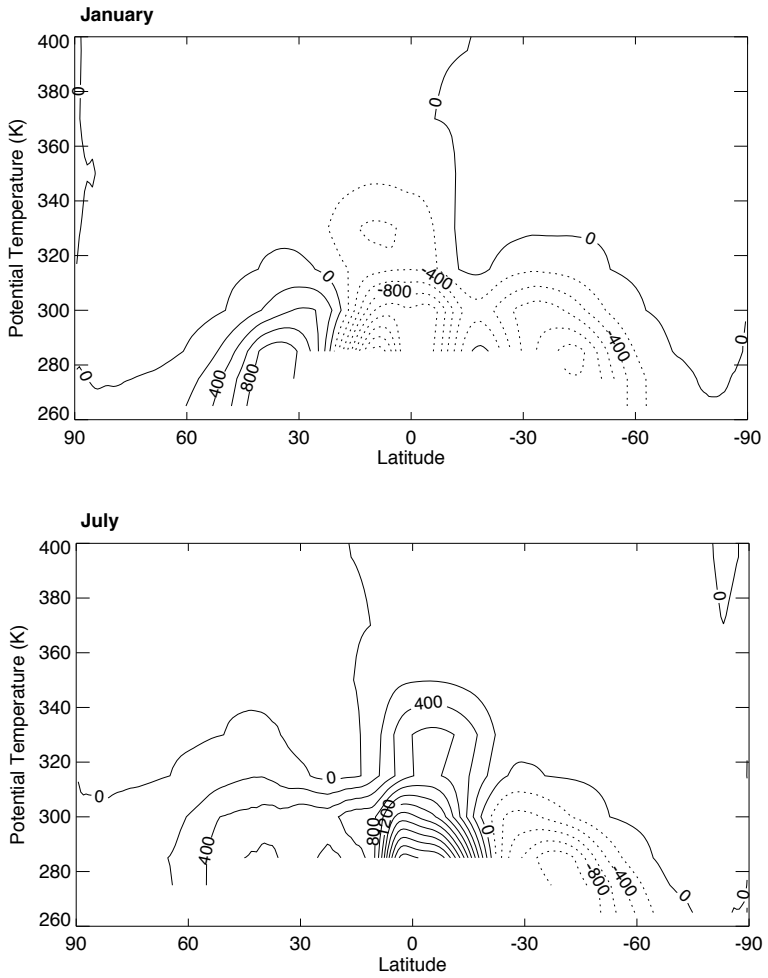


Figure 6.9: The stream function associated with the atmospheric circulation of total water, as seen in isentropic coordinates. Solid and dotted contours represent counterclockwise and clockwise circulations, respectively. The contour interval is 10^6 kg s^{-1}.

turbulence and convection is very nearly balanced by the downward flux due to precipitation.

6.4 Moist static energy transports

We now present a very similar analysis for the meridional transport of moist static energy. As a reminder, we showed in Chapter 4 that the moist static energy is approximately conserved under both moist and dry adiabatic processes, even when precipitation is occurring. Although the moist static energy is not even approximately equal to the total energy, the transport of moist static energy is very nearly equal to the transport of total energy.

Fig. 6.10 shows the latitude-height distribution of moist static energy. In the tropics,

and to some extent in the midlatitudes of the summer hemisphere, the moist static energy decreases upwards in the lower troposphere. Above the middle troposphere, or at any altitude near the poles, the water vapor mixing ratio is negligible. The moist static energy then reduces (approximately) to the dry static energy. As mentioned in Chapter 4, s normally increases upwards, but in the lower troposphere, where water vapor is plentiful, especially in the tropics, the upward decrease of the water vapor mixing ratio dominates, so that h decreases upwards. Above the middle troposphere, the water vapor mixing ratio is so small that $h \cong s$, and as a result h increases upward there. This is why the moist static energy has a minimum in the tropical middle troposphere, as shown in Fig. 6.10. The mid-tropospheric minimum is important and interesting, for reasons that are discussed in Chapter 8.

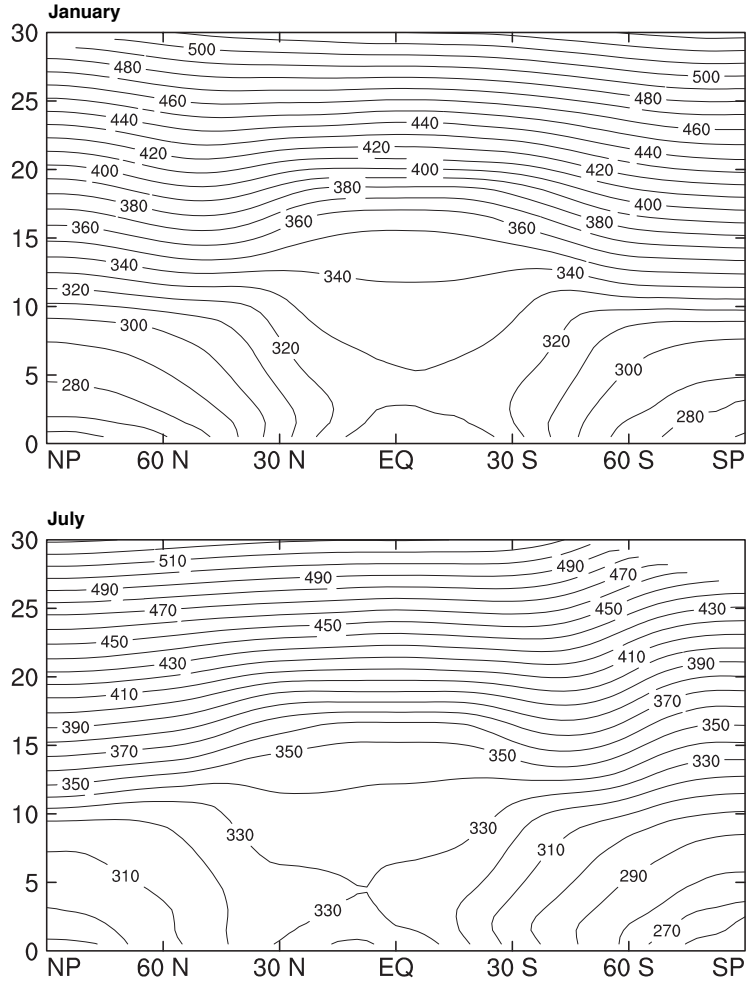


Figure 6.10: The observed latitude-height distribution of the zonally averaged moist static energy, in kJ kg^{-1}, as analyzed by ECMWF.

In isentropic coordinates, the conservation of moist static energy is expressed by

$$\frac{\partial}{\partial t}(\rho_\theta h) + \nabla\cdot(\rho_\theta \mathbf{v} h) + \frac{\partial}{\partial\theta}\left(\rho_\theta \dot{\theta} h + F_h + R\right) = 0\,. \tag{6.22}$$

As a reminder, F_h and R are the (positive upward) fluxes of moist static energy due to small-scale motions and radiation, respectively. Averaging (6.22) over longitude and time, we obtain

$$\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\rho_\theta v h}^{\lambda,t}\cos\varphi\right) + \frac{\partial}{\partial\theta}\left(\overline{\rho_\theta \dot{\theta} h}^{\lambda,t} + \overline{F_h}^{\lambda,t} + \overline{R}^{\lambda,t}\right) = 0 \tag{6.23}$$

Finally, we vertically integrate (6.23) through the entire atmospheric column:

$$\frac{\partial}{\partial\varphi}\left(\int_0^\infty \overline{\rho_\theta v h}^{\lambda,t}\, d\theta\cos\varphi\right) = a\cos\varphi\left(\overline{F_h}_S^{\lambda,t} + \overline{R}_S^{\lambda,t} - \overline{R}_\infty^{\lambda,t}\right)\,. \tag{6.24}$$

Here we have used $(\rho_\theta)_{\theta=0} = (\rho_\theta)_{\theta\to\infty} = 0$, and $(F_h)_{\theta\to\infty} = 0$.

Before presenting maps of the surface moist static energy flux, we start with Fig. 6.11, which shows maps of the surface dry static energy flux (also called the surface sensible heat flux). Compare with the latent heat flux maps shown in Fig. 6.4. Overall, the latent heat flux is much larger than the dry static energy flux. In northern winter, there are large dry static energy fluxes over the Gulf Stream and Kuroshio. These maps are averaged over all times of day. The dry static energy fluxes would be much larger at particular times in particular places, e.g., July afternoons in the Sahara desert. Negative surface dry static energy fluxes can be seen over Antarctica and Greenland.

The surface moist static energy flux can be obtained by adding the surface dry static energy heat flux to the surface latent heat flux; the result is shown in Fig. 6.12. The latent heat flux dominates.

The net atmospheric radiative cooling (ARC) is defined by

$$\mathrm{ARC} = -\left[\left(\overline{R}^{\lambda,t}\right)_S - \left(\overline{R}^{\lambda,t}\right)_\infty\right]\,. \tag{6.25}$$

A positive value of the ARC represents a net radiative cooling of the atmosphere. Fig. 6.13 shows maps of the ARC for January, July, and the annual mean. It is strongest in the winter hemisphere. In northern summer, the zonally averaged ARC is almost bimodal, with a near discontinuity close to the Equator, due to the strong upwelling infrared radiation from

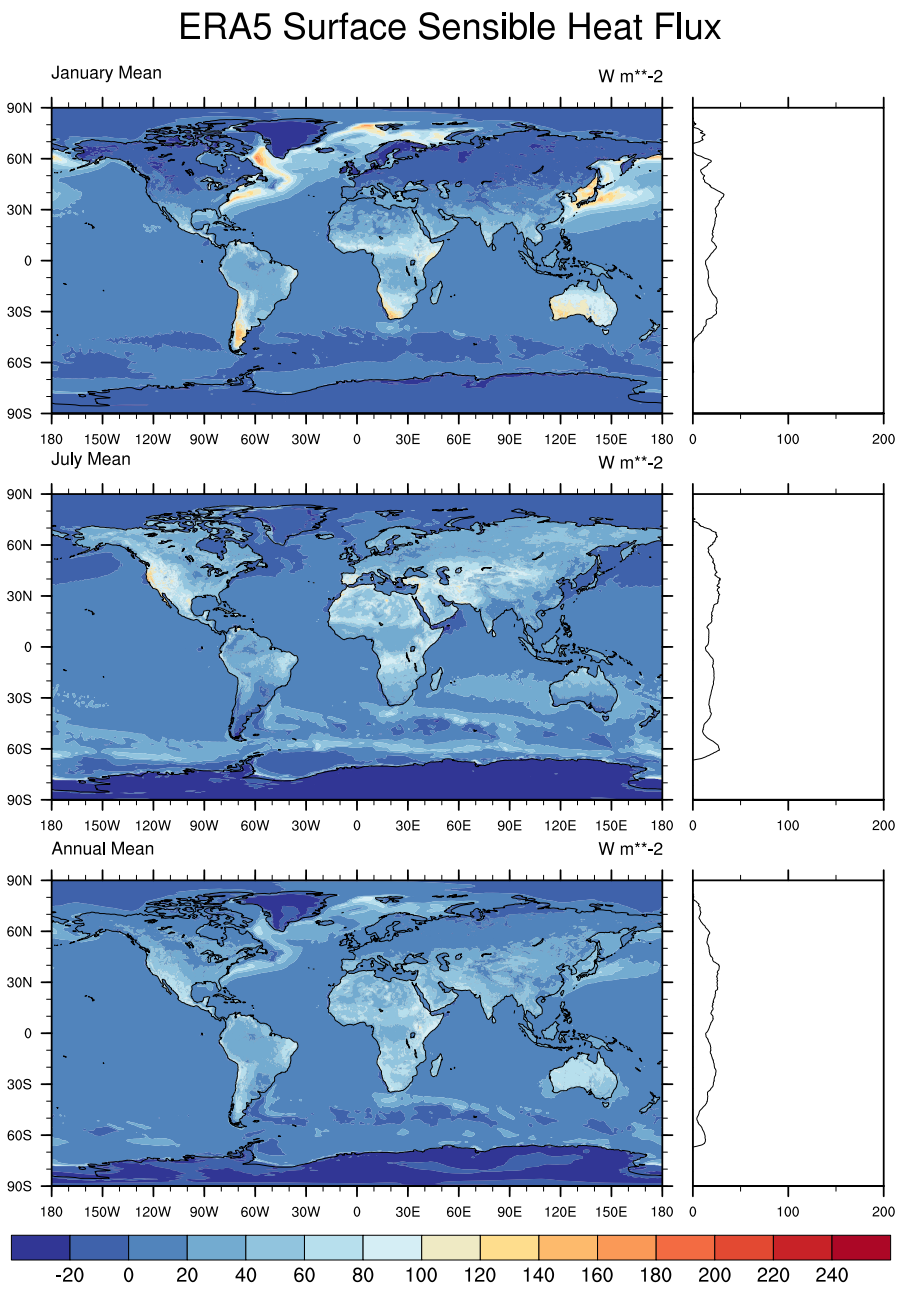


Figure 6.11: Maps of the surface dry static energy flux, for January, July, and the annual mean, based on ERA5. These data are not true observations, although they are influenced by observations.

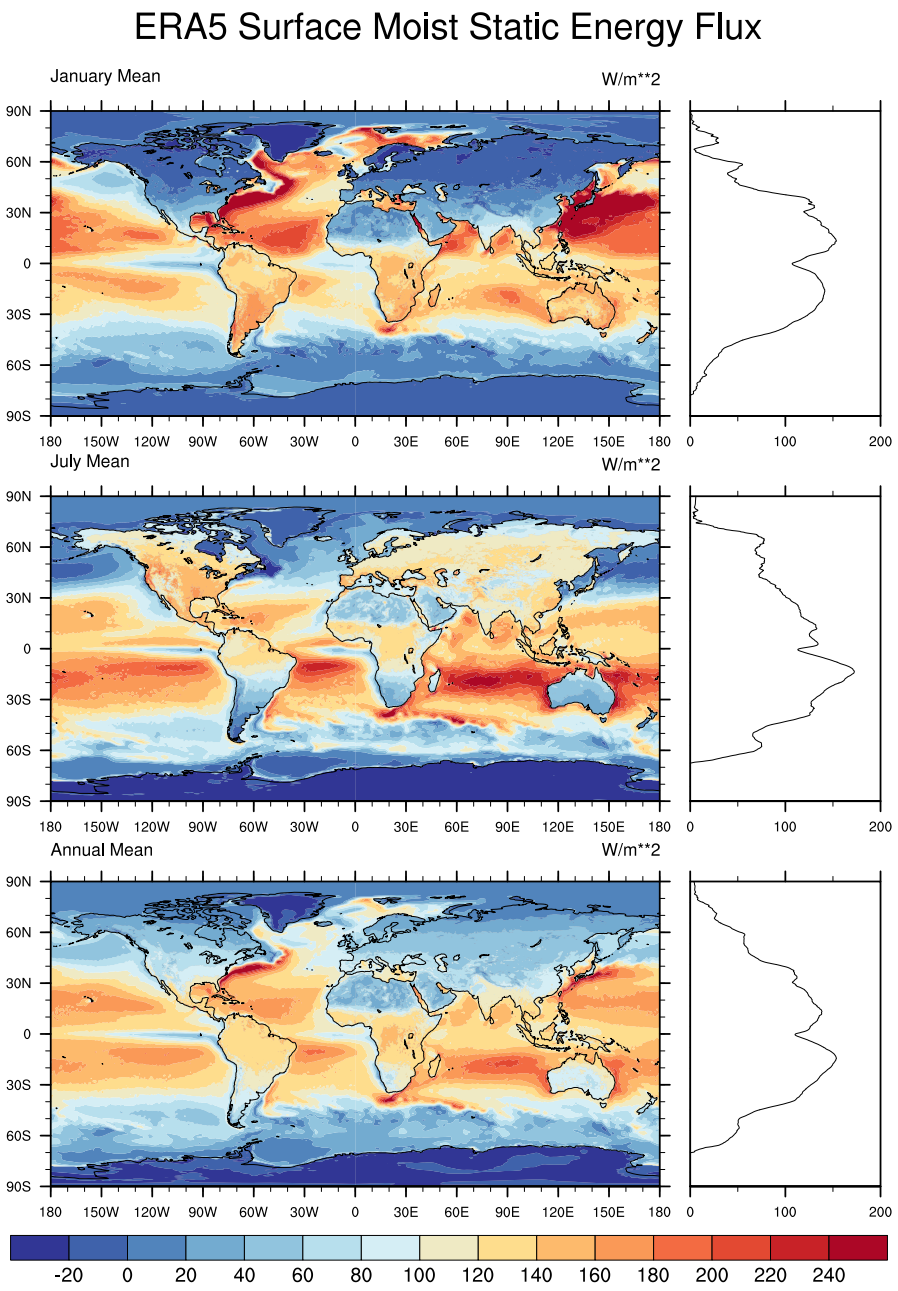


Figure 6.12: Maps of the surface moist static energy flux, for January, July, and the annual mean, based on ERA5. These data are not true observations, although they are influenced by observations.

the Northern Hemisphere continents. The strongest cooling tends to occur in regions of little precipitation, e.g., over the subtropical oceans of the winter hemisphere. As discussed in Chapter 2, the globally averaged ARC is nearly proportional to the globally averaged precipitation rate, but in the tropics the local radiative cooling rate is negatively correlated with the precipitation rate, because the high clouds associated with precipitating weather systems reduce the OLR.

Fig. 6.14 shows that on seasonal time scales the ARC is negatively correlated with the precipitation rate in the tropics, and positively correlated in middle latitudes. The reason is that in the tropics precipitating cloud systems reduce the ARC by blocking the OLR, while in middle latitudes precipitating cloud systems increase the ARC by increasing the downwelling infrared radiation at the surface. Further discussion is given by Naegele and Randall (2019).

To compute the vertically integrated northward transport of energy by the atmosphere, we have to add up the terms on the right-hand side of Eq. (6.22), i.e., $\left(\overline{F_h}^{\lambda,t}\right)_S + \left(\overline{R}^{\lambda,t}\right)_S - \left(\overline{R}^{\lambda,t}\right)_\infty$. We have used CERES data for $\left(\overline{R}^{\lambda,t}\right)_\infty$ and the ECMWF reanalysis for $\left(\overline{F_h}^{\lambda,t}\right)_S + \left(\overline{R}^{\lambda,t}\right)_S$. With this approach, the total (atmosphere plus ocean) transport is determined entirely by the CERES data, as it was in Chapter 2, the ocean transport is determined entirely by the ECMWF reanalysis, and the atmospheric transport is based on data from both sources.

Fig. 6.15 shows the implied net source of moist static energy for the atmosphere. The tropical atmosphere experiences a source of moist static energy, primarily due to the surface moist static energy flux (which is almost entirely due to surface evaporation), and the polar regions experience a sink due to radiative cooling. This will be shown in more detail below.

Fajber et al. (2023) emphasize that the surface latent heat flux is the most important source of atmospheric energy. It is balanced by the ARC. The annual and zonally averaged ARC varies only weakly with latitude, while the surface evaporation is strongly concentrated in the tropics and subtropics. For this reason, the meridional shape of the net atmospheric energy source is controlled by the meridional shape of the surface evaporation.

We now multiply both sides of (6.24) by $2\pi a$, and integrate with respect to latitude from the South Pole to an arbitrary latitude φ, to obtain

$$2\pi a \cos\varphi \int_0^\infty \overline{\rho_\theta v h}^{\lambda,t} \, d\theta = 2\pi a^2 \int_{-\frac{\pi}{2}}^{\varphi} \left[\left(\overline{F_h}^{\lambda,t}\right)_S + \left(\overline{R}^{\lambda,t}\right)_S - \left(\overline{R}^{\lambda,t}\right)_\infty \right] \cos\varphi' d\varphi' \, . \tag{6.26}$$

The quantity on the left-hand side of (6.26) is the total northward transport of moist static

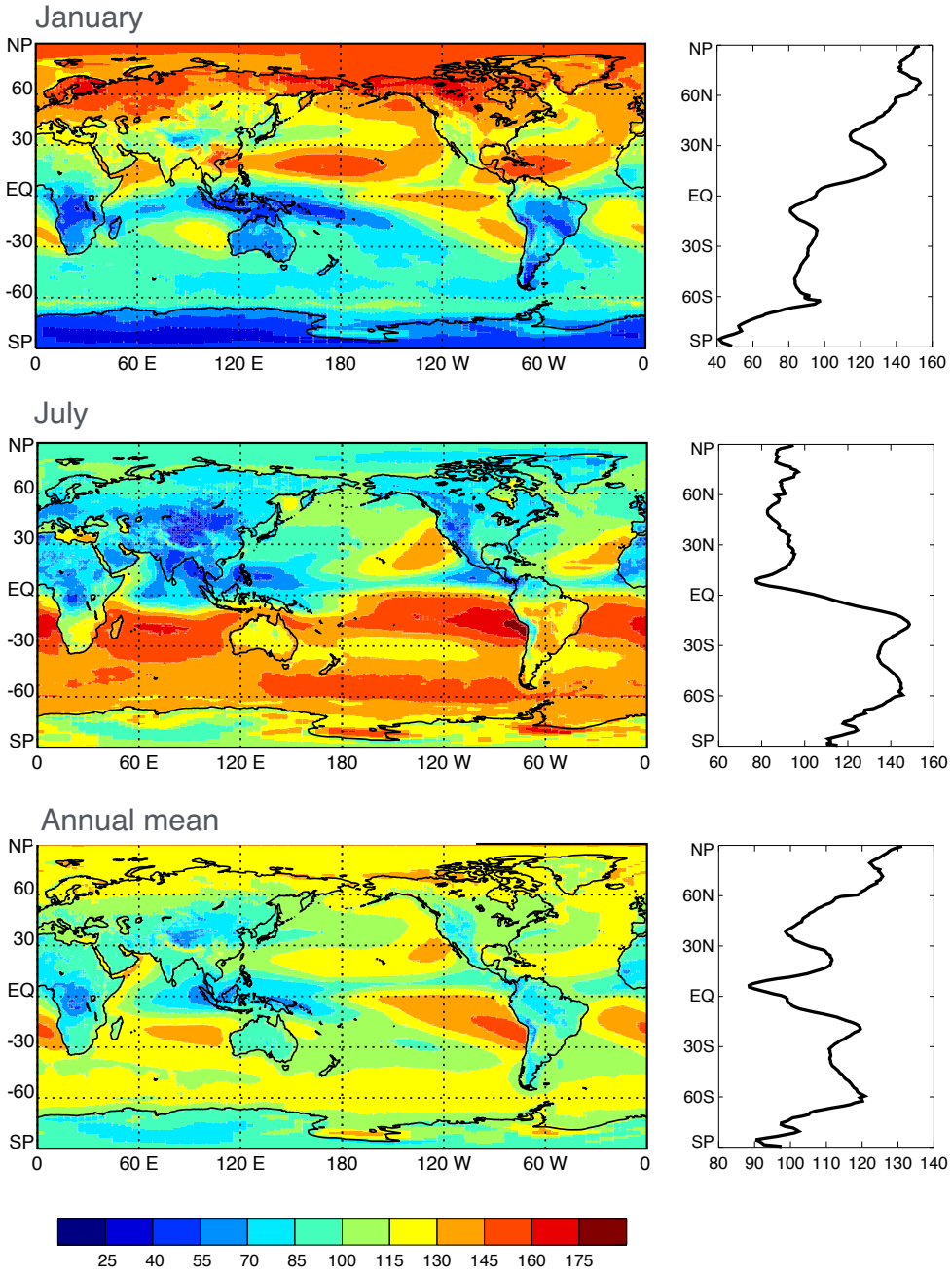


Figure 6.13: Maps of the atmospheric radiative cooling rate (ARC) for January and July, based on satellite data (Wielicki et al., 1996, 1998). The corresponding zonal averages are shown on the right.

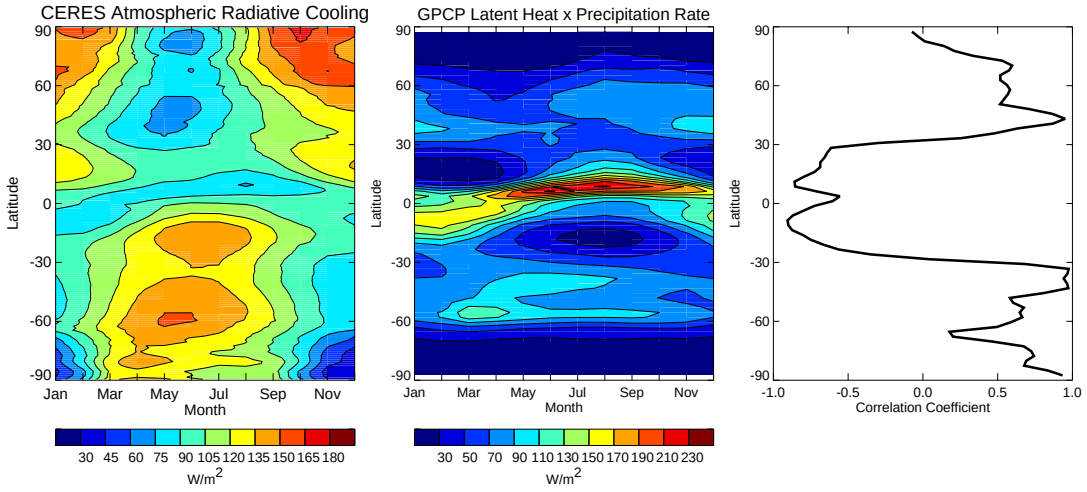


Figure 6.14: The left panel shows the seasonal cycle of the zonally averaged ARC, as inferred from satellite data. The center panel shows the seasonal cycle of the zonally averaged surface precipitation rate. The right panel shows the seasonal correlation of the two. Based on Naegele and Randall (2019).

energy across latitude φ. Compare (6.26) with (6.13). As usual, we correct to have a global mean of zero, and we set the northward transport to zero at the South Pole.

The total northward transport of moist static energy by the atmosphere is shown in the upper panel of Fig. 6.16, for January and July. Here we choose to show the ECMWF reanalysis product for the northward transport of total energy by the atmosphere, instead of the results of integrating (6.26). The northward transport by the atmosphere increases toward the north from about 40° S to 40° N, i.e., it diverges in that latitude belt, implying that the atmosphere is exporting moist static energy from this wide band of latitudes. The poleward transport shows a noticeable seasonal cycle in the Northern Hemisphere, with the strongest transport in the northern winter. The Southern Hemisphere transport shows less seasonal change.

The lower panel of Fig. 6.16 shows the annual mean energy transport by the oceans and atmosphere combined, computed from CERES data exactly as in Chapter 2. Also shown is the annual mean northward energy transport by the atmosphere, based on the ECMWF analysis. Finally, we show the ocean energy transport, obtained as the difference between the total transport and the atmospheric transport. The ocean energy transport is smaller than that by the atmosphere, but not negligible, especially in the Northern Hemisphere. There is insufficient data to determine directly from measurements of the ocean's temperature and currents, although the situation is improving (Roemmich et al., 2009; Jayne et al., 2017; McCarthy et al., 2015). Results analogous to those of Fig. 6.16 have also presented by VonderHaar and Oort (1973), Oort and Vonder Haar (1976), Carissimo et al. (1985), Savijärvi (1988), Masuda (1988). Trenberth and Caron (2001) provide a much more detailed discussion.

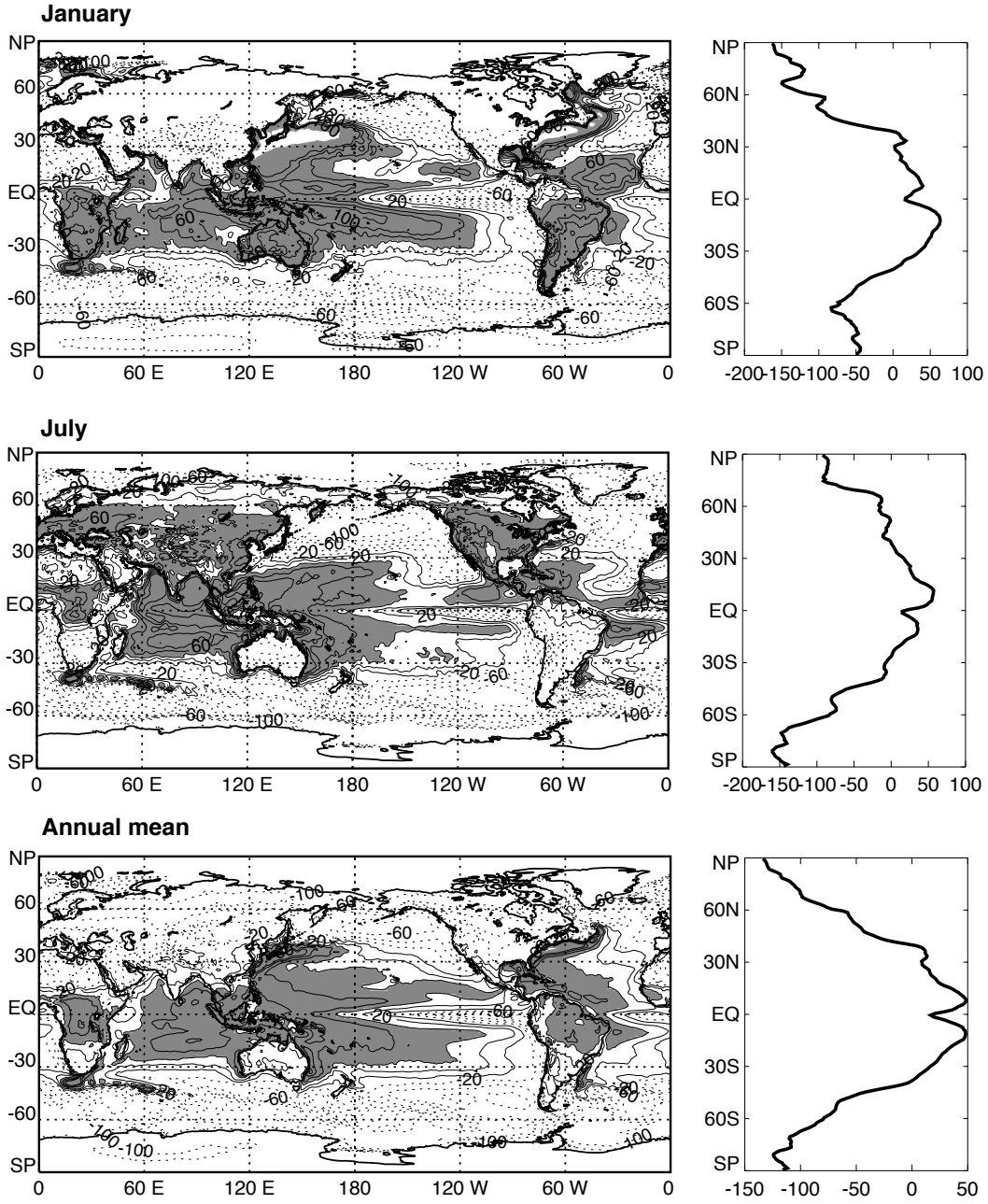


Figure 6.15: Maps of the net atmospheric source of moist static energy due to the combined effects of fluxes at the surface and the top of the atmosphere, for January, July, and the annual mean. The contour interval is 20 W m^{-2}, and values larger than 40 W m^{-2} are shaded. The corresponding zonal averages are shown on the right. These data are not true observations, although they are influenced by observations.

Fig. 6.17 shows the total, mean, bolus, and eddy meridional moist static energy fluxes. The mean and bolus fluxes make important contributions to the total flux, but the eddy

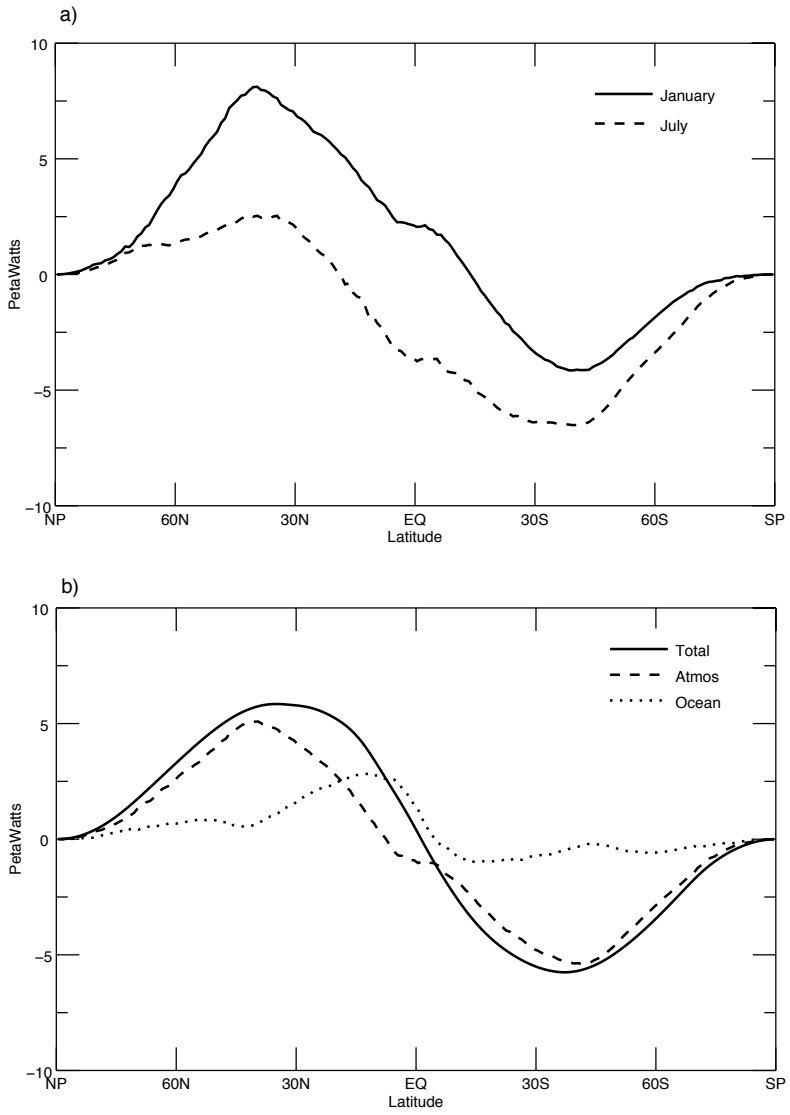


Figure 6.16: a) The vertically integrated northward transport of moist static energy by the atmosphere, for January and July. b) The annual-mean vertically integrated northward transport of moist static energy by the atmosphere. Also shown for comparison are the total transport by the atmosphere-ocean system, which was discussed in Chapter 2, and the difference between the total transport and the atmospheric transport, which can be interpreted as the transport by the oceans.

flux is small everywhere. The mean flux dominates in the tropics. The eddy flux of moist static energy is almost entirely due to the eddy flux of latent heat (i.e., water vapor), partly because dry static energy fluctuations on isentropic surfaces are small. The total moist static energy flux converges strongly on the 300 K surface in the middle latitudes of the winter hemisphere.

We can define a stream function, ψ_h, for the temporally and zonally averaged transport

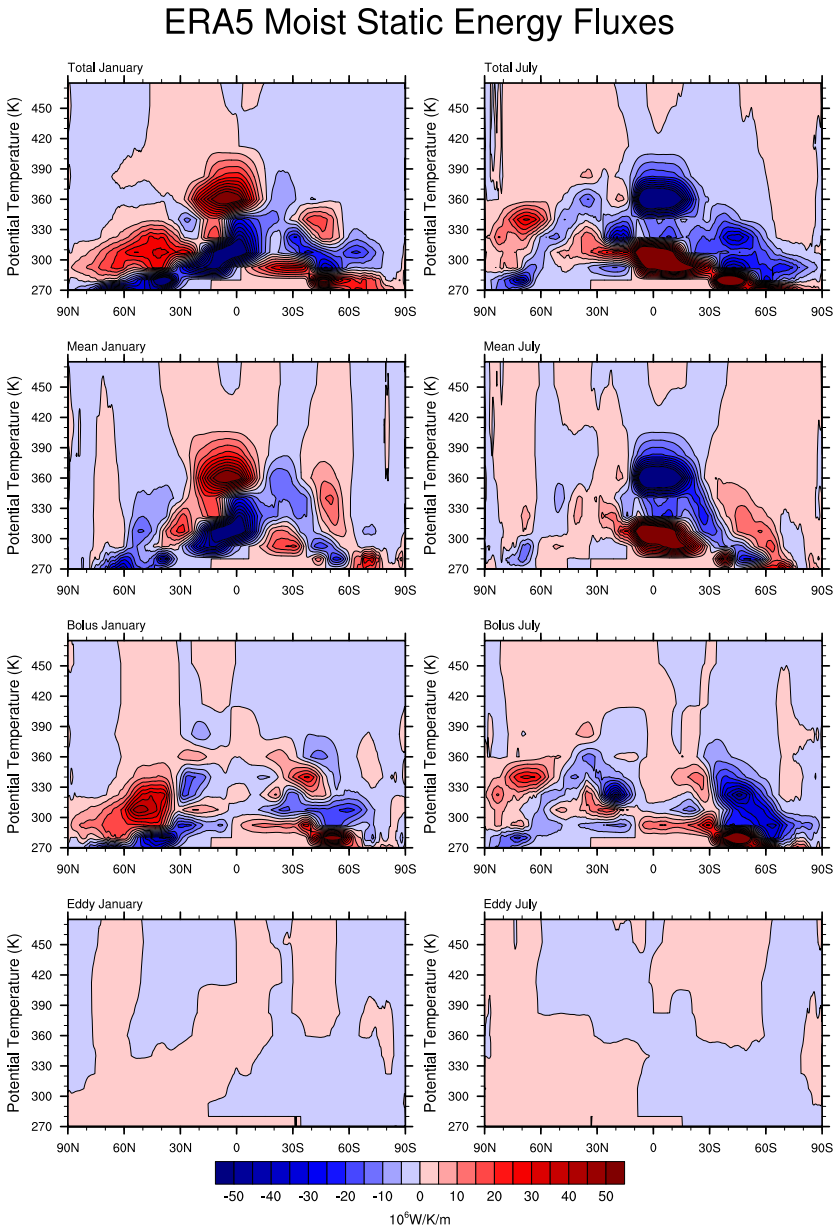


Figure 6.17: Isentropic meridional moist static energy mass fluxes for January (left) and July (right). The top row shows the total flux, the second row shows the flux associated with the zonally averaged meridional wind, the third row shows the bolus flux, and the bottom row shows the eddy flux. The contour interval is 0.5x10^6 W K^{-1} m^{-1}.

of moist static energy, using

$$\overline{\rho_\theta v \left(h - \overline{h}^{\mathrm{GM}}\right)}^{\lambda,t} 2\pi a \cos\varphi \equiv -\frac{\partial \psi_h}{\partial \theta} \,, \tag{6.27}$$

and

$$\left[\overline{\rho_\theta \dot{\theta} \left(h - \overline{h}^{\mathrm{GM}}\right)}^{\lambda,t} + \overline{F_h}^{\lambda,t} + \overline{R}^{\lambda,t}\right] 2\pi a^2 \cos\varphi \equiv \frac{\partial \psi_h}{\partial \varphi} \,. \tag{6.28}$$

By analogy with our earlier discussion of the meridional moisture fluxes, the meridional fluxes of h in (6.27) and (6.28) are expressed in terms of the difference $h - \overline{h}^{\mathrm{GM}}$, where $\overline{h}^{\mathrm{GM}}$ is the mass-weighted average of h over the entire atmosphere. There is a new twist, however. When we defined stream functions for the circulations of dry air and moisture, we set them to zero at the top of the atmosphere, on the grounds that there is no vertical mass flow there. We cannot do this with ψ_h, because radiation carries a flux of energy at the top of the atmosphere. The required upper boundary condition can be written as

$$R_\infty 2\pi a^2 \cos\varphi \equiv \frac{\partial}{\partial \varphi} (\psi_h)_\infty \,, \tag{6.29}$$

where R_∞ and $(\psi_h)_\infty$ are the net radiation and the stream function at the top of the atmosphere, respectively. As discussed in Chapter 2, the global mean of R_∞ is expected to be close to zero. When it is zero, Eq. (6.29) implies that $(\psi_h)_\infty$ takes the same value at both poles. An additional condition is needed, however, to determine what that value is. A simple possibility, used here, is to choose the constant so that the cosine-weighted meridional average of $(\psi_h)_\infty$ is zero. With this choice, $(\psi_h)_\infty$ will take both positive and negative values, depending on latitude.

Fig. 6.18 shows plots of ψ_h for January and July, and the annual mean. The upper boundary condition (6.29) strongly determines the solution. The isolines of the stream function are nearly vertical at all latitudes, indicating that the vertical flow of energy due to radiation is dominant. The stream function plots resemble a plot of the net downward radiation flux, as shown in Fig. 6.19. This shows the very strong role of radiation in the vertical transport of energy through the atmosphere. The zonally averaged radiative heating and cooling rates throughout the troposphere and stratosphere and shown in Fig. 6.20.

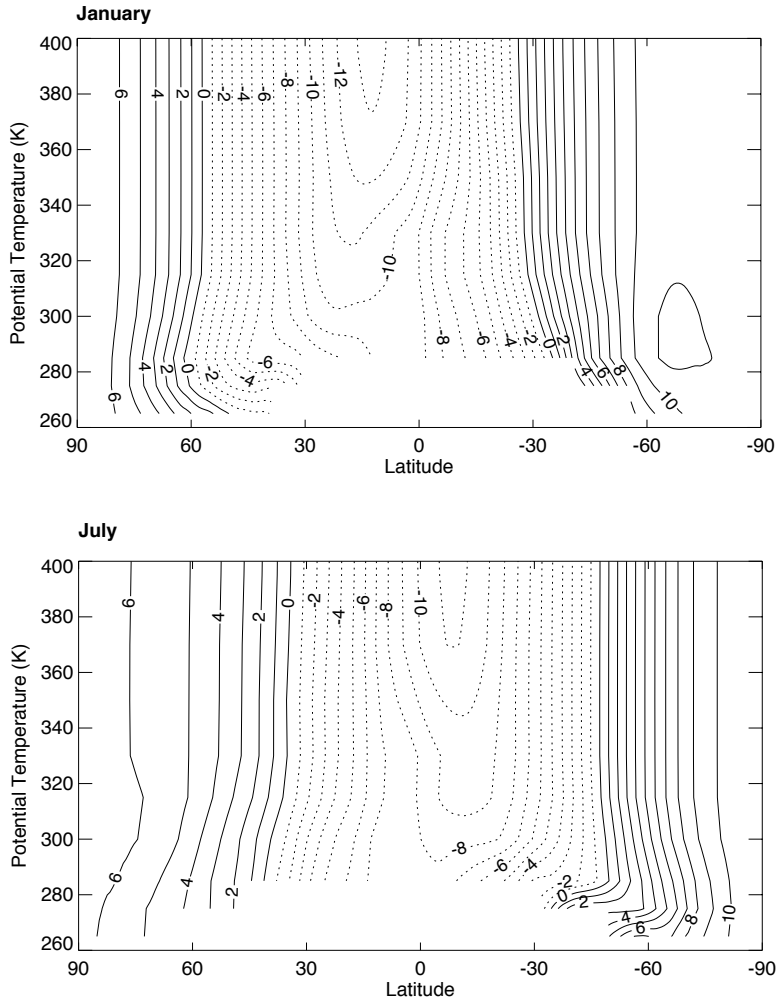


Figure 6.18: Stream function for the atmospheric transport of moist static energy through the atmosphere, for January and July. The contour interval is 1 PW.

6.5 Angular momentum transports

As discussed in Chapter 4, the component of the atmosphere's angular momentum vector that points in the direction of the Earth's axis of rotation is given by

$$M \equiv (\Omega a \cos\varphi + u)\, a \cos\varphi \tag{6.30}$$

per unit mass. Here we have used the Earth's radius as an approximation to the radial distance from the center of the Earth to a particle inside the atmosphere. The $\Omega a^2 \cos^2\varphi$ term of (6.30) represents the angular momentum associated with to the Earth's rotation,

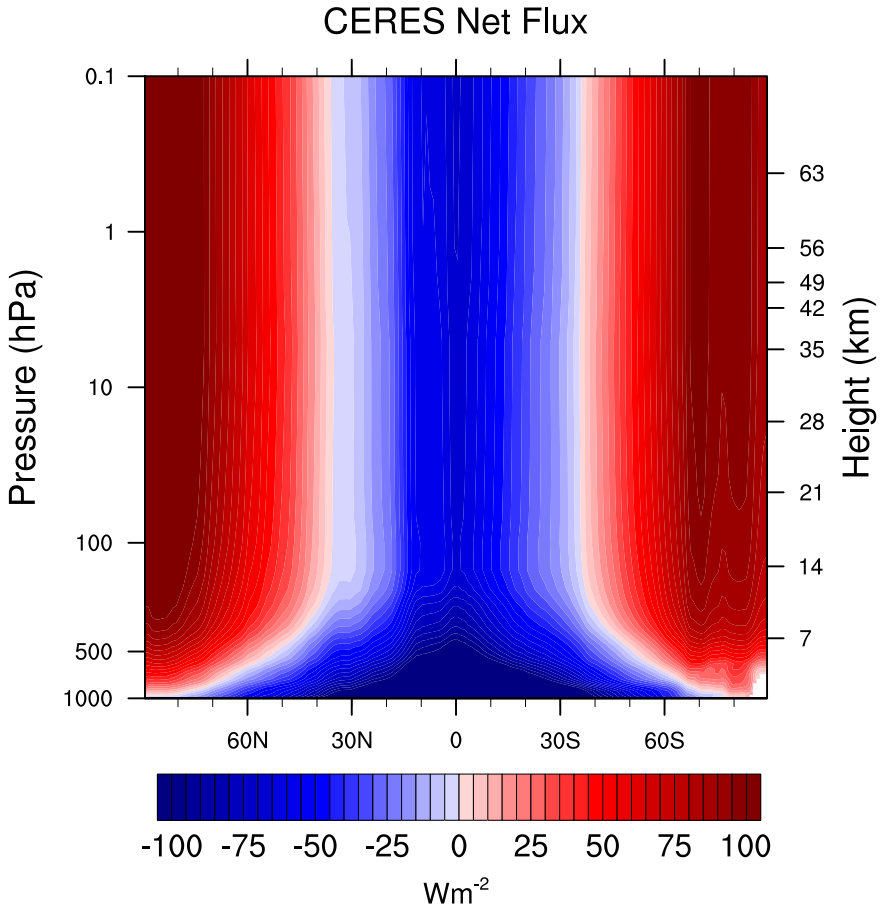


Figure 6.19: Zonally and annually averaged net radiation flux, based on CERES data. Negative values (blue) represent a downward net flux, and positive values (red) represent an upward net flux.

and the $ua\cos\varphi$ term represents the relative angular momentum due to the rotation of the atmosphere relative to the Earth's surface. The sum of the two is the absolute angular momentum. Fig. 6.21 shows the observed zonally averaged total angular momentum and relative angular momentum, both per unit mass, for January and July. The absolute angular momentum is generally about an order of magnitude larger than the relative angular momentum; it varies strongly with latitude, and only slightly with height. It is positive everywhere, and takes its largest values near the Equator. The plots of the zonally averaged relative angular momentum naturally resemble those of the zonally averaged zonal wind, which are presented in Chapter 3. In the tropical upper troposphere, the absolute angular momentum contours tilt or bulge slightly towards the winter pole, in both seasons. This suggests some tendency for the air flowing poleward in the main Hadley cell to conserve

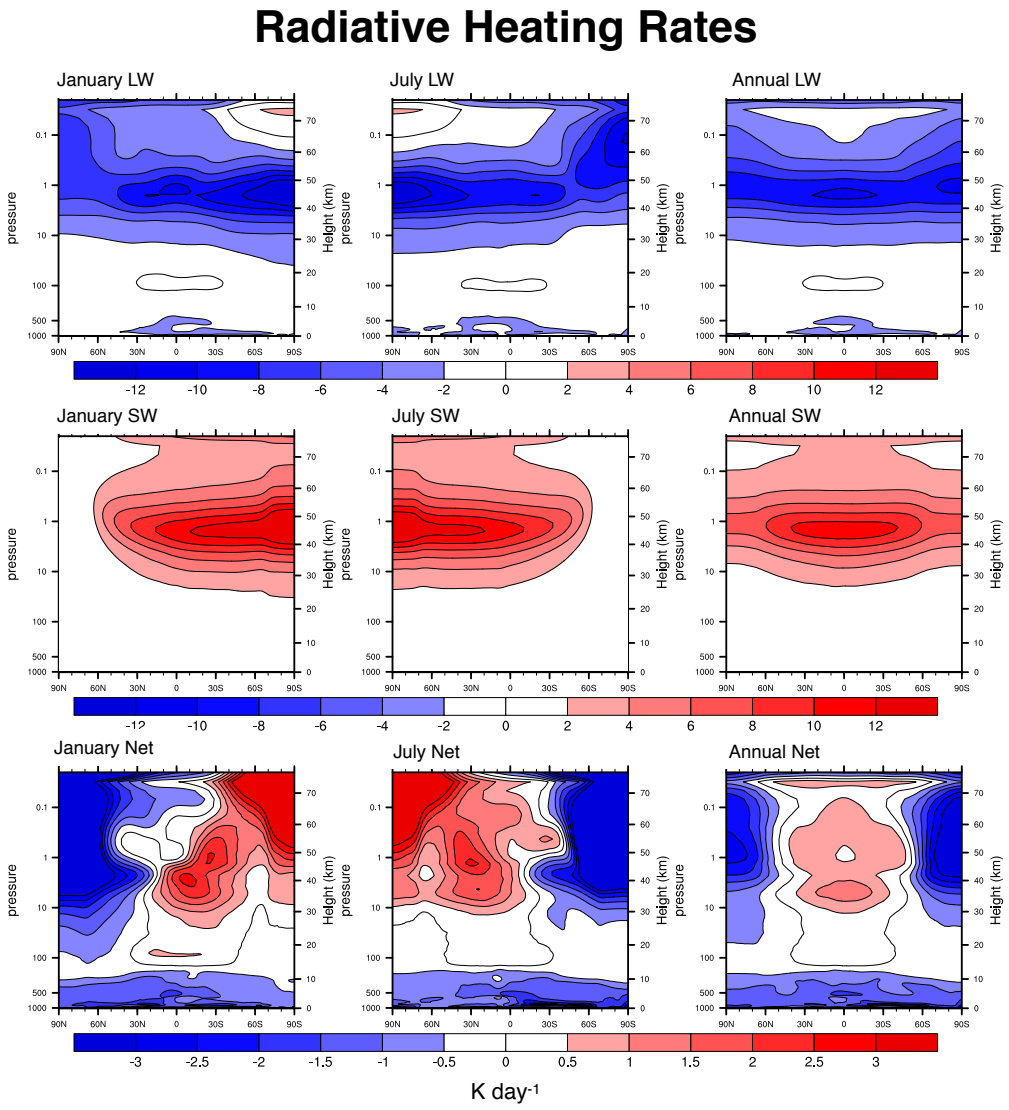


Figure 6.20: Latitude-height sections of longwave (LW), shortwave (SW) and net radiative heating rates for January, July, and the annual mean. The color bar for the net heating differs from the other two color bars. These data are from the ERA5 reanalysis, so they are not true measurements but they have been influenced by measurements.

its absolute angular momentum.

Early theories of the zonally averaged circulation (e.g. Held and Hou, 1980; Lindzen, 1990)] were based on the approximation of conservation of angular momentum in the upper branch of the Hadley cell – an approximation that was already well known to be not very accurate. As the air flows poleward in the upper branch of a Hadley cell, a tendency towards conservation of angular momentum implies a poleward increase of the westerlies, which are observed to take their maximum values in the jet streams near the latitudes

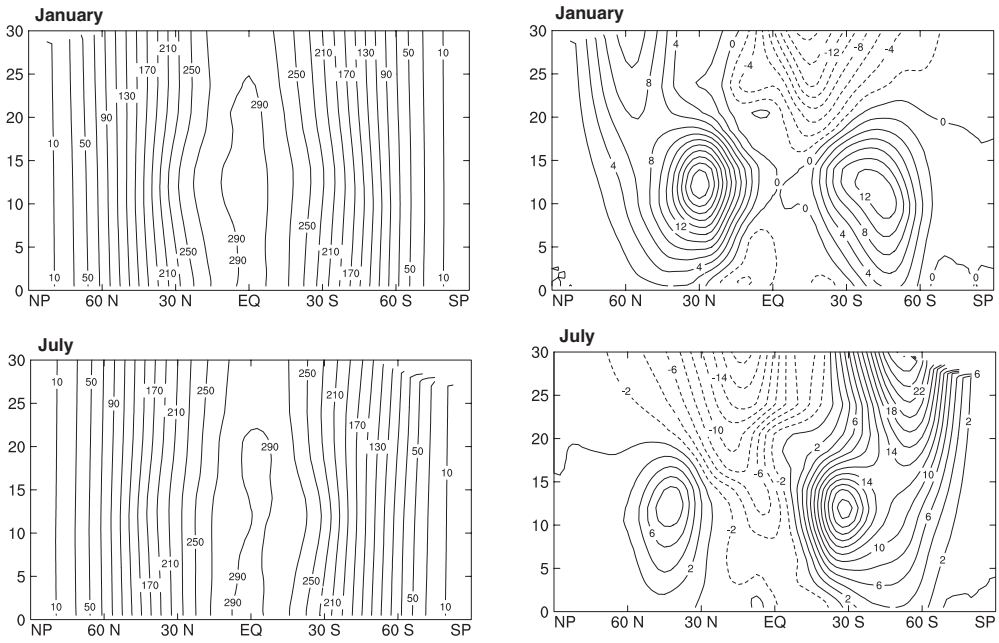


Figure 6.21: The observed absolute (left panels) and relative (right panels) atmospheric angular momentum per unit mass, for January and July. The units are 10^7 m^2 s^{-1}.

where the Hadley circulations stop. These subtropical jets can thus be interpreted as consequences of the poleward flow of air in the Hadley circulations. They are sometimes called Hadley-driven jets. In reality, however, they are much weaker than they would be if angular momentum were truly conserved in the poleward branches of the Hadley cells. Fig. 6.22 shows the zonal wind that would occur if a particle started on the equator with $u = 0$ and moved towards either pole while exactly conserving its angular momentum. Starting the particle from 10 degrees off the equator would only slightly change the plot. The zonal wind values shown at 30° north and south are much stronger than those observed in the subtropical jets. This implies that there are important sinks of angular momentum in the upper branches of the Hadley cells. Further discussion is given in Chapter 11; see also the reviews by Held and Hoskins (1985) and Schneider (2006).

In isentropic coordinates, conservation of angular momentum is expressed, in flux form, by

$$\frac{\partial}{\partial t}\left(\rho_\theta M\right) + \nabla\cdot\left(\rho_\theta \mathbf{v} M\right) + \frac{\partial}{\partial\theta}\left(\rho_\theta \dot{\theta} M + F_u a\cos\varphi\right) = -\rho_\theta \frac{\partial s}{\partial\lambda}\,. \tag{6.31}$$

Here F_u is the vertical flux of zonal momentum due to convection and turbulence, and

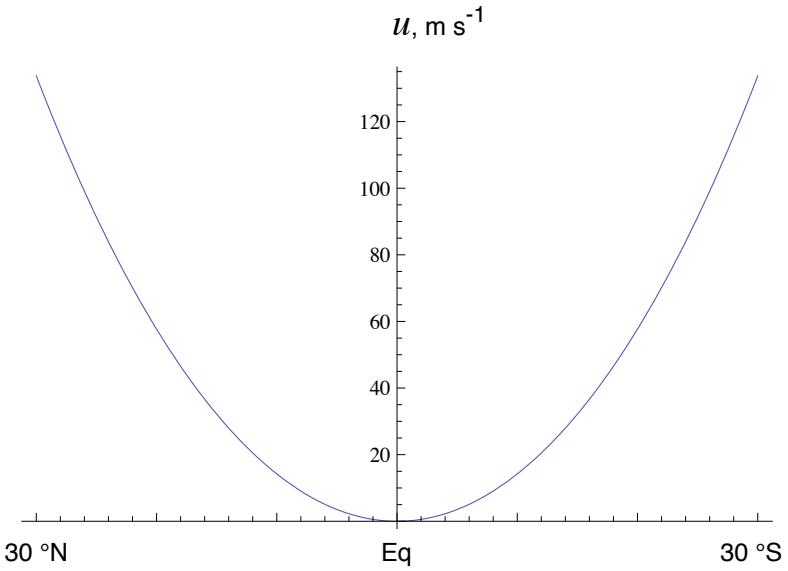


Figure 6.22: The hypothetical variation of the zonal wind with latitude under conservation of angular momentum, based on the assumption that $u = 0$ at the Equator.

$$\begin{aligned} s \equiv & c_p T + gz \\ = & \Pi\theta + gz \,, \end{aligned} \tag{6.32}$$

is the dry static energy (sometimes called the Montgomery potential), where

$$\Pi \equiv c_p \left(\frac{p}{p_0} \right)^{\kappa} \tag{6.33}$$

is the Exner function. As a reminder, the term on the right-hand side of (6.31) represents the effects of the zonal pressure-gradient force.

Applying a zonal average to (6.31), we obtain

$$\frac{\partial}{\partial t}\left(\overline{\rho_\theta M}^{\lambda}\right) + \frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\rho_\theta v M}^{\lambda}\cos\varphi\right) + \frac{\partial}{\partial\theta}\left(\overline{\rho_\theta \dot{\theta} M}^{\lambda} + \overline{F_u}^{\lambda} a\cos\varphi\right) = -\overline{\rho_\theta \frac{\partial s}{\partial\lambda}}^{\lambda} \,. \tag{6.34}$$

Referring back to (6.30), $\overline{\rho_\theta v M}^{\lambda,t}$ can be thought of as the sum of a flux associated with the "Earth angular momentum," $\Omega\,(a\cos\varphi)^2$, and a flux associated with the relative angular momentum, $ua\cos\varphi$. The Earth angular momentum is independent of longitude, and approximately independent of height, and time. Because $\Omega\,(a\cos\varphi)^2$ is large, the flux of Earth angular momentum is large at individual levels, especially in the tropics. Nevertheless, the *vertically integrated* flux of Earth angular momentum is negligible at every latitude in a sufficiently long time average, simply because, as discussed in Chapter 3 and shown in (6.4), the vertically integrated meridional mass flux is negligible in a sufficiently long time average.

The pressure-gradient term of (6.34) can be rewritten in a very interesting way, as follows. Using the hydrostatic equation in isentropic coordinates, i.e.,

$$\frac{\partial s}{\partial \theta} = \Pi\,, \tag{6.35}$$

we can write

$$\begin{aligned}
-\rho_\theta \frac{\partial s}{\partial \lambda} &= \frac{1}{g}\frac{\partial p}{\partial \theta}\frac{\partial s}{\partial \lambda} \\
&= \frac{1}{g}\frac{\partial}{\partial \theta}\left(p\frac{\partial s}{\partial \lambda}\right) - \frac{p}{g}\frac{\partial}{\partial \theta}\left(\frac{\partial s}{\partial \lambda}\right) \\
&= \frac{1}{g}\frac{\partial}{\partial \theta}\left(p\frac{\partial s}{\partial \lambda}\right) - \frac{p}{g}\frac{\partial}{\partial \lambda}\left(\frac{\partial s}{\partial \theta}\right) \\
&= \frac{1}{g}\frac{\partial}{\partial \theta}\left[p\frac{\partial}{\partial \lambda}\left(\Pi\theta + \phi\right)\right] - \frac{p}{g}\frac{\partial \Pi}{\partial \lambda} \\
&= \frac{\partial}{\partial \theta}\left(\theta\frac{p}{g}\frac{\partial \Pi}{\partial \lambda} + p\frac{\partial z}{\partial \lambda}\right) - \frac{p}{g}\frac{\partial \Pi}{\partial \lambda} \\
&= \frac{\theta}{g}\frac{\partial}{\partial \theta}\left(p\frac{\partial \Pi}{\partial \lambda}\right) + \frac{\partial}{\partial \theta}\left(p\frac{\partial z}{\partial \lambda}\right) \quad .
\end{aligned} \tag{6.36}$$

When we take the zonal mean of (6.36), the first term on the bottom line drops out (as you will show in Problem 1a at the end of this chapter), and we are left with

$$\boxed{-\overline{\rho_\theta \frac{\partial s}{\partial \lambda}}^{\lambda} = \frac{\partial}{\partial \theta}\left(\overline{p\frac{\partial z}{\partial \lambda}}^{\lambda}\right) = \frac{\partial}{\partial \theta}\left(\overline{p^*\frac{\partial z^*}{\partial \lambda}}^{\lambda}\right)\,.} \tag{6.37}$$

This result was first obtained by Klemp and Lilly (1978). Eq. (6.37) says that, in the zonal mean, the effect of the zonal pressure-gradient force on the mass-weighted angular momentum can be written as the divergence of a vertical flux of angular momentum, given by $\dfrac{\partial}{\partial\theta}\left(\overline{p^*\dfrac{\partial z^*}{\partial\lambda}}^{\lambda}\right)$. An interpretation is given later in this chapter.

Substituting (6.37) back into (6.34), and time-averaging gives the zonally averaged angular momentum equation in the form

$$\frac{\partial}{\partial t}\left(\overline{\rho_\theta M}^{\lambda}\right)+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\rho_\theta v M}^{\lambda,t}\cos\varphi\right)+\frac{\partial}{\partial\theta}\left(\overline{\rho_\theta\dot{\theta}M}^{\lambda,t}+\overline{F_u}^{\lambda,t}a\cos\varphi-\overline{p^*\frac{\partial z^*}{\partial\lambda}}^{\lambda,t}\right)=0\,. \tag{6.38}$$

Time averaging allows us to drop the tendency term, giving

$$\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\rho_\theta v M}^{\lambda,t}\cos\varphi\right)+\frac{\partial}{\partial\theta}\left(\overline{\rho_\theta\dot{\theta}M}^{\lambda,t}+\overline{F_u}^{\lambda,t}a\cos\varphi-\overline{p^*\frac{\partial z^*}{\partial\lambda}}^{\lambda,t}\right)=0\,. \tag{6.39}$$

Finally, vertical integration through the entire atmospheric column gives

$$\frac{\partial}{\partial\varphi}\left(\int_0^\infty\overline{\rho_\theta v M}^{\lambda,t}d\theta\cos\varphi\right)=a\cos\varphi\left[\left(\overline{F_u}^{\lambda,t}\right)_S a\cos\varphi-\overline{p_S^*\frac{\partial z_S^*}{\partial\lambda}}^{\lambda,t}\right]\,. \tag{6.40}$$

Here we have used $(\rho_\theta)_{\theta=0}=(\rho_\theta)_{\theta\to\infty}=0$, and $(F_u)_{\theta\to\infty}=0$.

The term $\overline{p_S^*\partial z_S^*/\partial\lambda}^{\lambda,t}$ on the right hand side of (6.40) represents the effects of "*mountain torque*." It vanishes if *either* p_S *or* z_S is independent of λ. Fig. 6.23 illustrates the mechanism that gives rise to mountain torque. Near a mountain range, there is a tendency for higher pressure on the upstream side and lower pressure on the downstream side, as shown in the figure. The spatial correlation shown in the figure need not occur at a particular moment or with respect to a particular mountain range, but it does hold in the time-averaged zonal mean. The pressure pattern shown in the figure gives rise to a "form drag" that is analogous to the aerodynamic drag on a car or other moving object, due to relatively high pressure on the upwind side and lower pressure on the downwind side. The

atmosphere and the mountain are exchanging momentum: The atmosphere tries to push the mountain downstream, and (as decreed by Isaac Newton) the atmosphere feels an equal and opposite force pushing it back in the upstream direction. As discussed later in this chapter, the force that the atmosphere exerts on the mountain can actually cause the Earth's rate of rotation to speed up or slow down slightly.

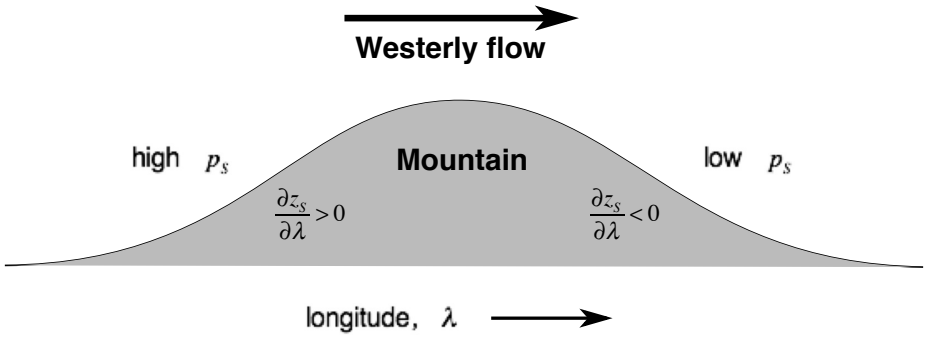


Figure 6.23: Sketch illustrating the concept of "mountain torque." High p_S occurs on the upstream side of the mountain, and low p_S on the downstream side.

In Problem 1b at the end of this chapter, you are asked to prove that the mountain torque would vanish if p_S were a function of z_S only, i.e., if the data in Fig. 3.1 fell exactly onto a single curve. This would be expected in the absence of all dynamical processes, i.e., if the atmosphere was "just sitting there," in a balanced, resting state.

As disussesd in Chapter 5, surface friction can be described using

$$(\mathbf{F}_\mathbf{v})_S = -\rho_S c_D |\mathbf{v}_S| \mathbf{v}_S \,, \tag{6.41}$$

which is another bulk aerodynamic formula. The minus sign in (6.41) means that the near-surface momentum flux opposes the surface wind. In particular, the zonal component of (6.41) is

$$(F_u)_S = -\rho_S c_D |\mathbf{v}_S| u_S \,, \tag{6.42}$$

which says that the flux of zonal momentum is negative, i.e., downward, when the zonal component of the surface wind is positive, i.e., westerly. Surface friction transfers positive angular momentum from the atmosphere to the ocean-solid Earth system in places where the surface winds are westerly, most importantly in the middle latitudes of the winter hemisphere. Conversely, positive angular momentum flows into the atmosphere in places where

the surface winds are easterly, most importantly in the tropics. The mountain torques generally have the same sign as the frictional torques.

Fig. 6.24 shows the total surface torque, the mountain torque, the frictional torque, and the "GWD torque", which is due to orographic gravity waves that are not resolved by the reanalysis. The mountain torque has been computed from the surface pressure and topography as resolved by the reanalysis. The frictional torque and the GWD torque are parameterized in the model that was used to create the reanalysis. The frictional torque dominates overall, but mountain torque is appreciable in the Northern Hemisphere. The parameterized torque due to unresolved orographic gravity waves is considerably smaller than the other contributions. In the tropics, the trade winds cause a strong flow of angular momentum into the atmosphere. In the middle latitudes of the Northern Hemisphere, both frictional drag and mountain torque remove angular momentum from the atmosphere. In the middle latitudes of the Southern Hemisphere, strong westerly winds deposit angular momentum in the oceans. The mountains of Greenland and Antarctica add angular momentum to the atmosphere.

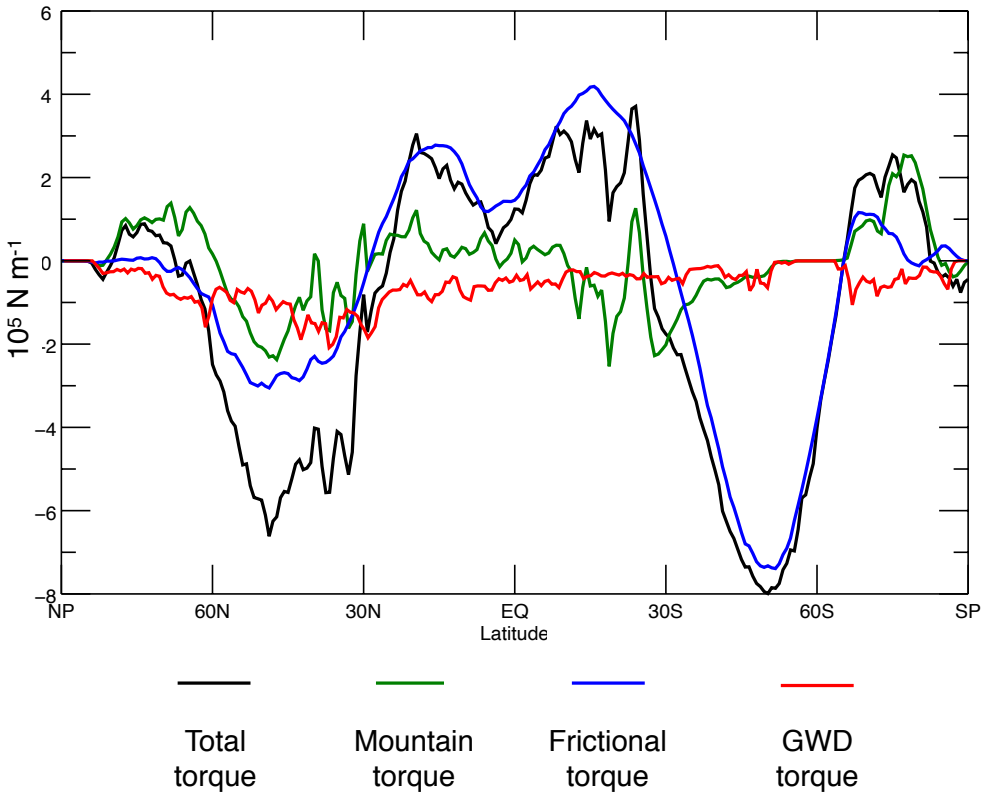


Figure 6.24: The annual mean of the zonally averaged total surface torque, the mountain torque, the frictional torque, and the "GWD torque", which is due to orographic gravity waves that are not resolved by the reanalysis, and are therefore parameterized in the forecast model. See text for details.

The Earth-atmosphere system can exchange angular momentum with other bodies, e.g., the Moon, through tidal torques. Such torques are causing the Moon to gain angular momentum from the Earth, so that the period of the Earth's rotation (the length of the day) and the radius of the Moon's orbit are both slowly increasing (e.g., Hide and Dickey, 1991). To the extent that we can neglect these slow proceses, the angular momentum of the Earth-atmosphere system must remain constant with time. Therefore, if the angular momentum of the atmosphere increases, the Earth's rotation must slow down, and vice versa. Evidence that this is really true is shown in Fig. 6.25. The data plotted in the figure indicate that changes in the rate of rotation of the solid Earth, i.e., the length of day, are highly correlated with changes in the angular momentum of the atmosphere. The length of day was measured by timing the Earth's rotation with respect to distant stars, while the atmospheric angular momentum was measured using radiosondes. The agreement shown is expected, but nevertheless amazing. See Agnew (2025) for a broader discussion of this subject.

The atmosphere of Venus is about 100 times more massive than the Earth's atmosphere, and angular momentum exchanges between the atmosphere and the surface of Venus cause much larger changes in the length of the Venusian day (Margot et al., 2021).

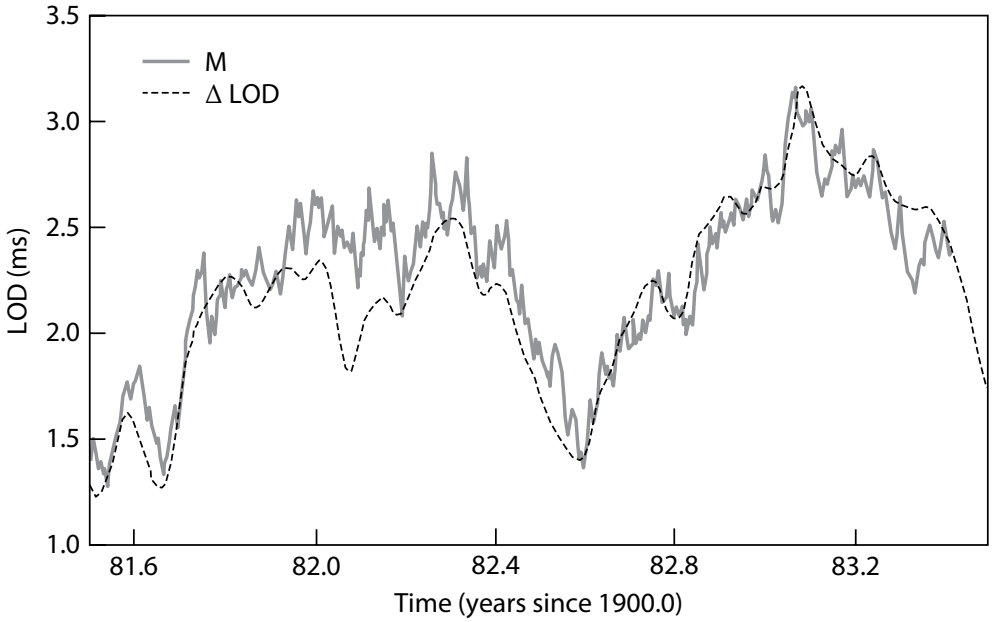


Figure 6.25: Changes in the length of day (LOD) in milliseconds as inferred from changes in the angular momentum of the atmosphere (grey solid line) and as determined from astronomical data (dotted line) over a four-year period starting in July 1981. The mean difference between the two quantities has been subtracted from the angular momentum curve. From Rosen et al. (1984).

Averaged over many years, the exchanges of angular momentum between the atmosphere and the solid Earth and oceans are expected to be very small. In equilibrium, there has to be be a mixture of easterly and westerly winds at the Earth's surface, distributed in

such a way that the net exchange of angular momentum between the atmosphere and the surface is near zero.

Multiply both sides of (6.38) by $2\pi a$, then integrate vertically through the entire atmosphere column, and also meridionally from the South Pole to an arbitrary latitude φ, to obtain

$$2\pi a \cos\varphi \int_0^\infty \overline{\rho_\theta v M}^{\lambda,t} \, d\theta = 2\pi a^2 \int_{-\pi/2}^{\varphi} \left[\left(\overline{F_u}^{\lambda,t} \right)_S a \cos\varphi' - \overline{p_S \partial z_S / \partial \lambda}^{\lambda,t} \right] \cos\varphi' d\varphi' \, . \tag{6.43}$$

The quantity on the left-hand side of (6.43) is the total northward transport of angular momentum across latitude φ. After correcting $\left[\left(\overline{F_u}^{\lambda,t} \right)_S a \cos\varphi - \overline{p_S \partial z_S / \partial \lambda}^{\lambda,t} \right]$ so that it has a global mean of zero, we obtain the vertically integrated angular momentum transport shown in Fig. 6.26, for January, July, and the annual mean. The angular momentum transport passes through zero near the Equator. It diverges in the tropics, and converges poleward of $\pm 30^\circ$, where the the subtropical jet streams are found. Angular momentum is carried from the tropics, where the surface easterlies extract it from the oceans and solid earth, to middle latitudes, where the surface westerlies put it back into the oceans and the solid earth. The transport is very weak poleward of $\pm 60^\circ$, because there are two factors of $\cos\varphi$ on the left-hand side of (6.43): one that is plainly visible outside the integral, and a second that is built into the definition of M.

To balance the poleward angular momentum transport by the atmosphere, there has to be a return flow of angular momentum from middle latitudes towards the Equator, carried by the ocean currents or the solid earth, or some combination of the two. Oort (1989) argues that the return flow occurs largely through stresses in the solid earth, rather than through the circulation of the oceans. These stresses do not necessarily involve movements of the solid earth, as in continental drift, because a rigid material (like a steel ball) can transmit forces from one place to another without undergoing any deformation. The modeling study of Bryan (1997) supports Oort's claim that meridional angular momentum transport by the oceans is small.

We have attributed the pressure-gradient term of (6.40) to the form drag on the topography of the Earth's surface. The pressure-gradient term of (6.38) has a very similar interpretation. It is a vertical exchange of angular momentum within the atmosphere by the form drag on "bumpy" isentropic surfaces – mountains made of air. We call it "isentropic form drag." The isentropic surfaces bulge upward where there are blobs of cold air at a given pressure level, and downward where there are blobs of warm air. From the perspective of isentropic coordinates, the upward flux of zonal momentum is associated with the pressure force, rather than with a covariance between the "vertical velocity" (which

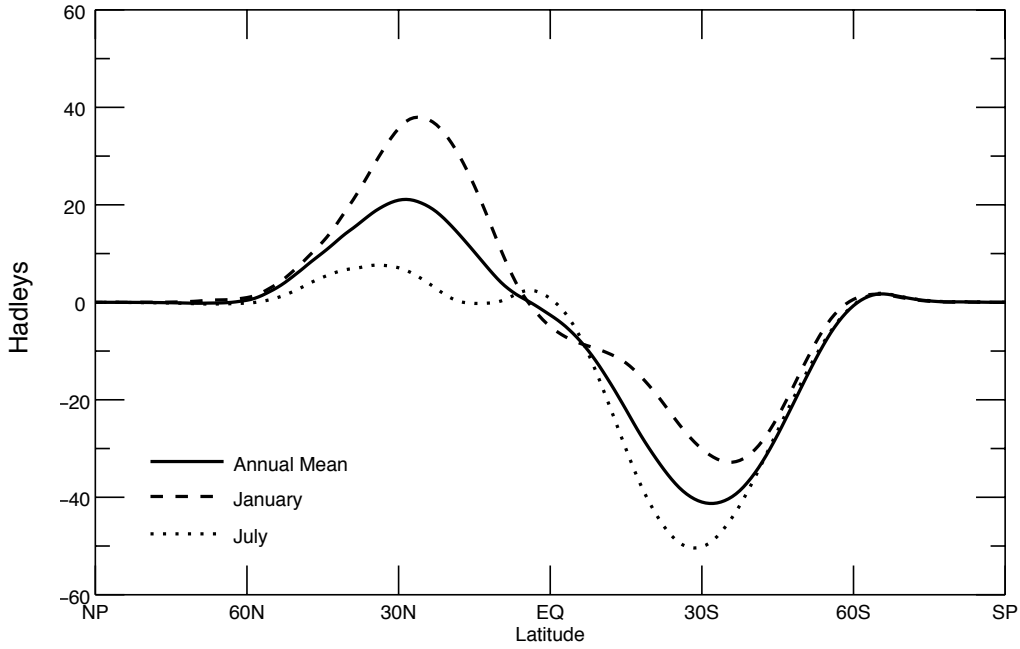


Figure 6.26: The vertically and zonally integrated meridional transport of angular momentum, for January, July, and the annual mean. A Hadley is 10^{18} Nt m.

vanishes in isentropic coordinates unless there is heating) and the zonal velocity. A layer of air confined between two isentropic surfaces will feel two momentum fluxes associated with the pressure force: one on its underside, and a second on its top side. It is the *difference* between these two forces that acts as a net force on the layer. That is why we see $-\dfrac{\partial}{\partial\theta}\left(\overline{p\dfrac{\partial z}{\partial\lambda}}^{\lambda,t}\right)$ in (6.38). The isentropic form drag is associated with eddies of many kinds, including baroclinic waves in middle latitudes, various equatorial waves, and gravity waves excited by diverse processes including flow over mountains, cumulus convection, and fronts. Further discussion is given in Chapter 11.

Fig. 6.27 shows the vertical flux of angular momentum due to isentropic form drag, for January and July. The flux is computed by evaluating $\overline{p\partial z/\partial\lambda}^{\lambda}$ using the reanalysis data. Comparison with Fig. 6.26 shows that the midlatitude surface sinks of angular momentum due to mountain torque and friction are supplied by the downward momentum transport via isentropic form drag. The flux is particularly strong in the winter hemisphere, and is produced by baroclinic waves.

Fig. 6.28 shows the total, mean, bolus, and eddy meridional fluxes of the absolute angular momentum. The Hadley cells produce large poleward fluxes in the upper troposphere, but they also produce large equatorward fluxes in the lower troposphere. These nearly cancel in a vertical average because, as mentioned earlier, the small value of the vertically integrated and zonally averaged meridional mass flux guarantees that the vertically

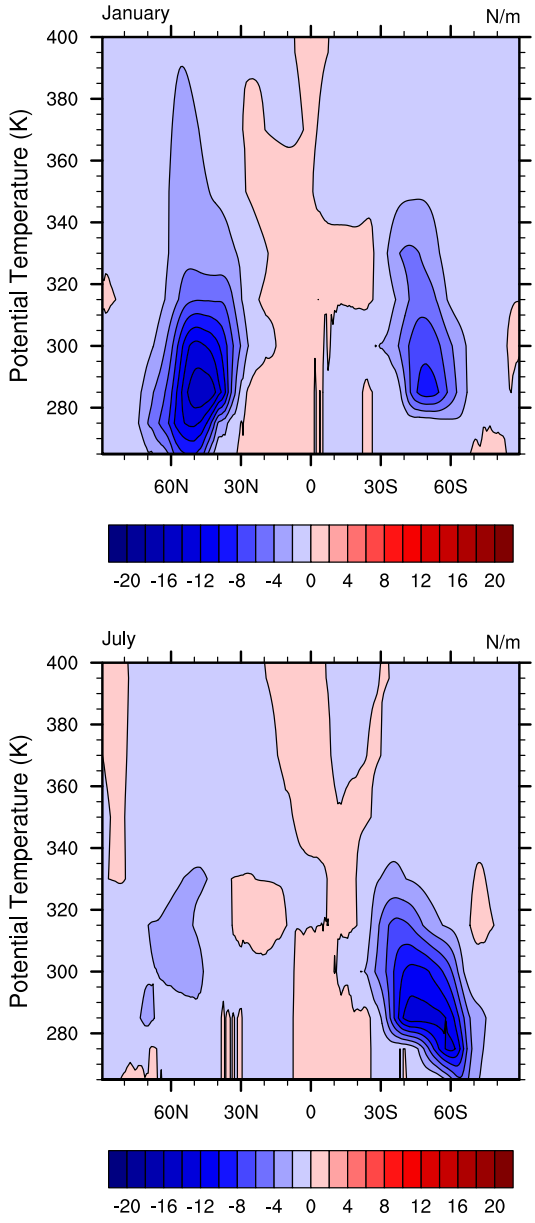


Figure 6.27: The vertical flux of angular momentum due to the isentropic form drag, for January and July. The contour interval is 10^6 Nt m^{-1}.

integrated and zonally averaged flux of Earth angular momentum is also small. We return to this point below. In both January and July, the total flux of angular momentum converges in midlatitudes in the winter hemisphere. The mean flux converges from the equatorward side, and the bolus flux converges from the poleward side. The eddy fluxes of angular momentum appear very weak in this figure, but as shown below they actually dominate after vertical integration. Fig. 6.29 is like Fig. 6.28, but modified to show the fluxes of relative

angular momentum only. The contour interval has been reduced by a factor of four. The importance of the eddy fluxes is apparent.

Eq. (6.34) can be used to define a stream function that depicts the zonally averaged flow of angular momentum through the atmosphere. We will call it ψ_M, and define it by

$$\overline{\rho_\theta v \left(M - \overline{M}^{\mathrm{GM}} \right)}^{\lambda,t} 2\pi a \cos\varphi \equiv -\frac{\partial \psi_M}{\partial \theta} \,, \tag{6.44}$$

and

$$\left[\overline{\rho_\theta \dot{\theta} \left(M - \overline{M}^{\mathrm{GM}} \right)}^{\lambda,t} - a\cos\varphi \overline{F_u}^{\lambda,t} \right] 2\pi a^2 \cos\varphi \equiv \frac{\partial \psi_M}{\partial \varphi} \,, \tag{6.45}$$

where $\overline{M}^{\mathrm{GM}}$ is the mass-weighted average of the angular momentum over the entire atmosphere. We can use (6.44) to determine ψ_M by vertical integration, given the meridional flux on the left-hand side and an upper or lower boundary condition. Since there is no angular momentum flux across the top of the atmosphere, we can choose $\psi_M = 0$ there. Fig. 6.30 shows plots of ψ_M for January and July. Angular momentum enters the atmosphere in the tropics, and is carried poleward, mainly towards the winter pole, by the upper branches of the Hadley cells. There is very strong downward angular momentum transport between 30 and 60 degrees off the Equator in the winter hemisphere. The stream function contours intersect the Earth's surface, indicating that angular momentum leaves the atmosphere there; compare with Fig. 6.24. Angular momentum converges meridionally near the 300 K isentropic surface in the same latitude band.

Further discussion of the role of angular momentum in the circulation of the atmosphere is given by Egger et al. (2007).

6.6 Segue

In this chapter, we have shown the meridional transports of mass, total moisture, moist static energy, and angular momentum. Using the implied stream function plots, we have also inferred the vertical transports of the same quantities.

What we have not discussed in detail yet is how the various meridional and vertical transports are actually produced. Most of the rest of the book is devoted to answering this very basic question, starting with Chapter 8, which discusses the key role of convection.

6.7 Problems

1. (a) Prove that the mountain torque vanishes if p_S depends only on z_S.

   (b) Prove that $\overline{p(\partial\Pi/\partial\lambda)}^{\lambda} = 0$.

2. (a) Consider an idealized atmosphere with a uniform surface pressure of 1000 hPa and a uniform angular velocity relative to the solid Earth, such that $u = u_0 \cos\varphi$, where $u_0 = 1$ m s^{-1}. What is the *relative* angular momentum of such an atmosphere?

   (b) The angular momentum of the solid Earth is about 7 x 10^{33} kg m^2 s^{-1}. Consider an atmosphere that is initially at rest with respect to the solid Earth. Suppose that the length of day decreases by one millisecond due to a transfer of angular momentum from the solid Earth to the atmosphere, and that after the transfer the atmosphere has a uniform angular velocity. What is the zonal wind at the Equator?

3. Derive an expression for the bolus mass flux in which the only dependent variable that appears is s. You may assume quasi-static and quasi-geostrophic dynamics.

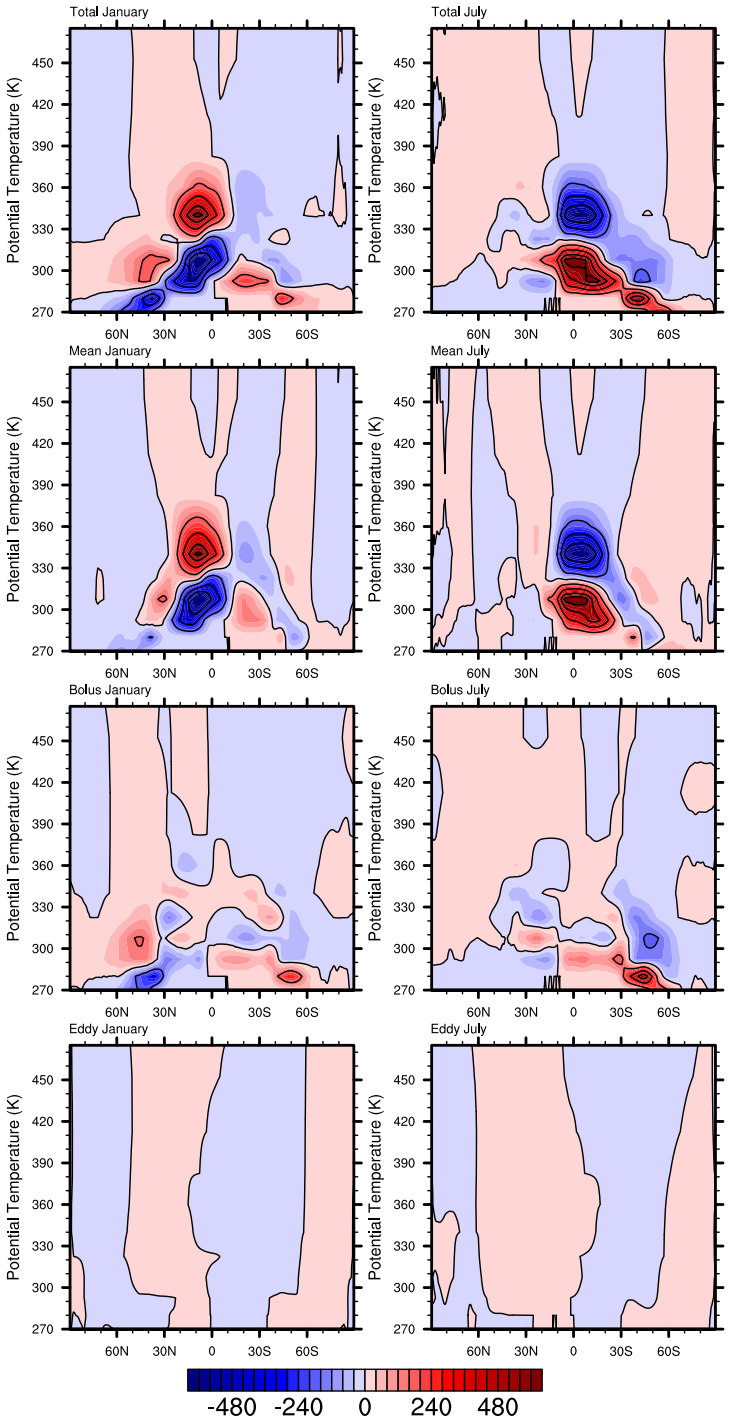


Figure 6.28: Isentropic meridional fluxes of absolute angular momentum for January (left) and July (right). The top row shows the total flux, the second row shows the flux associated with the zonally averaged meridional wind, the third row shows the bolus flux, and the bottom row shows the eddy flux. The contour interval is 20 x 10^9 Nt m^{-1} K^{-1}.

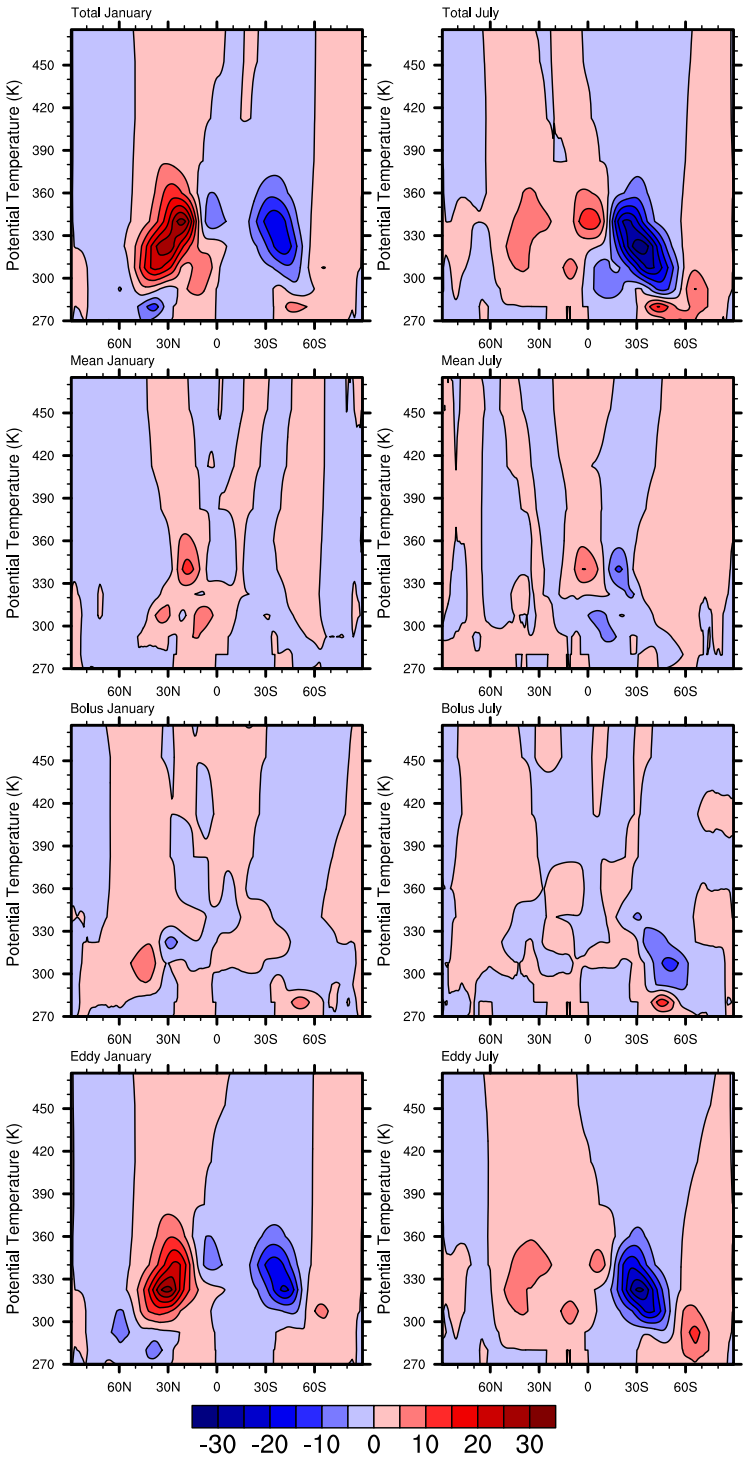


Figure 6.29: Like Fig. 6.28, but showing only the fluxes of *relative* angular momentum. The contour interval is 5 x 10^9 Nt m^{-1} K^{-1}, four times smaller than in Fig. 6.28.

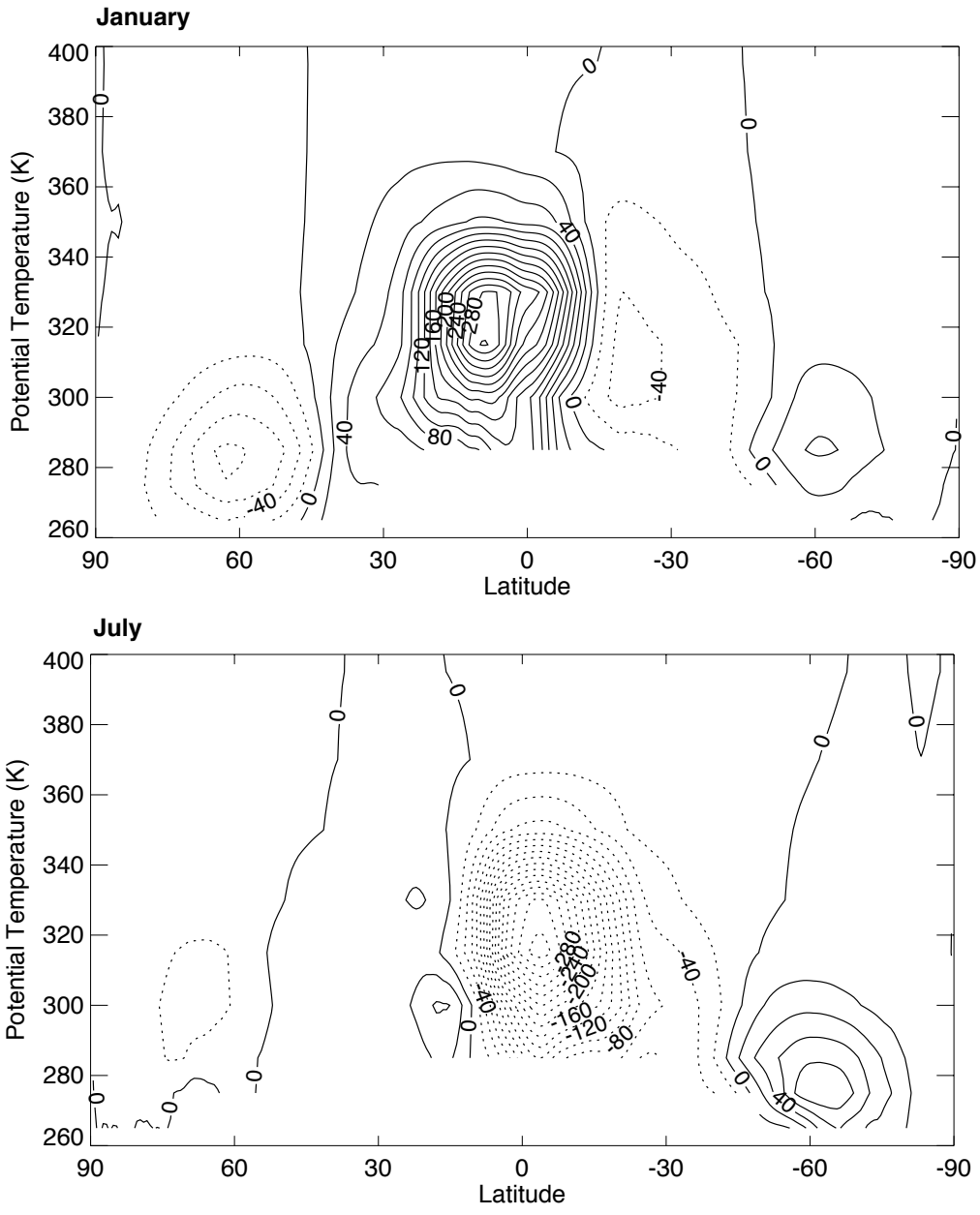


Figure 6.30: The stream functions for the zonally averaged angular momentum transport on isentropic surfaces, for January and July. The units are Hadleys.

Chapter 7

The energy cycle

7.1 The flow of energy across the atmosphere

It is useful to distinguish between horizontal and vertical fluxes of energy. With this in mind, we can rewrite the total energy equation, (4.94), as follows:

$$\begin{aligned}\frac{\partial}{\partial t}(\rho e_T) + \nabla_z \cdot [\rho \mathbf{v} e_T + p\mathbf{v} + \rho \mathbf{\Lambda}_H + \mathbf{R}_H + (\mathbf{F}_h)_H] \\ + \frac{\partial}{\partial z}\left[\rho w e_T + pw + \rho \Lambda_z + R_z + (F_h)_z\right] = 0\,.\end{aligned} \tag{7.1}$$

Here the subscripts H and z denote the "horizontal part" of a vector (i.e., a vector in the horizontal plane), and the (positive upward) vertical component of a vector, respectively. Vertically integrate (7.1) through the entire atmospheric column, using Leibniz' Rule to take the integrals inside the derivatives. The result can be written as

$$\begin{aligned}\frac{\partial}{\partial t}\left(\int_{z_S}^{\infty} \rho e_T dz\right) + \nabla_z \cdot \left[\int_{z_S}^{\infty} \rho \mathbf{v} e_T + p\mathbf{v} + \rho \mathbf{\Lambda}_H + \mathbf{R}_H + (\mathbf{F}_h)_H \, dz\right] \\ + (\rho e_T)_S \left(\frac{\partial z_S}{\partial t} + \mathbf{v}_S \cdot \nabla z_S - w_S\right) \\ = -p_S(\mathbf{v}_S \cdot \nabla z_S - w_S) - [(\rho \mathbf{\Lambda}_H)_S \cdot \nabla z_S - (\rho \Lambda_z)_S] \\ - \left\{[(\mathbf{F}_h)_H]_S \cdot \nabla z_S - \left[(F_h)_z\right]_S\right\} - (R_z)_\infty - [(\mathbf{R}_H)_S \cdot \nabla z_S - (R_z)_S]\,.\end{aligned} \tag{7.2}$$

Using (4.15), which is the boundary condition that no mass crosses the Earth's surface, we can simplify (7.2) to

$$\begin{aligned}\frac{\partial}{\partial t}\left(\int_{z_S}^{\infty}\rho e_T\,dz\right)+\nabla_z\cdot\left[\int_{z_S}^{\infty}\rho\mathbf{v}e_T+p\mathbf{v}+\rho\boldsymbol{\Lambda}_H+\mathbf{R}_H+(\mathbf{F}_h)_H\,dz\right]\\ =p_S\frac{\partial z_S}{\partial t}-\left[(\rho\boldsymbol{\Lambda}_H)_S\cdot\nabla z_S-(\rho\Lambda_z)_S\right]\\ -\left\{\left[(\mathbf{F}_h)_H\right]_S\cdot\nabla z_S-\left[(F_h)_z\right]_S\right\}-(R_z)_\infty-\left[(\mathbf{R}_H)_S\cdot\nabla z_S-(R_z)_S\right]\,.\end{aligned}\tag{7.3}$$

The terms on the right-hand side of (7.3) represent energy fluxes at the upper and lower boundaries. As pointed out earlier, the only flux of energy at the upper boundary is that due to radiation, denoted here by $-(R_z)_\infty$. For the lower boundary, there are terms representing pressure-work, frictional work, the flux of moist static energy, and radiation.

When $\partial z_S/\partial t = 0$, the pressure-work term of (7.3) vanishes. Over the ocean, however, the surface height fluctuates due to the passage of waves, which enables the atmosphere and ocean to exchange energy via pressure work. One effect of this is to add energy to the same waves that make the energy exchange possible. Even over land, the vegetation moves as the wind blows through it, so the pressure-work term can be non-zero. In addition, an earthquake or volcanic eruption can impart energy to the atmosphere through the pressure-work term; this is, of course, only episodically important.

The friction terms on the right-hand side of (7.3) represent the work done by surface drag which, as already discussed, is normally small compared to the other fluxes. As discussed in Chapter 6, the surface flux of moist static energy is a very important energy source for the atmosphere. Further discussion will be given in Chapter 8. Finally, as discussed in Chapter 6, the radiative energy flux is a major form of energy exchange at both the upper and lower boundaries of the atmosphere.

When we integrate (7.3) horizontally over the entire sphere, the horizontal flux divergence term vanishes, and we get

$$\begin{aligned}\frac{d}{dt}\left[\int_A\left(\int_{z_S}^{\infty}\rho e_T\,dz\right)\,dA\right]=\int_A p_S\frac{\partial z_S}{\partial t}\,dA-\int_A\left[(\rho\boldsymbol{\Lambda}_H)_S\cdot\nabla z_S-(\rho\Lambda_z)_S\right]\,dA\\ -\int_A\left\{\left[(\mathbf{F}_h)_H\right]_S\cdot\nabla z_S-\left[(F_h)_z\right]_S\right\}\,dA-\int_A(R_z)_\infty\,dA-\int_A\left[(\mathbf{R}_H)_S\cdot\nabla z_S-(R_z)_S\right]\,dA\,.\end{aligned}\tag{7.4}$$

Here $\int_A dA$ denotes a horizontal integral over the sphere. Eq. (7.4) shows that, in the absence of heating and friction, and when the height of the Earth's surface is independent of time,

$$\frac{d}{dt}\left[\int_A \left(\int_{z_S}^{\infty} \rho e_T dz\right) dA\right] = 0, \tag{7.5}$$

which means that the total energy of the atmosphere is invariant.

The pressure-work and frictional work terms of (7.4) remove energy from the global atmosphere, which means that in an overall sense the atmosphere does work on the lower boundary, rather than vice versa. It follows that, in a time average, the radiative and small-scale flux terms of (7.4) must act as an energy source. As discussed in Chapter 5, the rate at which the atmosphere does frictional work on the lower boundary is on the order of a few tenths of a watt per square meter. The rate at which pressure-work transfers energy between the atmosphere and the lower boundary is typically even smaller. As discussed in Chapter 2, the remaining individual terms on the right-hand side of (7.4) are typically larger by several orders of magnitude. It follows that these remaining terms must very nearly cancel in a time average, i.e.,

$$0 \cong -\int_A \left\{\overline{\left[(\mathbf{F}_h)_H\right]_S \cdot \nabla z_S - \left[(F_h)_z\right]_S}^t\right\} dA - \int_A \overline{(R_z)_\infty}^t dA - \int_A \left[\overline{(\mathbf{R}_H)_S \cdot \nabla z_S - (R_z)_S}^t\right] dA. \tag{7.6}$$

In this sense, the total "heating" of the atmosphere, given by the right-hand side of (7.6), is very nearly zero. Eq. (7.6) can be written more simply as

$$\int_V \nabla \cdot (\mathbf{R} + \mathbf{F}_h)\, dV \cong 0, \tag{7.7}$$

where $\int_V ()\, dV$ denotes an integral over the entire volume of the atmosphere.

7.2 The Earth's entropy budget

For any gas or liquid, the entropy per unit mass, ε, satisfies

$$T\frac{D\varepsilon}{Dt} = \frac{De}{Dt} + p\frac{D\alpha}{Dt}\,. \tag{7.8}$$

Comparing with the thermodynamic energy equation, (4.59), we see that

$$\rho T \frac{D\varepsilon}{Dt} = Q + \rho\delta \ . \tag{7.9}$$

Using the equation of state with (7.8), we can show that

$$\varepsilon = c_p \ln\left(\frac{T}{T_0}\right) - R \ln\left(\frac{p}{p_0}\right) \ . \tag{7.10}$$

Here T_0 and p_0 are suitable reference values that enter as "constants of integration." The reference values are needed because the argument of the logarithm must always be nondimensional, as discussed in Appendix B. There is a simple relationship between the entropy and the potential temperature, i.e.,

$$\varepsilon \equiv c_p \ln\left(\frac{\theta}{\theta_0}\right) \tag{7.11}$$

Here θ_0 is the potential temperature corresponding to T_0 and p_0. For saturated air containing liquid water, (7.11) must be replaced by

$$\varepsilon \cong c_p \ln\left(\frac{\theta}{\theta_0}\right) - \frac{Ll}{T} \ , \tag{7.12}$$

(e.g., Iribarne and Godson, 1981).

Eq. (7.9) can be rearranged as

$$\rho \frac{D\varepsilon}{Dt} = \frac{Q + \rho\delta}{T} \ . \tag{7.13}$$

Use continuity to rewrite (7.13) in flux form, then integrate over the entire mass of the atmosphere to obtain

$$\frac{d}{dt} \int\limits_V \rho\varepsilon \, dV = \int\limits_V \left(\frac{Q + \rho\delta}{T}\right) \, dV \ . \tag{7.14}$$

In an average over a sufficiently long time, (7.14) reduces to

$$\int_V \overline{\left(\frac{Q+\rho\delta}{T}\right)}^t dV = 0\,. \tag{7.15}$$

Because $\rho\delta \geq 0$, dissipation never decreases and normally increases the entropy. This allows us to convert (7.15) into an inequality by dropping the dissipation rate:

$$\boxed{\int_V \frac{\overline{Q}^t}{T}\, dV \leq 0\,.} \tag{7.16}$$

This important result means that, for the atmosphere as a whole, heating must act to decrease the entropy. In order for (7.16) to be satisfied, heating must occur, on the average, where the temperature is high, and cooling must occur, on the average, where the temperature is low. We must heat where it's hot and cool where it's cold. Then heating and cooling tend to increase temperature contrasts with time. One way to make this happen is to have heating in the tropics and near the surface, where the air is warm, and cooling near the poles and up high, where the air is cold. To see that this conclusion can be drawn from (7.16), divide Q and T into averages representing the volume integral in (7.15), denoted by overbars, and departures from the averages, denoted by primes:

$$Q = \overline{Q} + Q' \text{ and } T = \overline{T} + T'. \tag{7.17}$$

Then we can write

$$
\begin{aligned}
\overline{\left(\frac{Q}{T}\right)} &= \overline{\left(\frac{\overline{Q}+Q'}{\overline{T}+T'}\right)} \\
&= \overline{\frac{\overline{Q}\left(1+Q'/\overline{Q}\right)}{\overline{T}\left(1+T'/\overline{T}\right)}} \\
&\cong \frac{\overline{Q}}{\overline{T}}\overline{\left(1+Q'/\overline{Q}\right)\left(1-T'/\overline{T}\right)} \\
&= \frac{\overline{Q}}{\overline{T}}\overline{\left[1+\frac{Q'}{\overline{Q}}-\frac{T'}{\overline{T}}-\frac{Q'T'}{\overline{Q}\,\overline{T}}\right]} \\
&= \frac{\overline{Q}}{\overline{T}}\left(1-\frac{\overline{Q'T'}}{\overline{Q}\,\overline{T}}\right) \\
&= \frac{\overline{Q}}{\overline{T}}-\frac{\overline{Q'T'}}{\overline{T}^2}\,.
\end{aligned}
\tag{7.18}
$$

For $\overline{Q}=0$ [see (7.7)], we get

$$
\overline{\left(\frac{Q}{T}\right)} = -\frac{\overline{Q'T'}}{\overline{T}^2}\,. \tag{7.19}
$$

If we add energy where the temperature is already warm, and remove energy where the temperature is already cool, then $\overline{Q'T'} \geq 0$, so $\overline{Q/T}$ will be negative. A process that tends to "heat where it's hot, and cool where it's cold" acts to increase the temperature variance, i.e., the square of the departure of the temperature from its mean value, and it tends to decrease the entropy. We show later in this chapter that such a process also generates available potential energy.

We have concluded that heating and cooling tend to increase the temperature contrasts within the atmosphere. In order for the system to achieve a steady state, some other process must oppose this tendency. That process is energy transport, which on the average carries energy from warm regions to cool regions, e.g., from the tropics towards the polar regions, and from the warm surface towards the cold upper atmosphere. The energy transports by the global circulation tend to cool where the temperature is high (e.g., the tropical lower troposphere), and to warm where the temperature is low (e.g., near the poles). They tend to increase the entropy of the global atmosphere.

A final important point is that the global entropy budget is fundamentally different from the global energy budget. As discussed above, conservation of energy implies that in a time average the fluxes of energy into the system must be balanced by fluxes out. In contrast,

the Earth *makes* entropy via a wide variety of dissipative processes, including radiative, fluid dynamical, and biological processes. As a result, there is a net entropy flux out of the Earth system, via radiation. In a time average this outward flux of entropy is equal to the Earth's entropy production rate. It is therefore possible to use satellite measurements of the entropy flow across the top of the atmosphere to infer the Earth's entropy production rate. This has been done. For further discussion, see Stephens and O'Brien (1993).

7.3 The vertically integrated enthalpy

Edward Lorenz (see Fig. 7.1) was one of the most insightful atmospheric scientists of the twentieth century. In his first "famous" paper, Lorenz (1955) pointed out that, in a hydrostatically balanced atmosphere, the mass-weighted vertical integral of the enthalpy is equal to the sum of the mass-weighted vertical integrals of the internal and potential energies. To demonstrate this, we start from hydrostatics in the form

$$\frac{\partial p}{\partial z} = -\rho g \; . \tag{7.20}$$

The vertically integrated potential energy of the atmosphere, denoted by P, satisfies

$$\begin{aligned} P &\equiv \int_{z_S}^{\infty} (z - z_S)\, \rho g \, dz \\ &= -\int_{z_S}^{\infty} \left[(z - z_S) \frac{\partial p}{\partial z} \right] dz\} \\ &= -\int_{z_S}^{\infty} \left\{ \frac{\partial}{\partial z} \left[(z - z_S)\, p \right] - p \right\} dz \\ &= -\left\{ \left[(z - z_S)\, p \right]_{z=z_S}^{z=\infty} - \int_{z_S}^{\infty} \rho RT \, dz \right\} \\ &= \int_{z_S}^{\infty} \rho RT \, dz \; . \end{aligned} \tag{7.21}$$

Here z_S is the surface height, which we have used as the "reference level" where the potential energy is set to zero. Eq. (7.21) says that the column's total potential energy is proportional to its mass-weighted average temperature. The reason is that warmer air occupies a larger volume, so that a warmer column is "taller." This relation between integrals does *not* hold for the integrands; if it did, we would have $g\,(z - z_S) = RT$, which is obviously nonsense.



Figure 7.1: Prof. Edward N. Lorenz, who created the concept of available potential energy, and also published a lot of very fundamental research on various topics in both atmospheric science and the relatively new science of nonlinear dynamical systems.

Similarly, the vertically intergrated internal energy, I, is given by

$$I \equiv \int_{z_S}^{\infty} \rho c_v T dz \,, \tag{7.22}$$

and the vertically intergrated enthalpy is

$$H \equiv \int_{z_S}^{\infty} \rho c_p T dz \,, \tag{7.23}$$

It follows from (7.21) - (7.23) that

$$\boxed{H = P + I \,.} \tag{7.24}$$

Eq. (7.24) is useful because it says that the vertically integrated enthalpy (one thing) is equal to the sum of the vertically integrated potential energy and the vertically integrated internal energy (two things). One is simpler than two.

7.4 Available potential energy

We showed in section 7.1 above that, under dry adiabatic and frictionless processes, the total energy of the atmosphere is invariant. Using the notation defined above, this condition can be written as

$$\frac{d}{dt}(P + I + K) = 0 \,. \tag{7.25}$$

Here we take P and I to be integrals over the entire atmosphere, and similarly K represents the total kinetic energy of the entire atmosphere. It follows from (7.25) and (7.24) that

$$\boxed{\frac{d}{dt}(H + K) = 0 \text{ under dry adiabatic frictionless processes.}} \tag{7.26}$$

This is the starting point for Lorenz's theory of available potential energy. Imagine that we have the power to adiabatically and reversibly rearrange the mass of the atmosphere, moving parcels around at will. As the parcels move, their entropy and potential temperature do not change, but their enthalpy and temperature do change. Suppose that we are given a state of the atmosphere - a set of maps, if you like. Starting from this "given state," we move parcels around adiabatically and without friction until we find the unique state of the system that minimizes H. This means that we have reduced H as much as possible from its value in the given state. Because $H + K$ does not change, K is maximized in this special state, which Lorenz (1955) called the "reference state," and which we will call the "A-state."

In a problem at the end of this chapter you are asked to prove that the mass-integrated potential energy of the entire atmosphere is lower in the A-state than in the given state. This means that the "center of gravity" of the atmosphere descends as the atmosphere passes from the given state to the A-state. The potential energy that "disappears" in this process must be converted into kinetic energy, because the sum of H and K does not change. Therefore, we can say that an adiabatic process that reduces H will tend to increase K. Two very important processes of this type are convective instability and baroclinic instability.

In passing from the given state to the A-state we have

$$H_{\rm GS} \to H_{\rm min} \ , \tag{7.27}$$

and

$$K_{\rm GS} \to K_{\rm GS} + (H_{\rm GS} - H_{\rm min}) \equiv K_{\rm max} \ , \tag{7.28}$$

where subscript GS denotes the given state, and $H_{\rm min}$ is the value of H in the A-state. The non-negative quantity

$$\boxed{A = H_{\rm GS} - H_{\rm min} \geq 0} \tag{7.29}$$

is called the "available potential energy," or APE. The APE was first defined by Lorenz (1955). Eq. (7.29) is the disarmingly simple definition of the APE.

The APE is a property of the entire atmosphere. The literature contains studies in which the APE is computed, without rigorous justification, for a portion of the atmosphere, e.g., the Northern Hemisphere. An example is discussed later in this chapter.

The A-state is invariant under adiabatic processes, because it depends only on the probability distribution of θ over the mass, rather than on any particular spatial arrangement of the air. Therefore

$$\begin{aligned} \frac{dA}{dt} &= \frac{d}{dt}\left(H_{\mathrm{GS}} - H_{\min}\right) \\ &= \frac{dH_{\mathrm{GS}}}{dt} \,. \end{aligned} \tag{7.30}$$

Then (7.26) implies that

$$\boxed{\frac{d}{dt}\left(A+K\right) = 0 \,.} \tag{7.31}$$

The sum of the available potential and kinetic energies is invariant under adiabatic frictionless processes. Such processes only convert between A and K.

The APE of the A-state itself is obviously zero.

In order to compute A, we have to find the A-state and its (minimum) enthalpy. We can deduce the properties of the A-state as follows: There can't be any horizontal pressure gradients in the A-state, because if there were K could increase. It follows that the potential temperature must be constant on isobaric surfaces, which is of course equivalent to the statement that p is constant on θ surfaces. This means that both the variance of θ on isobaric surfaces and the variance of p on isentropic surfaces are measures of the APE.

For a similar reason, there can be no static instability in the A-state. From these considerations we conclude that, in the A-state,θ and T are uniform on each pressure-surface, or equivalently that p is uniform on each θ-surface, and also that θ does not decrease upward.

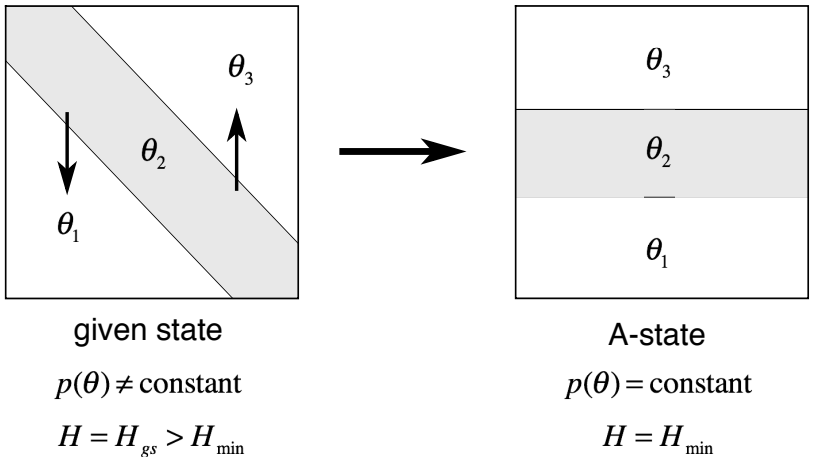
It would appear, and to a large extent it is true, that in passing from the given state to the A-state, no mass can cross an isentropic surface, because we allow only dry adiabatic processes for which $\dot{\theta} = 0$. This implies that $\overline{p}^{\theta}$, the average pressure on an isentropic surface, cannot change as we pass to the A-state. There is an important exception to this rule, however. If $\partial\theta/\partial z < 0$ in the given state, then the average pressure on an isentropic surface will be different in the A-state. This case of static instability is discussed below.

As shown by the heavy arrows in Fig. 7.2, warm air must rise (move to lower pressure) and cold air must sink (move to higher pressure) to pass from the given state to the A-state. This is what happens as APE is released; to reach the A-state, the isobaric surfaces flatten out to coincide with isentropic surfaces.

The concept of massless layers, discussed in Chapter 4, allows us to write

$$\int_0^{p_S} (\quad)\, dp = \int_0^{\infty} (\quad) \frac{\partial p}{\partial \theta} d\theta \ , \tag{7.32}$$

where $(\quad)$ can be anything. Note that the lower limit of integration on the right-hand side of (7.32) is zero. Eq. (7.32) will be used below. It is important for you to understand why it is true.



$$H_{gs} - H_{\min} \equiv APE$$

Figure 7.2: Sketch illustrating the transition from a given state to the A-state. Here p increases downward and we assume that $\theta_3 > \theta_2 > \theta_1$, which means that the atmosphere is statically stable.

A useful expression for the APE can be derived as follows. The total enthalpy is given by

$$\begin{aligned} H =& c_p p_0^{-\kappa} \int_M p^{\kappa} \theta \, dM \\ =& \frac{c_p a^2}{g p_0^{\kappa}} \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} \int_0^{p_S} p^{\kappa} \theta \cos\varphi \, dp \, d\lambda \, d\varphi \ . \end{aligned} \tag{7.33}$$

Integration by parts gives

$$H = \frac{c_p a^2}{(1+\kappa)\, g p_0^\kappa} \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} \int_0^{\infty} p^{1+\kappa} \cos\varphi \; d\theta \; d\lambda \; d\varphi \; . \tag{7.34}$$

Note that vertical integration is now with respect to θ rather than p, and that the lower limit of integration is $\theta = 0$. Let $\overline{p}^\theta$ be the average pressure on an isentropic surface, taking into account possible intersections with the ground. Recall that $\overline{p}^\theta$ is the same in the A-state as in the given state, so long as there are no regions of dry static instability. Then use of (7.34) in (7.29) gives

$$A = \frac{c_p a^2}{(1+\kappa)\, g p_0^\kappa} \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} \int_0^{\infty} \left[p^{1+\kappa} - \left(\overline{p}^\theta \right)^{1+\kappa} \right] \cos\varphi \; d\theta \; d\lambda \; d\varphi \; . \tag{7.35}$$

Note that (7.35) is valid only if $\partial\theta/\partial z \geq 0$ everywhere in the given state, because we have assumed that $\overline{p}^\theta$ is the same in the A-state as in the given state. Apart from that caveat, (7.35) is exact. The most general expression for the available potential energy is the definition (7.29).

Let $\overline{(\quad)}^\theta$ denote a global average on an isentropic surface, and let $p'(\theta)$ be the departure of p from its average on an isentropic surface, so that $p = \overline{p}^\theta + p'$, and $\overline{p'}^\theta = 0$. The binomial theorem tells us that

$$\begin{aligned} p^{1+\kappa}(\theta) &= \left(\overline{p}^\theta \right)^{1+\kappa} \left[1 + \frac{p'(\theta)}{\overline{p}^\theta} \right]^{1+\kappa} \\ &= \left(\overline{p}^\theta \right)^{1+\kappa} \left\{ 1 + (1+\kappa) \frac{p'(\theta)}{\overline{p}^\theta} + \frac{\kappa(1+\kappa)}{2!} \left[\frac{p'(\theta)}{\overline{p}^\theta} \right]^2 + \cdots \right\} . \end{aligned} \tag{7.36}$$

Lorenz averaged both sides of (7.36) to obtain

$$\overline{p^{1+\kappa}}(\theta) \cong \left(\overline{p}^\theta \right)^{1+\kappa} \left\{ 1 + \frac{\kappa(1+\kappa)}{2!} \overline{\left[\frac{p'(\theta)}{\overline{p}^\theta} \right]^2} \right\} . \tag{7.37}$$

and he showed that this is actually a pretty good approximation. Substitution of (7.37) into (7.35) gives

$$A \cong \frac{Ra^2}{2gp_0^\kappa} \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} \int_0^{\infty} \left(\overline{p}^\theta\right)^{1+\kappa} \left(\frac{p'}{\overline{p}^\theta}\right)^2 \cos\varphi \, d\theta \, d\lambda \, d\varphi \, . \tag{7.38}$$

This demonstrates that A is related to the variance of the pressure on isentropic surfaces.

Because he wanted to express his results in terms of temperature or potential temperature perturbations on isobaric surfaces, rather than pressure perturbations on isentropic surfaces, Lorenz also used

$$\begin{aligned} p'(\theta) \cong & \theta'(p) \frac{\partial p(\theta)}{\partial \theta} \\ \cong & \theta'(p) \frac{\partial \overline{p}^\theta}{\partial \theta} \\ \cong & \theta'(p) \left(\frac{\partial \overline{\theta}^p}{\partial p}\right)^{-1} , \end{aligned} \tag{7.39}$$

where, as before, $p'(\theta)$ represents the departure of p from $\overline{p}^\theta$, its global average on an isentropic surface, and $\theta'(p)$ represents the departure of θ from $\overline{\theta}^p$, its global average on a p-surface. Substitution of (7.39) into (7.38) gives

$$A \cong \frac{Ra^2}{2gp_0^\kappa} \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} \int_0^{p_S} \frac{\left(\overline{\theta}^p\right)^2}{(p^{1-\kappa})\left(-\frac{\partial \overline{\theta}^p}{\partial p}\right)} \left(\frac{\theta'}{\overline{\theta}^p}\right)^2 \cos\varphi \, dp \, d\lambda \, d\varphi \, . \tag{7.40}$$

Here the independent variable used for vertical integration has been changed from θ to p The limits of vertical integration have been changed accordingly, and a minus sign has been applied to the factor of $\partial \overline{\theta}^p / \partial p$ in the denominator. Eq. (7.40) says that the available potential energy is a weighted average of the variance of θ on a pressure surface. Variances are quite important in the study of the general circulation. The mechanisms that give rise to variances are discussed in some detail later in this chapter.

Finally, Lorenz used the hydrostatic equation in the form

$$-\frac{\partial\theta}{\partial p} = \frac{\kappa\theta}{p}\left(\frac{\Gamma_d - \Gamma}{\Gamma_d}\right) , \tag{7.41}$$

and also

$$\frac{\theta'}{\overline{\theta}^p} = \frac{T'}{\overline{T}^p} , \tag{7.42}$$

to rewrite (7.40) as

$$A = \frac{a^2}{2}\int_{-\pi/2}^{\pi/2}\int_{0}^{2\pi}\int_{0}^{p_S}\frac{\overline{T}^p}{(\Gamma_d - \Gamma)}\left(\frac{T'}{\overline{T}^p}\right)^2 \cos\varphi \, dp \, d\lambda \, d\varphi , \tag{7.43}$$

where

$$\Gamma \equiv -\frac{\partial T}{\partial z} \tag{7.44}$$

is the lapse rate, and

$$\Gamma_d = \frac{g}{c_p} \tag{7.45}$$

is the dry adiabatic lapse rate followed by a particle moving under dry adiabatic processes. The dry adiabatic lapse rate is approximately 9.8 K km^{-1}. Eq. (7.43) shows that the available potential energy is closely related to the variance of temperature on isobaric surfaces. The available potential energy also increases as the lapse rate increases, i.e., as the atmosphere becomes less statically stable in the dry sense.

Observations show that the APE is only about half a percent of $P + I$, and that it is comparable in magnitude to the total kinetic energy. Both the APE and the total kinetic energy are on the order of 10^6 - 10^7 J m^{-2}.

7.5 The gross static stability

The A-state defines a correspondence or mapping between θ and p. For each p there is one possible value of θ (the converse is not necessarily true). We can say that in the A-state, p and θ are perfectly correlated.

Consider the opposite limit, in which θ and p are completely uncorrelated. In this "S-state," all possible values of θ occur, with equal probability, for any given p. This would be the case if the θ-surfaces were vertical, so that $\partial\theta/\partial p = 0$, and the surface pressure is globally uniform. See Fig. 7.3. A globally uniform surface pressure seems reasonable enough in the absence of topography, but it is a very strange state when topography is present. Lorenz (personal communication, 2003) suggested an alternative definition of the S-state, in which the surface pressure is allowed to vary in a simple and natural way with the surface height (roughly as shown in Figure 3.1, but the (vertically uniform) potential temperature is geographically distributed so that it is uncorrelated with the surface elevation (and, therefore, uncorrelated with the surface pressure).

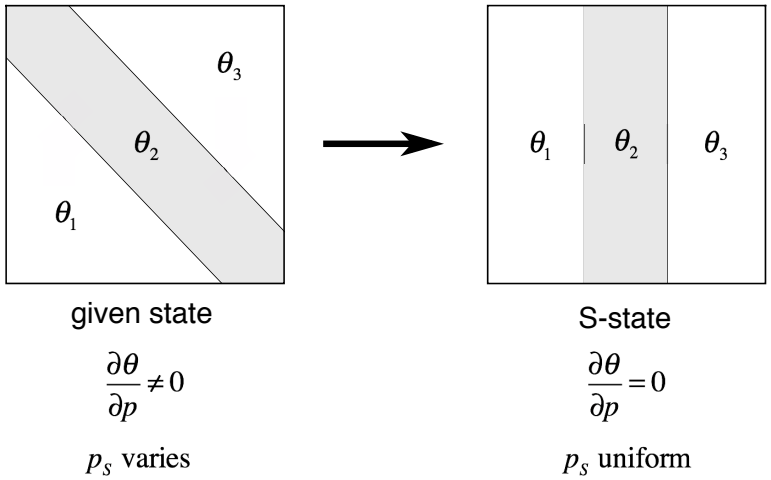


Figure 7.3: Sketch illustrating the transition from a given state to the S-state used to define the gross static stability.

We pass from the given state to the S-state by way of an adiabatic process; it follows that the S-state itself is invariant under adiabatic processes. The gross static stability is defined to be the enthalpy of the S-state minus the enthalpy of the given state, i.e.,

$$S \equiv H_{\rm SS} - H_{\rm GS} \ . \tag{7.46}$$

Here the subscript *SS* denotes the S-state. The total enthalpy of the S-state is given by

$$
\begin{aligned}
H_{\rm SS} =& c_p p_0^{-\kappa} \int\limits_M p^\kappa \theta dM \\
=& \frac{c_p a^2}{g p_0^\kappa} \int\limits_{-\pi/2}^{\pi/2} \int\limits_0^{2\pi} \int_0^{p_S} p^\kappa \theta \cos\varphi dp d\lambda d\varphi \\
=& \frac{c_p a^2}{g p_0^\kappa} \int\limits_{-\pi/2}^{\pi/2} \int\limits_0^{2\pi} \theta \left(\int_0^{p_S} p^\kappa \, dp \right) d\lambda \cos\varphi d\varphi \\
=& \frac{c_p a^2}{g p_0^\kappa} \int\limits_{-\pi/2}^{\pi/2} \int\limits_0^{2\pi} \theta \left(\frac{p_S^{1+\kappa}}{1+\kappa} \right) d\lambda \cos\varphi d\varphi \, .
\end{aligned}
\tag{7.47}
$$

At this point, we can consider each of the two definitions of the S-state that were mentioned above. If the surface pressure is globally uniform in the S-state, then we can simply replace p_S by $\overline{p}_S$ in (7.47), and write

$$
\begin{aligned}
H_{\rm SS} =& \frac{c_p a^2}{g p_0^\kappa} \int\limits_{-\pi/2}^{\pi/2} \int\limits_0^{2\pi} \theta \left[\frac{(\overline{p_S})^{1+\kappa}}{1+\kappa} \right] d\lambda \cos\varphi d\varphi \\
=& \frac{c_p a^2}{g p_0^\kappa} \left[\frac{(\overline{p_S})^{1+\kappa}}{1+\kappa} \right] \int\limits_{-\pi/2}^{\pi/2} \int\limits_0^{2\pi} \theta d\lambda \cos\varphi d\varphi \\
=& \frac{c_p 4\pi a^4}{g p_0^\kappa} \left[\frac{(\overline{p_S})^{1+\kappa} \overline{\theta}}{1+\kappa} \right] ,
\end{aligned}
\tag{7.48}
$$

where $\overline{p_S}$ is the globally averaged surface pressure, which is the same in the given state and the S-state, and $\overline{\theta}$ is the globally averaged potential temperature in the S-state, which is the same as the mass-averaged potential temperature in the given state.

Alternatively, if θ is uncorrelated with ${p_S}^{1+\kappa}$ in the the S-state, then it follows immediately from (7.47) that

$$
H_{\rm SS} = \frac{c_p 4\pi a^4}{g p_0^\kappa} \left[\frac{(\overline{p_S})^{1+\kappa} \overline{\theta}}{1+\kappa} \right] \, . \tag{7.49}
$$

We conclude that the two definitions of the S-state actually give exactly the same result for $H_{\rm SS}$.

Why is S called the "gross static stability?" To show the explanation, Eq. (7.47) can also be written as

$$H_{\rm SS} = \frac{c_p a^2}{g}\left(\frac{\tilde{p}^\kappa}{p_0^\kappa}\right)\int_{-\pi/2}^{\pi/2}\int_0^{2\pi}\left(\int_0^{p_S}\theta dp\right)\cos\varphi d\lambda d\varphi\,, \tag{7.50}$$

where

$$\tilde{p}^\kappa \equiv \frac{(\overline{p_S})^\kappa}{1+\kappa}\,. \tag{7.51}$$

In (7.50), the integral over pressure simply amounts to multiplication by $\overline{p_S}$, because θ is independent of height in the S-state. The integral of the potential temperature over the entire mass of the atmosphere must be exactly the same for the given state and the S-state, i.e.,

$$\frac{a^2}{g}\left[\int_{-\pi/2}^{\pi/2}\int_0^{2\pi}\left(\int_0^{p_S}\theta dp\right)\cos\varphi d\lambda d\varphi\right]_{\rm GS} = \frac{a^2}{g}\left[\int_{-\pi/2}^{\pi/2}\int_0^{2\pi}\left(\int_0^{\overline{p_S}}\theta dp\right)\cos\varphi d\lambda d\varphi\right]_{SS}\,. \tag{7.52}$$

This allows us to rewrite (7.50) as

$$H_{\rm SS} = \frac{c_p a^2}{g}\left(\frac{\tilde{p}^\kappa}{p_0^\kappa}\right)\left[\int_{-\pi/2}^{\pi/2}\int_0^{2\pi}\left(\int_0^{p_S}\theta dp\right)\cos\varphi d\lambda d\varphi\right]_{\rm GS}\,. \tag{7.53}$$

Now substitute (7.53) into (7.46), to obtain

$$S \equiv \frac{c_p a^2}{g}\left(\frac{\tilde{p}^\kappa}{p_0^\kappa}\right)\left[\int_{-\pi/2}^{\pi/2}\int_0^{2\pi}\left(\int_0^{p_S}\theta dp\right)\cos\varphi d\lambda d\varphi\right]_{\mathrm{GS}} - \frac{c_p a^2}{g p_0^\kappa}\int_{-\pi/2}^{\pi/2}\int_0^{2\pi}\int_0^{p_S} p^\kappa \theta\cos\varphi dp d\lambda d\varphi\,. \tag{7.54}$$

Eq. (7.54) can be rearranged to

$$S = \frac{c_p a^2}{g p_0^\kappa}\left\{\int_{-\pi/2}^{\pi/2}\int_0^{2\pi}\left[\int_0^{p_S}\left(\tilde{p}^\kappa - p^\kappa\right)\theta dp\right]\cos\varphi d\lambda d\varphi\right\}_{\mathrm{GS}}\,. \tag{7.55}$$

Integration by parts gives

$$S = \frac{c_p a^2}{g p_0^\kappa\,(1+\kappa)}\left\{\int_{-\pi/2}^{\pi/2}\int_0^{2\pi}\left[\int_0^{p_S}\left[\left(\overline{p_S}\right)^\kappa p - p^{1+\kappa}\right]\left(-\frac{\partial\theta}{\partial p}\right)dp\right]\cos\varphi d\lambda d\varphi\right\}_{\mathrm{GS}}, \tag{7.56}$$

which shows that S is a weighted average of $-\partial\theta/\partial p$; it is, therefore, a measure of the static stability, and this accounts for its name. Like the available potential energy, the gross static stability is defined only for the atmosphere as a whole.

The globally averaged surface pressure and the probability distribution of θ are invariant under adiabatic processes. This implies that the S-state is invariant as well. Because S is defined as the difference between the total enthalpy of the S-state and the total enthalpy of the given state, we can write

$$\frac{dS}{dt} = -\frac{dH}{dt}\,, \tag{7.57}$$

Since $d/dt\,(H+K) = 0$, it follows from (7.57) that

$$\frac{dS}{dt} = \frac{dK}{dt} \ , \tag{7.58}$$

This shows that when K is produced by conversion from APE, the gross static stability increases. In passing from the given state to the A-state, the available potential energy decreases, and the kinetic energy and gross static stability both increase. Isn't that what you would expect?

The observed state of the atmosphere is "in between" the A-state and the S-state.

7.6 How available potential energy is converted into kinetic energy

As discussed in Chapter 8, convection produces an upward flux of dry static energy, which means also an upward flux of θ and T. Suppose that the θ-profile in a particular atmospheric column is altered by a vertical flux of θ, i.e.,

$$\frac{\partial \theta}{\partial t} = g \frac{\partial F_\theta}{\partial p} \tag{7.59}$$

It follows that the enthalpy changes according to

$$c_p \frac{\partial T}{\partial t} = c_p g \frac{\partial}{\partial p} \left[\left(\frac{p}{p_0} \right)^\kappa F_\theta \right] - \frac{Rg}{p_0^\kappa} \left(\frac{F_\theta}{p^{1-\kappa}} \right) \ . \tag{7.60}$$

Integrating through the depth of the column, we find that the change of the total enthalpy of the column satisfies

$$\frac{\partial}{\partial t} \left(\int_0^{p_S} c_p T \, dp \right) = c_p \left[T_S \frac{\partial p_S}{\partial t} + g \left(\frac{p_S}{p_0} \right)^\kappa (F_\theta)_S \right] - \frac{Rg}{{p_0}^\kappa} \int_0^{p_S} \left(\frac{F_\theta}{p^{1-\kappa}} \right) dp \ . \tag{7.61}$$

The first term inside the square brackets on the right-hand side of (7.61) is zero if no mass is exchanged with neighboring columns. The second term inside the square brackets is diabatic, because it is proportional to an exchange of energy between the atmosphere and the lower boundary. The remaining term, $-\left(\frac{gR}{{p_0}^\kappa} \right) \int_0^{p_S} \frac{F_\theta}{p^{1-\kappa}} \, dp$, arises purely from

redistribution of θ within the column. The form of this redistribution term makes it clear that $F_\theta > 0$, i.e., an upward flux of θ, tends to reduce the total enthalpy of the column. It follows that the total potential energy of the column also decreases, or in other words that the "center of mass" of the column moves downward. The conclusion is that an adiabatic process that produces an upward flux of θ reduces the total enthalpy of the column, and so generates kinetic energy. This is relevant to both small-scale convection and large-scale "baroclinic eddies." Both produce an upward flux of θ, and so both lower the atmosphere's center of mass. Both convert available potential energy into kinetic energy.

7.7 Examples: The available potential energies of three simple systems

Available potential energy can arise in several different ways. We now consider three idealized examples, each of which illustrates a "source" of available potential energy in pure, unadulterated form.

7.7.1 Example #1: The APE associated with static instability

Consider two parcels of equal mass. In the given state, parcels with potential temperature θ_1 and θ_2 reside at pressures p_1 and p_2, respectively. We assume that $\theta_1 < \theta_2$ and $p_1 < p_2$, so that the given state is statically unstable. The enthalpy per unit mass of parcel i is $c_p \theta_i \left(p_i/p_0\right)^\kappa$. If the parcels are interchanged (or "swapped") so that parcel number 2 goes to pressure p_1 and vice versa, the change in the total enthalpy per unit mass is

$$\Delta H = c_p \left(\theta_1 - \theta_2\right) \left[\left(\frac{p_2}{p_0}\right)^\kappa - \left(\frac{p_1}{p_0}\right)^\kappa \right] , \tag{7.62}$$

which is negative. This implies that the total enthalpy has been reduced and so is minimized by the swap; the final state is the A-state, and the change in enthalpy given by (7.62) is minus the available potential energy of the system, per unit mass.

The process described above is an idealization of dry atmospheric convection, which transports potential temperature upward, increasing the static stability and lowering the atmosphere's center of mass. Lorenz (1978) generalized the concept of APE for a moist atmosphere, in which moist adiabatic processes are acknowledged to be, well, adiabatic. Randall and Wang (1992) showed that this moist APE can be used to define a generalized CAPE that represents the potential energy available for conversion into the kinetic energy of cumulus convection.

7.7.2 Example #2: The APE associated with meridional temperature gradients

The second example is intended to illustrate how APE can arise from a meridional temperature gradient. Consider an idealized planet, with no orography and a uniform surface pressure in the given state. Suppose that the potential temperature of the given state is a function of latitude only:

$$\theta_{\mathrm{GS}}(\mu) = \theta_0\left(1 - \Delta_H \mu^2\right) . \tag{7.63}$$

Here Δ is a constant, and $\mu \equiv \sin\varphi$. For realistic states $0 < \Delta_H < 1$. To obtain an expression for the APE, we need to compute the total enthalpies of the given state and the A-state, and subtract them. The first step is to find the A-state. Note that, in this idealized example, the given state is identical to the S-state.

The mass in a latitude belt of width $d\varphi$ is:

$$\begin{aligned} dm =& 2\left(2\pi a\cos\varphi\right)\left(\frac{p_S}{g}\right)\left(a\, d\varphi\right) \\ =& \frac{4\pi a^2 p_S}{g}\, d\mu \ . \end{aligned} \tag{7.64}$$

Here the leading factor of two is included because we have symmetry across the equator, so when we increment latitude we actually pick up mass from two "rings" of air, one in each hemisphere. The rate of change of θ as we add mass is:

$$\frac{d\theta}{dm} = \left(\frac{d\theta}{d\mu}\right)_{\mathrm{GS}} \left(\frac{dm}{d\mu}\right)^{-1} . \tag{7.65}$$

Combining (7.64) and (7.65), we find that

$$\frac{d\theta}{dm} = \frac{g}{4\pi a^2 p_S}\left(\frac{d\theta}{d\mu}\right)_{\mathrm{GS}} . \tag{7.66}$$

In the A-state, the θ-surfaces are flat, so that there is no meridional temperature gradient. Potentially warm air fills the upper portion of the domain, and potentially cool air fills the lower portion. In order to reach the A-state, the warm air in the given state has to rise

and spread out toward the poles, while the cold air descends and spreads out toward the Equator. Potential temperature is thus transported both upward and poleward.

To find the structure of the A-state, we start by writing down the increment of mass between two θ-surfaces, which is

$$dm = \frac{4\pi a^2}{g}\, dp\ . \tag{7.67}$$

or

$$dm = \frac{4\pi a^2}{g}\left(\frac{dp}{d\theta}\right)_{AS} d\theta\ . \tag{7.68}$$

from which it follows that

$$\left(\frac{d\theta}{dp}\right)_{AS} = \frac{4\pi a^2}{g}\frac{d\theta}{dm}\ . \tag{7.69}$$

Here the subscript "AS" denotes the A-state. Next, we substitute from (7.66) into (7.69). We can do this because the distribution of θ over the mass must be the same in the A-state as in the given state. The result is

$$\left(\frac{d\theta}{d\sigma}\right)_{AS} = \left(\frac{d\theta}{d\mu}\right)_{\rm GS}\ , \tag{7.70}$$

where

$$\sigma \equiv \frac{p}{p_S} \tag{7.71}$$

is a normalized pressure, usually called the "sigma coordinate." I'm using the symbol σ here because it is conventional, even though the same symbol was used in Chapter 8 to denote the fractional area covered by convective updrafts.

Care is needed with Eq. (7.70). The left-hand-side refers to the distribution of θ with pressure in the A-state. The right-hand-side refers to the distribution of θ with μ in the

given state. Keep in mind that, in this idealized example, θ does not vary with pressure in the given state, and it does not vary with μ in the A-state.

From (7.71), we see that σ plays the same role in the A-state as μ plays in the given state. Also recall that θ must increase upward in the A-state. Referring back to (7.63), we conclude that

$$\theta_{\text{AS}} = \theta_0 \left(1 - \Delta_H \sigma^2\right) \ . \tag{7.72}$$

which is the desired formula for the distribution of potential temperature in the A-state. It can be verified that (7.72) gives the correct values of θ at $p = 0$ and $p = p_S$.

Now all we have to do to find the APE is work out the enthalpy in the given state and the A-state, and subtract. In the given state, the potential temperature is independent of height, so (7.63) gives

$$\begin{aligned} H_{\text{GS}} =& \int_{-1}^{1} \int_{0}^{p_S} 2\pi a^2 c_p \theta_0 \left(1 - \Delta_H \mu^2\right) \left(\frac{p}{p_0}\right)^{\kappa} \frac{dp}{g} d\mu \\ =& 4\pi a^2 c_p \theta_0 \left(\frac{p_S}{g}\right) \left(\frac{p_S}{p_0}\right)^{\kappa} \left(\frac{1 - \frac{1}{3}\Delta_H}{1+\kappa}\right) \ . \end{aligned} \tag{7.73}$$

Similarly, the total enthalpy of the A-state is

$$\begin{aligned} H_{\text{min}} =& \int_{-1}^{1} \int_{0}^{p_S} 2\pi a^2 c_p \theta_0 \left[1 - \Delta_H \left(\frac{p}{p_S}\right)^2\right] \left(\frac{p}{p_0}\right)^{\kappa} \frac{dp}{g} d\mu \\ =& 4\pi a^2 c_p \theta_0 \left(\frac{p_S}{g}\right) \left(\frac{p_S}{p_0}\right)^{\kappa} \left[\left(\frac{1}{1+\kappa}\right) - \left(\frac{\Delta_H}{3+\kappa}\right)\right] \ . \end{aligned} \tag{7.74}$$

Finally, we obtain

$$A = H_{\text{GS}} - H_{\text{min}} = 4\pi a^2 c_p \theta_0 \Delta_H \left(\frac{p_S}{g}\right) \left(\frac{p_S}{p_0}\right)^{\kappa} \left[\frac{2\kappa}{3(3+\kappa)(1+\kappa)}\right] \ . \tag{7.75}$$

Note that A is proportional to Δ_H, as might be expected.

Now consider a meridional transport process:

$$\begin{aligned}\frac{\partial \theta_{\mathrm{GS}}}{\partial t} &= -\frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(F_\theta \cos\varphi\right) \\ &= -\frac{1}{a}\frac{\partial}{\partial \mu}\left(F_\theta\sqrt{1-\mu^2}\right) .\end{aligned} \tag{7.76}$$

Here F_θ is a meridional flux of potential temperature, which we regard as given. We assume that F_θ is independent of height and longitude and symmetrical about the Equator. It is necessary that

$$F_\theta = 0 \text{ at both poles.} \tag{7.77}$$

Assume that the surface pressure does not change with time at any latitude. What is the time rate of change of the APE associated with this meridional redistribution of potential temperature?

To answer this question, first note that the time rate of change of the total enthalpy of the given state is

$$\begin{aligned}\frac{\partial H_{\mathrm{GS}}}{\partial t} =& \int_{-1}^{1}\int_{0}^{p_S} 2\pi a^2 c_p \left(\frac{\partial \theta}{\partial t}\right)\left(\frac{p}{p_0}\right)^\kappa \frac{dp}{g} \\ =& \frac{2\pi a^2 c_p}{(1+\kappa)}\left(\frac{p_S}{g}\right)\left(\frac{p_S}{p_0}\right)^\kappa \int_{-1}^{1}\frac{\partial \theta}{\partial t}d\mu \\ =& -\frac{2\pi a^2 c_p}{(1+\kappa)}\left(\frac{p_S}{g}\right)\left(\frac{p_S}{p_0}\right)^\kappa \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(F_\theta\cos\varphi\right)\cos\varphi d\varphi \\ =& 0 .\end{aligned} \tag{7.78}$$

Here we have used the facts that $\partial\theta/\partial t$ is independent of height and that p_S is independent of latitude, and we have substituted from (7.76). According to (7.78), the specified transport process has no effect on the total enthalpy of the given state. The reason is that the average on each pressure surface is unchanged, and it follows that the average temperature on each pressure surface is unchanged.

The meridional transport process can, however, alter the total enthalpy of the A-state. Here again there is a possibility of confusion. As already mentioned, the specified meridional transport process does not alter the average value of θ on a pressure surface. This statement seems to imply that the process is isentropic, and we already know that isentropic processes do not alter the A-state. The entropy is proportional to $\ln\theta/\theta_0$, however, and the average value of $\ln\theta/\theta_0$ *is* altered by the transport process. Another point of view is that generally speaking the specified transport process is not reversible. For example, if F_θ is a downgradient flux due to diffusive mixing, then it could, in principle, homogenize θ throughout the entire atmosphere. This process would clearly be irreversible. Following such homogenization, the A-state would be the same as the (homogenized) given state and would, therefore, be different from the A-state found above.

From (7.78), it follows that

$$\frac{dA}{dt} = -\frac{dH_{\min}}{dt} \ . \tag{7.79}$$

Now recall that, based on comparison of (7.63) and (7.72), σ plays the same role in the A-state as μ plays in the given state. In particular, $\sigma = 1$ (the surface) corresponds to (the pole), and $\sigma = 0$ (the "top of the atmosphere") corresponds to (the Equator). Our goal is to determine the time-rate of change of θ in the A-state at a particular instant, namely, the time when the distribution of θ satisfies (7.63) [and (7.72)], and the time-rate-of-change of θ in the given state satisfies (7.76). We can find the time-rate-of-change of θ in the A-state at this instant by simply by going to our expression for the time-rate-of-change of θ in the given state, and replacing μ by σ, everywhere. The time-rate-of-change of θ in the A-state thus satisfies

$$\frac{\partial\theta_{\rm AS}}{\partial t} = -\frac{1}{a}\frac{\partial}{\partial p_*}\left(F_\theta\sqrt{1-p_*^2}\right) \ . \tag{7.80}$$

Again there is a possibility of confusion. We have specified that F_θ is not a function of height, although of course it does depend on latitude. It would thus appear that we can pull out of the derivative in (7.80), but this is not correct. The reason is that, when we replaced μ by σ we also replaced the μ-dependence of by a corresponding σ-dependence. Thus, in (7.80), F_θ should be regarded as a function of σ, but not as a function of latitude! This is understandable, because F_θ is acting to change $\theta_{AS}(\sigma)$. As an example, suppose that F_θ is symmetric about the Equator, and poleward in both hemispheres, and that $\Delta_H > 0$ so that the poles are in fact colder than the tropics. Then F_θ tends to warm the poles and cool the tropics, reducing Δ_H and, we expect, reducing A. As the tropics cool and the poles warm

in the given state, the A-state evolves in a corresponding way, so that θ_{AS} cools aloft and warms at the lower levels.

We now write the time-rate-of-change of the total enthalpy in the A-state as

$$\begin{aligned}\frac{\partial H_{\rm GS}}{\partial t} =& \int_{-1}^{1}\int_{0}^{p_S} 2\pi a^2 c_p \left(\frac{\partial\theta}{\partial t}\right)\left(\frac{p}{p_0}\right)^{\kappa}\frac{dp}{g} \\ =& \frac{2\pi a^2 c_p}{(1+\kappa)}\left(\frac{p_S}{g}\right)\left(\frac{p_S}{p_0}\right)^{\kappa}\int_{-1}^{1}\frac{\partial\theta}{\partial t}d\mu \\ =& -\frac{2\pi a^2 c_p}{(1+\kappa)}\left(\frac{p_S}{g}\right)\left(\frac{p_S}{p_0}\right)^{\kappa}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(F_\theta\cos\varphi\right)\cos\varphi d\varphi \\ =& 0\ .\end{aligned} \tag{7.81}$$

We cannot do the integral on the last line of (7.81), because the form of F_θ has not been specified. The last step would be to substitute (7.81) into (7.79).

The example above is relevant to the process by which mid-latitude baroclinic eddies transport potential temperature upward and poleward, reducing the meridional temperature gradient and increasing the static stability.

7.7.3 Example #3: The APE associated with surface pressure variations

The third example is designed to illustrate that APE can occur in the presence of surface pressure gradients, even when there are no potential temperature gradients; this is analogous to the APE of shallow water with a non-uniform free-surface height. To explore this possibility in a simple framework, consider a planet with an atmosphere of uniform potential temperature, $\theta = \theta_0$. The surface pressure is given as a function of longitude and latitude. For simplicity we assume that the Earth's surface is flat, although a similar but more complicated analysis can be developed for the case of arbitrary surface topography.

We begin with (7.29), the basic definition of the APE. The total enthalpy satisfies

$$H = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_{0}^{2\pi}\int_{0}^{p_S} c_p\theta\left(\frac{p}{p_0}\right)^{\kappa}\frac{dp}{g}a\cos\varphi\, d\lambda\, ad\varphi\ . \tag{7.82}$$

Because $\theta = \theta_0$ everywhere in this example, we can simplify (7.82) considerably:

$$H = \frac{c_p \theta_0 a^2}{g\,(1+\kappa)\,p_0^{\kappa}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{0}^{2\pi} {p_S}^{1+\kappa} \cos\varphi \, d\lambda \, d\varphi \, . \tag{7.83}$$

Eq. (7.83) can be applied to both the given state and the A-state, so that the APE is given by

$$A = \frac{c_p \theta_0 a^2}{g\,(1+\kappa)\,p_0^{\kappa}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{0}^{2\pi} \left[{(p_S)_{\rm GS}}^{1+\kappa} - {(p_S)_{AS}}^{1+\kappa} \right] \cos\varphi \, d\lambda \, d\varphi \, . \tag{7.84}$$

Before we can evaluate the double integrals in (7.84), we must substitute for $[p_S(\lambda,\varphi)]_{\rm GS}$ and $[p_S(\lambda,\varphi)]_{AS}$. The former is assumed to be known. Our problem thus reduces to finding $[p_S(\lambda,\varphi)]_{AS}$. This is very straightforward, because (since by assumption there are no mountains) the surface pressure is globally uniform in the A-state, and equal to the globally averaged surface pressure in the given state. We denote this globally averaged surface pressure by $\overline{p_S}$, and rewrite (7.81) as

$$A = \frac{c_p \theta_0 a^2}{g\,(1+\kappa)\,p_0^{\kappa}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{0}^{2\pi} \left[{(p_S)_{\rm GS}}^{1+\kappa} - (\overline{p_S})^{1+\kappa} \right] \cos\varphi \, d\lambda \, d\varphi \, . \tag{7.85}$$

As an exercise, prove that (7.85) gives $A \geq 0$.

This example can be generalized to include topography.

7.8 Where do variances come from?

Eq. (7.40) shows that the available potential energy is closely related to the spatial variance of potential temperature on pressure surfaces. We now examine a conversion process that couples the variance associated with the meridional gradient of the zonally averaged potential temperature with the eddy variance of potential temperature. This same process is closely related to the conversion between the zonal available potential energy, A_Z, and eddy available potential energy, A_E. We show later in this section that the eddy potential temperature variance interacts with the meridional gradient of the zonally averaged potential temperature through

$$\frac{\partial}{\partial t}\left(\frac{1}{2} \overline{\theta^{*2}}^{\lambda} \right) \sim \frac{-\overline{\theta^* v^*}^{\lambda}}{a} \frac{\partial \overline{\theta}^{\lambda}}{\partial \varphi} \, . \tag{7.86}$$

The term shown in the right-hand side of (7.86) can be called the "meridional gradient-production term." The full equation for $\frac{1}{2}\overline{\theta^{*2}}^{\lambda}$ contains several additional terms, which will be discussed below.

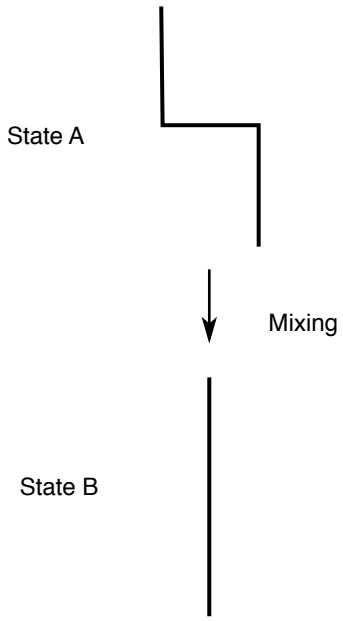


Figure 7.4: A simple example to explain the idea of gradient production. State B is obtained by homogenizing State A. The average of the square decreases in passing from State A to State B.

To gain an intuitive understanding of the gradient production terms, consider the simple example illustrated in Fig. 7.4. State A consists of two latitude belts of equal mass, each with a different (but uniform) θ. State B is obtained by homogenizing State A, without changing the average. Consider the average of the square of θ, for each state. For State A,

$$\overline{\theta^2} = \frac{1}{2}\left(\theta_1^2 + \theta_2^2\right) \ . \tag{7.87}$$

Here the overbar denotes an average over both latitude belts. For State B,

$$\begin{aligned}
\overline{\theta}^2 &= \left[\frac{1}{2}\left(\theta_1 + \theta_2\right)\right]^2 \\
&= \frac{1}{4}\left(\theta_1^2 + 2\theta_1\theta_2 + \theta_2^2\right) \\
&= \frac{1}{2}\left(\theta_1^2 + \theta_2^2\right) - \frac{1}{4}\left(\theta_1^2 - 2\theta_1\theta_2 + \theta_2^2\right) \cdot \\
&= \frac{1}{2}\left(\theta_1^2 + \theta_2^2\right) - \frac{1}{4}\left(\theta_1 - \theta_2\right)^2 \\
&\leq \frac{1}{2}\left(\theta_1^2 + \theta_2^2\right) \ .
\end{aligned} \tag{7.88}$$

This demonstrates that mixing reduces the average of the square. This appears as a "dissipation" of the *variance of the zonal mean*. We will show later that a process that decreases the variance of the zonal mean also increases the *eddy variance*, i.e., the variance associated with departures from the zonal mean.

To see where (7.86) comes from, start from the conservation equation for potential temperature, which can be written out in spherical coordinates as

$$\frac{\partial\theta}{\partial t}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda}(u\theta)+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}(v\theta\cos\varphi)+\frac{\partial}{\partial p}(\omega\theta)=\dot{\theta}\,. \tag{7.89}$$

We will also need mass continuity, in the form

$$\frac{1}{a\cos\varphi}\frac{\partial u}{\partial\lambda}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}(v\cos\varphi)+\frac{\partial\omega}{\partial p}=0\,. \tag{7.90}$$

Zonally average (7.89) and (7.90), to obtain

$$\frac{\partial\overline{\theta}^{\lambda}}{\partial t}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v\theta}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial p}\left(\overline{\omega\theta}^{\lambda}\right)=\overline{\dot{\theta}}^{\lambda}\,, \tag{7.91}$$

$$\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v}^{\lambda}\cos\varphi\right)+\frac{\partial\overline{\omega}^{\lambda}}{\partial p}=0\,. \tag{7.92}$$

Subtracting (7.91) and (7.92) from (7.89) and (7.90), respectively, we find that

$$\begin{aligned}
&\frac{\partial\theta^*}{\partial t}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda}(u\theta)+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left\{\left(v\theta-\overline{v\theta}^{\lambda}\right)\cos\varphi\right\}+\frac{\partial}{\partial p}\left(\omega\theta-\overline{\omega\theta}^{\lambda}\right)\\
=&\frac{\partial\theta^*}{\partial t}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda}\left(u^*\overline{\theta}^{\lambda}+\overline{u}^{\lambda}\theta^*+u^*\theta^*\right)+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\left(v^*\overline{\theta}^{\lambda}+\overline{v}^{\lambda}\theta^*+v^*\theta^*-\overline{v^*\theta^*}^{\lambda}\right)\cos\varphi\right]\\
&+\frac{\partial}{\partial p}\left(\omega^*\overline{\theta}^{\lambda}+\overline{\omega}^{\lambda}\theta^*+\omega^*\theta^*-\overline{\omega^*\theta^*}^{\lambda}\right)\\
=&\frac{\partial\theta^*}{\partial t}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda}\left(\overline{u}^{\lambda}\theta^*\right)+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\left(\overline{v}^{\lambda}\theta^*\right)\cos\varphi\right]+\frac{\partial}{\partial p}\left(\overline{\omega}^{\lambda}\theta^*\right)\\
&+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda}(u^*\theta^*)+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}(v^*\theta^*\cos\varphi)+\frac{\partial}{\partial p}(\omega^*\theta^*)\\
&+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda}\left(u^*\overline{\theta}^{\lambda}\right)+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(v^*\overline{\theta}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial p}\left(\omega^*\overline{\theta}^{\lambda}\right)\\
&-\frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda}\left(\overline{v^*\theta^*}^{\lambda}\cos\varphi\right)-\frac{\partial}{\partial p}\left(\overline{\omega^*\theta^*}^{\lambda}\right)\\
=&\dot{\theta}^*\,,
\end{aligned} \tag{7.93}$$

$$\frac{1}{a\cos\varphi}\frac{\partial u^*}{\partial\lambda}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(v^*\cos\varphi\right)+\frac{\partial\omega^*}{\partial p}=0\,. \tag{7.94}$$

To obtain the first equality of (7.93), we have used

$$\begin{aligned} v\theta &= \left(\overline{v}^\lambda+v^*\right)\left(\overline{\theta}^\lambda+\theta^*\right) \\ &= \overline{v}^\lambda\overline{\theta}^\lambda+v^*\overline{\theta}^\lambda+\overline{v}^\lambda\theta^*+v^*\theta^*\,, \end{aligned} \tag{7.95}$$

$$\overline{v\theta}^\lambda=\overline{v}^\lambda\overline{\theta}^\lambda+\overline{v^*\theta^*}^\lambda\,, \tag{7.96}$$

$$v\theta-\overline{v\theta}^\lambda=v^*\overline{\theta}^\lambda+\overline{v}^\lambda\theta^*+v^*\theta^*-\overline{v^*\theta^*}^\lambda\,, \tag{7.97}$$

and so on. We can use (7.92) and (7.94) to rewrite (7.93) as follows:

$$\begin{aligned} &\left(\frac{\partial}{\partial t}+\frac{\overline{u}^\lambda}{a\cos\varphi}\frac{\partial}{\partial\lambda}+\frac{\overline{v}^\lambda}{a}\frac{\partial}{\partial\varphi}+\overline{\omega}^\lambda\frac{\partial}{\partial p}\right)\theta^* \\ &+\left(\frac{u^*}{a\cos\varphi}\frac{\partial}{\partial\lambda}+\frac{v^*}{a}\frac{\partial}{\partial\varphi}+\omega^*\frac{\partial}{\partial p}\right)\theta^*+\frac{v^*}{a}\frac{\partial\overline{\theta}^\lambda}{\partial\varphi}+\omega^*\frac{\partial\overline{\theta}^\lambda}{\partial p} \\ &=\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^*\theta^*}^\lambda\cos\varphi\right)+\frac{\partial}{\partial p}\left(\overline{\omega^*\theta^*}^\lambda\right)+\dot{\theta}^*\,. \end{aligned} \tag{7.98}$$

Multiplying (7.98) by θ^*, and using (7.94) again, we obtain:

$$\left(\frac{\partial}{\partial t}+\frac{\overline{u}^{\lambda}}{a\cos\varphi}\frac{\partial}{\partial\lambda}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial}{\partial\varphi}+\overline{\omega}^{\lambda}\frac{\partial}{\partial p}\right)\left(\frac{1}{2}\theta^{*2}\right)$$
$$+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda}\left[u^{*}\left(\frac{1}{2}\theta^{*2}\right)\right]+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[v^{*}\left(\frac{1}{2}\theta^{*2}\right)\cos\varphi\right]+\frac{\partial}{\partial p}\left[\omega^{*}\left(\frac{1}{2}\theta^{*2}\right)\right]$$
$$=\theta^{*}\left[\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^{*}\theta^{*}}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial p}\left(\overline{\omega^{*}\theta^{*}}^{\lambda}\right)\right]-\frac{v^{*}\theta^{*}}{a}\frac{\partial\overline{\theta}^{\lambda}}{\partial\varphi}-\omega^{*}\theta^{*}\frac{\partial\overline{\theta}^{\lambda}}{\partial p}+\dot{\theta}^{*}\theta^{*}\,. \tag{7.99}$$

Zonally averaging (7.99) gives

$$\left(\frac{\partial}{\partial t}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial}{\partial\varphi}+\overline{\omega}^{\lambda}\frac{\partial}{\partial p}\right)\left(\frac{1}{2}\overline{\theta^{*2}}^{\lambda}\right)$$
$$+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\overline{v^{*}\left(\frac{1}{2}\theta^{*2}\right)}^{\lambda}\cos\varphi\right]+\frac{\partial}{\partial p}\left[\overline{\omega^{*}\left(\frac{1}{2}\theta^{*2}\right)}^{\lambda}\right] \tag{7.100}$$
$$=-\frac{\overline{v^{*}\theta^{*}}^{\lambda}}{a}\frac{\partial\overline{\theta}^{\lambda}}{\partial\varphi}-\overline{\omega^{*}\theta^{*}}^{\lambda}\frac{\partial\overline{\theta}^{\lambda}}{\partial p}+\overline{\dot{\theta}^{*}\theta^{*}}^{\lambda}\,.$$

Finally, we can use (7.92) to rewrite (7.99) in flux form:

$$\boxed{\begin{aligned}&\frac{\partial}{\partial t}\left(\frac{1}{2}\overline{\theta^{*2}}^{\lambda}\right)+\underbrace{\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v}^{\lambda}\frac{1}{2}\overline{\theta^{*2}}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial p}\left(\overline{\omega}^{\lambda}\frac{1}{2}\overline{\theta^{*2}}^{\lambda}\right)}_{\text{advection by the mean flow}}\\&\underbrace{+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^{*}\frac{1}{2}\theta^{*2}}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial p}\left(\overline{\omega^{*}\frac{1}{2}\theta^{*2}}^{\lambda}\right)}_{\text{eddy transport}}\\&=\underbrace{-\frac{\overline{v^{*}\theta^{*}}^{\lambda}}{a}\frac{\partial\overline{\theta}^{\lambda}}{\partial\varphi}-\overline{\omega^{*}\theta^{*}}^{\lambda}\frac{\partial\overline{\theta}^{\lambda}}{\partial p}}_{\text{gradient production}}+\overline{\dot{\theta}^{*}\theta^{*}}^{\lambda}\,.\end{aligned}} \tag{7.101}$$

According to (7.101), $\overline{\theta^{*2}}^{\lambda}$ can change due to advection by the mean meridional circulation, or due to transport by the eddies themselves, or due to "gradient production."

Eq. (7.101) governs the eddy variance at a particular latitude. The meridional and vertical gradients of $\overline{\theta}^{\lambda}$ also contribute to the *global* variance of θ. To show this, we start by using (7.92) to rewrite (7.91) as

$$\left(\frac{\partial}{\partial t}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial}{\partial\varphi}+\overline{\omega}^{\lambda}\frac{\partial}{\partial p}\right)\overline{\theta}^{\lambda}=-\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^{*}\theta^{*}}^{\lambda}\cos\varphi\right)-\frac{\partial}{\partial p}\left(\overline{\omega^{*}\theta^{*}}^{\lambda}\right)+\overline{\dot{\theta}}^{\lambda}\ . \tag{7.102}$$

Multiplication by $\overline{\theta}^{\lambda}$ gives

$$\left(\frac{\partial}{\partial t}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial}{\partial\varphi}+\overline{\omega}^{\lambda}\frac{\partial}{\partial p}\right)\frac{1}{2}\left(\overline{\theta}^{\lambda}\right)^{2}=-\frac{\overline{\theta}^{\lambda}}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^{*}\theta^{*}}^{\lambda}\cos\varphi\right)-\overline{\theta}^{\lambda}\frac{\partial}{\partial p}\left(\overline{\omega^{*}\theta^{*}}^{\lambda}\right)+\overline{\theta}^{\lambda}\overline{\dot{\theta}}^{\lambda}\ . \tag{7.103}$$

This can be rearranged to

$$\begin{aligned}&\left(\frac{\partial}{\partial t}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial}{\partial\varphi}+\overline{\omega}^{\lambda}\frac{\partial}{\partial p}\right)\frac{1}{2}\left(\overline{\theta}^{\lambda}\right)^{2}\\&=-\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\theta}^{\lambda}\overline{v^{*}\theta^{*}}^{\lambda}\cos\varphi\right)-\frac{\partial}{\partial p}\left(\overline{\theta}^{\lambda}\overline{\omega^{*}\theta^{*}}^{\lambda}\right)+\frac{\overline{v^{*}\theta^{*}}^{\lambda}}{a}\frac{\partial\overline{\theta}^{\lambda}}{\partial\varphi}+\overline{\omega^{*}\theta^{*}}^{\lambda}\frac{\partial\overline{\theta}^{\lambda}}{\partial p}+\overline{\theta}^{\lambda}\overline{\dot{\theta}}^{\lambda}\ .\end{aligned} \tag{7.104}$$

Converting back to flux form, we find that

$$\boxed{\begin{aligned}&\frac{\partial}{\partial t}\left[\frac{1}{2}\left(\overline{\theta}^{\lambda}\right)^{2}\right]+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\overline{v}^{\lambda}\frac{1}{2}\left(\overline{\theta}^{\lambda}\right)^{2}\cos\varphi\right]+\frac{\partial}{\partial p}\left[\overline{\omega}^{\lambda}\frac{1}{2}\left(\overline{\theta}^{\lambda}\right)^{2}\right]\\&=-\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\theta}^{\lambda}\overline{v^{*}\theta^{*}}^{\lambda}\cos\varphi\right)-\frac{\partial}{\partial p}\left(\overline{\theta}^{\lambda}\overline{\omega^{*}\theta^{*}}^{\lambda}\right)+\frac{\overline{v^{*}\theta^{*}}^{\lambda}}{a}\frac{\partial\overline{\theta}^{\lambda}}{\partial\varphi}+\overline{\omega^{*}\theta^{*}}^{\lambda}\frac{\partial\overline{\theta}^{\lambda}}{\partial p}+\overline{\theta}^{\lambda}\overline{\dot{\theta}}^{\lambda}\ .\end{aligned}} \tag{7.105}$$

When we add (7.105) and (7.101), the gradient production terms cancel. This shows that those terms represent a "conversion" between $\frac{1}{2}\left(\overline{\theta}^{\lambda}\right)^{2}$ and $\frac{1}{2}\overline{\theta^{*2}}^{\lambda}$. Note that $\overline{\theta^{2}}=$

$\left(\overline{\theta}^{\lambda}\right)^2 + \overline{\theta^{*2}}^{\lambda}$, i.e., the zonal average of the square is the sum of the square of the zonal average and the square of the departure from the zonal average. Similarly, $\overline{\theta\dot{\theta}}^{\lambda} = \overline{\theta}^{\lambda}\overline{\dot{\theta}}^{\lambda} + \overline{\dot{\theta}_*\theta_*}^{\lambda}$. We find that the total variance, $\frac{1}{2}\overline{\theta^2}^{\lambda}$, satisfies:

$$\begin{aligned}\frac{\partial}{\partial t}\left(\frac{1}{2}\overline{\theta^2}^{\lambda}\right) + \frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\left(\overline{v}^{\lambda}\frac{1}{2}\overline{\theta^2}^{\lambda} + \overline{v^*\frac{1}{2}\theta^{*2}}^{\lambda} + \overline{\theta}^{\lambda}\overline{v^*\theta^*}^{\lambda}\right)\cos\varphi\right] \\ + \frac{\partial}{\partial p}\left(\overline{\omega}^{\lambda}\frac{1}{2}\overline{\theta^2}^{\lambda} + \overline{\omega^*\frac{1}{2}\theta^{*2}}^{\lambda} + \overline{\theta}^{\lambda}\overline{\omega^*\theta^*}^{\lambda}\right) = \overline{\theta\dot{\theta}}^{\lambda}\,.\end{aligned} \tag{7.106}$$

The gradient production terms have no effect on the total variance. The only source term is due to heating. Finally, meridional and vertical integration of (7.106) shows that the variance budget for the entire atmosphere is expressed by

$$\frac{d}{dt}\int_M \left(\frac{1}{2}\theta^2\right) dM = \int_M \theta\dot{\theta}\, dM\,. \tag{7.107}$$

This shows that the total variance of θ changes only due to the covariance between θ and $\dot{\theta}$.

7.9 Generation of available potential energy, and its conversion into kinetic energy

Earlier we derived (7.43), Lorenz's approximate expression for the available potential energy of a statically stable atmosphere in terms of the temperature variance on isobaric surfaces, which is repeated here for convenience:

$$A = \frac{a^2}{2}\int_{-\pi/2}^{\pi/2}\int_0^{2\pi}\int_0^{p_S} \frac{\overline{T}^p}{(\Gamma_d - \Gamma)}\left(\frac{T'}{\overline{T}^p}\right)^2 \cos\varphi\, dp\, d\lambda\, d\varphi\,. \tag{7.108}$$

Recall that in this equation an $\overline{(\quad)}^p$ represents a global mean on a pressure surface, and a prime denotes a departure from that isobaric average. The APE is (approximately) an integral of the variance of the temperature about its global mean on isobaric surfaces. In the last section, we derived an equation for the time rate of change of the potential temperature

variance. We now work out an approximate equation for the time rate of change of A due to generation, dissipation, and conversion to or from kinetic energy.

We can show that, for any two quantities α and β,

$$\overline{(\alpha\beta)}^p = \overline{\left(\overline{\alpha}^\lambda \overline{\beta}^\lambda + \overline{\alpha^* \beta^*}^\lambda\right)}^p \ . \tag{7.109}$$

and

$$\begin{aligned} \overline{\alpha'\beta'}^p =& \overline{(\alpha\beta)}^p - \overline{\alpha}^p \overline{\beta}^p \\ =& \overline{\left(\overline{\alpha}^\lambda \overline{\beta}^\lambda + \overline{\alpha^* \beta^*}^\lambda\right)}^p - \overline{\alpha}^p \overline{\beta}^p \ . \end{aligned} \tag{7.110}$$

As a special case of (7.110), the variance of an arbitrary quantity about its global mean is given by

$$\overline{\alpha'^2}^p = \overline{\left[\left(\overline{\alpha}^\lambda\right)^2 + \overline{(\alpha^*)^2}^\lambda\right]}^p - (\overline{\alpha}^p)^2 \ . \tag{7.111}$$

Using this notation, we can approximate the expression for the available potential energy given by (7.108) as

$$\begin{aligned} A \cong& 2\pi a^2 \int_0^{p_S} \frac{\overline{T}^p}{\left(\Gamma_d - \overline{\Gamma}^p\right)} \left(\frac{\overline{T'^2}^p}{\overline{T}^{p2}}\right) dp \\ =& 2\pi a^2 \int_0^{p_S} \frac{\overline{T}^p}{\left(\Gamma_d - \overline{\Gamma}^p\right)} \left(\frac{\overline{\theta'^2}^p}{\overline{\theta}^{p2}}\right) dp \ . \end{aligned} \tag{7.112}$$

We now work out an equation for dA/dt, based on (7.112). The global means of (7.89) and (7.90) are

$$\frac{\partial \overline{\theta}^p}{\partial t} + \frac{\partial \overline{(\omega\theta)}^p}{\partial p} = \overline{\dot{\theta}}^p \ , \tag{7.113}$$

and

$$\frac{\partial \overline{\omega}^p}{\partial p} = 0 \,, \tag{7.114}$$

respectively. Since $\omega = 0$ at $p = 0$, it follows from (7.114) that

$$\overline{\omega}^p = 0 \text{ for all } p \,. \tag{7.115}$$

This allows us to write

$$\begin{aligned} \overline{(\omega\theta)^p} =& \overline{\omega}^p \overline{\theta}^p + \overline{\omega'\theta'}^p \\ =& \overline{\omega'\theta'}^p \,. \end{aligned} \tag{7.116}$$

Area-weighted integration of (7.106) over all latitudes, at a given pressure-level, leads to

$$\begin{aligned} & \frac{\partial}{\partial t} \frac{1}{2} \overline{\left[\left(\overline{\theta}^\lambda \right)^2 + \overline{\theta^{*2}}^\lambda \right]}^p \\ + \frac{\partial}{\partial p} & \overline{\left\{ \overline{\omega}^\lambda \frac{1}{2} \left[\left(\overline{\theta}^\lambda \right)^2 + \overline{\theta^{*2}}^\lambda \right] + \overline{\omega^* \frac{1}{2} \theta^{*2}}^\lambda + \overline{\theta}^\lambda \overline{\omega^* \theta^*}^\lambda \right\}}^p \\ & = \overline{\left(\overline{\theta}^\lambda \dot{\overline{\theta}}^\lambda + \overline{\dot{\theta}^* \theta^*}^\lambda \right)}^p \,. \end{aligned} \tag{7.117}$$

Here we ignore, as usual, the complications arising from the fact that some pressure surfaces intersect the Earth's surface. From (7.111), we see that

$$\frac{\partial \overline{T'^2}^p}{\partial t} = \frac{\partial}{\partial t} \overline{\left[\left(\overline{\theta}^\lambda \right)^2 + \overline{\theta^{*2}}^\lambda \right]}^p - 2\overline{\theta}^p \frac{\partial \overline{\theta}^p}{\partial t} \,. \tag{7.118}$$

Substituting into (7.118) from (7.113) and (7.117), and using (7.110) and (7.116), we find that

$$\begin{aligned}
\frac{\partial}{\partial t}\left(\frac{\overline{T'^2}^p}{2}\right) &= -\frac{\partial}{\partial p}\overline{\left\{\overline{\omega}^\lambda \frac{1}{2}\left[\left(\overline{\theta}^\lambda\right)^2 + \overline{\theta^{*2}}^\lambda\right] + \overline{\omega^*\frac{1}{2}\theta^{*2}}^\lambda + \overline{\theta}^\lambda\overline{\omega^*\theta^*}^\lambda\right\}}^p + \overline{\theta}^p\frac{\partial\overline{(\omega\theta)}^p}{\partial p} \\
&\quad + \left[\overline{\left(\overline{\theta}^\lambda\overline{\dot{\theta}}^\lambda + \overline{\dot{\theta}^*\theta^*}^\lambda\right)}^p - \overline{\theta}^p\overline{\dot{\theta}}^p\right] \\
&= -\frac{\partial}{\partial p}\overline{\left\{\overline{\omega}^\lambda \frac{1}{2}\left[\left(\overline{\theta}^\lambda\right)^2 + \overline{\theta^{*2}}^\lambda\right] + \overline{\omega^*\frac{1}{2}\theta^{*2}}^\lambda + \overline{\theta}^\lambda\overline{\omega^*\theta^*}^\lambda - \overline{\theta}^p(\omega\theta)\right\}}^p \\
&\quad - \overline{\omega\theta}^p\frac{\partial\overline{\theta}^p}{\partial p} + \overline{\theta'\dot{\theta}'}^p \\
&= -\frac{\partial}{\partial p}\overline{\left\{\overline{\omega}^\lambda \frac{1}{2}\left[\left(\overline{\theta}^\lambda\right)^2 + \overline{\theta^{*2}}^\lambda\right] + \overline{\omega^*\frac{1}{2}\theta^{*2}}^\lambda + \left[\left(\overline{\theta}^\lambda\overline{\omega^*\theta^*}^\lambda\right)^p - \overline{\theta}^p(\omega\theta)^p\right]\right\}}^p \\
&\quad - \overline{\omega'\theta'}^p\frac{\partial\overline{\theta}^p}{\partial p} + \overline{\theta'\dot{\theta}'}^p\,.
\end{aligned} \tag{7.119}$$

It is not hard to recognize the various things that are going on in this equation. Vertical transport is clearly visible, as are gradient production (in the vertical) and "Heat where it's hot, cool where it's cold." Use of (7.41) in the approximate form

$$-\frac{\partial\overline{\theta}^p}{\partial p} \cong \frac{\kappa\overline{\theta}^p}{p}\left(\frac{\Gamma_d - \overline{\Gamma}^p}{\Gamma_d}\right) \tag{7.120}$$

allows us to rewrite (7.119) as

$$\begin{aligned}
\frac{\partial}{\partial t}\left(\frac{\overline{T'^2}^p}{2}\right) &= -\frac{\partial}{\partial p}\overline{\left\{\overline{\omega}^\lambda \frac{1}{2}\left[\left(\overline{\theta}^\lambda\right)^2 + \overline{\theta^{*2}}^\lambda\right] + \overline{\omega^*\frac{1}{2}\theta^{*2}}^\lambda + \left[\left(\overline{\theta}^\lambda\overline{\omega^*\theta^*}^\lambda\right)^p - \overline{\theta}^p\,\overline{\omega\theta}^p\right]\right\}}^p \\
&\quad + \overline{\omega'\theta'}^p\frac{\kappa\overline{\theta}^p}{p}\left(\frac{\Gamma_d - \overline{\Gamma}^p}{\Gamma_d}\right) + \overline{\theta'\dot{\theta}'}^p\,.
\end{aligned} \tag{7.121}$$

Vertical integration of (7.121) gives

$$\frac{d}{dt}\left(\int_0^{p_S}\frac{\overline{T'^2}^p}{2}\,dp\right)=\int_0^{p_S}\overline{\omega'\theta'}^p\frac{\kappa\overline{\theta}^p}{p}\left(\frac{\Gamma_d-\overline{\Gamma}^p}{\Gamma_d}\right)dp+\int_0^{p_S}\overline{\theta'\dot{\theta}'}^p dp\,. \tag{7.122}$$

Here we write a total time derivative, d/dt, because we have now integrated over all three space variables. With the use of (7.112), Eq. (7.122) becomes

$$\frac{dA}{dt}=4\pi a^2\int_0^{p_S}\left(\frac{\overline{\omega'\theta'}^p}{\overline{\theta}^p}\right)\frac{\kappa\overline{T}^p}{p\Gamma_d}\,dp+4\pi a^2\int_0^{p_S}\frac{\overline{\theta'\dot{\theta}'}^p\overline{T}^p}{\overline{\theta}^{p2}\left(\Gamma_d-\overline{\Gamma}^p\right)}\,dp\,. \tag{7.123}$$

Finally, we use

$$\frac{\theta'}{\overline{\theta}^p}=\frac{\alpha'}{\alpha^p}=\frac{T'}{\overline{T}^p}\,, \tag{7.124}$$

so that

$$\begin{aligned}\frac{dA}{dt}=&4\pi a^2\int_0^{p_S}\left(\frac{\overline{\omega'\alpha'}^p}{\overline{\alpha}^p}\right)\frac{\kappa\overline{T}^p}{p\Gamma_d}\,dp+4\pi a^2\int_0^{p_S}\frac{\overline{\theta'\dot{\theta}'}^p\overline{T}^p}{\left(\overline{\theta}^p\right)^2\left(\Gamma_d-\overline{\Gamma}^p\right)}\,dp\\ =&\frac{4\pi a^2}{g}\int_0^{p_S}\overline{\omega'\alpha'}^p\,dp+4\pi a^2\int_0^{p_S}\frac{\overline{\theta'\dot{\theta}'}^p\overline{T}^p}{\left(\overline{\theta}^p\right)^2\left(\Gamma_d-\overline{\Gamma}^p\right)}\,dp\,.\end{aligned} \tag{7.125}$$

To obtain the second line of (7.125), we have used $\Gamma_d=g/c_p$.

Recall that C represents conversion between kinetic and non-kinetic energy. By inspection of (7.125) we can identify the rate of conversion of KE into APE as

$$\boxed{C=4\pi a^2\int_0^{p_S}\left(\frac{\overline{\omega'\alpha'}^p}{\overline{\alpha}^p}\right)\frac{\kappa\overline{T}^p}{p\Gamma_d}\,dp\,.} \tag{7.126}$$

Not surprisingly, the rate of conversion involves the product $\omega\alpha$, which we first discussed in Chapter 4. Similarly, the rate of generation or destruction of APE by heating is

$$\boxed{G \cong 4\pi a^2 \int_0^{p_S} \frac{\left(\overline{\frac{\theta'\dot{\theta}'}{\overline{\theta}^p}}\right)\left(\frac{\overline{T}^p}{\overline{\theta}^p}\right)}{\left(\Gamma_d - \overline{\Gamma}^p\right)}\, dp\,.} \tag{7.127}$$

Note that G will be positive if we "heat where it's hot and cool where it's cold." This should also sound familiar. A heating field that generates APE destroys entropy.

We can summarize (7.125) as

$$\boxed{\frac{dA}{dt} = C + G\,.} \tag{7.128}$$

Similarly, we can show that

$$\boxed{\frac{dK}{dt} = -C - D\,.} \tag{7.129}$$

where K is the globally integrated kinetic energy (e.g., in Joules), and D is the globally integrated dissipation rate. We can also conclude that C is negative from the fact that $D > 0$. There must be a net conversion of APE into K, in order to replenish the K that is destroyed by dissipation. Because this conversion depletes the APE, the rate of generation of APE must be positive. In other words, energy must flow as shown in Fig. 7.5.

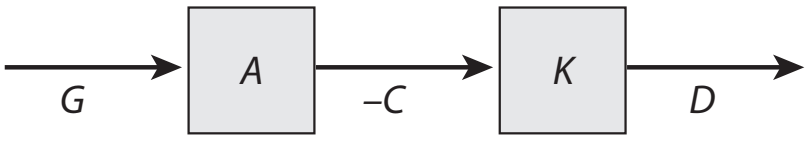


Figure 7.5: Sketch illustrating the flow of energy through the global atmospheric circulation. Generation produces APE, which is converted to KE, which is dissipated.

7.10 The governing equations for the eddy kinetic energy, zonal kinetic energy, and total kinetic energy

We now present a discussion of the eddy kinetic energy, zonal kinetic energy, and total kinetic energy equations. The derivations of these equations follow methods similar to those used to derive the conservation equation for the potential energy variance. The details are given in Appendix G.

We define the eddy kinetic energy per unit mass by

$$KE \equiv \frac{1}{2}\left[\overline{(u*)^2}^{\lambda} + \overline{(v*)^2}^{\lambda}\right] . \tag{7.130}$$

We can show that it satisfies the following equation:

$$\begin{aligned}
&\frac{\partial}{\partial t}KE \\
&+ \frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left\{\overline{v}^{\lambda}KE\cos\varphi + \frac{1}{2}\left[\overline{v^*\left(u^{*2}+v^{*2}\right)}^{\lambda}\right] + \overline{v^*\phi^*}^{\lambda}\cos\varphi\right\} \\
&+ \frac{\partial}{\partial p}\left\{\overline{\omega}^{\lambda}KE + \frac{1}{2}\left[\overline{\omega^*\left(u^{*2}+v^{*2}\right)}^{\lambda}\right] + \overline{\omega^*\phi^*}^{\lambda}\right\} \\
&- \frac{\overline{u^*v^*}^{\lambda}}{a}\frac{\partial\overline{u}^{\lambda}}{\partial\varphi} - \frac{\overline{v^*v^*}^{\lambda}}{a}\frac{\partial\overline{v}^{\lambda}}{\partial\varphi} - \overline{u^*\omega^*}^{\lambda}\frac{\partial\overline{u}^{\lambda}}{\partial p} - \overline{\omega^*v^*}^{\lambda}\frac{\partial\overline{v}^{\lambda}}{\partial p} \\
&= \left(-\overline{u}^{\lambda}\overline{u^*v^*}^{\lambda} + \overline{u^*u^*}^{\lambda}\overline{v}^{\lambda}\right)\frac{\tan\varphi}{a} \\
&- \overline{\omega^*\alpha^*} \\
&+ \overline{u^*g\frac{\partial F_u{}^*}{\partial p}}^{\lambda} + \overline{v^*g\frac{\partial F_v{}^*}{\partial p}}^{\lambda} .
\end{aligned} \tag{7.131}$$

Gradient production terms appear on the fourth line of (7.131). They represent the conversion between the kinetic energy of the mean flow and that of the eddies. This conversion is in the sense of increasing the eddy kinetic energy when the eddy momentum flux is "down the gradient," i.e. when it is from higher mean momentum to lower mean momentum. The term represents eddy kinetic energy generation from eddy available potential energy, while the terms involving represent the effects of "pressure work."

The appearance of the metric terms in (7.131) may be somewhat surprising. They arise because we have defined "eddies" in terms of departures from the zonal mean, so that

a particular latitude-longitude coordinate system is implicit in the very definition of KE. Obviously there cannot be any metric terms in the equation for the total kinetic energy per unit mass, which we denote by K.

Define the zonal kinetic energy by

$$KZ \equiv \frac{1}{2}\left[\left(\overline{u}^{\lambda}\right)^2 + \left(\overline{v}^{\lambda}\right)^2\right] . \tag{7.132}$$

and note that

$$\overline{K}^{\lambda} = KZ + KE . \tag{7.133}$$

All three quantities in Eq. (7.133) are independent of longitude. The zonal kinetic energy satisfies

$$\begin{aligned}
&\frac{\partial}{\partial t}KZ \\
&+ \frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\left(\overline{v}^{\lambda}KZ + \overline{u}^{\lambda}\overline{u^*v^*}^{\lambda} + \overline{v}^{\lambda}\overline{v^*v^*}^{\lambda} + \overline{v}^{\lambda}\overline{\phi}^{\lambda}\right)\cos\varphi\right] \\
&+ \frac{\partial}{\partial p}\left(\overline{\omega}^{\lambda}KZ + \overline{u}^{\lambda}\overline{\omega^*u^*}^{\lambda} + \overline{v}^{\lambda}\overline{\omega^*v^*}^{\lambda} + \overline{\omega}^{\lambda}\overline{\phi}^{\lambda}\right) \\
&= \frac{\overline{u^*v^*}^{\lambda}}{a}\frac{\partial\overline{u}^{\lambda}}{\partial\varphi} + \frac{\overline{v^*v^*}^{\lambda}}{a}\frac{\partial\overline{v}^{\lambda}}{\partial\varphi} + \overline{\omega^*u^*}^{\lambda}\frac{\partial\overline{u}^{\lambda}}{\partial p} + \overline{\omega^*v^*}^{\lambda}\frac{\partial\overline{v}^{\lambda}}{\partial p} \\
&+ \left(\overline{u}^{\lambda}\overline{u^*v^*}^{\lambda} - \overline{v}^{\lambda}\overline{u^*u^*}^{\lambda}\right)\frac{\tan\varphi}{a} \\
&- \overline{\omega}^{\lambda}\overline{\alpha}^{\lambda} \\
&+ \overline{u}^{\lambda}g\frac{\partial\overline{F_u}^{\lambda}}{\partial p} + \overline{v}^{\lambda}g\frac{\partial\overline{F_v}^{\lambda}}{\partial p} .
\end{aligned} \tag{7.134}$$

The "gradient production" terms of (7.131) appear with the opposite sign in (7.134). They represent conversions between KE and KZ. Adding the equations for KE and KZ gives the equation for the zonally averaged total kinetic energy:

$$
\begin{aligned}
&\frac{\partial \overline{K}^{\lambda}}{\partial t} \\
&+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left\{\left(\overline{v}^{\lambda}\overline{K}^{\lambda}+\frac{1}{2}\left[\overline{v^{*}\left(u^{*2}+v^{*2}\right)}^{\lambda}\right]+\overline{u}^{\lambda}\overline{u^{*}v^{*}}^{\lambda}+\overline{v}^{\lambda}\overline{v^{*}v^{*}}^{\lambda}+\overline{v}^{\lambda}\overline{\phi}^{\lambda}+\overline{v^{*}\phi^{*}}^{\lambda}\right)\cos\varphi\right\} \\
&+\frac{\partial}{\partial p}\left(\overline{\omega}^{\lambda}\overline{K}^{\lambda}+\frac{1}{2}\left[\overline{\omega^{*}\left(u^{*2}+v^{*2}\right)}^{\lambda}\right]+\overline{u}^{\lambda}\overline{\omega^{*}u^{*}}^{\lambda}+\overline{v}^{\lambda}\overline{\omega^{*}v^{*}}^{\lambda}+\overline{\phi}^{\lambda}\overline{\omega}^{\lambda}+\overline{\omega^{*}\phi^{*}}^{\lambda}\right) \\
&=-\overline{\omega}^{\lambda}\overline{\alpha}^{\lambda}-\overline{\omega^{*}\alpha^{*}} \\
&+\overline{u^{*}g\frac{\partial F_{u}{}^{*}}{\partial p}}^{\lambda}+\overline{v^{*}g\frac{\partial F_{v}{}^{*}}{\partial p}}^{\lambda}+\overline{u}^{\lambda}g\frac{\partial \overline{F_{u}}^{\lambda}}{\partial p}+\overline{v}^{\lambda}g\frac{\partial \overline{F_{v}}^{\lambda}}{\partial p}\,.
\end{aligned}
\tag{7.135}
$$

As expected, the metric terms have cancelled; they cannot affect the zonally averaged total kinetic energy.

7.11 Observations of the energy cycle

Arpe et al. (1986) discussed the observed energy cycle of the atmosphere, based on ECMWF analyses. They wrote the following equations for the energy cycle:

$$
\frac{d}{dt}KZ=-\sum_{m}CK\left(m\right)+CZ-DZ\,, \tag{7.136}
$$

$$
\frac{d}{dt}AZ=-\sum_{m}CA\left(m\right)-CZ+GZ\,. \tag{7.137}
$$

A direct mean meridional circulation, such as the Hadley Circulation, converts AZ into KZ, and so is associated with positive values of CZ. The indirect Ferrel Cells contribute negatively to CZ. Arpe et al. (1986) wrote the equations for the eddy kinetic energy and eddy available potential energy as

$$
\frac{d}{dt}KE\left(m\right)=CK\left(m\right)+LK\left(m\right)+CE\left(m\right)-DE\left(m\right)\,, \tag{7.138}
$$

and

$$\frac{d}{dt}AE(m) = CA(m) + LA(m) - \sum_m CE(m) + \sum_m GE(m) \ . \tag{7.139}$$

Here m is the zonal wave number. The eddy kinetic energy and eddy available potential energy are defined as functions of m, and the contributions for the individual waves have been worked out in the analysis of the data. The terms "$LK(m)$" and "$LA(m)$" represent wave-wave interactions, due to nonlinear processes. For example, if we have a "kinetic energy cascade" from lower wave numbers to higher wave numbers, then $LK(m)$ will represent a flow of energy from larger scales to smaller scales.

Summing (7.138) and (7.139) leads to

$$\frac{d}{dt}KE = \sum_m CK(m) + \sum_m LK(m) + \sum_m CE(m) - \sum_m DE(m) \ , \tag{7.140}$$

$$\frac{d}{dt}AE = \sum_m CA(m) + \sum_m A(m) - \sum_m CE(m) + \sum_m GE(m) \ . \tag{7.141}$$

Adding (7.136) and (7.140) leads to

$$\frac{d}{dt}KZ + KE = CZ + \sum_m LK(m) + \sum_m CE(m) - \sum_m DE(m) - DZ \ . \tag{7.142}$$

Similarly, adding (7.137) and (7.141) gives

$$\frac{d}{dt}AZ + ZE = \sum_m A(m) - \sum_m CE(m) + \sum_m GE(m) + GZ \ . \tag{7.143}$$

Baroclinic instability of the zonal flow is represented by the combination of positive CA and positive CE; the first of these represents the conversion of AZ to AE, and the second represents the conversion of AE to KE; the net effect is thus conversion of AZ to KE. When AZ is converted to AE, an eddy temperature variance is created by conversion from the global temperature variance associated with the meridional gradient of the zonally averaged temperature. This is the gradient production process analyzed earlier in this chapter. When AE is converted to KE, warm air rises and cool air sinks, potential temperature is transported upward, and the center of gravity of the atmosphere is lowered.

Mathematically, CK is represented by the gradient production term in the eddy kinetic energy equation. There is a tendency for eddies to pump momentum into the jet, increasing its strength. This is represented by negative values of CK.

It is important to understand the meaning of each term of (7.136) through (7.139), as outlined above. It is also important, however, to notice that certain terms are *not* present, which means that certain processes do not exist. For example, there is no process that directly converts AZ into $KE(m)$. Such conversion can occur only indirectly, in two steps, i.e., first $AZ \rightarrow AE(m)$, and then $AE(m) \rightarrow KE(m)$.

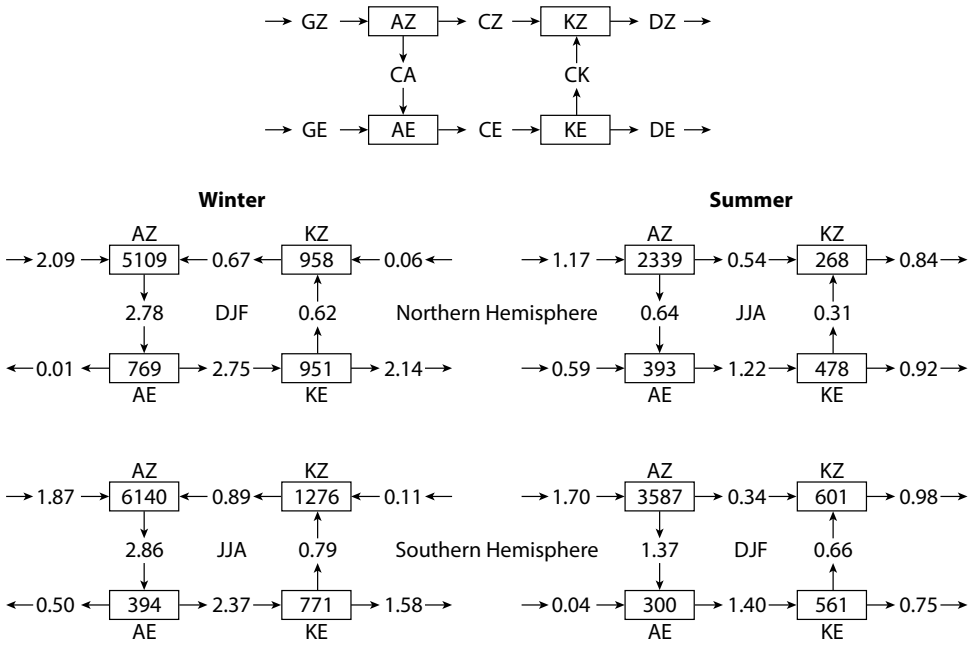


Figure 7.6: Energy cycle during winter (left) and summer (right) for both hemispheres. Integrals between 1000 and 50 hPa are presented. Data are calculated from 00 GMT initialized analyses. Energy amounts are given in kJ m^{-2} and conversion rates in W m^{-2}. From Arpe et al. (1986).

Fig. 7.6 shows the observed energy cycle for (northern) winter and summer, and for the two hemispheres separately. The figure is arranged so that the "summer hemispheres" are on the right (for both seasons) and the "winter hemispheres" are on the left. The numbers in the boxes represent amounts of energy, and the numbers on the arrows between boxes represent energy conversions, or processes that generate or destroy energy. Note that Arpe et al. (1986) defined the APE for the Northern and Southern Hemispheres separately; as discussed earlier, this is not strictly correct.

The arrows leading into AZ and AE from the left represent generation (an arrow leading out simply represents negative generation), and the arrows leading out of KE and KZ to the right represent dissipation. These arrows can also represent interactions between

the hemispheres. For example we see an arrow leading into KZ from the right for the Northern Hemisphere in winter, apparently indicating a physically impossible negative dissipation rate, but actually representing a gain of KZ in the Northern Hemisphere via energy exchanges with the Southern Hemisphere.

Fig. 7.6 shows that, especially in the Northern Hemisphere, the energy flows of the atmosphere are much more vigorous in winter than in summer. Note that AZ is several times larger than AE or KZ or KE. For the Northern Hemisphere in winter, AZ is strongly generated; this energy is converted to AE, leading to a production of KE via baroclinic instability; the eddies act to increase KZ by transporting angular momentum into the jet stream. Simmons and Hoskins (1978) found that this happens as occluding baroclinic eddies near the end of their life cycles. To maintain thermal wind balance with the stronger jet, KZ is converted back to AZ by strengthening the meridional temperature gradient. Meanwhile KE and KZ are both dissipated. This suggests that the mean meridional circulation is overall "indirect," i.e. we do not see a net conversion of AZ into KZ as we would if the direct Hadley circulation were dominating the energetics.

Note, however, that in the summer hemisphere (in both DJF and JJA) the mean meridional circulation is direct, and the eddies are much less active. From the point of view of energetics, then, the winter hemisphere is dominated by eddy processes, while the summer hemisphere is dominated by the mean meridional circulation. Of the four conversion processes CA, CZ, CK, and CE, only the hemispheric values of CZ change sign seasonally; the others fluctuate in magnitude but not in sign.

In all cases, KE is supplied from AE; baroclinic instability is the dominant mechanism for generating eddies. Fig. 7.7 shows the annual cycles of energy conversions and amounts. The various panels show the zonal means and the Northern and Southern Hemispheres separately. The global means are relatively constant throughout the year, while the individual hemispheres show large seasonal cycles. Within each hemisphere, the winter is much more active, in all respects, than the summer. CZ changes sign seasonally. Recall that when CZ is positive the MMC is direct overall, and that when CK is positive the eddies are deriving kinetic energy from the jet, thus tending to weaken it (rather than acting to increase the kinetic energy of the jet). The figure makes it clear that the energetics of the summer and winter hemispheres are drastically different. The global AZ is a maximum shortly after the two solstices and a minimum shortly after the two equinoxes.

Fig. 7.8 shows the vertical and meridional distributions of the zonally averaged eddy kinetic energy for January and August. The Northern Hemisphere shows a strong seasonal cycle, while the Southern Hemisphere does not. Wave numbers 10 and above play little role.

Fig. 7.1 again shows the mean annual cycles of the energy amounts and conversions for the two hemispheres, but this time the information is given for several distinct wave number groups. Wave numbers 10-15 are quite unimportant, while wave numbers 4-9 tend to be the most active energetically, as would be expected from the theory of baroclinic instability.

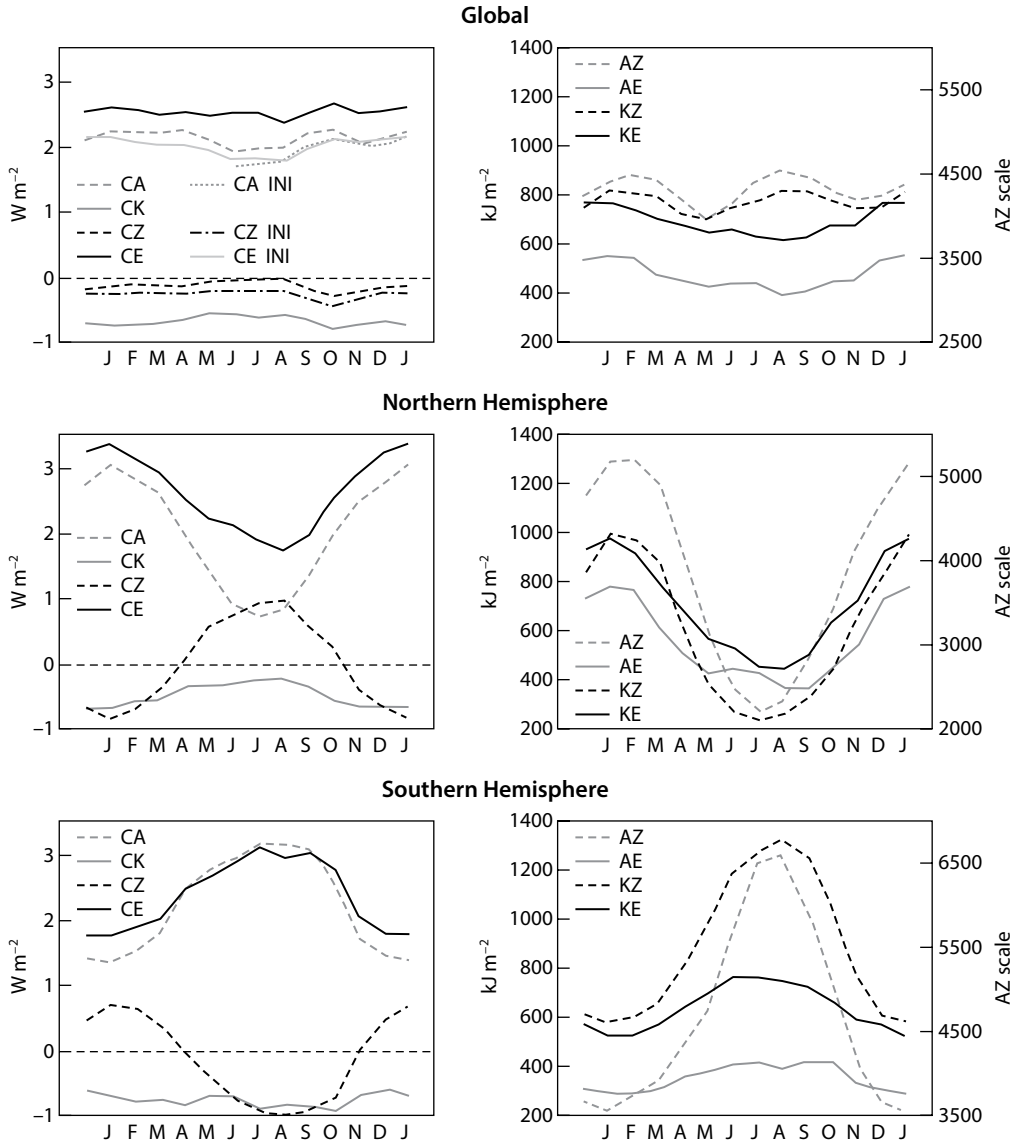


Figure 7.7: Mean annual cycles of energy conversions and amounts for global and hemispheric averages. Data are calculated from 12 hour forecasts and for global averages. Data from initialized analyses are given as well (thinner curves). In the top three panels, the zero line is drawn for reference only. Note that the lower panels have two vertical scales. From Arpe et al. (1986).

7.12 The role of heating

We have seen that heating generates APE only when it occurs where the temperature is warm. Heating at cold temperatures actually destroys APE, because it reduces temperature contrasts in the atmosphere. In many cases, heating is in fact a response to a lowering of the temperature. For example, large surface sensible heat fluxes heat the cold air rushing from

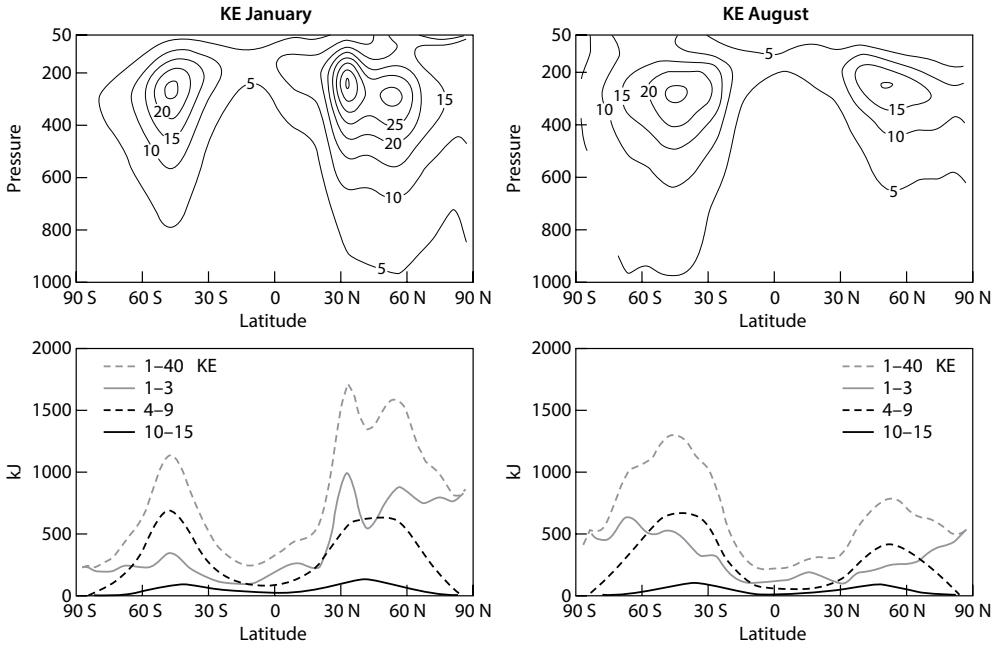


Figure 7.8: Vertical and meridional distribution of zonal mean eddy kinetic energy KE together with contributions by wavenumber groups to vertical integrals in January and August. Units in the cross-sections are J (m^2 Pa)$^{-1}$ = 100 kJ (m^2 bar)$^{-1}$. From Arpe et al. (1986).

the continents out over warm ocean currents in the winter. Similarly, deep moist convection heats the upper troposphere when cooling occurs aloft, e.g. because of large-scale lifting or the horizontal advection of cold air aloft. Such examples illustrate that heating does not necessarily promote a more vigorous circulation.

7.13 Moist available energy

Lorenz (1978, 1979) extended the concept of available potential energy by allowing moist adiabatic processes to occur during the transition to the A-state. In order to do this, he had to replace the dry enthalpy, c_pT, by a moist enthalpy, which is approximately given by $c_pT - Ll$.

As before, it can be shown that the integral of the moist enthalpy over the whole atmosphere, added to the integral of the kinetic energy over the whole atmosphere, is invariant under moist adiabatic, frictionless processes; in other words, (7.26) still applies, with a suitable re-definition of H as the vertical integral of $c_pT - Ll$. This idea can be used to generalize the concept of convective available potential energy, which was discussed in Chapter 8 (Randall and Wang, 1992).

With the conventional "dry available energy," which we have been calling the available

potential energy up to now, processes that involve phase changes can either create or destroy A. In the case of moist available energy, phase changes have no direct effect, but the surface fluxes of moisture due to evaporation and precipitation can be quite important (Lorenz, 1979). The concept of moist available energy has not been developed much up to now. It is ripe for further investigation.

7.14 Segue

The analysis of Arpe et al. (1986) reminds us of the wide range of eddy scales that are simultaneously at work in the general circulation the atmosphere. Chapter 10 gives an overview of the many types of eddies that are found in the Earth's atmosphere. All of these eddies undergo their life cycles in the presence of the same mean meridional circulation. The complicated nonlinear interactions among the eddies are the subject of Chapter 11, in which we view the general circulation as a kind of large-scale turbulence.

7.15 Problems

1. Show that, for an arbitrary process,

$$\begin{aligned} \alpha dp &= \theta d\Pi, \\ Td\varepsilon &= \Pi d\theta, \\ c_p T &= \Pi\theta + \text{constant} \end{aligned} \qquad . \tag{7.144}$$

2. A process that mixes (i.e., homogenizes) the potential temperature increases the entropy. The following two exercises illustrate this, in slightly different ways:

   (a) Consider two parcels of equal mass. Parcel number one has potential temperature

$$\theta_1 = \theta_0 + \Delta\theta \ , \tag{7.145}$$

   and parcel number two has potential temperature

$$\theta_2 = \theta_0 - \Delta\theta \ , \tag{7.146}$$

   where

$$0 \leq \Delta\theta < \theta_0 \, . \tag{7.147}$$

Show that, for a given value of θ_0, the combined entropy of the two parcels is maximized if they have the same potential temperature.

(b) Consider a process that mixes potential temperature, i.e.,

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x}\left(K \frac{\partial \theta}{\partial x} \right) \, . \tag{7.148}$$

where $K \geq 0$ is a diffusion coefficient, acting over a domain that is either periodic or else closed in the sense that $K\partial\theta/\partial x = 0$ on the boundaries. Prove that this process causes the domain-averaged entropy to increase with time.

3. Starting from

$$\int_V p \frac{D\alpha}{Dt} \rho dV > 0 \, , \tag{7.149}$$

show that the entropy, ε, satisfies

$$\int_V T \frac{D\varepsilon}{Dt} \rho dV > 0 \, , \tag{7.150}$$

and that

$$\int_V \Pi \frac{D\theta}{Dt} \rho dV > 0 \, . \tag{7.151}$$

Discuss the physical meaning of these inequalities.

4. Prove that the mass-integrated potential energy of the entire atmosphere is smaller in the A-state than in the given state.

5. Prove that

$$\frac{\partial \theta}{\partial p} = -\frac{\kappa\theta}{p}\left(\frac{\Gamma_d - \Gamma}{\Gamma_d}\right) . \tag{7.152}$$

This is used to obtain (7.43) from (7.40).

6. Prove that

$$\int_0^{p_S} p^\kappa \theta dp = \frac{1}{1+\kappa}\int_0^\infty p^{1+\kappa} d\theta . \tag{7.153}$$

Note that on the right-hand side the lower limit of integration is zero. Eq. (7.153) is used to derive (7.34) from (7.33).

7. Prove that the gross static stability cannot be negative. State any assumptions.

8. As discussed in Chapter 4, the shallow water equations are:

$$\frac{\partial h}{\partial t} + \nabla \cdot (h\mathbf{V}) = 0 , \tag{7.154}$$

$$\frac{\partial \mathbf{V}}{\partial t} + (\zeta + f)\,\mathbf{k} \times \mathbf{V} + \nabla\left[K + g\,(h + h_T)\right] = 0 \tag{7.155}$$

where $\zeta \equiv \mathbf{k} \cdot (\nabla \times \mathbf{V})$, $K \equiv \frac{1}{2}(\mathbf{V} \cdot \mathbf{V})$, h is the depth of the water, and h_T is the height of the lower boundary.

   (a) Show that the available potential energy of the system, per unit area, is $A = \frac{1}{2} g \left[\overline{(h_{fs})^2} - \overline{(h_{fs})}^2\right]$, where $h_{fs} \equiv h + h_T$ is the height of the free surface, and the bar represents an average over the whole domain.

   (b) Prove that

$$\frac{d}{dt}\left(A + \overline{hK}\right) = 0 . \tag{7.156}$$

9. (a) For the example given in the discussion beginning with Eq. (7.63), calculate the variance of θ for both the given state and the A-state, and demonstrate that the two variances are equal.

   (b) Continuing the example, assume that

$$F_\theta = -\frac{D}{a}\frac{\partial \theta}{\partial \varphi} \,, \tag{7.157}$$

   where D is a positive diffusion coefficient. Derive expressions for the time rates of change of the APE and the potential temperature variance, valid at the instant when $\theta(\varphi)$ satisfies Eq. (7.63).

   (c) Suppose that we add a "heating" term to the potential temperature equation, so that

$$\frac{\partial \theta}{\partial t} = -\frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(F_\theta \cos\varphi\right) + Q(\varphi) \ . \tag{7.158}$$

   Continue to use the form of F_θ given in part b) above. Find and plot the form of $Q(\varphi)$ needed to maintain a steady state. Find the global mean of $Q(\varphi)$. Find the covariance of $Q(\varphi)$ and θ. Discuss the roles of the diffusion and heating terms in the equation for the variance of θ.

10. Suppose that

$$\frac{d}{dt}AZ = GZ - \frac{AZ}{\tau_Z} \cong 0 \,. \tag{7.159}$$

$$\frac{d}{dt}KE = CE - \frac{KE}{\tau_E} \cong 0 \,. \tag{7.160}$$

    Here the time scales τ_Z and τ_E are used to represent dissipative processes. Estimate the values of τ_Z and τ_E in units of days, by using the numerical values given in Fig. 7.6. Compare GZ and CE with the actual rates of change of AZ and KE as shown in Fig. 7.7.

11. In passing from a given state to the A-state, the average of the pressure on an isentropic surface remains the same. Does the average of θ on an isobaric surface also remain the same? Explain why or why not, as clearly as you can.

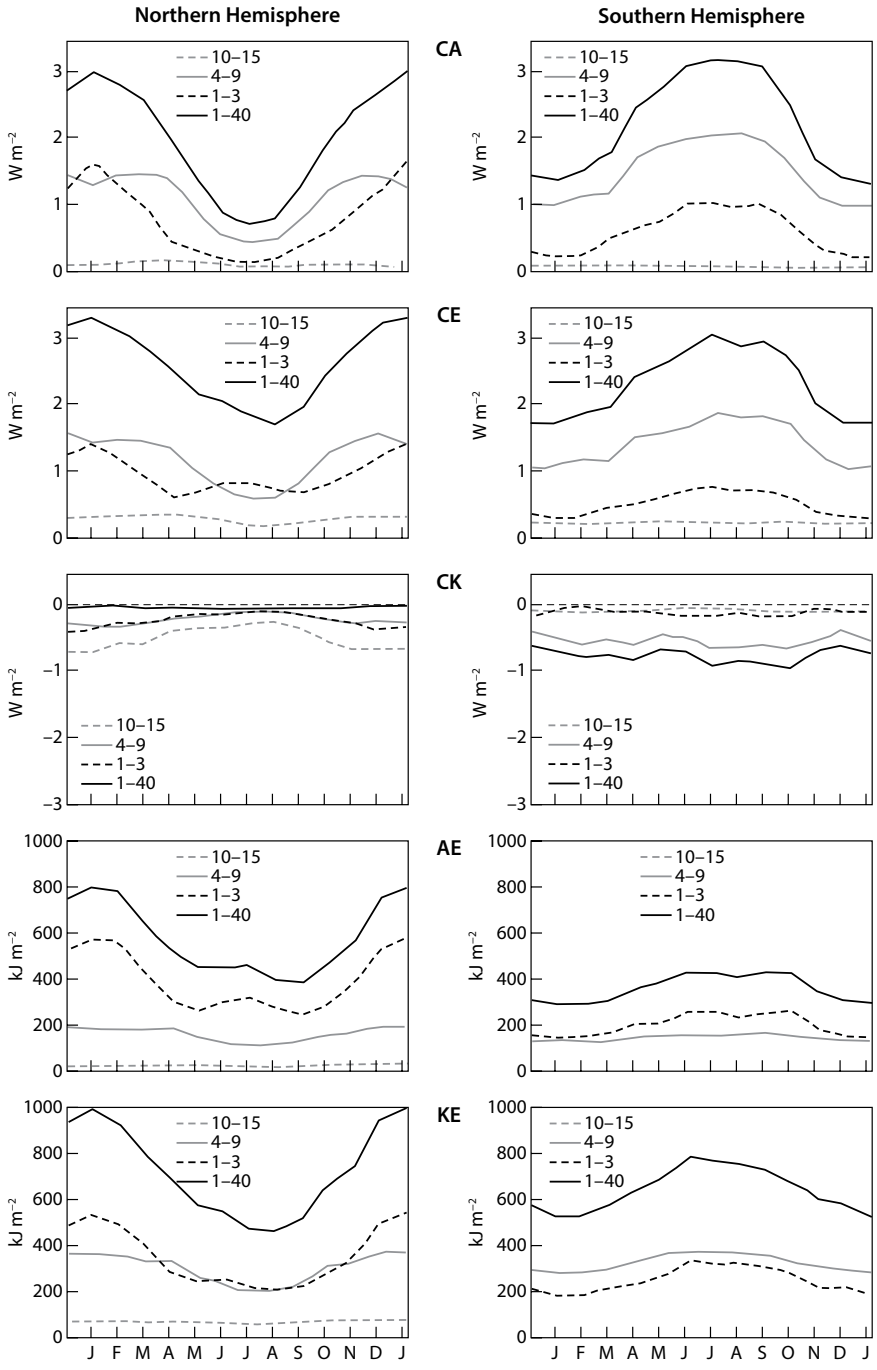


Figure 7.9: Mean annual cycles of energy amounts and conversions in both hemispheres and contributions from wavenumber groups calculated from 12 hour forecasts. Arpe et al. (1986).

Chapter 8

The roles of moist convection

8.1 What convection is, and what it does

In atmospheric science, the term "convection" refers to an important type of *buoyancy-driven* circulation, which gives rise to thermals in the boundary layer and cumulus clouds in the free troposphere above. In the older literature, the term "natural convection" was sometimes used instead. Convection involves advection, but convective motion is always buoyancy-driven. The terms "advection" and "convection" are sometimes mixed up, but typically not by atmospheric scientists.

Moist convection is important for the global circulation of the atmosphere in (at least) four ways:

- As discussed in Chapter 3 and again later in this chapter, moist convection is a key mechanism for upward energy transport .
- Convection also transports moisture and momentum, as well as various chemical constituents.
- As hypothesized by Manabe and Wetherald (1967), convection acts to prevent the lapse rate in convectively active regions from exceeding a value close to the moist adiabatic lapse rate.
- Convection produces a large fraction of the Earth's precipitation.
- Convection leads to the formation of radiatively important stratiform clouds, especially in the upper troposphere in regions of deep convection.

It is no exaggeration to say that in order to understand the global circulation we must understand the interactions between the global circulation and moist convection. There is an enormous literature on this subject, but many issues remain unresolved, and some have hardly been confronted yet (Randall et al., 2003). Current ideas are that cumulus convection produces its effects on the large-scale stratification by transporting mass vertically (via the

"cumulus mass flux"), and that the intensity of convection is regulated by the processes that act to produce convective instability; these include radiative cooling of the air relative to the temperature of the lower boundary, surface fluxes of sensible and latent heat, and the effects of both horizontal and vertical advection of temperature and moisture.

8.2 The basics of dry and moist convection

8.2.1 What is buoyancy?

The equation of vertical motion, (4.54), can be written in simplified form as

$$\frac{Dw}{Dt} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g \, . \tag{8.1}$$

Suppose that the potential temperature is given. A form of the hydrostatic equation is

$$\frac{\partial \Pi_{\rm qs}}{\partial z} \equiv -\frac{g}{\theta} \, . \tag{8.2}$$

Here the subscript qs denotes a quasi-static value. As indicated, Eq. (8.2) is the *definition* of $\Pi_{\rm qs}$. It can be used to compute $\Pi_{\rm qs}(z)$ given $\theta(z)$ and a boundary condition. We can then define the quasi-static pressure, $p_{\rm qs}$, by

$$p_{\rm qs} \equiv p_0 \left(\frac{\Pi_{\rm qs}}{c_p} \right)^{1/\kappa} \, . \tag{8.3}$$

In this way, we can compute the quasi-static pressure, $p_{\rm qs}(z)$, from $\theta(z)$. Once $p_{\rm qs}$ is known, the quasi-static density can be defined using

$$\rho_{\rm qs} \equiv -\frac{1}{g}\frac{\partial p_{\rm qs}}{\partial z} \, . \tag{8.4}$$

This is just another definition; we have *not* assumed hydrostatic balance, because the actual pressure and density can differ from $p_{\rm qs}$ and $\rho_{\rm qs}$. In fact, we define the non-hydrostatic "perturbation" pressure and density, δp and $\delta \rho$, by

$$p \equiv p_{\rm qs} + \delta p \tag{8.5}$$

and

$$\rho \equiv \rho_{\rm qs} + \delta\rho \ , \tag{8.6}$$

respectively. Notice that $\delta\theta = 0$.

With this preparation, we can now substitute (8.5) and (8.6) into (8.1), to obtain

$$\frac{Dw}{Dt} = -\frac{1}{(\rho_{\rm qs}+\delta\rho)}\frac{\partial}{\partial z}\left(p_{\rm qs}+\delta p\right) - g \ . \tag{8.7}$$

With the use of (8.4), this can be rewritten as

$$\begin{aligned}
\frac{Dw}{Dt} &= -\frac{1}{(\rho_{\rm qs}+\delta\rho)}\left(-\rho_{\rm qs}g + \frac{\partial\delta p}{\partial z}\right) - g \\
&= -\frac{1}{(\rho_{\rm qs}+\delta\rho)}\left[-\rho_{\rm qs}g + \frac{\partial\delta p}{\partial z} + (\rho_{\rm qs}+\delta\rho)\,g\right] . \\
&= -\frac{1}{(\rho_{\rm qs}+\delta\rho)}\left(\frac{\partial\delta p}{\partial z} + g\delta\rho\right)
\end{aligned} \tag{8.8}$$

Up to this point, we have not made any approximations. It is almost universal practice to neglect $\delta\rho$ in the denominator of (8.8). This leads to

$$\frac{Dw}{Dt} = -\frac{1}{\rho_{\rm qs}}\frac{\partial\delta p}{\partial z} - g\frac{\delta\rho}{\rho_{\rm qs}} \ . \tag{8.9}$$

The g term of (8.9) shows that a negative value of $\delta\rho$ favors an upward acceleration of the particle.

In the following subsections, we explore three types of buoyancy-driven convection, starting with the simplest.

8.2.2 Dry convection

Before discussing the moist convection that gives rise to cumulus clouds, we briefly discuss dry convection. For simplicity, we will use the linearized Boussinesq system of equations with a single horizontal dimension, x, and no mean flow. We neglect rotation and friction. The equation set is

$$\frac{\partial u'}{\partial t} = -\frac{1}{\rho_0}\frac{\partial p'}{\partial x} \,, \tag{8.10}$$

$$\frac{\partial w'}{\partial t} = -\frac{1}{\rho_0}\frac{\partial p'}{\partial z} + \frac{g}{c_p \overline{T}} s' \,, \tag{8.11}$$

$$\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} = 0 \,, \tag{8.12}$$

$$\frac{\partial s'}{\partial t} = -w' \frac{\partial \overline{s}}{\partial z} \,. \tag{8.13}$$

Here ρ_0 is a constant density, overbars denote horizontal averages, and primes denote departures from the horizontal averages. The gravity term of (8.11) represents the effects of buoyancy, which we have written in terms of the dry static energy perturbation. We assume for simplicity that $\partial \overline{s}/\partial z$, which appears in (8.13), is independent of height.

The perturbation pressure, p' can be eliminated by forming the vorticity equation. The horizontal vorticity vector in Cartesian coordinates is shown in Eq. (4.195); for the present case we need only the y-component, which is

$$\eta_y = \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \,. \tag{8.14}$$

The linearized equation that governs η_y is

$$\frac{\partial \eta_y'}{\partial t} = -\frac{g}{c_p \overline{T}} \frac{\partial s'}{\partial x} \,. \tag{8.15}$$

As mentioned above, the perturbation pressure has dropped out.

Eqs. (8.12) - (8.15) form a closed system for the four unknowns η_y', u', w', and s'. Elimination of u' between (8.14) and (8.12) gives

$$\nabla^2 w' = -\frac{\partial \eta_y'}{\partial x} \,, \tag{8.16}$$

where $\nabla^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial z^2$ is the two-dimensional Laplacian operator. [1] We use (8.16) to eliminate η_y in (8.15). The result is

$$\frac{\partial}{\partial t}\nabla^2 w' = \frac{g}{c_p \overline{T}} \frac{\partial^2 s'}{\partial x^2} \,. \tag{8.17}$$

The two equations (8.13) and (8.17) involve only w' and s'. We choose to eliminate s', which gives an equation involving only w':

$$\frac{\partial^2}{\partial t^2}\nabla^2 w' = -\frac{g}{c_p \overline{T}} \frac{\partial \overline{s}}{\partial z} \frac{\partial^2 w'}{\partial x^2} \,. \tag{8.18}$$

We look for a separable solution, of the form

$$w'(x,\, z,\, t) = \mathrm{Re}\left\{\hat{w}\, \exp\left(ikx + imz + \sigma t\right)\right\} \,. \tag{8.19}$$

where $\hat{w}$ is a complex constant; k and m are horizontal and vertical wave numbers, respectively; and σ can be either real or imaginary. Substituting (8.19) into (8.18), we find that for $\hat{w} \neq 0$

$$\sigma^2 = -\frac{g}{c_p \overline{T}} \frac{\partial \overline{s}}{\partial z} \left(\frac{1}{1 + m^2/k^2}\right) \,. \tag{8.20}$$

[1] Eq. (8.16) says that if we know $\partial \eta_y/\partial x$ we can solve an elliptic equation to obtain the vertical velocity, w'. This is actually quite interesting and important, but it's not what we are going to do here.

For $\partial\overline{s}/\partial z < 0$, σ is real, and there is an exponentially growing solution with $\sigma > 0$; this is dry convective instability. We say that

$$\frac{\partial \overline{s}}{\partial z} < 0 \text{ is the criterion for dry convective instability .} \tag{8.21}$$

For a given value of $\partial\overline{s}/\partial z$, the growth rate, σ, is maximized for $m^2/k^2 \to 0$, which means circulations that are tall and narrow. It can be inferred from (8.13) that in the exponentially growing solution, with $\sigma > 0$ and $\partial\overline{s}/\partial z < 0$, $w'(t)$ and $s'(t)$ have the same sign for all time, so that $\overline{w's'} > 0$, i.e., convection transports dry static energy upward. We already showed in Chapter 7 that an upward temperature flux tends to lower the atmosphere's center of gravity, i.e., it reduces the total potential energy of the atmospheric column. The reduction in potential energy coincides with a generation of convective kinetic energy through the work done by the buoyancy force, so that total energy is conserved. The generation of convective kinetic energy through an upward flux of dry static energy can be seen directly by multiplying both sides of the equation of vertical motion, (8.11), by w', and averaging, to obtain

$$\boxed{\frac{\partial}{\partial t}\left(\frac{1}{2}\overline{w'^2}\right) = -\frac{w'}{\rho_0}\overline{\frac{\partial p'}{\partial z}} + \frac{g}{c_p\overline{T}}\overline{w's'}} \; . \tag{8.22}$$

This equation shows that buoyancy supplies the kinetic energy of the convection.

The other case is $\partial\overline{s}/\partial z > 0$, so that σ is imaginary and the solutions are oscillatory in time; these are gravity waves. Their frequency is $\sqrt{(g/c_p\overline{T})\partial\overline{s}/\partial z} > 0$, which is is called the Brunt-Väisällä frequency. Using the analysis given above as a starting point, you should be able to show that $\overline{w's'} = 0$ for a gravity wave. This means that gravity waves do not transport dry static energy. As discussed in Chapter 11, they do transport momentum, and this is quite important for the global circulation.

The analysis above shows that, in the absence of phase changes, convection and gravity waves are mutually exclusive; they cannot occur in the same place at the same time. We will see later that this conclusion does not necessarily apply when phase changes are allowed.

Before moving on, we need to do one more thing. It is easy to show that

$$\frac{\partial s}{\partial z} = c_p\left(\Gamma_d - \Gamma\right) \; , \tag{8.23}$$

The temperature decreases upward at the dry adiabatic lapse rate when the dry static energy (or potential temperature) is independent of height. Use of (8.23) allows us to rewrite (8.13) as

$$\frac{\partial T'}{\partial t} = w' \left(\overline{\Gamma} - \Gamma_d \right) \ . \tag{8.24}$$

and re-state the criterion for dry convective instability as

$$\Gamma > \Gamma_d \ . \tag{8.25}$$

8.2.3 Convection in uniformly saturated air

Up to this point, we have considered dry adiabatic motion. To analyze moist convection, we will assume saturated moist adiabatic motion. As discussed in Chapter 4, the moist static energy, h, is approximately conserved under moist adiabatic processes. We replace (8.13) by

$$\frac{\partial h'}{\partial t} = -w' \frac{\partial \overline{h}}{\partial z} \ . \tag{8.26}$$

For saturated motion, the moist static energy must be equal to the saturation moist static energy,

$$h_{\rm sat} \equiv c_p T + gz + L q_{\rm sat}(T,p) \ , \tag{8.27}$$

where the saturation mixing ratio depends, as indicated, on both temperature and pressure. We can rewrite (8.26) as

$$\frac{\partial h'_{\rm sat}}{\partial t} = -w' \frac{\partial \overline{h}_{\rm sat}}{\partial z} \ . \tag{8.28}$$

Next, we have to relate the buoyancy term of (8.15) to $h'_{\rm sat}$. In nature, the buoyancy is influenced not only by temperature, but also by water vapor (the "virtual temperature"

effect) and condensed water ("condensate loading"). Here we include only the effects of temperature, to avoid unnecessary complications. At fixed height, and at approximately fixed pressure, perturbations of $h_{\rm sat}$ and s are related by

$$h'_{\rm sat} \cong s' \, (1+\gamma) \ , \tag{8.29}$$

where, in the linearization,

$$\gamma \equiv \frac{L}{c_p} \left(\frac{\partial q_{\rm sat}}{\partial T} \right)_p \tag{8.30}$$

is evaluated using the mean-state temperature and pressure. The nondimensional parameter γ is positive and of order one. We can write

$$\frac{s'}{c_p \overline{T}} = \frac{h'_{\rm sat}}{c_p \overline{T} \, (1+\gamma)} \ . \tag{8.31}$$

Substitution of (8.31) into (8.15) gives

$$\frac{\partial \eta'_y}{\partial t} = - \frac{g}{(1+\gamma) \, c_p \overline{T}} \frac{\partial h'_{\rm sat}}{\partial x} \ . \tag{8.32}$$

From this point, the argument is exactly the same as for dry convection, and we find that

$$\sigma^2 = - \frac{g}{(1+\gamma) \, c_p \overline{T}} \frac{\partial \overline{h_{\rm sat}}}{\partial z} \ , \tag{8.33}$$

which implies that

$$\frac{\partial \overline{h_{\rm sat}}}{\partial z} < 0 \text{ is the criterion for convective instability of a saturated atmosphere .} \tag{8.34}$$

Compare with (8.21).

The criterion for moist convective instability can also be expressed in terms of the moist adiabatic lapse rate, which is given by

$$\begin{aligned} \Gamma_m &\equiv -\left(\frac{\partial \overline{T}}{\partial z}\right)_{\text{moist adiabatic}} \\ &\cong \Gamma_d \left[\frac{1+\frac{L q_{\text{sat}}(T,p)}{R_d T}}{1+\frac{L^2 q_{\text{sat}}(T,p)}{c_p R_v T^2}}\right] . \end{aligned} \tag{8.35}$$

Eq. (8.35) is derived in Appendix F. In the square brackets of (8.35), the denominator is larger than the numerator, so $\Gamma_m < \Gamma_d$, although $\Gamma_m \to \Gamma_d$ at cold temperatures. For example, with a pressure of 1000 hPa and a temperature of 288 K, we find that $\Gamma_m = 4.67$ K km^{-1} (see Fig. 8.1). As the temperature increases, the moist adiabatic lapse rate decreases. As discussed in Chapter 14, this is important for climate change. For saturated motion, we can rewrite (8.13) as

$$\frac{\partial T'}{\partial t} = w' \left(\overline{\Gamma} - \Gamma_m\right) . \tag{8.36}$$

According to (8.36), the criterion for moist convective instability is $\overline{\Gamma} > \Gamma_m$. Compare (8.36) with (8.24). Both of these equations will be used later.

8.2.4 Conditional (or cumulus) instability

We have considered convection in both dry air and uniformly saturated air. Cumulus convection, which is extremely important for the global circulation of the atmosphere, involves saturated updrafts surrounded by unsaturated air. It is more complicated and also more interesting than dry convection and uniformly saturated convection.

If a parcel of air containing watervapor is lifted adiabatically from near the surface, it will eventually become saturated due to the cooling caused by adiabatic expansion. Prior to reaching its lifting condensation level, both the dry static energy and the moist static energy of the parcel will be conserved. Suppose that, once the lifting-condensation level has been exceeded, the moist static energy of the now-cloudy parcel remains constant, i.e., that

$$\frac{\partial h_c}{\partial z} = 0 . \tag{8.37}$$

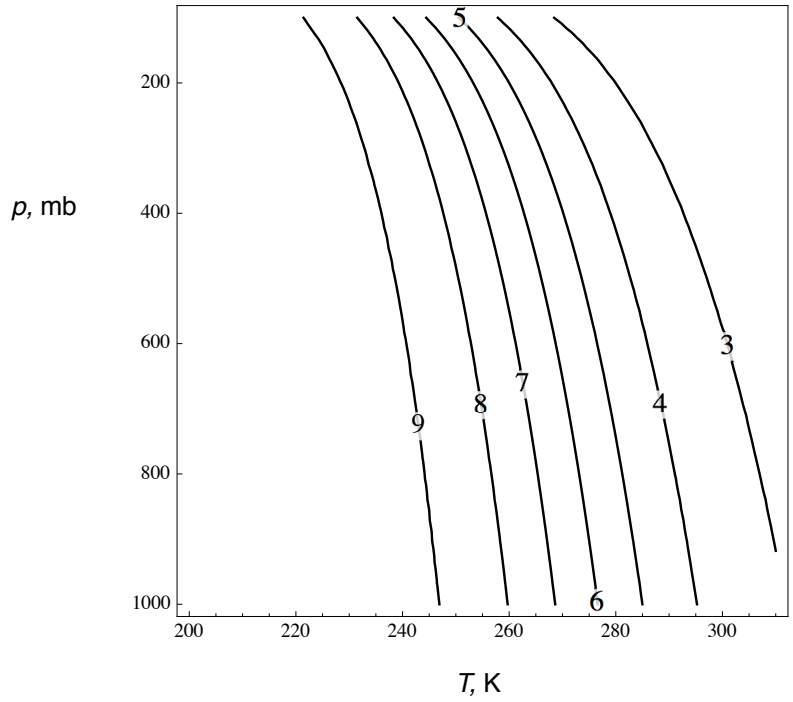


Figure 8.1: A plot of the moist adiabatic lapse rate, Γ_m, in K km^{-1}, as a function of temperature and pressure.

This will be the case if the effects of radiation and mixing with the environment are negligible; such an in-cloud sounding is represented by the red, vertical line in Fig. 8.2. The dry static energy of the cloudy air will increase upward due to latent heat release. The liquid water mixing ratio will increase, and the water vapor mixing ratio will correspondingly decrease. Under moist adiabatic processes, the total mixing ratio, $q_v + l$, will not change. If precipitation occurs, $q_v + l$, will decrease upward.

Conservation of h and $q_v + l$ implies that

$$s_l \equiv h - L(q_v + l) \tag{8.38}$$

is also conserved. We refer to s_l as the "liquid water static energy." It is conserved under moist adiabatic processes. Precipitation from a parcel is not a moist adiabatic process, because it involves the removal of mass from the parcel. Precipitation can change the value of s_l, but it does not change the value of h. This means that h is "more conservative" than s_l. Both variables are useful.

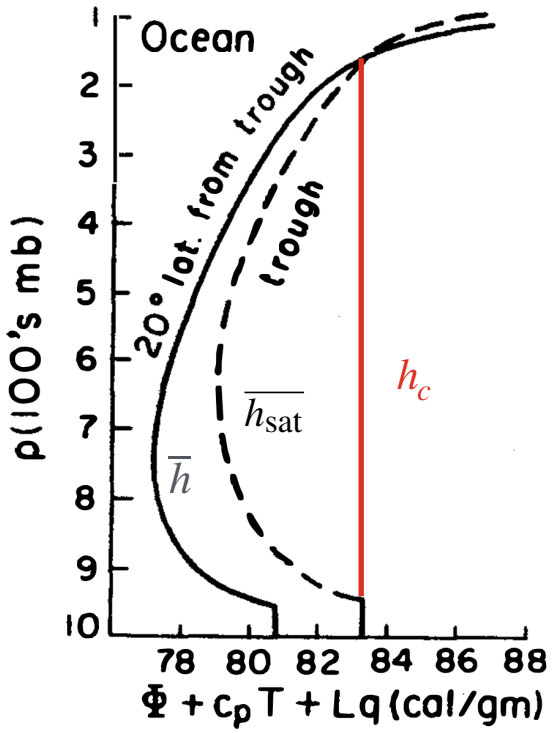


Figure 8.2: A figure from Riehl and Malkus (1958), showing typical tropical soundings of the moist static energy in the ITCZ (dashed curve) and the subtropics (solid curve). Note the archaic units. Compare with Figure 6.10. The red vertical line has been added to represent the moist static energy of a cumulus updraft that originates near the surface and rises without mixing with the surrounding air.

Using (8.29), we can write the temperature difference between the cloudy air and its environment at the same level as

$$c_p\left(T_c-\overline{T}\right)=\frac{\left(h_{\rm sat}\right)_c-\overline{h_{\rm sat}}}{1+\gamma}\,, \tag{8.39}$$

where a subscript c denotes the cloudy air, and an overbar denotes an area average. Because the updraft air is asssumed to be saturated, we can replace $\left(h_{\rm sat}\right)_c$ by h_c, giving

$$c_p\left(T_c-\overline{T}\right)=\frac{h_c-\overline{h_{\rm sat}}}{1+\gamma}\,. \tag{8.40}$$

This shows that the buoyancy of the cloudy air, as measured by the difference between its

temperature and the temperature of the environment, is proportional to the difference between the moist static energy of the cloudy air and the *saturation* moist static energy of the environment. Recall, however, that we are considering a parcel that is lifted adiabatically from near the surface, conserving its moist static energy. This means that h_c is equal to the low-level moist static energy of the sounding.

In order for the non-entraining updraft to become positively buoyant as it rises, we need

$$\frac{\partial \overline{h_{\rm sat}}}{\partial z} < 0 \,, \tag{8.41}$$

so that the $\overline{h_{\rm sat}}$ curve will fall to the left of the vertical line that represents h_c. Eq. (8.41) involves only the temperature profile; it is the criterion for conditional instability (i.e., cumulus instability) in terms of temperature. Of course, we also need enough water vapor to make the updraft saturated. That is the "condition" for conditional instability. As and shown in (8.34), Eq. (8.41) is also the condition for neutral stability in uniformly saturated air. Note, however, that we have derived (8.41) without assuming that the air is saturated; it applies even when the relative humidity of the air outside the cloudy updraft is less than 100%. More discussion is given later.

The cloudy updraft will reach its level of neutral buoyancy where $h_c = \overline{h_{\rm sat}}$. If the level of neutral buoyancy is high and cold, then $\overline{h_{\rm sat}} \cong \overline{h} \cong s$. In nature, the convective updrafts can overshoot their levels of neutral buoyancy, but as a first approximation we can assume that the updraft cloud tops are *at* the level of neutral buoyancy for air lifted from the boundary layer, i.e.,

$$\overline{h} \text{ at the cloud top} \cong \overline{h} \text{ in the boundary layer} \tag{8.42}$$

in regions where deep convection is active. Most deep convection reaches its level of neutral buoyancy several kilometers below the tropopause. To have positive gross moist stability, the Hadley circulation must penetrate higher than the level where (8.42) is satisfied.

8.3 Upward energy transport by deep convection

Riehl and Malkus (1958, see Fig. 8.3) argued from the observed energy balance and vertical structure of the tropical atmosphere that deep, penetrative cumulus convection is the primary mechanism for upward energy transport in the tropics. They began by estimating the mass circulation across a latitude 10° on the winter side of the Inter-Tropical Convergence Zone (ITCZ), which they called the "Equatorial Trough Zone." Recall that the

"body" of the main solstitial Hadley Cell lies in the winter hemisphere. Riehl and Malkus neglected the mass transport across the boundary of the ITCZ on the summer side. They then attempted to evaluate the lateral energy transports on the winter side of the ITCZ, as functions of height. They considered transports of internal energy, potential energy, and latent energy. Their estimates of the various quantities are summarized in Table 8.1.



Figure 8.3: Joanne Malkus (later Joanne Simpson), and Herbert Riehl. The photo was taken at approximately the time of their collaboration on Riehl and Malkus (1958), and was graciously provided by Joanne Simpson.

Because the low-level inflow is warm and wet, while the upper level outflow is cold and dry, both internal and latent energy flow in to the ITCZ; nevertheless there is a net loss of total energy due to the export of potential energy in the elevated outflow layer. *The net export of energy by the meridional flow implies that other processes supply a net energy input to the ITCZ.* This is one of the key conclusions of their paper.

8.4 The gross moist stability

Neelin and Held (1987) considered the moist static energy budget of the ITCZ from a perspective very similar to that of Riehl and Malkus (1958). They reasoned as follows: In a time average, the vertically integrated moist static energy budget of a column of air is expressed by

$$g^{-1}\nabla\cdot\left(\int_0^{p_S}\mathbf{v}h\,dp\right) = -\left(N_S - N_{\mathrm{T}}\right)\ , \tag{8.43}$$

Table 8.1: Lateral energy transports on the poleward side of the ITCZ, based on Table 3 of Riehl and Malkus (1958). The units have been converted to conform with modern standards.

δp, 100s hPa	v, m s^{-1}	M_0, 10^{13} g s^{-1}	s, J g^{-1}	sM_0, PW	Lq, J g^{-1}	LqM_0, PW	hM_0, PW	Eddy moisture transport, PW	Total energy transport out of the ITCZ, PW
10-9	-1.3	-5.2	301.5	-15.6	37.6	-2.0			
9-8	-1.1	-4.4	305.7	-13.4	27.6	-1.2			
8-7	-0.4	-1.6	311.6	-4.9	18.4	-0.3			
7-6	0.0	0.0	317.8	0.0	11.3	0.0			
6-5	0.0	0.0	323.3	0.0	7.1	0.0			
5-4	0.3	1.2	329.1	3.9	4.2	0.0			
4-3	0.6	2.4	335.4	8.0	2.1	0.0			
3-2	1.3	5.2	340.8	17.7	0.8	0.0			
2-1.25	0.8	2.4	348.4	8.3	0.0	0.0			
10-5				-33.9		-3.5	-37.4	0.7	-36.7
5-1.25				37.9		0.1	38.0	0.0	38.0
10-1.25				4.0		-3.4	0.6	0.7	1.3

where N is the net *downward* flux of energy due to radiation, turbulence, and convection, and subscripts T and S denote the top of the atmosphere and the surface, respectively. We have neglected horizontal variations of the surface pressure. Similarly, mass continuity gives

$$\nabla \cdot \left(\int_0^{p_S} \mathbf{v} dp \right) = 0 \,. \tag{8.44}$$

They divided the column into upper and lower portions, and wrote

$$
\begin{aligned}
p_S^{-1}\nabla\cdot\left(\int_0^{p_S}\mathbf{v}h\,dp\right) =& \nabla\cdot(\mathbf{v}h)_u+\nabla\cdot(\mathbf{v}h)_l \\
=& -g\left(\frac{N_S-N_{\rm T}}{p_S}\right) ,
\end{aligned}
\tag{8.45}
$$

$$
\begin{aligned}
p_S^{-1}\nabla\cdot\left(\int_0^{p_S}\mathbf{v}\,dp\right) =& \nabla\cdot\mathbf{v}_u+\nabla\cdot\mathbf{v}_l \\
=& 0 .
\end{aligned}
\tag{8.46}
$$

The upper and lower portions of the column correspond to the bottom two "summary" rows of Table 8.1. In the tropics, horizontal variations of h are weak on large scales, so that it is useful to *define* h_u and h_l by

$$
\begin{aligned}
\nabla\cdot(\mathbf{v}h)_u \equiv& h_u(\nabla\cdot\mathbf{v}_u) \\
=& -h_u(\nabla\cdot\mathbf{v}_l) ,
\end{aligned}
\tag{8.47}
$$

and

$$
\nabla\cdot(\mathbf{v}h)_l \equiv h_l(\nabla\cdot\mathbf{v}_l) , \tag{8.48}
$$

from which it follows that

$$
\nabla\cdot(\mathbf{v}h)_u+\nabla\cdot(\mathbf{v}h)_l = -(\nabla\cdot\mathbf{v}_l)(h_u-h_l) . \tag{8.49}
$$

We can also show from (8.45)-(8.49) that

$$
\nabla\cdot\mathbf{v}_l = \frac{g}{p_S}\left(\frac{N_S-N_T}{h_u-h_l}\right) . \tag{8.50}
$$

This implies that low-level convergence ($\nabla \cdot \mathbf{v}_l < 0$) must occur where the column is gaining energy ($N_S - N_T < 0$), provided that $h_u - h_l > 0$. Neelin and Held named the quantity $h_u - h_l$ the "gross moist stability." According to this definition, if the (lower) troposphere is moistened, without changing the temperature profile, the gross moist stability tends to decrease. As discussed later in detail, adding moisture to the lower troposphere tends to increase the convective available potential energy (CAPE). A decrease in the gross moist stability is consistent with an increase in the CAPE.

Eq. (8.49) shows that low-level convergence of mass (i.e., $\nabla \cdot (\mathbf{v}h)_l < 0$, as in the ITCZ) leads to a net vertically integrated divergence of moist static energy (i.e., $\nabla \cdot (\mathbf{v}h)_u + \nabla \cdot (\mathbf{v}h)_l > 0$) when the gross moist stability is positive. This must be the average state of affairs in the ITCZ, because, as Riehl and Malkus showed, the ITCZ exports moist static energy towards higher latitudes.

On the other hand, in a region where the gross moist stability is negative, low-level convergence leads to a net vertically integrated convergence of moist static energy. That would tend to make the gross moist stability even more negative, suggesting the interesting possibility of a positive feedback in the presence of a negative gross moist stability: Low-level convergence leads to a moistening of the column, which increases the CAPE, which leads to more intense convection, which favors more low-level convergence.

In summary, according to this line of reasoning, the sign of the gross moist stability determines whether a column that is experiencing low-level convergence imports or exports energy. As a matter of terminology, this is consistent with the fact that, for a dry column, the (always positive) dry static stability determines how much energy the column exports.

Note, however, that there is an issue of causality. We can alternatively reason as follows: In a time average, a column can only export energy if it experiences a net diabatic energy source. If the column gains energy diabatically through the combination of radiation and the surface moist static energy flux, then in a time average it must export energy through the horizontal winds. In such a case, the gross moist stability must be positive. Conversely, if the column loses energy diabatically through the combination of radiation and the surface moist static energy flux, then in a time average it must import energy through the horizontal winds. Therefore its gross moist stability must be negative. From this point of view, the net diabatic source or sink of energy for the column determines *both* the sign of the gross moist stability *and* whether the column exports or imports energy. In my opinion, it is a mistake to think that the sign of the gross moist stability *determines* whether the ITCZ exports or imports energy. The opposite is true.

According to (8.50), the pattern of the gross moist stability is closely linked to the pattern of low-level convergence, for a given distribution of $N_S - N_T$. As discussed later [see (8.42)], we expect the gross moist stability to be small where cumulus convection is active.

The term "gross moist stability" has been redefined in multiple ways in various publications that followed that of Neelin and Held (1987). If you run across the term in a paper, check carefully to see what definition is being used.

8.5 The tight coupling of heating and vertical motion in the tropics

The ideas discussed above fit well with the work of Charney (1963), which was already mentioned in Chapter 3. Recall Charney's explanation for the fact that horizontal temperature gradients tend to be weak in the tropics. As discussed in Chapter 4, the thermodynamic energy equation can be written as

$$c_p \left(\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T + \omega \frac{\partial T}{\partial p} \right) = \omega \alpha + Q \, . \tag{8.51}$$

Here the vertical coordinate is pressure, and Q is the heating rate per unit mass. An averaging interval of one month is sufficient to make the time-rate-of-change term of(8.51) fairly small. If we also invoke Charney's ideas to neglect the horizontal advection of temperature, we find that

$$\boxed{\overline{\omega \frac{\partial s}{\partial p}}^{t} \cong \overline{Q}^{t} \quad \text{in the tropics} \, .} \tag{8.52}$$

Here s is the dry static energy, and the hydrostatic equation has been used to combine the two terms of (8.51) that involve ω. Recall that s normally increases upwards. The vertical motion term of (8.52) therefore represents cooling ($-\omega \partial s / \partial p < 0$) when the air is rising ($\omega < 0$) and warming ($-\omega \partial s / \partial p > 0$) when the air is sinking ($\omega > 0$). Eq. (8.52) simply says that the only way to balance tropical heating or cooling is through vertical motion; there must be a very strong correspondence between the patterns of tropical vertical motion and heating, such that tropical rising motion occurs almost exclusively where latent and radiative heating are active, i.e., where $Q > 0$, and tropical sinking motion occurs almost exclusively where radiative cooling is dominant, i.e., where $Q < 0$.

These conclusions apply not only in the tropics, but also in the middle latitudes in summer, simply because horizontal temperature gradients tend to be weak there as well. In the middle-latitudes in winter, the horizontal advection term of (8.51) can be dominant.

8.6 Radiative-convective equilibrium

The study of Riehl and Malkus showed that in the tropics (and also in the moist convective regions of the summer hemisphere middle latitudes) upward transport of energy is due to small-scale convection rather than vertical advection by the large-scale vertical motion. Recall that the brightness temperature of the Earth corresponds to a level in the middle troposphere; the brightness temperature of the tropical convective regions corresponds to the actual temperature in the upper troposphere. We can consider that convection transports energy upward, to the middle or upper troposphere, where radiation can take over and carry the energy on out to space.

We can imagine an idealized system in which radiation and convection (including turbulence) are the only processes acting, and large-scale motion, including large-scale vertical motion, is absent. Note that the ITCZ is not a radiative-convective system because it does include large-scale vertical motion. The time-averaged properties of a radiative-convective system are called "*radiative-convective equilibrium,*" which can be abbreviated as RCE. Although RCE does not exist in the real world, it can be simulated using models.

Starting in the 1960s, there have been many modeling studies of RCE. A model that can represent RCE includes physical parameterizations that suffice to determine the time rates of change of the temperature and moisture within the atmospheric column and at the Earth's surface, omitting the effects of large-scale circulations. Time-averaged results from such a model are interpreted as simulations of RCE.

The physical ingredients of a model that can simulate RCE include parameterizations of radiation, convection, turbulence, and the processes that determine the change of the surface temperature and moisture. For the atmosphere, we write

$$\rho c_p \frac{\partial T}{\partial t} = \rho LC - \frac{\partial F_s}{\partial z} + Q_R \ . \tag{8.53}$$

As a reminder, L is the latent heat of condensation, C is the condensation rate, and F_s is the dry static energy flux (a.k.a. the sensible heat flux). We use Q_R to denote the radiative heating rate. Horizontal and vertical advection are deliberately omitted in (8.53), because the purpose of a radiative-convective model is to help us to understand what the atmosphere would look like in their absence. In order to use (8.53), we have to determine the condensation rate and the convective fluxes. This will entail consideration of the moisture budget. In addition, the vertical distributions of water vapor and clouds are needed to determine Q_R, which satisfies

$$Q_R = \frac{\partial}{\partial z}(S - R) \ , \tag{8.54}$$

Here S is the net solar radiation (positive down), and R is the net terrestrial radiation (positive up).

We also include an energy budget for the Earth's surface:

$$C_g \frac{\partial T_S}{\partial t} = N_S\left(T_S\right) \ . \tag{8.55}$$

Here C_g is the effective "heat capacity" of the surface, which was discussed in Chapter 2; T_S is the surface temperature, and

$$\begin{aligned} N_S \equiv & S_S - R_S - \left(F_h\right)_S \\ = & S_S + \left(R\downarrow\right)_S - \sigma T_S^4 - \left(F_h\right)_S \ . \end{aligned} \tag{8.56}$$

denotes the net downward vertical flux of energy due to turbulence, convection, and radiation. Eq. (8.55) can be used to determine T_S, provided that $N_S\left(T_S\right)$ is known. The value of C_g determines how rapidly the surface temperature changes in response to a given value of N_S. When C_g is large, T_S changes slowly. In the limit $C_g \to 0$, T_S adjusts instantaneously, so as to keep $N_S = 0$.

The balance requirements that must be satisfied in RCE can be stated as follows: As discussed in Chapter 2, in a time average there can be no net radiative energy flux at the top of the (globally averaged) atmosphere. We impose that requirement in RCE:

$$N_T = S_T - R_T = 0 \ . \tag{8.57}$$

Here N_T is also positive downward. The atmospheric column must be in energy balance, so

$$N_S = N_T \ . \tag{8.58}$$

From (8.57) - (8.58), it follows that, in equilibrium, the Earth's surface must also be in energy balance:

$$N_S = 0 \, . \tag{8.59}$$

In fact, the net vertical energy flux must be zero at every level throughout the atmosphere.

In a series of studies during the 1960s, Syukuro Manabe (Figure 8.4) and his colleagues investigated the degree to which pure radiative equilibrium and/or RCE can explain the observed vertical distribution of temperature. The most famous of these studies is described by Manabe and Wetherald (1967). Their work led to major advances in our understanding of the vertical structure of the atmosphere, and also the water-vapor feedback that plays a major role in climate variability. A very interesting first-person account of the research is given by Manabe and Broccoli (2020).



Figure 8.4: Syukuo Manabe, a pioneer of global atmospheric modeling and studies of anthropogenic climate change. He was awarded a share of the 2021 Nobel Prize in physics.

In today's world, RCE can be simulated using a wide variety of numerical models that include the effects of radiation, turbulence, cumulus convection, and stratiform clouds. These modern models simulate the distribution of both water vapor and condensed water. Some of them can also simulate the photochemical processes that produce ozone in the stratosphere. A recent study of RCE is discussed by Wing et al. (2020).

During the 1960s, models were much less mature. In particular, not much was known about how to include the effects of cumulus convection and other moist physics. Manabe and Wetherald (1967) chose not to explicitly represent moist processes in their model; instead, they considered two alternative assumptions to determine the vertical distribution of moisture (discussed below), and adopted fairly drastic but empirically justified simplifying assumptions to determine the effects of latent heat release and moist convection on the atmospheric temperature profile. Manabe and Wetherald (1967) began by specifying

- the composition of the dry air, including ozone and carbon dioxide, as a function of height;

- the albedo of the Earth's surface;
- the vertical distribution of water vapor, using two different approaches as discussed below; and
- the vertical distribution of radiatively active clouds, as a function of height.

The vertical distribution of ozone is important for the radiation calculation, because it is the absorption of solar radiation by ozone that accounts for the upward increase of temperature in the stratosphere; without ozone, there would be no stratosphere. Manabe and colleagues did not attempt to compute the vertical distribution of ozone; instead they prescribed it according to observations.

The vertical distribution of water vapor has a powerful effect on the radiation budgets of both the atmosphere and the surface. In order to predict the vertical distribution of water vapor, a model must include representations of turbulence and cumulus convection. Manabe and colleagues did not attempt this. Instead, they prescribed the vertical distribution of water vapor using two alternative methods. The first is fixed specific humidity, which they called "fixed absolute humidity," and the second is fixed relative humidity. The specified vertical distributions of specific humidity and relative humidity were both based on observations similar to those shown in Chapter 3.

More modern simulations of RCE do include a moisture budget, in which water vapor is introduced by evaporation from the sea surface, convection and turbulent diffusion carry the moisture upward, and precipitation removes it. An example is given later.

Manabe and colleagues used a time-marching method and some simplifying assumptions to find equilibrium solutions. The time-marching method appears to describe the temporal approach to equilibrium, but some of the assumptions used rule out that interpretation. The time-marching is simply an iterative method to find an equilibrium solution. Initial conditions were given for

- the temperature of the atmosphere as a function of height, and
- the temperature of the surface.

From the vertical profiles of temperature, water vapor, ozone, and cloudiness, it is possible to calculate the solar and terrestrial radiative energy fluxes and the net radiative cooling of the atmosphere at each level. Manabe and colleagues used a simple (but good for the time) radiative transfer model to compute the net radiative cooling as a function of height, and also the net radiative fluxes at the surface and at the top of the atmosphere. So far, so good.

How can we determine the surface moist static energy flux, $(F_h)_S$? In principle, we could use the bulk aerodynamic formula, as discussed in Chapter 6, but there are drawbacks to that approach:

- Over the ocean, $(F_h)_S$ depends on the low-level wind speed, which can't be calculated in their simple model. It is best to specify as few things as possible, so this option is not attractive.
- Over land, the surface latent heat flux, which contributes to N_S, also depends on the soil moisture content and vegetation cover, among other things. This is more complexity than we are willing to include in the model.

As one simple possibility, we could assume that on each time step the net radiative cooling of the atmosphere is equal to the net radiative warming of the surface, i.e.,

$$\text{ARC} = S_S - R_S \tag{8.60}$$

Eq. (8.60) is exactly true if the net radiation at the top of the atmosphere is zero (see the tables in Chapter 2), so it must hold in equilibrium, but we assume that it also applies during the approach to equilibrium. From (8.59), (6.25), and (8.60), it follows that the surface flux of moist static energy must be whatever it takes to balance the net radiative cooling of the atmospheric column:

$$(F_h)_S = \text{ARC} \, . \tag{8.61}$$

Again, this is required in equilibrium, but there is no physical reason why it has to be true during the approach to equilibrium. Eq. (8.61) means that the atmospheric column as a whole is in energy balance, even though imbalances occur at particular levels inside the column during the approach to equilibrium. Using (8.61), an updated surface temperature can be determined by time-stepping (8.55) with (8.55).

An RCE state can be found using the computational procedure summarized in Fig. 8.5. Initial conditions are given for the vertical profile of atmospheric temperature. After the temperature profile has been modified by radiation, on each time step, the resulting temperature profile is checked for convective stability. Manabe and Wetherald (1967) assumed that moist convective instability exists if the temperature decreases upward more rapidly than 6.5 K km^{-1}. If instability is found, the temperature profile is adjusted so as to restore a lapse rate of 6.5 K km^{-1}. For the case of fixed relative humidity, the vertical distribution of water vapor is then corrected to take into account the temperature change.

In the model, the vertical structure of the sounding is represented by a "stack" of layers. The adjustment was done between pairs of neighboring layers. Removing the instability between one pair of layers can create instability with respect to neighboring layers, so to completely remove all instability it is necessary to apply the procedure iteratively. This

pairwise adjustment between neighboring layers ignores the observed fact that cumulus clouds can penetrate through the entire troposphere.

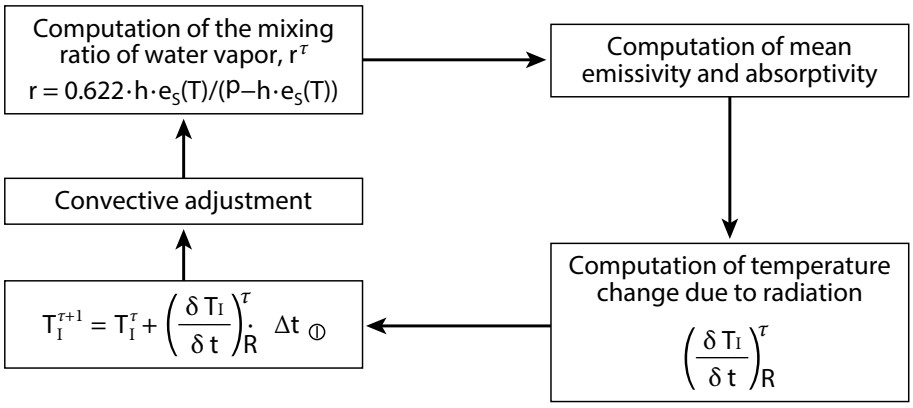


Figure 8.5: Flow chart for the solution of the radiative-convective model of Manabe and Wetherald (1967).

The assumption that the lapse rate "adjusts" to 6.5 K km^{-1} is based on the *physical hypothesis that convection acts to prevent the lapse rate from becoming much steeper than the moist adiabatic lapse rate*. Based on the discussion given earlier in the chapter, the convective adjustment of the temperature could have been designed to make $\partial \overline{h_{sat}}/\partial z = 0$ in the adjusted state. Fig. 8.1 shows that in radiative-convective equilibirum the moist adiabatic lapse rate is 6.5 K km^{-1} in the lower troposphere when the temperature is about 275 K. In the tropics, the lower-tropospheric moist adiabatic lapse rate is closer to 4 K km^{-1}.

The key physical idea here is that *radiation tries to make the sounding unstable, and convection responds by removing the instability. This means that the intensity of the convective activity is determined by the rate at which radiation tries to destabilize the column*. In equilibrium, there is a balance between radiative destabilization and convective stabilization. Further discussion is given later in this chapter.

Fig. 8.6, which is taken from Manabe and Wetherald (1967), shows equilibrium solutions for three cases:

- pure radiative equilibrium of the clear atmosphere with a given distribution of relative humidity,
- pure radiative equilibrium of the clear atmosphere with a given distribution of absolute humidity, and
- radiative-convective equilibrium of the atmosphere with a given distribution of relative humidity.

Pure radiative equilibria exhibit a troposphere and a stratosphere, with the tropopause at a

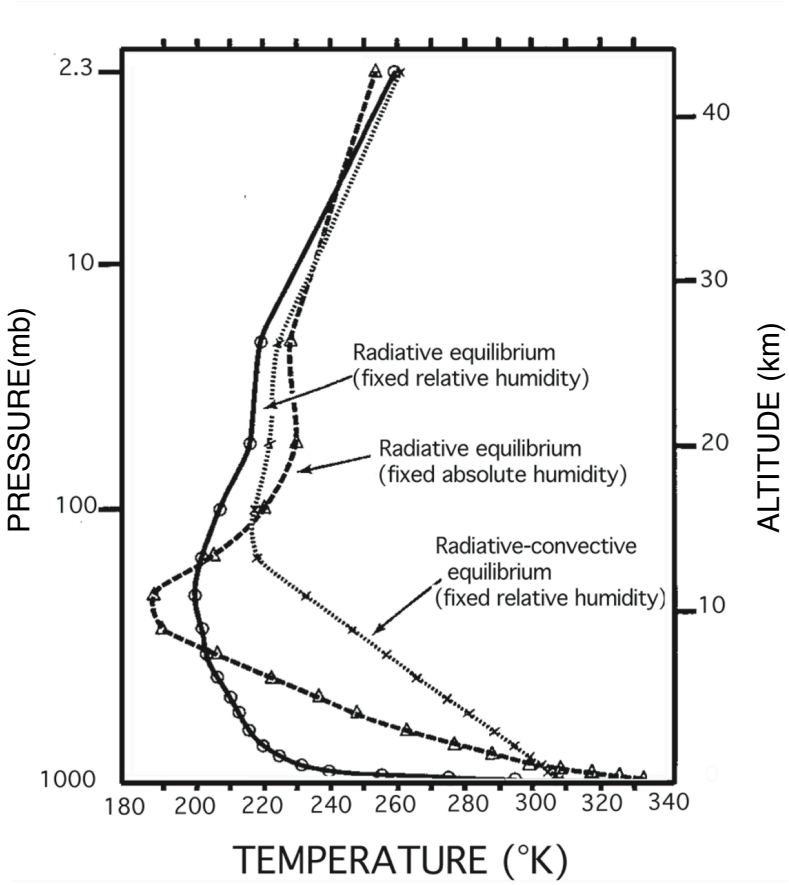


Figure 8.6: Solid line, radiative equilibrium of the clear atmosphere with a given distribution of relative humidity; dashed line, radiative equilibrium of the clear atmosphere with a given distribution of absolute humidity; dotted line, radiative convective equilibrium of the atmosphere with a given distribution of relative humidity. From Manabe and Wetherald (1967).

fairly realistic height. An unrealistic aspect of both of the radiative equilibria is that the lower troposphere is convectively unstable, even for dry convection. In the RCE calculations, this instability is assumed to be removed by convection. RCE with fixed relative humidity is amazingly realistic, considering that the model ignores all large-scale dynamical processes.

Manabe et al. (1965) used a more general version of his "moist convective adjustment" concept as a parameterization of cumulus convection in the early versions of the GFDL (Geophysical Fluid Dynamics Laboratory) global circulation model (GCM). The model predicted water vapor, of course. The parameterization was activated when the simulated relative humidity reached 100 percent and the simulated lapse rate exceeded the simulated moist adiabatic lapse rate. The adjustment removed the instability by re-setting the actual lapse rate to the moist adiabatic lapse rate, and mixing water vapor between the adjusted layers. For the adjusted state, any water in excess of saturation was assumed to precipitate

out instantaneously. Both energy and total water were conserved by the moist convective adjustment. Note that the adjustment was used even though the simulated sounding was continually evolving as the model marched forward in time. This means that the adjustment was not to an equilibrium. Further discussion is given later in this chapter.

Fels (1985) reported the results of similar RCE calculations on the sphere, but his model included a representation of photochemistry, and he emphasized the structure of the stratosphere and mesosphere, which are often referred to as the "middle atmosphere." Because the middle atmosphere has a very stable stratification, convection is not active there, and it is not clear that the representation of convection in the troposphere of his model had any significant influence on the results for the middle atmosphere, which are shown in Fig. 8.7. The observed state of the summer stratosphere (see panel b of Fig. 8.7) resembles that predicted by the model, but the winter polar stratosphere simulated by the model is much too cold. Also, the mesosphere of the real world is warm near the winter pole and cold near the summer pole, while the model predicts just the opposite. The differences between the observations and the model results can be attributed to the effects of atmospheric dynamics, which are neglected in the model. Obviously, the motions make quite a difference in the winter middle atmosphere, where they must transport energy poleward in order to account for the differences between the observations and the results of the radiative-convective model. This will be discussed in detail later. It appears that atmospheric motions have little effect on the thermal structure of the stratosphere in summer, however. This is a clue that the summer stratosphere is close to a state of radiative equilibrium.

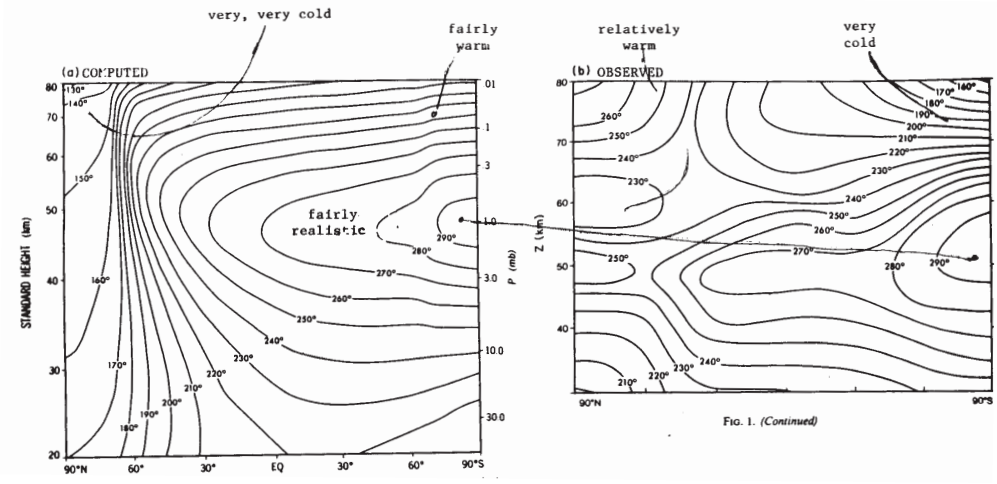


Figure 8.7: (a) Zonal mean temperatures for 15 January calculated by using a time-marched radiative-convective-photochemical model. (b) Zonal mean temperatures for January. From Fels (1985).

8.7 Some comments on radiative-convective equilibrium

As originally conceived, RCE means coupling radiation and convection (optionally including the effects of turbulence and radiatively active clouds), but without large-scale circulations. The RCE studies discussed above demonstrate that the vertical structure of the atmosphere is strongly controlled by radiation and convection, at least in the tropics. Today, RCE studies are being used to seek basic understanding in an idealized framework. There is no question that RCE is an interesting and useful concept, but some caveats are needed:

First of all, large-scale motions do occur in the real world, so RCE can only be simulated as an idealization of reality. There is no observational basis to say what RCE looks like. In the real tropical atmosphere, large-scale rising motion can enhance convection, increasing the precipitation rate beyond that expected in RCE, and large-scale sinking motion can suppress convection, cutting off precipitation entirely. We can compare our RCE simulations with observations from places in the real world, e.g., the tropical Western Pacific ocean, and the comparison may be informative, but there is no a priori reason to expect agreement.

Second, a simulation of RCE must be referred to a specific "domain" or region. The horizontal width of the domain matters. Studies of RCE typically assume a medium-sized domain, e.g., 100 km on a side, but in many cases this assumption is not made explicit. If the domain is very small, e.g., one square meter, then it is not physically possible for cumulus convection to occur inside. If the domain is very large, e.g., the whole global domain, then we are really talking about the global circulation of the atmosphere, including its many and varied weather systems.

There are now many examples of RCE calculations based on numerical models with grid spacings fine enough to explicitly simulate the growth and decay of individual clouds. For example, using such a model, Khairoutdinov and Emanuel (2013) showed that, when rotation is included, simulations of RCE over a sufficiently warm prescribed SST, and with a sufficiently large domain, give rise to the formation of tropical cyclones. At what point do simulations of RCE become simulations of weather systems?

8.8 Mass fluxes

8.8.1 Origin story

The mass-flux concept was created for use in a parameterization of cumulus convection by Arakawa (1969). It was used in an observational study by Yanai et al. (1973). It was used in a theory of shallow cumulus convection by Betts (1973). It was explored in more detail by Arakawa and Schubert (1974), and has been very widely used since their paper appeared.

8.8.2 The simplest mass flux model

The simplest mass flux model consists of just two "subdomains" within a grid cell: The first subdomain consists of an "updraft" rising *relative to the mean vertical motion*, and the second consists of a "downdraft" sinking *relative to the mean vertical motion* (Randall, 1987). The area-averaged vertical motion, $\overline{w}$, can be written as

$$\overline{w} = \sigma_{\rm up} w_{\rm up} + (1 - \sigma_{\rm up}) \, w_{\rm dn} \ , \tag{8.62}$$

where $\sigma_{\rm up}$ is the fractional area covered by the updraft, and $w_{\rm up}$ and $w_{\rm dn}$ are the vertical velocities of the updraft and downdraft, respectively. By rearranging (8.62), we can show that

$$w'_{\rm up} \equiv w_{\rm up} - \overline{w} = (1 - \sigma_{\rm up}) \, (w_{\rm up} - w_{\rm dn}) \ , \tag{8.63}$$

and

$$w'_{\rm dn} \equiv w_{\rm dn} - \overline{w} = -\sigma_{\rm up} \, (w_{\rm up} - w_{\rm dn}) \ . \tag{8.64}$$

Now consider an arbitrary scalar, A. Let $A_{\rm up}$ and $A_{\rm dn}$ be the values of A associated with the updrafts and downdrafts, respectively, so that

$$\overline{A} = \sigma_{\rm up} A_{\rm up} + (1 - \sigma_{\rm up}) \, A_{\rm dn} \ . \tag{8.65}$$

The upward flux of A is then given by

$$F_A = \rho \left[\sigma_{\rm up} \, (w_{\rm up} - \overline{w}) \left(A_{\rm up} - \overline{A} \right) + (1 - \sigma_{\rm up}) \, (w_{\rm dn} - \overline{w}) \left(A_{\rm dn} - \overline{A} \right) \right] \ . \tag{8.66}$$

Eq. (8.66) is just a straightforward transcription of the Reynolds-averaging formula $F_A = \rho \overline{w' A'}$. An interpretation of (8.66) is that an updraft mass flux given by $M_{\rm up} \equiv \rho \, \sigma_{\rm up} \, (w_{\rm up} - \overline{w})$ carries $A_{\rm up} - \overline{A}$, while a downdraft mass flux given by $M_{\rm dn} \equiv \rho \, (1 - \sigma_{\rm up}) \, (w_{\rm dn} - \overline{w})$ carries $A_{\rm dn} - \overline{A}$. Notice that each mass flux includes an areal weight.

Using (8.63), (8.64), and similar relations for $A_{\rm up}-\overline{A}$ and $A_{\rm dn}-\overline{A}$, we can rewrite (8.66) as

$$\begin{aligned} F_A =& \rho \left\{ \sigma_{\rm up} \left[(1-\sigma_{\rm up})^2 (w_{\rm up}-w_{\rm dn}) (A_{\rm up}-A_{\rm dn}) \right] \right. \\ & \left. + (1-\sigma_{\rm up}) \left[\sigma_{\rm up}^2 (w_{\rm up}-w_{\rm dn}) (A_{\rm up}-A_{\rm dn}) \right] \right\} \\ =& \rho \sigma_{\rm up} (1-\sigma_{\rm up}) (w_{\rm up}-w_{\rm dn}) (A_{\rm up}-A_{\rm dn}) \left[(1-\sigma_{\rm up}) + \sigma_{\rm up} \right] \\ =& \rho \sigma_{\rm up} (1-\sigma_{\rm up}) (w_{\rm up}-w_{\rm dn}) (A_{\rm up}-A_{\rm dn}) \ . \end{aligned} \tag{8.67}$$

The quantity $\rho \sigma_{\rm up} (1-\sigma_{\rm up}) (w_{\rm up}-w_{\rm dn})$ will appear in the flux of *any* scalar, so it is convenient to give it a name. We call it

$$M_c \equiv \rho \sigma_{\rm up} (1-\sigma_{\rm up}) (w_{\rm up}-w_{\rm dn}) \ , \tag{8.68}$$

In the important special case $\sigma_{\rm up} \ll 1$, we get

$$A_{\rm dn} \cong \overline{A} \ \text{ for } \ \sigma_{\rm up} \ll 1 \quad \text{provided that } \ A_{\rm up}-A_{\rm dn} \text{ is less than or comparable to } \overline{A} \ . \tag{8.69}$$

This means that the sinking air has almost the same value of A as the domain-average of A, simply because the sinking air occupies almost the entire area. Note, however that $w_{\rm up}-w_{\rm dn} \gg \overline{w}$ is very possible, because for $\sigma_{\rm up} \ll 1$ the updrafts may be much stronger than $\overline{w}$. Therefore *we cannot assume* that $w_{\rm dn} \cong \overline{w}$ for $\sigma_{\rm up} \ll 1$. We can, however, write

$$w_{\rm up}-w_{\rm dn} \cong w_{\rm up} \quad \text{for} \quad \sigma_{\rm up} \ll 1 \ . \tag{8.70}$$

Then we can simplify (8.67) to

$$F_A \cong \rho \sigma_{\rm up} w_{\rm up} (A_{\rm up}-A_{\rm dn}) \quad \text{for} \quad \sigma_{\rm up} \ll 1 \ . \tag{8.71}$$

This shows that

$$M_c \cong \rho \sigma_{\rm up} w_{\rm up} \quad \text{for} \quad \sigma_{\rm up} \ll 1 \; . \tag{8.72}$$

We can also write

$$M_c \cong M_{\rm up} \quad \text{for} \quad \sigma_{\rm up} \ll 1 \; . \tag{8.73}$$

Using similar methods, we can show that the second and third moments of w are given by

$$\overline{w'^2} = \sigma_{\rm up} \left(1 - \sigma_{\rm up}\right) \left(w_{\rm up} - w_{\rm dn}\right)^2 \tag{8.74}$$

and

$$\overline{w'^3} = \sigma_{\rm up} \left(1 - \sigma_{\rm up}\right) \left(1 - 2\sigma_{\rm up}\right) \left(w_{\rm up} - w_{\rm dn}\right)^3 \; , \tag{8.75}$$

respectively. Suppose that we know $\overline{w}$, $\overline{w'^2}$, and $\overline{w'^3}$. Randall et al. (1992) showed that

$$\sigma_{\rm up} = \frac{1}{2} - \frac{{\rm Sk}\left(w\right)}{2\sqrt{4 + {\rm Sk}\left(w\right)^2}} \; , \tag{8.76}$$

and

$$M_c = \frac{\rho \overline{w'^2}}{\sqrt{4 + {\rm Sk}\left(w\right)^2}} \; , \tag{8.77}$$

where

$${\rm Sk}\left(w\right) \equiv \overline{w'^3} \; / \left(\overline{w'^2}\right)^{3/2} \tag{8.78}$$

is the skewness of w. This means that $\overline{w}$, $\overline{w'^2}$, and $\overline{w'^3}$ can be used to determine $\sigma_{\rm up}$ and M_c.

8.8.3 Generalization

The mass flux method can be generalized by allowing many subdomains, each with its own mass flux. Using Reynolds averaging, we can write

$$\begin{aligned} F_h = \rho\overline{w'h'} =& \rho \sum_{j=1}^{J} \left[\sigma_j \left(w_j - \overline{w}\right)\left(h_j - \overline{h}\right)\right] \\ =& \sum_{j=1}^{J} \left[M_j \left(h_j - \overline{h}\right)\right] \ , \end{aligned} \tag{8.79}$$

where σ_j is a fractional area [2] such that

$$\sum_{j=1}^{N} \sigma_j = 1 \ , \tag{8.80}$$

h is an arbitrary scalar, and

$$M_j \equiv \rho\sigma_j \left(w_j - \overline{w}\right) \tag{8.81}$$

is a mass flux. M_j is usually called the "convective" mass flux in the context of cumulus parameterization, but mass fluxes can be used to describe all vertical fluxes, including those due to mechanically driven turbulence. Eq. (8.79) demonstrates that mass fluxes arise purely from Reynolds averaging, and that they involve no approximation in the limit $J \to \infty$. In principle we could define a mass flux for each square millimeter of a grid cell. This is not a good idea because so much detail is (presumably) unnecessary, and because it would be very impractical to deal with such a large number of mass fluxes. In practice, of course, J must be finite.

The second line of (8.79) can be used to calculate $\rho\overline{w'h'}$ if the M_j and h_j can be determined somehow. Historically (e.g., Arakawa and Schubert, 1974), plume models have been used to calculate the h_j, and "mass-flux closures" have been used to determine the M_j. As a result, there is a tendency to associate the mass-flux method with plume models, but Eq. (8.79) shows that mass fluxes can be defined without any reference to plumes.

[2] Not to be confused with the growth rate discussed earlier in this chapter

8.8.4 EDMF

Are the subdomains of the mass-flux model featureless and dynamically inactive? Of course not. Many studies (Soares et al., 2004; Neggers, 2006; Köhler, 2006; Siebesma et al., 2007; Witek et al., 2011; Han et al., 2016; Tan et al., 2018; Teixeira et al., 2023) have recognized this by including diffusive fluxes that occur inside some or all of the subdomains, usually in the "environment" of the updrafts. This approach is called "Eddy Diffusion with Mass Fluxes," or EDMF .

8.9 Large-scale budgets



Figure 8.8: Prof. Akio Arakawa, cruising along in mid-lecture. He made several major contributions to atmospheric science, including a theory of cumulus convection and a widely used numerical method for solving the equation of motion.

The effects of convection on the global circulation were analyzed bu Arakawa and Schubert (1974, hereafter AS; see Fig. 8.8). Their theory was developed partly to advance understanding, and partly for use in simulating heating and drying rates with numerical models. Let an overbar denote an average over an area large enough to contain a good sample of cumulus clouds, but small enough so that it spans only part of a larger-scale weather system (Figure 8.9). As pointed out by AS, the existence of such an area is a fundamental assumption of their theory.

The area-averaged budget equations for mass, dry static energy, water vapor mixing ratio, and liquid water mixing ratio can be written as:

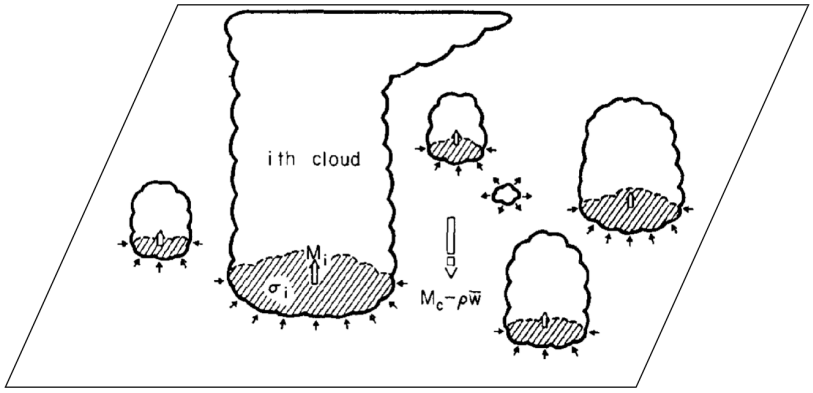


Figure 8.9: A "box" large enough to contain a good sample of clouds, but small compared to a large-scale weather system. From Arakawa and Schubert (1974).

$$0 = -\nabla\cdot(\rho\overline{\mathbf{v}}) - \frac{\partial(\rho\overline{w})}{\partial z}\,, \tag{8.82}$$

$$\rho\frac{\partial\overline{s}}{\partial t} = -\rho(\overline{\mathbf{v}}\cdot\nabla)\overline{s} - \rho\overline{w}\frac{\partial\overline{s}}{\partial z} + \overline{Q_R} + \rho L\overline{C} - \frac{\partial F_s}{\partial z}\,, \tag{8.83}$$

$$\rho\frac{\partial\overline{q_v}}{\partial t} = -\rho(\overline{\mathbf{v}}\cdot\nabla)\overline{q_v} - \rho\overline{w}\frac{\partial\overline{q_v}}{\partial z} - \rho\overline{C} - \frac{\partial F_{q_v}}{\partial z}\,, \tag{8.84}$$

$$\rho\frac{\partial\overline{l}}{\partial t} = -\rho(\overline{\mathbf{v}}\cdot\nabla)\overline{l} - \rho\overline{w}\frac{\partial\overline{l}}{\partial z} + \rho\overline{C} - \frac{\partial F_l}{\partial z} - \overline{\chi}\,, \tag{8.85}$$

$$\rho\frac{\partial\overline{\mathbf{v}}}{\partial t} = -\rho(\overline{\mathbf{v}}\cdot\nabla)\overline{\mathbf{v}} - \rho\overline{w}\frac{\partial\overline{\mathbf{v}}}{\partial z} - \nabla\overline{p} - \rho f\mathbf{k}\times\overline{\mathbf{v}} - \frac{\partial\mathbf{F_v}}{\partial z}\,. \tag{8.86}$$

Here ρ is the density of the air, which is presumed to be quasi-constant at each height; s is the dry static energy; q_v is the water vapor mixing ratio; w is the vertical velocity; $\overline{\mathbf{v}}$ is the horizontal velocity vector; and χ is the rate at which liquid water is converted into precipitation, which then falls out and so acts as a sink of l. The vertical "eddy fluxes,"

$F_s = \rho\overline{ws} - \rho\overline{w}\,\overline{s}$, $F_{q_v} = \rho\overline{wq_v} - \rho\overline{w}\,\overline{q_v}$, $F_l = \rho\overline{wl} - \rho\overline{w}\,\overline{l}$, and $\mathbf{F_v} \equiv \rho\overline{w\mathbf{v}} - \rho\overline{w}\,\overline{\mathbf{v}}$ can in principle represent quite a variety of physical processes, but here we assume for simplicity that above the boundary layer the fluxes are due only to the convective updrafts and downdrafts associated with cumulus convection.

8.10 Updrafts and environment

AS used a very simple cumulus cloud model to formulate the eddy fluxes that appear in (8.83) - (8.85) in terms of a convective mass flux and the differences between the in-cloud and environmental soundings. The cloud model was also used to formulate the net condensation rate, C, per unit mass flux. AS allowed the possibility that clouds of many different "types" coexist; here a cloud type can be roughly interpreted as a cloud size category. We now briefly explain the AS parameterization, in a simplified way.

As a first step, we divide the domain into an arbitrary number of subdomains, each having a characteristic fractional area , a characteristic vertical velocity , and corresponding characteristic values of the moist static energy, dry static energy, water vapor mixing ratio, and all of the other variables of interest. Some of the subdomains represent cloudy updrafts or downdrafts, while others could represent mesoscale subdomains or the broad "environment" of the clouds. The environment is expected to have small but non-zero vertical velocity, usually downward, and it is expected to cover a large fraction of the total area.

We assume for simplicity that the effects of mesoscale organization and convective-scale downdrafts can be neglected, so that the cloudy layer consists of concentrated convective updrafts of various sizes and intensities, embedded in a broad uniform environment. We use a superscript tilda to denote an environmental value, and a subscript c to denote the collective properties of the cloudy updrafts. Then (8.80) simplifies to

$$\sigma_c + \widetilde{\sigma} = 1 \,, \tag{8.87}$$

where

$$\sigma_c \equiv \sum_{\text{all clouds}} \sigma_j \tag{8.88}$$

is the total fractional area covered by all of the convective updrafts, and $\widetilde{\sigma}$ is the fractional area of the environment. Similarly,

$$\sigma_c w_c + \widetilde{\sigma}\widetilde{w} = \overline{w} \,, \tag{8.89}$$

and

$$\sigma_c h_c + \widetilde{\sigma}\widetilde{h} = \overline{h} \,, \tag{8.90}$$

where we define

$$w_c \equiv \frac{1}{\sigma_c} \sum_{\text{all clouds}} \sigma_j w_j \,, \tag{8.91}$$

and

$$h_c \equiv \frac{1}{\sigma_c} \sum_{\text{all clouds}} \sigma_j h_j \,. \tag{8.92}$$

8.11 Why the updraft fractional area is small

It is observed that

$$\sigma_c \ll 1 \text{ and } \widetilde{\sigma} \cong 1 \,. \tag{8.93}$$

This means that the cumulus updrafts occupy only a very small fraction of the area. A simple explanation for this important fact was given by Jakob (Bjerknes, 1938, ; Fig. 8.10). Suppose that at a certain time the temperature is horizontally uniform, with lapse rate Γ. Consider temperature changes due to adiabatic vertical motion only. As discussed earlier, in a cloudy subregion, the temperature satisfies

$$\frac{\partial T_c}{\partial t} = w_c \left(\Gamma - \Gamma_m \right) \,, \tag{8.94}$$

while in a clear subregion,

$$\frac{\partial \widetilde{T}}{\partial t} = \widetilde{w}\left(\Gamma - \Gamma_d\right) \ . \tag{8.95}$$

If the sounding is conditionally unstable, then

$$\Gamma_d > \Gamma > \Gamma_m \ . \tag{8.96}$$

Suppose that the cloudy air is rising and the environmental air is sinking, so that $w_c > 0$ and $\widetilde{w} < 0$. Comparing (8.96) with (8.94) and (8.95), we see that when the sounding is conditionally unstable both T_c and $\widetilde{T}$ will tend to increase with time. The buoyancy of the updraft is proportional to $T_c - \widetilde{T}$. An increase in T_c will tend to increase the buoyancy of the cloudy air, but an increase in $\widetilde{T}$ will tend to decrease the buoyancy. Which effect wins out? The answer depends on the value of σ_c.

Using the definition of the mean vertical motion, we find that

$$w_c = \overline{w} + \left(1 - \sigma_c\right)\left(w_c - \widetilde{w}\right) \ , \tag{8.97}$$

$$\widetilde{w} = \overline{w} - \sigma_c \left(w_c - \widetilde{w}\right) \ . \tag{8.98}$$

Subtracting (8.95) from (8.94), and substituting from (8.97) and (8.98), we find that

$$\begin{aligned} \frac{\partial}{\partial t}\left(T_c - \widetilde{T}\right) =& w_c\left(\Gamma - \Gamma_m\right) - \widetilde{w}\left(\Gamma - \Gamma_d\right) \\ =& \overline{w}\left(\Gamma_d - \Gamma_m\right) + \left(w_c - \widetilde{w}\right)\left[\left(1 - \sigma_c\right)\left(\Gamma - \Gamma_m\right) + \sigma_c\left(\Gamma - \Gamma_d\right)\right] \ . \end{aligned} \tag{8.99}$$

Consider the quantity $\left[\left(1 - \sigma_c\right)\left(\Gamma - \Gamma_m\right) + \sigma_c\left(\Gamma - \Gamma_d\right)\right]$, which multiplies $\left(w_c - \widetilde{w}\right)$ on the second line of (8.99). Inspection shows that $\left[\left(1 - \sigma_c\right)\left(\Gamma - \Gamma_m\right) + \sigma_c\left(\Gamma - \Gamma_d\right)\right]$ is maximized for $\sigma_c \to 0$. The physical interpretation is simple. With a conditionally unstable sounding, saturated rising motion is aided by positive buoyancy created through condensation, while unsaturated sinking motion must fight against the dry-stable stratification. The rate of temperature increase in the updraft is proportional to the updraft speed, while the rate of



Figure 8.10: Jakob Bjerknes (1897-1975) was born in Sweden. Both Jakob and his father, Vilhelm Bjerknes, were highly accomplished meteorologists. Jakob founded the UCLA Department of Meteorology (now the Department of Atmospheric and Oceanic Sciences. From `https://en.wikipedia.org/wiki/Jacob_Bjerknes`.

temperature increase in the downdraft is proportional to the downdraft speed. Therefore, *convection is favored by rapid rising motion in the cloudy region, and slow sinking motion in the clear region, both of which can be achieved, for a given value of* $(w_c - \widetilde{w})$*, by making the updraft narrow and the downdraft broad.* This is why $\sigma_c \ll 1$.

With this simple idea, Bjerknes (1938) explained the observed smallness of σ_c, which is important because it means that even in the regions where deep convection is most active there is a lot of clear sky where no phase changes occur, and the mean relative humidity in such regions can be well below 100%. Under some conditions, convective clouds are capable of "tunneling through" deep layers of (mostly) unsaturated air. Cloud processes would be relatively simple if they involved only uniform cloudiness over large regions. Nature's preference for narrow saturated updrafts in clear environments makes the interaction of moist convection with the global circulation a more subtle (and interesting) problem than it would otherwise be.

8.12 How small a small updraft area simplifies things

In view of (8.93), we can write (8.90) as

$$\widetilde{h} \cong \overline{h} \, . \tag{8.100}$$

It is *not* true, however, that $\widetilde{w} \cong \overline{w}$, because the cumulus updrafts are typically several orders of magnitude stronger than the large-scale vertical motions, i.e.,

$$w_j \gg \overline{w} \, . \tag{8.101}$$

Similarly, it is not true in general that $\widetilde{l} \cong \overline{l}$, because there may be no liquid water at all in the environment of the convective clouds. Using (8.101), we can approximate M_j by

$$M_j \cong \rho \sigma_j w_j \, . \tag{8.102}$$

The approximations (8.100) - (8.102) are expected to fail with small domains (as in high-resolution models), for which $\overline{w}$ can easily be comparable to or even larger than w_j.

8.13 Large-scale budgets again

From this point, we simplify the discussion by considering only one type of convective cloud, whose properties are denoted by subscript c. We write

$$F_s \cong M_c \left(s_c - \overline{s} \right) \, , \tag{8.103}$$

$$F_{q_v} \cong M_c \left[\left(q_v \right)_c - \overline{q_v} \right] \, , \tag{8.104}$$

$$F_l \cong M_c \left(l_c - \widetilde{l} \right) \, . \tag{8.105}$$

Here s_c, $(q_v)_c$, and l_c are the in-cloud dry static energy, water vapor mixing ratio, and liquid water mixing ratio, respectively, and

$$M_c \equiv \rho \sigma_c w_c \ . \tag{8.106}$$

In (8.105), we allow the possibility of liquid water in the environment of the cumulus clouds, but we do not assume that $\widetilde{l} \cong \overline{l}$, because it is quite possible that the only liquid present is in the convective updrafts. It is also possible, however, that the environment is filled or partially filled with stratiform cloud, perhaps created by earlier detrainment from the convective clouds. Finally we need

$$\overline{s} = (1-\sigma_c)\,\widetilde{s} + \sigma_c s_c \ , \tag{8.107}$$

$$\overline{q_v} = (1-\sigma_c)\,\widetilde{q}_v + \sigma_c (q_v)_c \ , \tag{8.108}$$

$$\overline{l} = (1-\sigma_c)\,\widetilde{l} + \sigma_c l_c \ , \tag{8.109}$$

$$\overline{C} = (1-\sigma_c)\,\widetilde{C} + \sigma_c C_c \ , \tag{8.110}$$

$$\overline{\chi} = (1-\sigma_c)\,\widetilde{\chi} + \sigma_c \chi_c \ . \tag{8.111}$$

With these results, we can now rewrite (8.83) - (8.85) as

$$\rho \frac{\partial \overline{s}}{\partial t} = -\rho \left(\overline{\mathbf{v}} \cdot \nabla\right) \overline{s} - \rho \overline{w} \frac{\partial \overline{s}}{\partial z} + Q_R + \rho L \left(\widetilde{C} + \sigma_c C_c\right) - \frac{\partial}{\partial z} \left[M_c \left(s_c - \overline{s}\right)\right] \ , \tag{8.112}$$

$$\rho\frac{\partial\overline{Q}_v}{\partial t}=-\rho\left(\overline{\mathbf{v}}\cdot\nabla\right)\overline{Q}_v-\rho\overline{w}\frac{\partial\overline{Q}_v}{\partial z}-\rho\left(\widetilde{C}+\sigma_c C_c\right)-\frac{\partial}{\partial z}\left\{M_c\left[\left(q_v\right)_c-\overline{Q}_v\right]\right\}\ , \tag{8.113}$$

$$\rho\frac{\partial\overline{l}}{\partial t}=-\rho\left(\overline{\mathbf{v}}\cdot\nabla\right)\overline{l}-\rho\overline{w}\frac{\partial\overline{l}}{\partial z}+\rho\left(\widetilde{C}+\sigma_c C_c\right)-\frac{\partial}{\partial z}\left[M_c\left(l_c-\widetilde{l}\right)\right]-\left(\widetilde{\chi}+\sigma_c\chi_c\right)\ . \tag{8.114}$$

The convective condensation rate, C, appears in all three of these equations, as would be expected.

8.14 A simple cumulus cloud model

To go further, we need to know the soundings inside the updrafts. For this purpose, a simple cumulus cloud model is required. We assume that all cumulus clouds originate from the boundary layer, carrying the mixed-layer properties upward. The mass flux changes with height according to

$$\frac{\partial M_c\left(z\right)}{\partial z}=E\left(z\right)-D\left(z\right)\ . \tag{8.115}$$

Here E is the entrainment rate, and D is the detrainment rate. The in-cloud profile of moist static energy, $h_c\left(z\right)$, is governed by

$$\begin{aligned}\frac{\partial}{\partial z}\left[M_c\left(z\right)h_c\left(z\right)\right]=&E\left(z\right)\widetilde{h}\left(z\right)-D\left(z\right)h_c\left(z\right)\\ \cong&E\left(z\right)\overline{h}\left(z\right)-D\left(z\right)h_c\left(z\right)\ .\end{aligned} \tag{8.116}$$

If we replace h by 1, then (8.116) reduces to eq:cloud model continuity. There are no source or sink terms in (8.116) because the moist static energy is unaffected by phase changes and/or precipitation processes, and we neglect radiative effects. By combining (8.115) and (8.116), we can show that

$$\frac{\partial h_c\left(z\right)}{\partial z}=\frac{E\left(z\right)}{M_c}\left[\overline{h}\left(z\right)-h_c\left(z\right)\right]\ . \tag{8.117}$$

This shows that h_c is affected by entrainment, which dilutes the cloud with environmental air, but not by detrainment, which expels the cloud's own moist static energy at each level.

Similarly, we can write

$$\frac{\partial}{\partial z}\left(M_c s_c\right) = E\overline{s} - Ds_c + \rho\sigma_c LC_c \ . \tag{8.118}$$

$$\frac{\partial}{\partial z}\left[M_c (q_v)_c\right] = E\overline{q_v} - D(q_v)_c - \rho\sigma C_c \ . \tag{8.119}$$

$$\frac{\partial}{\partial z}\left(M_c l_c\right) = E\widetilde{l} - Dl_c + \rho\sigma_c C_c - \chi_c \ . \tag{8.120}$$

A simple microphysical model is needed to determine χ_c, i.e. to determine how much of the condensed water is converted to precipitation, and the fate of the precipitation. The role of convectively generated precipitation, which drives convective downdrafts and moistens the lower troposphere by evaporating as it falls, is actually an important issue, but it will not be discussed here.

8.15 Compensating subsidence

By using (8.118)-(8.120), the large-scale budget equations can be rewritten in a very interesting way, as follows. First, consider the dry static energy. We write

$$\frac{\partial}{\partial z}\left[M_c\left(s_c - \overline{s}\right)\right] = \frac{\partial}{\partial z}\left(M_c s_c\right) - M_c\frac{\partial \overline{s}}{\partial z} - \overline{s}\frac{\partial M_c}{\partial z} \ . \tag{8.121}$$

Substitute from (8.115) and (8.118) into (8.121), to obtain

$$\begin{aligned}\frac{\partial}{\partial z}\left[M_c\left(s_c - \overline{s}\right)\right] &= \left(E\overline{s} - Ds_c + \rho L\sigma_c C_c\right) - M_c\frac{\partial \overline{s}}{\partial z} - \overline{s}\left(E - D\right) \\ &= -M_c\frac{\partial \overline{s}}{\partial z} - D\left(s_c - \overline{s}\right) + \rho L\sigma_c C_c \ .\end{aligned} \tag{8.122}$$

This allows us to rewrite (8.112) as

$$\rho \frac{\partial \overline{s}}{\partial t} = -\rho \left(\overline{\mathbf{v}} \cdot \nabla \right) \overline{s} - \rho \overline{w} \frac{\partial \overline{s}}{\partial z} + \overline{Q_R} + \rho L \widetilde{C} + M_c \frac{\partial \overline{s}}{\partial z} + D \left(s_c - \overline{s} \right) \ . \tag{8.123}$$

The last two terms on the right-hand side of (8.123) represent the cumulus effects, and the first of these in particular is quite interesting. It "looks like" an advection term. It represents the warming of the environment due to the downward advection of air from above, with higher dry static energies, by the environmental sinking motion that compensates for the rising motion in the cloudy updraft. The environmental sinking motion is often called "compensating subsidence," because it compensates for the concentrated rising motion in the saturated updrafts. Up moist, down dry.

The role of compensating subsidence can be seen more explicitly by combining the two "vertical advection" terms of (8.113), and using the definition of $\overline{w}$, to obtain

$$\boxed{\rho \frac{\partial \overline{s}}{\partial t} = -\rho \left(\overline{\mathbf{v}} \cdot \nabla \right) \overline{s} - \widetilde{M} \frac{\partial \overline{s}}{\partial z} + \overline{Q_R} + \rho L \widetilde{C} + D \left(s_c - \overline{s} \right) .} \tag{8.124}$$

where

$$\widetilde{M} \equiv \rho \overline{w} - M_c \ . \tag{8.125}$$

is the environmental mass flux. Why does $\widetilde{M}$ appear in (8.124)? The reason is that the environmental subsidence is modifying $\widetilde{s}$, but $\widetilde{s} \cong \overline{s}$. The last term on the right-hand side of (8.124) represents the effects of detrainment. You may be surprised to see that the cumulus condensation rate does not appear in (8.123) or (8.124). The reason is that condensation inside the updraft cannot directly warm the environment. Since almost the entire area is the environment, condensation in the updrafts does not, to any significant degree, directly affect the area-averaged dry static energy. Instead, the effects of condensation are felt indirectly, through the compensating subsidence term. The physical role of condensation, then, is to maintain the buoyancy of the convective updraft that drives the compensating subsidence, which in turn warms the environment. This is how condensation warms *indirectly*. Note that the vertical profile of the warming due to compensating subsidence is expected to be different from the vertical profile of the convective condensation rate.

In a similar way, the water vapor budget equation can be rewritten as

$$\boxed{\rho \frac{\partial \overline{q_v}}{\partial t} = -\rho \left(\overline{\mathbf{v}} \cdot \nabla\right) \overline{q_v} - \widetilde{M} \frac{\partial \overline{q_v}}{\partial z} - \rho \widetilde{C} + D\left[\left(q_v\right)_c - \overline{q_v}\right] .} \tag{8.126}$$

Eq. (8.126) describes the convective drying in terms of compensating subsidence, which brings down drier air from aloft. Detrainment of water vapor, liquid, and ice from the convective clouds can tend to moisten the environment, although this is not entirely straightforward, for reasons discussed in Chapter 10.

Finally, similar methods can be used to write the liquid water budget equation as

$$\boxed{\rho \frac{\partial \bar{l}}{\partial t} = -\rho \left(\overline{\mathbf{v}} \cdot \nabla\right) \bar{l} - \rho \overline{w} \frac{\partial \bar{l}}{\partial z} + \rho \overline{C} + M_c \frac{\partial \widetilde{l}}{\partial z} + D\left(l_c - \widetilde{l}\right) - \left[\left(1 - \sigma_c\right) \widetilde{\chi} + \sigma_c \chi_c\right] .} \tag{8.127}$$

In (8.127), we cannot combine the two "vertical advection" terms, because one of them involves $\partial \bar{l} / \partial z$, while the other involves $\partial \widetilde{l} / \partial z$. Detrained liquid (or ice) can persist in the form of stratiform "anvil" and cirrus clouds.

Within the limits of applicability of the assumptions used above, most importantly the assumption that $\sigma_c \ll 1$, (8.124) and (8.126) - (8.127) are equivalent to (8.83) - (8.85). We will use (8.124) and (8.126) in Chapter 10.

8.16 The vertical distribution of the mass flux

As discussed earlier, the buoyancy of the cloudy air at height z, is approximately given by

$$B\left(z\right) \cong T_c - \overline{T} \sim \frac{1}{c_p} \left[h_c\left(z\right) - \overline{h_{\text{sat}}}\left(z\right)\right] , \tag{8.128}$$

where h_{sat} is the saturation moist static energy. Because more rapidly entraining clouds lose their buoyancy at lower levels, in effect the cloud types differ according to their cloud-top heights, for a given sounding. The cloud top occurs at level $\hat{z}$, where

$$B\left(\hat{z}\right) \cong 0 , \tag{8.129}$$

This equation can be used to find $\hat{z}$, after the in-cloud sounding has been determined using Eqs. (8.115) - (8.120). In order to find the in-cloud sounding, we need to know the entrainment rate.

AS assumed that

$$E = \lambda M_c \ , \tag{8.130}$$

where λ, which is called the *fractional* entrainment rate. λ has the units of inverse length, is assumed to be a constant (with height) for each cloud type. A spectrum of cloudscan be defined by allowing λ to take values in the range $0 < \lambda < \infty$. Larger values of λ mean stronger entrainment, $\lambda = 0$ means no entrainment, and $\lambda < 0$ has no physical meaning and so is not allowed. For a given sounding, clouds with smaller values of λ (weaker fractional entrainment) will have higher tops.

For simplicity, AS assumed that detrainment occurs only at cloud top. This means that the mass flux jumps discontinuously to zero at cloud top. Below the cloud-top, we have entrainment but no detrainment, so that (8.115) reduces to

$$E = \lambda M_c \ , \tag{8.131}$$

Combining (8.130) and (8.131), and using the assumption that λ is constant with height for each cloud type, we find that

$$M_c \left(z, \lambda \right) = M_B(\lambda) \exp \left(\lambda z \right) \ , \tag{8.132}$$

where $M_B(\lambda)$ is the "cloud-base mass flux distribution function." We define a normalized mass flux, denoted by $\eta \left(\lambda, z \right)$; the normalization is in terms of the cloud-base mass flux:

$$M_c \left(z, \lambda \right) \equiv M_B \left(\lambda \right) \eta \left(\lambda, z \right) \ , \tag{8.133}$$

By definition, $\eta \left(\lambda, z_B \right) = 1$, where z_B is the cloud-base height.

For each λ in the range $0 < \lambda < \infty$, we can calculate the in-cloud sounding. For some values of λ, there may be no positive buoyancy at any level. Such clouds "do not exist." For

values of λ that do allow positive buoyancy, the cloud-top level is determined by finding the height where they buoyancy decreases to zero.

There are various technical complications involving the shape of the sounding. Those will not be discussed here.

At this point, we have the equations we need to determine the rates at which cumulus convection warms and dries the large-scale state, if we can determine $M_B(\lambda)$, which is a measure of convective intensity. An additional physical idea is needed to determine $M_B(\lambda)$.

8.17 What determines the intensity of the convection?

To determine the intensity of convective activity, AS proposed a "quasi-equilibrium" hypothesis, according to which the convective clouds quickly convert whatever moist convective available potential energy is present in convectively active atmospheric columns into convective kinetic energy. The starting point for the quasi-equilibrium closure is the recognition that cumulus convection occurs as a result of moist convective instability, in which the potential energy of the mean state is converted into the kinetic energy of cumulus convection.

AS defined the "cloud work function," A, for a cumulus subensemble, as a vertical integral of the buoyancy of the cloud air with respect to the large-scale environment:

$$A(\lambda) = \int_{z_B}^{z_D(\lambda)} \frac{g}{c_p \overline{T}(z)} \eta(z,\lambda) \left[s_c(z,\lambda) - \overline{s}(z) \right] dz \,, \tag{8.134}$$

Here $z_D(\lambda)$ is the height of the detrainment level for cloud type λ. From (8.124) we see that the function is a property of the large-scale environment. A positive value of means that a cloud with fractional entrainment rate can convert the potential energy of the mean state into convective kinetic energy. For $\lambda = 0$, the cloud work function is equivalent to the convective available potential energy (CAPE), as conventionally defined.

Numerical models use the conservation equations for thermodynamic energy and moisture to predict $\overline{s}(z)$ and $\overline{q_v}(z)$, from which $A(\lambda)$ can be determined; therefore, these models indirectly predict $A(\lambda)$. By taking the time derivative of (8.134), and using the conservation equations for thermodynamic energy and moisture, AS derived an equation that can be written in simplified form as

$$\frac{dA(\lambda)}{dt} = J(\lambda) M_B(\lambda) + F(\lambda) \,, \tag{8.135}$$

The $J(\lambda)M_B(\lambda)$ term of (8.135) represents all of the terms involving convective processes, each of which turns out to be proportional to $M_B(\lambda)$. The term actually represents an integral over cloud types, and is written here as a product merely to simplify the discussion. The quantity $J(\lambda)$ symbolically represents the kernel of the integral, which is a property of the large-scale sounding; see AS for details. The $J(\lambda)M_B(\lambda)$ term of (8.135) tends to reduce $A(\lambda)$, because cumulus convection stabilizes the environment, so that $J(\lambda)$ is usually negative. Keep in mind that an equation like (8.135) holds for each cumulus subensemble.

The $F(\lambda)$ term of (8.135) represents what AS called the "large-scale forcing," i.e., the rate at which the cloud work function tends to increase with time due to a variety of processes including:

- horizontal and vertical advection by the mean flow;
- the surface turbulent fluxes of sensible and latent heat, and the rate of change of the planetary boundary-layer depth;
- radiative heating and cooling;
- precipitation and turbulence in stratiform clouds.

Note that some of these "forcing" processes, e.g., those involving boundary-layer turbulence and stratiform clouds, are themselves parameterized processes that may involve fluctuations on small spatial scales; for this reason it seems inappropriate to describe the collection of processes that contribute to as "large-scale;" a better term would be "non-convective."

In equilibrium, with $dA/dt = 0$, the terms on the right-hand side of (8.135) must add up to zero. AS considered a "quasi-equilibrium" of the cloud work function, in which the terms on the right-hand side of (8.135) *almost* add up to zero. They proposed that

$$\boxed{\frac{d}{dt}A(\lambda) = JM_B(\lambda) + F(\lambda) \cong 0 \text{ when } F(\lambda) > 0\,.} \tag{8.136}$$

Eq. (8.136) means that the moist convective instability generated by the forcing, $F(\lambda)$, is very rapidly consumed by cumulus convection, i.e., the two terms on the right-hand side of (8.136) approximately balance each other. In a steady-state situation, this balance is of course trivially satisfied, by definition. The physical content of (8.136) is, therefore, the assertion that near-balance is maintained even when is varying with time, provided that the variations of are sufficiently slow. The cumulus ensemble thus closely follows the lead of the forcing, like a defensive basketball player (the convection) playing man-to-man against an offensive player (the forcing). Keep in mind, however, that the forcing depends on the large-scale circulation, which is strongly affected by the convection, just as the play of an offensive basketball player is strongly affected by the moves of his or her defensive

opponent. We should not imagine that the forcing is "given" and that the convection just meekly responds to it. The convection and the forcing evolve together according to the rules defined by the combination of large-scale dynamics and cloud dynamics.

QE of the cloud work function does not imply QE of the temperature or QE of the water vapor. What it does imply is that the temperature and water vapor change together, in a coupled way, so that the cloud work function is in QE. The coupling is of course due to the convection. But why should the cloud work function be in QE?

The QE approximation is expected to hold if $\tau_{\rm LS}$, the time scale for changes in $F(\lambda)$, is much longer than the "adjustment time," $\tau_{\rm adj}$, required for the convection to consume the available CAPE; this allows the convection to keep up with the changes in $F(\lambda)$. AS introduced the concept of $\tau_{\rm adj}$ by asking what would happen if a conditionally unstable initial sounding were modified by cumulus convection, without any forcing to maintain the CAPE over time; they asserted that the CAPE would be consumed by the convection (i.e. converted into convective kinetic energy) on a time-scale that they defined as and estimated to be on the order of a few hours. Just such an unforced convective situation has been numerically simulated by Soong and Tao (1980) and others, using high-resolution cloud models; their results are consistent with the scenario of AS. If the adjustment time is on the order of 10^3 to 10^4 s, then use of (8.136) is justified, as an approximation, for the simulation of "weather" with time scale on the order of one day or longer, i.e., at least one order of magnitude longer than $\tau_{\rm adj}$.

By using (8.136), together with

$$|JM_B| \sim \frac{A}{\tau_{\rm adj}} \,, \tag{8.137}$$

AS found that

$$A \sim \tau_{\rm adj} F \ll \tau_{\rm LS} F \,, \tag{8.138}$$

where $\tau_{\rm LS}$ is the time-scale on which the forcing itself is varying, i.e., on which the large-scale weather situation is varying. Eq. (8.138) means that the cloud work function is "small" compared to $\tau_{\rm LS}F$, which is the value that the cloud work function would take if the forcing acted without opposition over a time scale $\tau_{\rm LS}$. Although we should expect to see day-to-day variations of A, we should not expect to see values as large as $\tau_{\rm LS}F$. Eq. (8.138) means that A is trapped in the range of values between zero (since by definition cannot be negative) and $\tau_{\rm adj}F$. In this sense, A is "close to zero" (see also Xu and Emanuel, 1989).

Based on the analysis above, it can be asserted that the cloud work function (or the CAPE) "is quasi-invariant with time," i.e. that

$$\frac{dA}{dt} \cong 0 \,, \tag{8.139}$$

which is a short-hand form of Eq. (8.136); and that "the CAPE is small" in convectively active regimes, i.e.

$$A \cong 0 \,, \tag{8.140}$$

which is a short-hand form of Eq. (8.138). A sounding for which $A \cong 0$ tends to follow a saturated moist adiabat, throughout the depth of the convective layer. This provides a rationalization for the assumption of Manabe and Wetherald (1967) that the lapse rate cannot exceed the moist adiabatic lapse rate, which they approximated as 6.5 K km^{-1}.

Unfortunately, because (8.139) and (8.140) are short-hand forms, they are subject to misinterpretation. For example, data like those shown in Fig. 8.11 are sometimes viewed as being inconsistent with QE. It is natural to wonder how the CAPE be described as "quasi-invariant" when it is observed to undergo such "large" changes. This point of view appears to be based on the tacit assumption that when we say that the changes in the CAPE are "small," we mean that they are small compared with the time average of the CAPE. In fact, however, it should be clear from the preceding discussion that this is not what is meant at all. Instead, we mean that the changes in the CAPE are small compared to those that would occur if the convection were somehow suppressed while the non-convective processes continued to increase the CAPE with time. Eq. (8.136) does not imply that $A(\lambda)$ is invariant from day to day, and the observed day-to-day changes in the CAPE, such as those shown in Fig. 8.11, are not in conflict with QE. What QE does imply is that the day-to-day changes in the CAPE that we actually see are much smaller than they would be if the negative convective term of (8.136) could somehow be suppressed, so that the positive (under disturbed conditions) forcing term could have its way with the sounding.

Eq. (8.136) can be used to solve for the convective mass flux as a function of cloud type, λ. After discretization, this leads to a system of linear equations (Lord, 1982). Although the system is linear, the mass flux distribution function, $M_B(\lambda)$, is required to be non-negative for all λ. This cannot be guaranteed without making additional assumptions (e.g., Hack et al., 1984). An alternative approach that avoids these difficulties was proposed by Randall and Pan (1993) and Pan and Randall (1998).

The QE hypothesis has been observationally tested by Arakawa and Schubert (1974), Lord and Arakawa (1980), Lord (1982), Kao and Ogura (1987), Arakawa and Chen (1986),

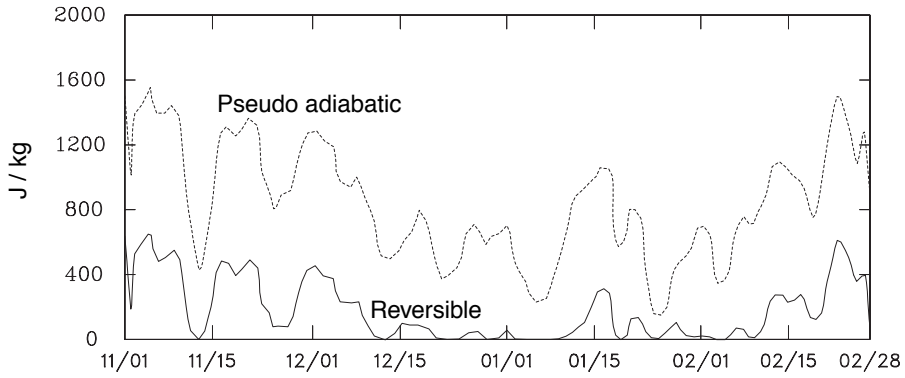


Figure 8.11: Observed time sequence of the CAPE during TOGA COARE, over the Intensive Flux Array.

Grell et al. (1991), Wang and Randall (1994), and Cripe and Randall (2001), among others, using both tropical and midlatitude data. In addition, idealized tests of QE using a high-resolution cloud ensemble model have been performed by Xu and Arakawa (1992) and Jones and Randall (2011).

For further discussions of cumulus parameterization see the collections of essays edited by Emanuel and Raymond (1993) and Smith (2013), as well as the articles by Arakawa (2004) and Randall et al. (2003); Randall (2013).

In Chapter 10, we will discuss the role of convection in tropical eddies.

8.18 Conditional symmetric instability

Deep convection occurs in both the tropics and middle latitudes. The convective instability can interact with inertial instability, and in fact there are strong analogies between the two types of instability. Bennetts and Hoskins (1979) and Emanuel (1979, 1982, 1983) pointed out that when the air is saturated with water vapor the criterion for inertial instability can be expressed in terms of angular momentum variations along surfaces of constant equivalent potential temperature (or moist static energy), rather than surfaces of constant dry potential temperature. For saturated motion, instability can occur when the angular momentum decreases poleward along surfaces of constant saturation moist static energy. Another way of saying the same thing is that, for saturated motion, instability can occur when the saturation moist static energy decreases upward along surfaces of constant angular momentum.

To show how this works, we now discuss inertial instability and its connections to convection. To avoid unnecessary complications, we will also assume that there are no variations with longitude. This can be justified, at least in this pedagogical example, because the physics of inertial instability relate to spatial variations of the angular momentum, which occur mainly with latitude. Inertial instability is sometimes called “symmetric instability”

in the atmospheric science literature. This is not meant to suggest that inertial instability has to be or is intrinsically independent of longitude. It can occur, for example, in the vicinity of a local maximum of the zonal wind, i.e., a "jet max."

We use θ coordinates with the quasi-static approximation, and assume that there are no variations in the longitudinal direction, and no heating. Since we are assuming that there is no heating, particles move along isentropic surfaces, and so do not experience buoyancy forces. Consider the components of the horizontal momentum equation in the form

$$\frac{Du}{Dt} - \left(2\Omega + \frac{u}{a\cos\varphi}\right) v \sin\varphi = 0 \,, \tag{8.141}$$

$$\frac{Dv}{Dt} + \left(2\Omega + \frac{u}{a\cos\varphi}\right) u \sin\varphi + \Omega^2 a \cos\varphi \sin\varphi = 0 \,, \tag{8.142}$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \frac{v}{a}\frac{\partial}{\partial \varphi} \,, \tag{8.143}$$

The $\Omega^2 a \cos\varphi \sin\varphi$ term of (8.142) represents the centrifugal acceleration. The metric terms (which arise from the use of spherical coordinates, as discussed in Chapter 4) are also included in (8.141) and (8.142). As discussed in Chapter 4, Eq. (8.141) implies that angular momentum is conserved, i.e.,

$$\frac{DM}{Dt} = 0 \,, \tag{8.144}$$

where

$$M \equiv a\cos\varphi \left(u + \Omega a \cos\varphi\right) \,. \tag{8.145}$$

Using (8.145), we can rewrite (8.142) in the more compact form

$$\frac{Dv}{Dt} + \frac{M^2 \sin\varphi}{a^3 \cos^3\varphi} = -\frac{1}{a}\frac{\partial s}{\partial \varphi} \,. \tag{8.146}$$

Eqs. (8.144) and (8.146) are pleasingly simple. In (8.146), the Coriolis, metric, and centrifugal terms have all been combined into a single term that is proportional to M^2.

We now linearize the system about a basic state in which the flow is purely zonal and in gradient-wind balance, so that

$$\overline{v}^{\lambda} = 0 \,, \tag{8.147}$$

and

$$\frac{\left(\overline{M}^{\lambda}\right)^2 \sin\varphi}{a^3 \cos^3\varphi} = -\frac{1}{a}\frac{\partial \overline{s}^{\lambda}}{\partial \varphi} \,. \tag{8.148}$$

The linearized versions of (8.146) and (8.144) can be written as

$$\frac{\partial v'}{\partial t} + \left(\frac{2\overline{M}^{\lambda} \sin\varphi}{a^3 \cos^3\varphi}\right) M' = -\frac{1}{a}\frac{\partial s'}{\partial \varphi} \,, \tag{8.149}$$

$$\left(\frac{\partial M'}{\partial t}\right)_{\theta} + \frac{v'}{a}\left(\frac{\partial \overline{M}^{\lambda}}{\partial \varphi}\right)_{\theta} = 0 \,. \tag{8.150}$$

To investigate inertial stability and instability, we neglect the pressure-gradient term on the right-hand side of (8.149), which plays only a secondary role; the dry static energy fluctuations on an isentropic surface are quite small. Assuming solutions of the form $(v', M') = \left(\widehat{v}, \widehat{M}\right) e^{\sigma t}$, our system reduces to

$$\sigma \widehat{v} + \left(\frac{2\overline{M}^{\lambda} \sin\varphi}{a^3 \cos^3\varphi}\right) \widehat{M} = 0 \,, \tag{8.151}$$

$$\sigma\widehat{M} + \frac{v'}{a}\left(\frac{\partial \overline{M}^{\lambda}}{\partial \varphi}\right)_{\theta} = 0\,. \tag{8.152}$$

For nontrivial solutions, we need

$$\begin{aligned}\sigma^2 &= \left(\frac{2\overline{M}^{\lambda}\sin\varphi}{a^3\cos^3\varphi}\right)\frac{1}{a}\left(\frac{\partial \overline{M}^{\lambda}}{\partial \varphi}\right)_{\theta} \\ &= \frac{\sin\varphi}{a^4\cos^3\varphi}\left[\frac{\partial}{\partial\varphi}\left(\overline{M}^{\lambda}\right)^2\right]_{\theta}\,.\end{aligned} \tag{8.153}$$

According to (8.153), the system is inertially stable if the angular momentum decreases towards the pole along isentropic surfaces, and inertially unstable if the angular momentum increases towards the pole along isentropic surfaces. In a neutrally stable state, the angular momentum is constant along isentropic surfaces. These conclusions apply to both hemispheres.

The criterion for inertial instability is often expressed in terms of vorticity rather than angular momentum. Here's how that works. In spherical coordinates, the vertical component of the absolute vorticity is given by

$$\zeta_{\theta} + f = \frac{1}{a\cos\varphi}\left(\frac{\partial v}{\partial\lambda}\right)_{\theta} - \frac{1}{a\cos\varphi}\left[\frac{\partial}{\partial\varphi}\left(u\cos\varphi\right)\right]_{\theta} + 2\Omega\sin\varphi\,. \tag{8.154}$$

On the other hand, the meridional derivative of the angular momentum is

$$\frac{1}{a}\left(\frac{\partial \overline{M}^{\lambda}}{\partial\varphi}\right)_{\theta} = \left[\frac{\partial}{\partial\varphi}\left(u\cos\varphi\right)\right]_{\theta} - 2\Omega a\cos\varphi\sin\varphi\,. \tag{8.155}$$

Comparing (8.154) and (8.155), we see that

$$\zeta_{\theta} + f = \frac{-1}{a^2\cos\varphi}\left(\frac{\partial \overline{M}^{\lambda}}{\partial\varphi}\right)_{\theta} \quad \text{for purely zonal flow}\,. \tag{8.156}$$

This allows us to rewrite (8.153) as

$$\sigma^2 \cong \frac{-2\overline{M}^{\lambda}\sin\varphi}{a^2\cos^2\varphi}\left(\overline{\zeta_\theta}^{\lambda}+f\right) \quad \text{for purely zonal flow .} \tag{8.157}$$

We can safely assume that $\overline{M}^{\lambda} > 0$, except possibly very close to the poles. It follows that the sign of σ^2 is determined by the sign of $-\sin\varphi\left(\overline{\zeta}_{\theta}^{\lambda}+f\right)$. Inertial instability occurs when $\sigma^2 > 0$. In the Northern Hemisphere, where $\sin\varphi > 0$, the criterion for instability is $\overline{\zeta_\theta}^{\lambda} + f < 0$, and in the Southern Hemisphere it is $\overline{\zeta_\theta}^{\lambda} + f > 0$. *In either hemisphere, inertial instability occurs when the absolute vorticity along isentropic surfaces has the "wrong" sign.* In middle latitudes, this situation can occur on small scales ($\sim$ 100 km) on the anticyclonic (i.e., equatorward) side of a westerly jet maximum. Since $\overline{\zeta}^{\lambda} + f$ passes through zero near the Equator, inertial instability can easily occur there, when absolute vorticity is advected across the Equator (e.g., Tomas and Webster, 1997).

When the zonal wind does not vary with height, surfaces of constant angular momentum are vertical. When the zonal wind does vary with height, angular momentum surfaces are tilted in the latitude-height plane. This can occur, for example, below a jet stream or in the vicinity of a front. It is possible for the saturation moist static energy to decrease along such tilted surfaces even when it does not decrease upward. This is called conditional symmetric instability (CSI), a term that is now widely used in the forecasting community. Emanuel (1983) argued that CSI is relevant to extratropical squall lines that form in regions of strong vertical wind shear.

8.19 Segue

The observed vertical structure of the atmosphere is controlled, to a remarkable degree, by diabatic processes. For example, the observed height of the tropopause is approximately that predicted by radiative-convective equilibrium models, which completely ignore the effects of large-scale circulation systems. The message is that although large-scale dynamics is important, it is strongly constrained by radiation and convection. At the same time, the heating due to radiation, convection, and boundary-layer turbulence is strongly controlled by the global circulation. It is impossible to understand the circulation without understanding the heating, and vice versa. Further discussion is given in later chapters.

Chapters 2 and 6 presented evidence that much of the upward transport of energy and moisture are carried out by small-scale convection. The current chapter outlines the mechanisms by which convection produces these transports. The energetics of the convection itself played a key role in the discussion.

8.20 Problems

1. Refer to Table 8.1.

   (a) Assume that the area covered by the ITCZ is 4 x 10^{13} m^2. Using the numbers given in the table, estimate the *vertically integrated* source of moist static energy in the ITCZ, in W m^{-2}. Explain your answer.

   (b) Assume that the moist static energy source is uniform with pressure through the whole depth of the troposphere, i.e., the "heating rate" is the same at all levels. Based on the numbers given in the table, what is the total upward flux of moist static energy across the 500 hPa surface, in W m^{-2}? Explain your answer.

2. Using the methods of Section 8.2.2,

   (a) Prove that dry convection transports dry static energy upward, i.e., $\overline{w's'} > 0$, where the overbar represents an average over one horizontal wavelength.

   (b) Prove that for gravity waves $\overline{w's'} = 0$.

3. Consider an idealized Hadley circulation in which the rising branch occupies fractional area σ_{up} and the sinking branch occupies fractional area $1 - \sigma_{up}$. Assume that the horizontally averaged vertical mass flux and the horizontally averaged heating rate per unit mass are both zero, i.e.,

$$\rho \sigma_{up} w_{up} + \rho (1 - \sigma_{up}) w_{dn} = 0 , \tag{8.158}$$

   and

$$\sigma_{up} Q_{up} + (1 - \sigma_{up}) Q_{dn} = 0 . \tag{8.159}$$

   Here Q is the heating rate per unit mass, and the subscripts up and dn denote the rising and sinking branches, respectively. Also assume that the temperature is horizontally uniform at each level, for the reasons given by Charney (1963). Show that

$$\sigma_{up} = \frac{-Q_{dn}}{Q_{up} - Q_{dn}} , \tag{8.160}$$

   and

$$\overline{M_{\rm HC}} = \frac{-\rho}{\partial \bar{s}/\partial z} \left(\frac{Q_{\rm up} Q_{\rm dn}}{Q_{\rm up} - Q_{\rm dn}} \right) , \tag{8.161}$$

where

$$\begin{aligned} \overline{M_{\rm HC}} \equiv & \rho \, \sigma_{\rm up} w_{\rm up} \\ = & - \rho \, (1 - \sigma_{\rm up}) \, w_{\rm dn} \, . \end{aligned} \tag{8.162}$$

The value of $\overline{M_{\rm HC}}$ in the middle troposphere is a measure of the strength of the Hadley circulation.

Chapter 9

Circulation regimes and cloud regimes

9.1 How the surface winds influence the tropical and subtropical SST

As discussed in Chapter 2, the property of the oceans that most strongly influences the atmosphere is the sea surface temperature, which affects the upward longwave radiation, the sensible heat flux, and the latent heat flux. Fig. 9.1 shows the distribution of the sea surface temperature in the tropical and subtropical Pacific. Relatively cold water is found on the eastern side of the basin, and along the equator. This pattern is driven in part by the near-surace winds, which can drive ocean currents by pushing on the water. The currents are typically very slow (centimeters per second) compared to the usual near-surface wind speeds, so for practical purposes the oceans can be considered to be at rest when air-sea momentum exchanges are considered.

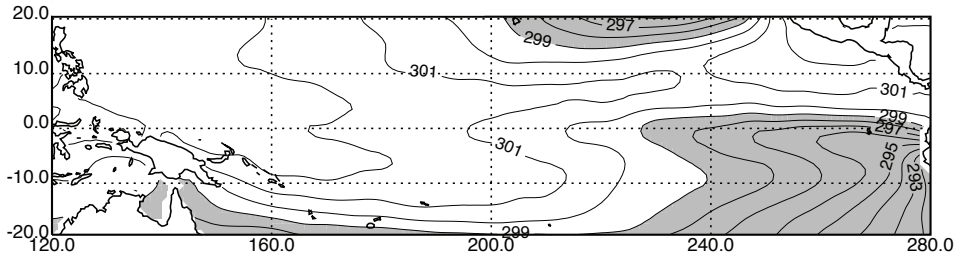


Figure 9.1: Climatological distribution of sea-surface temperature in the tropical Pacific Ocean. Note the cold water along the equator and throughout most of the eastern side of the ocean basin, and the warm water on the western side of the basin and stretching across the basin north of the equator.

The immediate effect of the surface wind stress is to push the ocean water in the same direction as the near-surface wind. The Coriolis acceleration then turns the current towards

the right (relative to the wind) in the Northern Hemisphere, and towards the left (relative to the wind) in the Southern Hemisphere. This has interesting and important consequences for the ocean circulation.

The prevailing westerlies in middle latitudes induce an equatorward drift superimposed on a general eastward current in the upper ocean, and the prevailing easterlies in the tropics tend to cancel out this equatorward drift, while driving the surface currents back towards the west. The circulation of the upper oceans thus takes the form of a pair of huge "gyres," one in each hemisphere. The poleward currents, such as the Gulf Stream and Kuroshio, tend to be warmer than average at a given latitude, and the equatorward currents, such as the California Current and the Humboldt Current, tend to be cooler than average at a given latitude.

The east-to-west trade winds along the Equator, where $f = 0$, lead to surface currents away from the Equator in both hemispheres (i.e., towards the right of the wind north of the Equator, and towards the left of the wind south of the Equator). This drives upwelling along the Equator. This is why the surface waters are colder along the Equator than they are on either side. The cold equatorial waters are called the "cold tongue." It stretches from South America westward to near 160° E.

The equatorward winds along the west coasts of the continents drive equatorward currents, which carry cool water. Both the directions of the currents (from the poles) and the upwelling favor cold water, which is what is observed. Consider a surface wind blowing parallel to a coastline in the Northern Hemisphere, with the coastline on its left side, as in the California high in July. The surface current moves to the right of the wind, i.e., away from the coast. This drives coastal upwelling, which leads to cool surface water. Such upwelling occurs near each of the subtropical highs, in both hemispheres. This is one reason why the subtropical highs tend to occur over cool water.

Pushing by the trade winds drives a westward current in the upper ocean, which causes warm water to pile up on the western side of each basin. The SST maximum in the tropical western Pacific is known as the Warm Pool. The sea water flows back to the east very close to the Equator and slightly below the surface, in what is called the Equatorial Undercurrent. There is also a near-surface west-to-east return flow called the North Equatorial Counter Current.

9.2 Tropical and subtropical cloud regimes

Fig. 9.2 shows three cloud regimes that are important in the tropics and subtropics. The subtropical *marine stratocumulus regime* is shown on the right side of the figure, in a region of large-scale subsidence. The sea surface temperatures are relatively cool in such regimes. Towards the left, we enter the *tradewind cumulus regime*, which has weaker subsidence and warmer sea surface temperatures. Finally on the left side of the figure we reach the *deep cumulus regime*, characterized by warm sea surface temperatures and large-scale rising

motion.

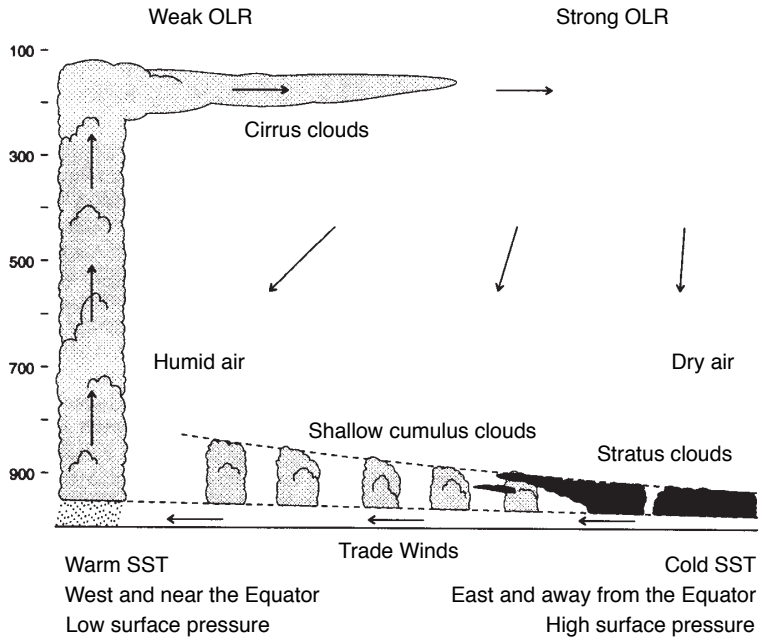


Figure 9.2: Schematic diagram summarizing the relationships among the cloud regimes depicted in Fig. 9.4, and how they fit into the mean meridional circulation. Adapted from Schubert et al. (1995).

For reasons discussed in Chapter 8, the lapse rate of the deep convective regime is close to the moist adiabatic lapse rate. As discussed in Chapter 3, however, the horizontal temperature gradients are weak throughout the tropics and subtropics, for reasons discussed by Charney (1963). This means that the lapse rate in the subtropics, above the trade inversion, must be nearly the same as the lapse rate in the tropics.

For the reasons given by Charney (1963), we can write an approximate thermodynamic balance for the descending branch of the Hadley cell as

$$\omega \frac{\partial s}{\partial p} = Q_R \, . \tag{9.1}$$

Here Q_R is the radiative cooling rate. We see from (9.1) that the speed of the large-scale sinking motion in the subtropics is essentially determined by the requirement of thermodynamic balance. Typical clear-sky tropical tropospheric radiative cooling rates are on the order of 2 K day^{-1} (Fig. 9.3, which corresponds to a few hundred meters of sinking per day. As the air descends, it is gradually cooled by radiation. As a result, the potential temperature of the air in the subtropical free atmosphere decreases downward; in other words, it increases upward.

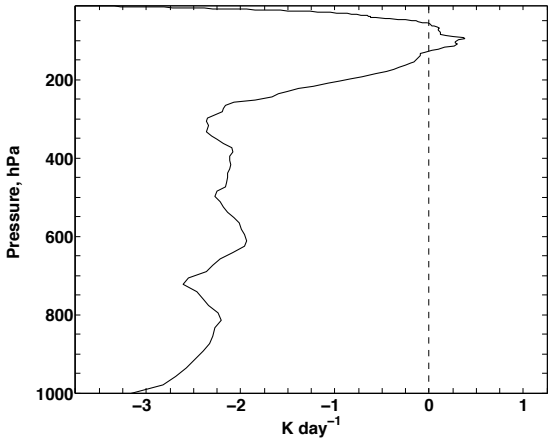


Figure 9.3: Typical tropical clear-sky infrared radiative cooling rate based on a standard tropical sounding.

9.3 Representative soundings

Fig. 9.4 shows the observed vertical structure of the atmosphere for

- Darwin, Australia, a very tropical place;
- Porto Santo Island in the Atlantic tradewind regime; and
- San Nicolas Island, in the subtropical marine stratocumulus regime off the coast of southern California.

The quantities plotted are the dry static energy, s, the moist static energy, h, and the saturation moist static energy, which is defined by

$$h_{\rm sat} \equiv s + Lq_{\rm sat} \ , \tag{9.2}$$

where $q_{\rm sat}$ is the saturation mixing ratio. Quite generally, we have

$$s \leq h \leq h_{\rm sat} \ . \tag{9.3}$$

The panels on the right side of the figure cover both the troposphere and the lower stratosphere, while those on the left zoom in on the lower troposphere to show more detail. All values are divided by c_p to give units in K.

Consider the tropical (i.e., Darwin) sounding shown in Fig. 9.4. You should compare the Darwin sounding with Figure 8.2 and also with the zonally averaged moist static energy

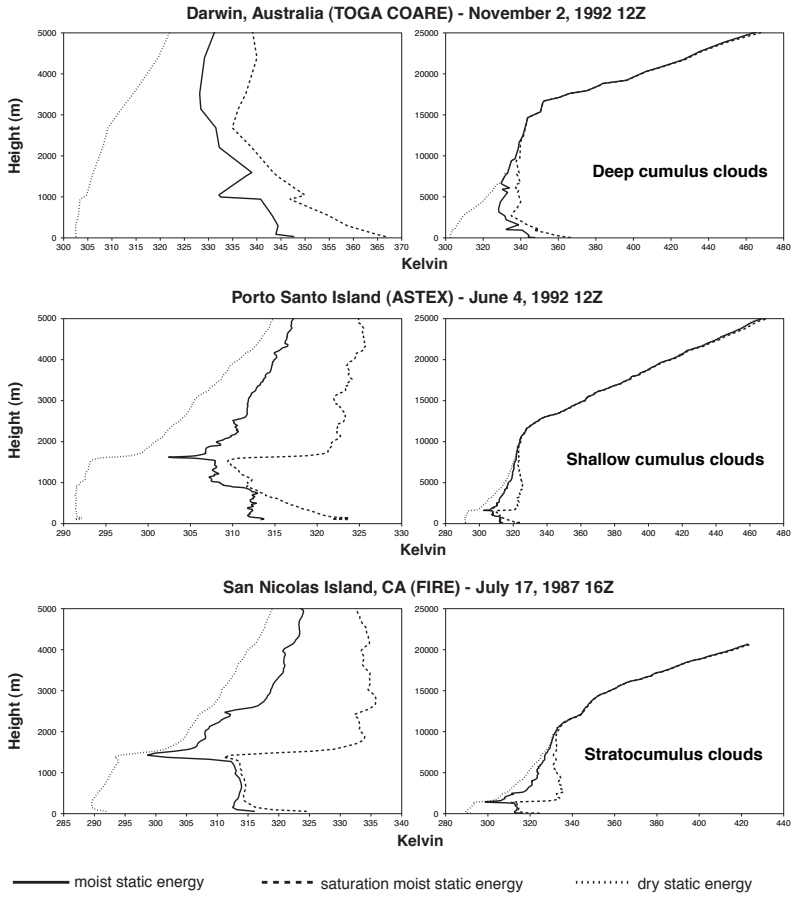


Figure 9.4: Representative observed soundings for Darwin, Australia, a very tropical place; Porto Santo Island in the Atlantic tradewind regime; and San Nicolas Island, in the subtropical marine stratocumulus regime off the coast of southern California. The panels on the right cover both the troposphere and the lower stratosphere, while those on the left zoom in on the lower troposphere to show more detail. All values are divided by c_p to give units in K.

profile shown in Figure 6.10. Near the surface, $h_{\rm sat} > h$, indicating that the air is unsaturated. If we lift a parcel adiabatically from near the surface, its moist static energy follows a straight, vertical line in the diagram, starting from a near-surface value. At the same time, the environmental saturation moist static energy decreases upward. After rising a kilometer or so, the vertical line representing the moist static energy traced out by adiabatic parcel ascent from the surface will be to the right of the observed sounding of saturation moist static energy, so that $h_c - \overline{h_{\rm sat}} > 0$. According to (8.40), the parcel will then be positively buoyant, if it is saturated. We can thus tell, simply by looking at Fig. 9.4, that the Darwin sounding is conditionally unstable. The positive buoyancy of the lifted parcel will continue upward until the vertical line representing constant moist static energy again crosses over to the left side of the curve representing the environmental saturation moist static energy. For the Darwin soundings, this occurs at about the 15 km level, near the tropopause. We thus

expect deep cumulus convection to occur in this sounding, although the mere existence of a conditionally unstable sounding is not enough, in itself, to show that cumulus convection will be significantly active.

Another interesting aspect of the Darwin sounding is that the saturation moist static energy is nearly uniform with height throughout most of the troposphere. Recall that in an atmosphere that is neutrally stable with respect to saturated moist convection, $\partial \overline{h_{\rm sat}}/\partial z = 0$. The Darwin sounding is very humid and close to neutrally stable for moist convection.

The subtropical "tradewind" sounding is also conditionally unstable, but only through a shallow layer. The tradewind convective layer is capped by a very strong temperature inversion, and the water-vapor mixing ratio decreases strongly upward through this "trade inversion." The middle troposphere is much drier in the tradewind sounding than in the tropical sounding.

The subtropical marine stratocumulus sounding is not conditionally unstable at all. It shows evidence of cloudiness in the lowest kilometer. The cloud layer is capped by a very strong inversion, essentially similar to the trade inversion, but residing at a lower level. Marine stratocumulus regimes occur in several places around the world, typically in association with subtropical highs. See Fig. 9.5.

9.4 Entrainment across the cloud-top inversion

In the tradewind cumulus and stratocumulus regimes, the sinking air passes through the trade inversion. This process is called "entrainment." How does it happen? It is remarkable, for example, that the average mixing ratio of the air suddenly increases from perhaps 1 g kg^{-1} above the stratocumulus to 6 or 7 g kg^{-1} below. It is the same air, after all; how does it suddenly become so moist? The answer is that the convective vertical motions transport moisture upwards from the sea surface, where it is used to rapidly moisten the sinking air. The air also cools rapidly as it descends through the inversion. This cooling is produced by a combination of concentrated radiative cooling near the cloud tops, cooling due to the evaporation of liquid water carried upward and deposited at the cloud-top, and a downward flux of sensible heat. (Lilly, 1968, Fig. 9.6) was the first to explain the process.

A macroscopic view of the entrainment process is as follows. Let A be an arbitrary scalar, satisfying a conservation equation that can be written in "flux form" as

$$\frac{\partial(\rho A)}{\partial t} + \nabla\cdot(\rho \mathbf{v} A) + \frac{\partial(\rho w A)}{\partial z} = -\frac{\partial F_A}{\partial z} + S_A \tag{9.4}$$

where F_A is the upward turbulent flux of A, bars are omitted on the mean quantities, and is a source or sink of A, per unit volume. Integrating (9.4) from just below to just above the inversion, and using Leibniz' rule, we get

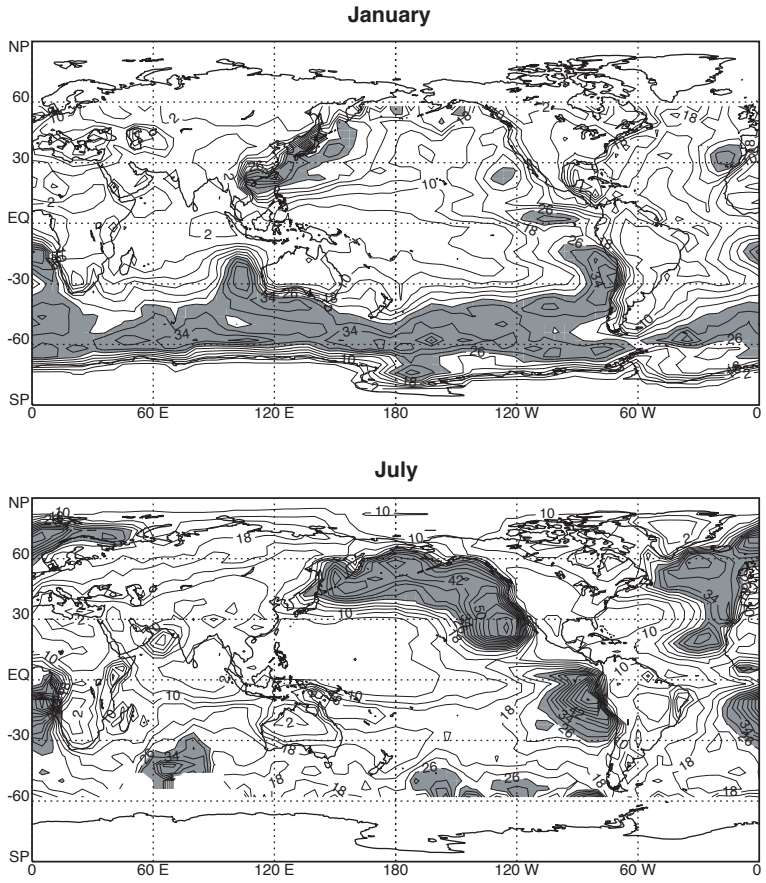


Figure 9.5: Observed locations of low-level stratus clouds, including the one west of California, as observed by ISCCP, the International Satellite Cloud Climatology Project. Cloud amounts higher than 26% are shaded.

$$\begin{aligned} &\frac{\partial}{\partial t}\left(\int_{z_B-\varepsilon}^{z_B+\varepsilon}\rho A dz\right)-\Delta(\rho A)\frac{\partial z_B}{\partial t}+\nabla\cdot\left(\int_{z_B-\varepsilon}^{z_B+\varepsilon}\rho\mathbf{v}A dz\right)\\ &-\Delta(\rho\mathbf{v}A)\cdot\nabla z_B+\Delta(\rho wA)=(F_A)_B+\int_{z_B-\varepsilon}^{z_B+\varepsilon}S_A dz\,, \end{aligned} \tag{9.5}$$

Here we have used the notation $\Delta(.)\equiv(.)_{z=z_B+\varepsilon}-(.)_{z=z_B-\varepsilon}\equiv(.)_{B+}-(.)_B$, and henceforth subscripts $B+$ and B denote levels just above and just below the inversion, respectively. In writing (9.5), we have assumed that $(F_A)_{B+}=0$, since level $B+$ is above the boundary layer's top.

The $\partial/\partial t$ and $\nabla\cdot$ terms of (9.5) can be neglected for $\varepsilon\to 0$. With this assumption, and

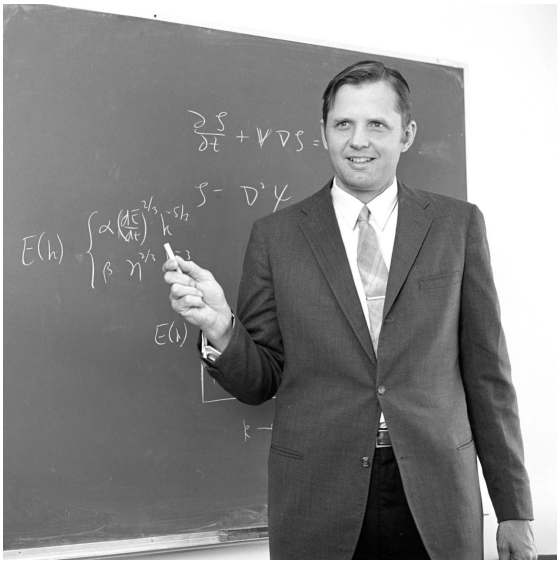


Figure 9.6: Douglas Lilly, who did important work on a wide range of topics, including cumulus convection, numerical methods, gravity waves, and stratocumulus clouds.

for $A \equiv 1$, (9.5) reduces to mass conservation in the form

$$\rho_{B+}\left(\frac{\partial z_B}{\partial t} + \mathbf{v}_{B+}\cdot\nabla z_B - w_{B+}\right) = \rho_B\left(\frac{\partial z_B}{\partial t} + \mathbf{v}_B\cdot\nabla z_B - w_B\right) \equiv E\ , \tag{9.6}$$

where E is the downward mass flux across the inversion. In essence, (9.6) says that the flow of mass is continuous across the boundary layer top, i.e., no mass is created or destroyed between levels B and $B+$. We interpret E as the mass flux due to the turbulent entrainment of free atmospheric air across the inversion. With the definition of E as given by (9.6), we can simplify (9.5) to

$$-E\Delta A = (F_A)_B + \int_{z_B-\varepsilon}^{z_B+\varepsilon} S_A dz\ . \tag{9.7}$$

For $S_A = 0$, (9.7) simply says that the total flux of A must be continuous across the inversion. Notice that for $\Delta A \neq 0$, a mass flux across the inversion is generally associated with the

convergence of a turbulent flux of A at level B. This flux convergence changes the A of entering particles from A_{B+} to A_B.

As a simple example, consider the moistening of the air as it moves down across the inversion. The dry entrained air is moistened by an upward moisture flux that converges "discontinuously" at level B. This is described by

$$-\Delta q_T E = \left(F_{q_T}\right)_B , \tag{9.8}$$

which is a special case of (9.7). Here $q_T \equiv q_v + l$ is the total water mixing ratio.

9.5 The Walker Circulation

The Walker Circulation, named by Bjerknes (1966), is an east-west overturning of the atmosphere above the tropical Pacific Ocean, with rising motion on the west side, over the so-called "Warm Pool," and sinking motion on the east side. The Walker Circulation can be viewed as a thermally excited stationary eddy. Although the Walker Circulation is driven by the east-to-west sea surface temperature gradient, it also helps to maintain that gradient through the mechanisms discussed above. For this reason, the Walker Circulation is best understood as a coupled ocean-atmosphere phenomenon. It undergoes strong interannual variability. Fig. 9.7 is a schematic illustration of the Walker circulation and its relation to the surface wind field in the southern Hemisphere, taken from the book by Philander (1990). The equatorward flow just west of South America can be viewed as the inflow to the ITCZ (which is generally north of the Equator in this region), and so it is in a sense a portion of the lower branch of the Hadley circulation.

The Hadley Circulation is defined in terms of zonal averages, so a particle participating in the Hadley Circulation cannot "escape" by moving to a different longitude. In contrast, a particle participating in the east-west Walker Circulation can escape by moving off to a different latitude. For this reason, we should not think of the Walker Circulation as a closed "race track." The Hadley and Walker Circulations are closely linked. For example, a parcel may travel westward across the tropical Pacific in the lower branch of the Walker Circulation, ascend to the tropopause over the Warm Pool, and then move both poleward and eastward away from the Warm Pool, possibly descending in the subtropical eastern Pacific. It can then join the trades, and repeat its westward and Equatorward journey through the boundary layer.

Bjerknes (1969) theorized that the cool, dry air of the trade winds is heated and moistened as it moves westward until it finally undergoes large-scale moist-adiabatic ascent over the Warm Pool. If there were no mass exchange with adjacent latitudes, a simple circulation would develop in which the flow is easterly at low levels and westerly at upper levels.

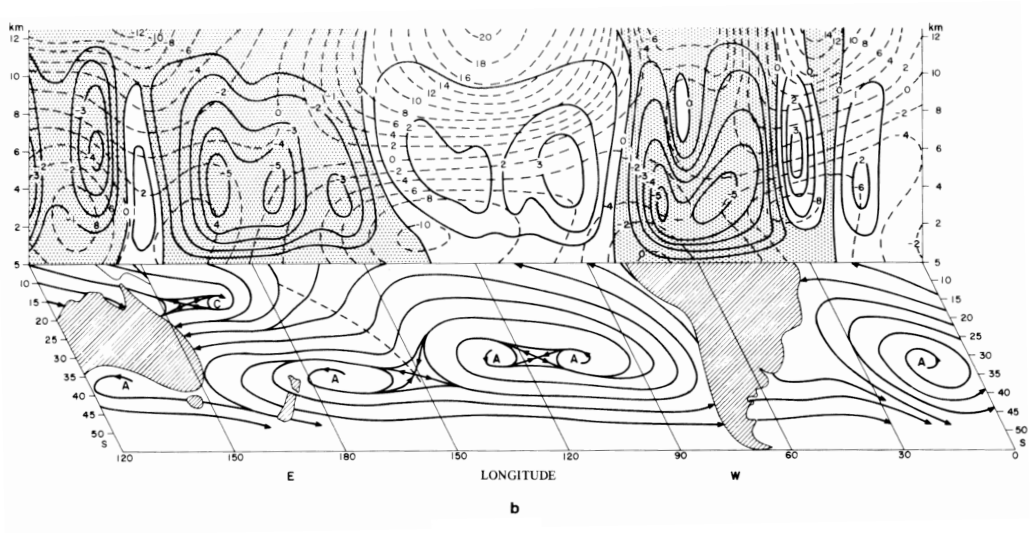


Figure 9.7: The Walker and Hadley circulations. The foreground shows the streamlines of the lower-tropospheric flow for January. The vertical cross section in the background shows the vertical motion (solid lines) and the zonal wind component at 5° S (dashed lines). From Philander (1990). Reprinted with permission from Elsevier.

When meridional mass exchange is considered, this simple picture has to be altered, because absolute angular momentum is exported to adjacent latitudes. Under steady-state conditions, the flux divergence of angular momentum at the equator must be balanced by an easterly surface wind stress. Thus surface easterlies on the equator are stronger than those imposed by the Walker circulation. The net result is that a thermally driven Walker cell is imposed on a background of easterly flow. The intensity of the easterlies depends on the strength of the angular momentum flux divergence.

Fig. 9.8 shows the observed longitude-height cross section of the zonal wind, for January. Fig. 9.9 shows that the 1000-hPa winds above the tropical Pacific (between 10° N and 10° S) have an easterly component in both solstitial seasons. Easterly flow occurs along and near the Equator, west of about 90° W. In January, the easterly component is particularly strong above the central equatorial Pacific, and convergence is evident along the ITCZ near 8° N. In July, there is strong cross-equatorial flow in the eastern Pacific, the convergence zone has moved to about 10°N. At the latitude of the ITCZ, the tradewind flow extends to the east of Central America during both seasons.

Lindzen and Nigam (1987) used a linear model to analyze how SST gradients force low-level winds and convergence in the tropics, in the presence of friction. Linearizing about a state of rest, they found a pressure field that qualitatively resembles the observations, although the wind speeds were unrealistically strong. Neelin et al. (1998) show that the model used by Lindzen and Nigam (1987) is very similar to that of Gill (1980).

Newell et al. (1996) compared water-vapor data from the Upper Atmosphere Research

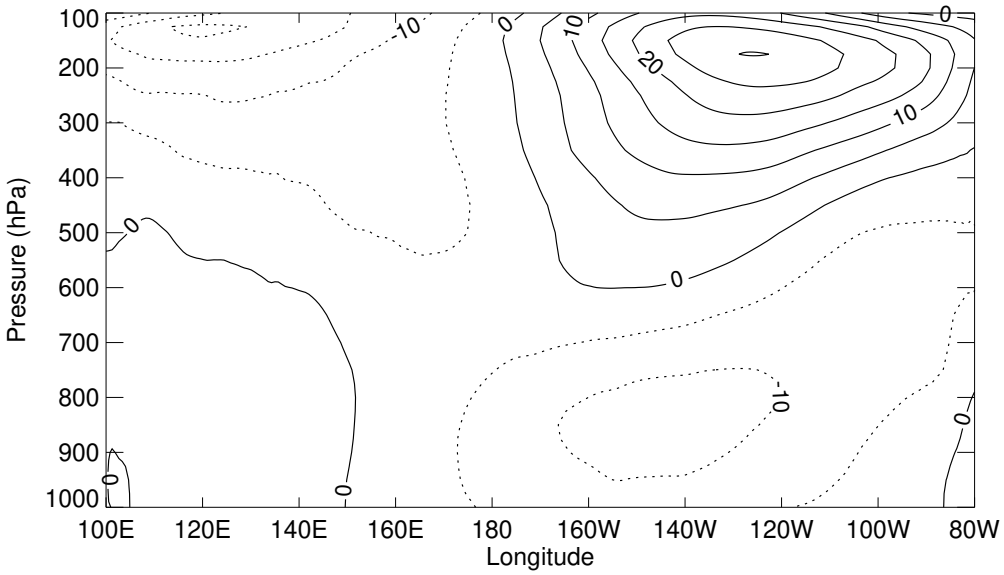


Figure 9.8: The longitude-height cross sections of the zonal wind along the equator, for January, as analyzed by ECMWF. The contour interval is 5 m s^{-1}.

Satellite (UARS) with upper-air wind data from the ECMWF reanalysis dataset to deduce horizontal and vertical motions in the tropical atmosphere. Their results indicate regions of strong ascending motion over the western Pacific Warm Pool and the South Pacific Convergence Zone. The main regions of sinking motion, which are located off South America and extend westward to the dateline just south of the equator, exhibit little seasonal movement. For comparison, Fig. 9.10 shows the vertical velocity fields at 300 hPa for January and July. The South Pacific Convergence Zone is clearly evident in the January data, with a large region of ascending motion that extends southeastward from 145° E to 160° W. Sinking motion straddles the Equator and extends eastward from 160° E. During the northern summer, the Intertropical Convergence Zone is well developed north of the Equator. The general pattern is one in which ascending motion dominates over the tropical western Pacific, while sinking motion occurs over the tropical central and eastern Pacific. Easterlies extend across the equatorial Pacific from South America to 170° W. West of 160° E, the low-level winds are very weak along the Equator, but easterlies span the Pacific at 5° S and 5° N.

Fig. 9.11 completes the picture, showing the upper branch of the Walker circulation. West of the dateline, the zonal winds over the equator are easterly. Upper-level westerly flow occurs to the east of the rising motion. During July 1989, the upper-level flow above the equatorial Pacific ocean is entirely from the east. In the Northern Hemisphere, weak

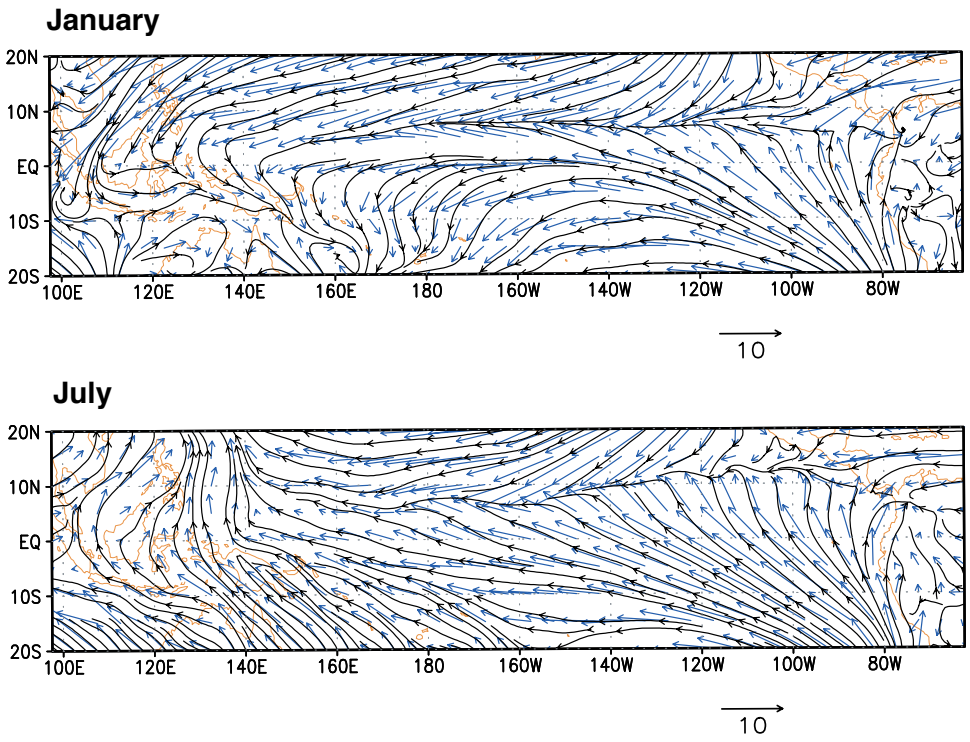


Figure 9.9: Streamlines and horizontal wind vectors for the tropical Pacific at 1000 hPa for January 1989 and July 1989. The units are m s^{-1}.

westerly flow appears between 170° W and 140° W poleward of 15° N. In the Southern Hemisphere, a westerly component of the wind exists south of 5° S to the east of the dateline. An interpretation is that the Walker circulation has migrated into the Southern Hemisphere. A reexamination of Fig. 9.10b indicates that sinking motion is confined mainly to the Southern Hemisphere, and extends as far west as 165° E.

The Walker Circulation is is closely tied to east-west sea surface temperature gradients that are influenced by the Walker Circulation itself, so it can be viewed as a coupled atmosphere-ocean phenomenon Bjerknes (1966, 1969).

The distribution of tropical and subtropical cloudiness is influenced by the large-scale spatial pattern of the SST, rather than the local SST. The tropical Pacific Warm Pool is a region of intense deep convection and heavy rainfall. In Fig. 9.12, regions in which the outgoing longwave radiation (OLR) is less than 225 W m-2 can be identified as areas of frequent convection Webster (1994). The OLR threshold corresponds to a monthly mean emission temperature of 250 K. Due to longwave trapping by optically thick anvil clouds, which are produced by deep convection, the OLR is reduced and threshold values of OLR can therefore be used as surrogates to infer the presence of deep convection. From the figure, we see that deep convection occurs throughout the Warm Pool, and in the South

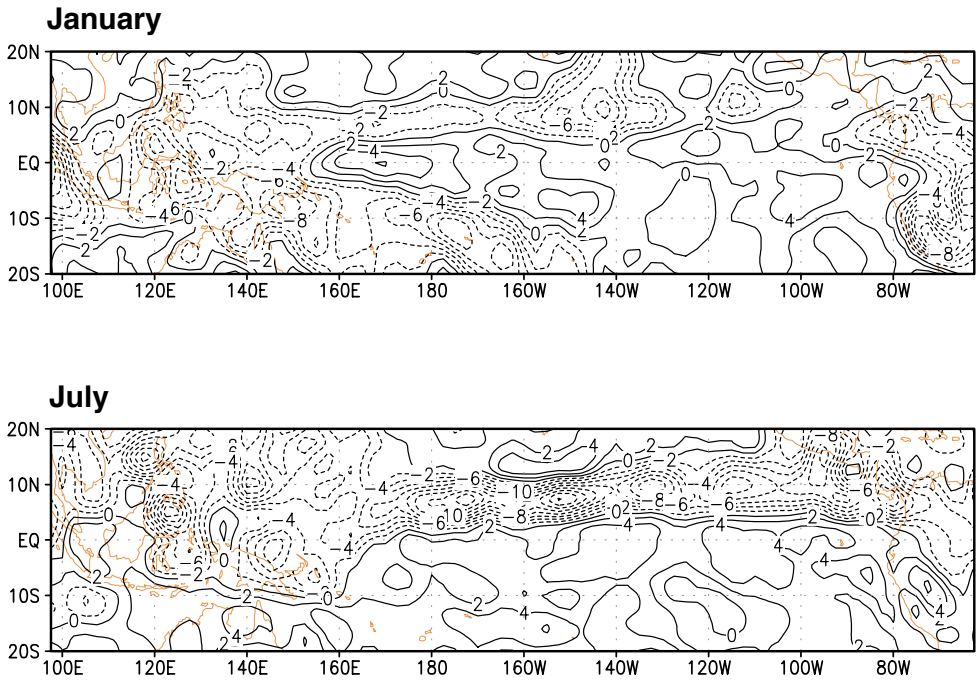


Figure 9.10: Contour plot of mean vertical velocity at 300 hPa (units are 10^{-2} Pa s^{-1}) from the ECMWF reanalysis dataset for January and July 1989. The contour interval is 2 × 10^{-2} Pa s^{-1}; negative contours are dashed.

Pacific convergence zone (SPCZ). On the other hand, the OLR is generally larger than 275 W m-2 across the equatorial cold tongue, indicating that convection is infrequent there.

The high, cold, and sometimes bright clouds of the Warm-Pool region limit the radiative cooling of the atmosphere over the Warm Pool, but they also limit the solar warming of the ocean. Ramanathan and Collins (1993) hypothesized that cirrus clouds act as a thermostat to regulate tropical SST. They used Earth Radiation Budget Experiment (ERBE) data to deduce the inter-relationships among shortwave and longwave cloud radiative forcings and radiative forcing of the clear atmosphere. They emphasized that the shortwave effects of clouds dominate over the longwave effects in regulating SST. According to their hypothesis, as SST increases, the cloud albedo increases. According to their idea, the atmosphere warms as a result of longwave cloud radiative effects, stronger latent-heat release by convection, and a stronger SST gradient over the tropical Pacific. This warming leads to an amplification of the large-scale flux convergence of moisture. The process continues until the reflectivity clouds increases sufficiently to cool the surface. A criticism of their study is that changes in the strength of the ocean and atmosphere circulations were not included. Nevertheless it undoubtedly true that the blocking of shortwave radiation by deep cloud systems tends to limit the sea-surface temperature in the Warm Pool.

As seen in Fig. 9.13, in the eastern tropical Pacific, stratus clouds in the boundary layer

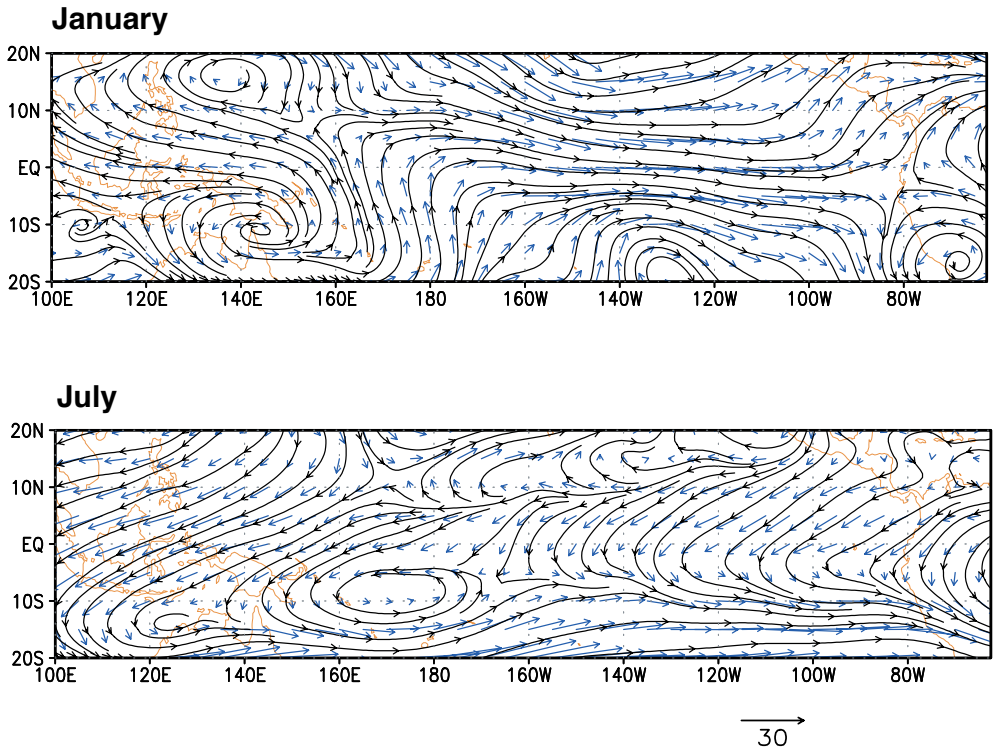


Figure 9.11: Streamlines and horizontal wind vectors for the tropical Pacific at 200 hPa for January 1989 and July 1989.

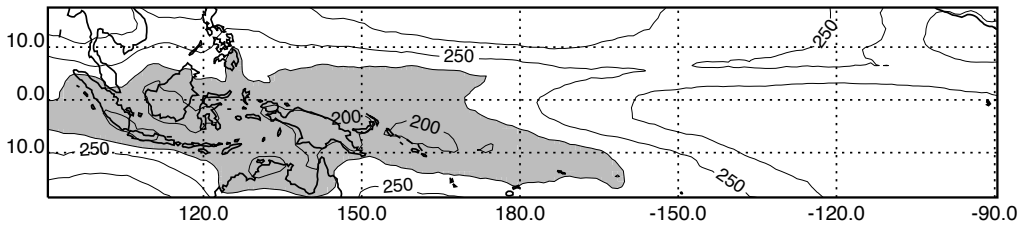


Figure 9.12: Tropical OLR for January averaged over 1985 to 1988. The contour interval is 25 W m^{-2}.

intercept sunlight and strongly reduce the heat flux into the ocean below (e.g., Hartmann et al., 1992). In this way, the atmosphere helps to maintain the cooler SSTs of the eastern Pacific. Stratus clouds form preferentially over cold water (Klein and Hartmann, 1993), so a positive feedback is at work here (Ma et al., 1996). Latent heat exchange between the ocean and atmosphere is influenced by the surface relative humidity and the surface winds. For fixed relative humidity and SST, the evaporative cooling of the ocean increases as the surface wind stress increases. The winds also influence the SST distribution by generating

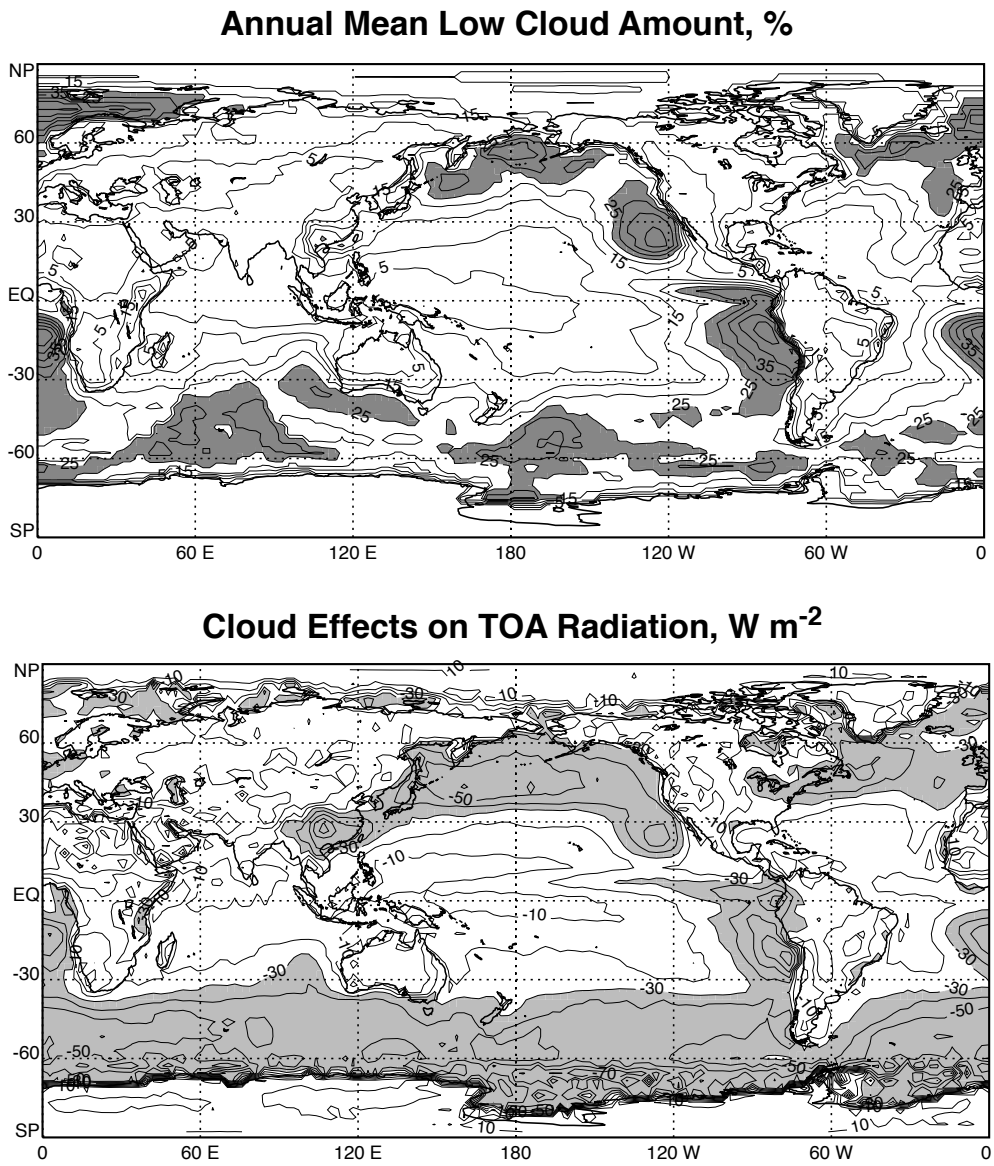


Figure 9.13: Observed annual-mean low-cloud amount (upper panel) and the net effects of clouds on the Earth's radiation budget (lower panel). The contour interval for low cloudiness is 5%; values greater than 25% are shaded. The contour interval for the lower panel is 10 W m^{-2}; values more negative than 30 W m^{-2} are shaded. Negative values in the lower panel indicate a cooling; that is, shortwave reflect

cold-water upwelling in the eastern Pacific, and along the equator in the eastern and central Pacific.

The driving force behind the Walker circulation is the zonally varying heating that is balanced by zonally varying adiabatic heating/cooling due to sinking/rising motions. Over

the warm waters of the western Pacific, latent heat release due to intense convection and radiative warming of the atmospheric column are balanced by the adiabatic cooling associated with rising motion (Webster, 1987). Over the eastern tropical Pacific, where the SST is relatively cold, convection is infrequent, and so a balance between radiative cooling and subsidence prevails.

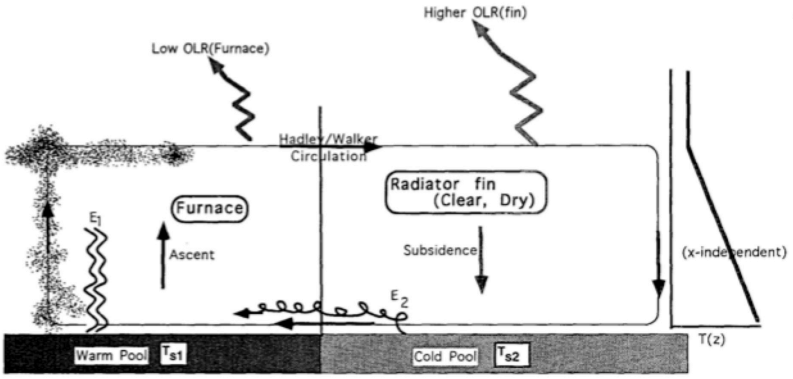


Figure 9.14: Pierrehumbert's (1995) schematic representation of the "furnace/radiator-fin" model of tropical circulation. The symbols E and TS represent the evaporation rate and SST, respectively. The subscript 1 denotes the warm pool or "furnace," and the subscript 2 denotes the cold pool or "radiator fin." American Meteorological Society. Used with permission.

Pierrehumbert (1995; hereafter P95) presented an influential two-box model of the Hadley/Walker circulation. Fig. 9.14 presents a schematic of his "furnace/radiator-fin" model. The model includes separate energy budgets for its Cold-Pool and Warm-Pool "boxes." The SSTs of the Cold Pool and Warm Pool are assumed to be those that give energy balance for each box of the model atmosphere and for the Cold-Pool ocean. Surface energy balance for the Warm Pool was not explicitly included in the model. A vertically and horizontally uniform lapse rate was assumed, and the free-tropospheric temperature profile was assumed to be uniform across the tropics. The radiating temperature of the Cold-Pool free atmosphere was assumed to be the air temperature of the middle troposphere. The solution was obtained by first computing the net energy flux at the top of the Warm-Pool atmosphere for a given SST and relative humidity profile. The positive net radiative flux at top of the atmosphere over the Warm-Pool was assumed to be balanced by horizontal energy transport to the Cold Pool. The Cold-Pool SST and radiating temperature were then computed by assuming that the net diabatic cooling must balance the energy imported from the Warm Pool.

The mass flux of the Hadley-Walker circulation was assumed to be that required to give a balance between adiabatic warming by dry subsidence and the net radiative cooling of the Cold-Pool region. It can be shown that the horizontal heat transport by the Warm-Pool atmosphere is proportional to the diabatic cooling of the Cold-Pool atmosphere and to the ratio of Cold-Pool area and Warm-Pool area. This area ratio is a prescribed parameter of the model. P95 showed that for very small values of the Cold-Pool emissivity, the Warm-

Pool SST increases without limit, because the Cold Pool cannot radiate enough energy to balance the energy absorbed in the Warm Pool. Because the Warm Pool controls the temperature profile, its equilibrium SST must decrease as the Cold-Pool radiating temperature decreases. As the Cold-Pool emissivity increases, the Warm-Pool cools off. The simulated Cold-Pool and Warm-Pool SSTs resemble the present-day climate for a range of conditions. The diagnosed mass flux is realistic.

A weakness of Pierrehumbert's model is that it fails to account for cloud-radiative effects. Miller (1997) extended Pierrehumbert's model by including the radiative effects of low clouds in the Cold-Pool region. He constructed a three-box model with an updraft region, a Warm Pool region, and a Cold Pool region. The model includes energy- and moisture-balance equations for the boundary layer and free troposphere, and a surface energy budget for each of the three boxes. Miller (1997) assumed, following Bjerknes (1938), that the updraft region occupies a small area, He also assumed that the lapse rate of the Warm Pool region is moist adiabatic, and that the free-tropospheric temperature is horizontally uniform across both the Warm and Cold Pool regions (Charney, 1963). His main finding is that low clouds act to reduce the tropical SST.

9.6 Segue

Add text here

Chapter 10

A taxonomy of eddies

10.1 Not all eddies are waves

The purpose of this chapter is to describe the observed climatological distribution of eddy activity, and to offer some theories of the mechanisms that produce both stationary and transient eddies. Eddies are important aspects of the circulation in their own right. They are also of interest because they can affect the zonally averaged flow by producing fluxes and flux convergences. Chapter 11 discusses effects of the eddies on the zonally averaged flow.

As discussed near the beginning of Chapter 6, we use the term "eddies" for all circulation features that vary with longitude. Eddies appear in the winds, temperatures, geopotential heights, surface pressure, water vapor mixing ratio, and all of the other fields that characterize the circulation. Eddies can be produced by a variety of mechanisms (Fig. 10.1), including flow over mountains, baroclinic and barotropic instability, and (possibly convective) heating that is localized at particular longitudes.

Virtually all waves are eddies, but only a subset of eddies are waves: *A wave can be defined as a process that transports energy and momentum without transporting mass.* Consistent with this definition, a particle does not undergo any net displacement during the passage of a small-amplitude wave; it moves in a more-or-less circular path, and returns to its starting point. For example, a waiting surfer gently bobs up and down as small-amplitude waves pass by. Because waves do not transport mass, they cannot spatially redistribute any property that is "attached" to the mass and travels along with it. The surfer doesn't go anywhere. In the atmosphere, quantities that are carried with the mass and so are *not* redistributed by waves include the mixing ratios of water vapor and other minor atmospheric constituents, the potential temperature (under dry adiabatic processes), and the potential vorticity.

We know, however, that a surfer can catch a ride on a large-amplitude wave. This illustrates the fact that large-amplitude (often called "finite-amplitude") waves can transport

mass, and therefore can also transport any intensive property that is "attached" to the mass. Beyond some limit, a large-amplitude wave is not really a wave at all. Finite-amplitude waves are doomed to "break," however; they self-destruct into turbulence. As discussed in Section 4.15.5, mass can also be transported by traveling vortices (e.g., Willoughby, 1998; Provenzale, 1999; Zhang et al., 2014). In summary, some eddies transport mass, but waves don't.



Figure 10.1: A cartoon that appeared in MOREL (1973). The man on the diving board is Jule Charney. The other three people are presumably M.I.T. graduate students of the early 1970s.

The atmosphere supports a very wide variety of wave motions, including Rossby waves, Kelvin waves, inertia-gravity waves, mixed Rossby-gravity waves and of course sound waves. Wavelengths range from the circumference of the Earth, to the width of North America, to a few millimeters. Periods range from thousandths of a second for some sound waves to minutes for some gravity waves to days for baroclinic waves to weeks for some Rossby waves (Williams et al., 2017). The spectral distributions of energy and fluxes over these various scales are complicated and interesting.

Rossby waves, which arise from conservation of potential vorticity on the sphere, are often called "planetary waves," because they depend for their existence on rotation in the presence of spherical geometry, and so are characteristic of planetary atmospheres and oceans, although of course they also occur in stars. Rossby waves propagate westward relative to the mean flow, so that they can be stationary (with respect to the surface) in a westerly regime. They can be excited in many ways, including interactions of the mean flow with mountains, convective events, and instabilities of various kinds. The most energetic Rossby waves have very large horizontal scales. To the extent that they are excited at low levels, their energy propagation can be upward. Upward-propagating Rossby waves are believed to play an important role in sudden stratospheric warmings. Meridionally propagating Rossby waves carry angular momentum between the tropics and middle latitudes. Further discussion is given in Chapter 11.

Gravity waves were discussed briefly in Chapter 8. They depend on the action of the buoyancy force under stable stratification. They occur on many scales. Gravity waves with sufficiently long wavelengths are significantly influenced by rotation; these are called inertia-gravity waves. They can be produced by many mechanisms, including topographic forcing and convection. As discussed near the beginning of Chapter 11, vertically propagating gravity waves produce important vertical momentum transports that affect the large-

scale circulation, especially in the stratosphere and above. Today, the effects of gravity waves on the global circulation is a very active area of research.

The tropical atmosphere is home to two unique classes of equatorially trapped waves. These are the mixed Rossby-gravity waves, also called Yanai waves, which propagate westward; and the Kelvin waves, which propagate eastward. As described in Chapter 11, these two types of waves are thought to drive the Quasi-Biennial Oscillation, although inertia-gravity waves are now known to be important too. Kelvin waves are also believed to play important roles in other major tropical phenomena discussed later in this chapter. Observations and theory also deal with equatorially trapped Rossby waves, which are for the most part similar to mid-latitude Rossby waves.

10.2 Free and forced oscillations of a thin atmosphere on a rotating sphere

10.2.1 The basic state and the linearized governing equations



Figure 10.2: A portrait of Pierre-Simon Laplace, who lived from 1749 to 1827. From `http://en.wikipedia.org/wiki/Pierre-Simon_Laplace`.

We begin with a general introduction to small-amplitude wave motions on the sphere, based on the work of Pierre-Simon Laplace (Fig. 10.2), a French mathematician and astronomer whose name is attached to many important ideas. His remarkably prescient study dealt with the free and forced oscillations of a thin atmosphere on a spherical planet. The work is still very relevant today. Here we summarize Laplace's analysis. More detailed discussions are given by Chapman and Lindzen (1969) and Lindzen (1990).

Laplace considered a spherical planet without mountains, and with a highly idealized basic state:

$$\begin{gathered} \overline{\mathbf{v}} = 0, \quad \overline{\omega} = 0, \\ \frac{\partial \overline{\phi}}{\partial p} = -\overline{\alpha}, \quad p\overline{\alpha} = R\overline{T}(p), \\ p_S = p_0 = \text{constant} \end{gathered} \tag{10.1}$$

Here we are using pressure as our vertical coordinate, and $\overline{T}$ is an arbitrary function of p only. This simple basic state has no meridional temperature gradient and no mean flow. It is, of course, a balanced, self-consistent solution of the equations.

The linearized governing equations in spherical coordinates are

$$\frac{\partial u'}{\partial t} = (2\Omega \sin\varphi)\, v' - \frac{1}{a\cos\varphi}\frac{\partial \phi'}{\partial \lambda}\,, \tag{10.2}$$

$$\frac{\partial v'}{\partial t} = -(2\Omega \sin\varphi)\, u' - \frac{1}{a}\frac{\partial \phi'}{\partial \varphi}\,, \tag{10.3}$$

$$\frac{1}{a\cos\varphi}\left[\frac{\partial u'}{\partial \lambda} + \frac{\partial}{\partial \varphi}\left(v'\cos\varphi\right)\right] + \frac{\partial \omega'}{\partial p} = 0\,, \tag{10.4}$$

$$\frac{\partial}{\partial p}\left(\frac{\partial \phi'}{\partial t}\right) + S_p \omega' = -\kappa\frac{Q'}{p}\,. \tag{10.5}$$

Here

$$S_p \equiv -\frac{\overline{\alpha}}{\overline{\theta}}\frac{\partial \overline{\theta}}{\partial p} \tag{10.6}$$

is the static stability (first defined in Chapter 4), which is assumed to depend only on p; $\kappa \equiv R/c_p$ as usual; and Q is the heating rate. Friction has been neglected in the momentum equations.

The perturbation geopotential, ϕ', is given by

$$\phi' = gz' + \phi'_{\mathrm{ET}}(\lambda, \varphi, t)\,, \tag{10.7}$$

where $\phi_{\rm ET}(\lambda,\varphi,t)$ is the external gravitational tidal potential, due to the Moon and/or Sun, which was discussed in Chapter 4. As mentioned in Chapter 4, the variation of $\phi'_{\rm ET}$ with p is neglected, on the grounds that the atmosphere is thin compared to the distances to the Sun and Moon.

The solutions, and also the externally imposed thermal and gravitational forcing, are assumed to have the separable form

$$\begin{bmatrix} u' \\ v' \\ \omega' \\ \phi' \\ Q' \\ \phi'_{\rm ET} \end{bmatrix} = \sum_n \left\{ \begin{bmatrix} U_n^{\sigma,m}(p) \\ V_n^{\sigma,m}(p) \\ W_n^{\sigma,m}(p) \\ Z_n^{\sigma,m}(p) \\ J_n^{\sigma,m}(p) \\ G_n^{\sigma,m} \end{bmatrix} \Theta_n^{\sigma,m}(\varphi) \right\} \exp\left[i \left(\underbrace{m\lambda + \sigma t}_{phase} \right) \right]. \tag{10.8}$$

where the $\Theta_n^{\sigma,m}(\varphi)$ are as-yet-undetermined functions of latitude only, m is the zonal wave number, assumed to be non-negative, and σ is the frequency. With this convention,

$$\sigma < 0 \rightarrow \text{eastward moving, and } \sigma > 0 \rightarrow \text{westward moving.} \tag{10.9}$$

Eq. (10.8) expresses each field in terms of a function of pressure, multiplied by a function of latitude (the same function in all cases), multiplied by trigonometric functions of longitude and time.

The superscripts σ and m simply denote the particular frequency and zonal wave number associated with each mode. The subscript n is sometimes called the "wave type." It is related to the meridional structure of the solution in a way that will be explained later. There is a solution for each value of n, and the total solution is obtained by summing over n.

As mentioned above, at this point, we do not know what meridional structures are represented by the $\Theta_n^{\sigma,m}(\varphi)$. They are called Hough functions, and will be discussed below. It can be shown that the set $\{\Theta_n^{\sigma,m}(\varphi)\}$ for all n is complete for $-\pi/2 \le \varphi \le \pi/2$. This means that an expansion in terms of Hough functions can represent any meridional structure.

Using (10.8), we can rewrite (10.2) – (10.5) as

$$i\sigma u' - (2\Omega \sin\varphi)\, v' = -\frac{1}{a\cos\varphi} im\phi' \,, \tag{10.10}$$

$$(2\Omega \sin\varphi)\, u' + i\sigma v' = -\frac{1}{a}\frac{\partial \phi'}{\partial \varphi} \,, \tag{10.11}$$

$$\frac{1}{a\cos\varphi}\left[imu' + \frac{\partial}{\partial\varphi}\left(v'\cos\varphi\right)\right] + \frac{\partial \omega'}{\partial p} = 0 \,, \tag{10.12}$$

$$i\sigma\frac{\partial\phi'}{\partial p} + S_p\omega' = -\kappa\frac{Q'}{p} \,. \tag{10.13}$$

Now define the nondimensional frequency

$$\nu \equiv \frac{\sigma}{2\Omega} \,, \tag{10.14}$$

and introduce

$$\mu \equiv \sin\varphi \tag{10.15}$$

as a proxy for φ, so that $d\mu = \cos\varphi\, d\varphi$. Using these definitions, Eqs. (10.10) and (10.11) can be solved for u' and v' in terms of ϕ':

$$u' = \frac{\cos\varphi}{2\Omega a\left(\nu^2 - \mu^2\right)}\left(\mu\frac{\partial}{\partial\mu} - \frac{\nu m}{1-\mu^2}\right)\phi' \,, \tag{10.16}$$

and

$$v' = \frac{i\cos\varphi}{2\Omega a\left(\nu^2 - \mu^2\right)}\left(\nu\frac{\partial}{\partial\mu} - \frac{\mu m}{1-\mu^2}\right)\phi' \,. \tag{10.17}$$

The next step is to use (10.16) and (10.17) to eliminate u' and v' in (10.12). The result can be written as

$$\frac{i\nu}{2\Omega a^2} F\left(\phi'\right) + \frac{\partial \omega'}{\partial p} = 0\,. \tag{10.18}$$

Here F is a complicated linear operator, defined by

$$\begin{aligned} F \equiv & \left(\frac{m}{\nu^2-\mu^2}\right)\left(\frac{\mu}{\nu}\frac{\partial}{\partial\mu} - \frac{m}{1-\mu^2}\right) + \frac{\partial}{\partial\mu}\left[\left(\frac{1-\mu^2}{\nu^2-\mu^2}\right)\frac{\partial}{\partial\mu}\right] - \frac{\partial}{\partial\mu}\left[\frac{\mu m}{(\nu^2-\mu^2)\,\nu}\right] \\ = & \frac{\partial}{\partial\mu}\left[\left(\frac{1-\mu^2}{\nu^2-\mu^2}\right)\frac{\partial}{\partial\mu}\right] - \left(\frac{1}{\nu^2-\mu^2}\right)\left[\frac{m}{\nu}\left(\frac{\nu^2+\mu^2}{\nu^2-\mu^2}\right) + \frac{m^2}{1-\mu^2}\right]\,. \end{aligned} \tag{10.19}$$

On the second line of (10.19), all derivatives have been collected in the first term. Note that F involves both first and second derivatives with respect to μ.

At this point we have a system of two equations, namely (10.13) and (10.18), for the two unknowns ϕ' and ω'. Elimination of ϕ' gives an equation that can be solved for the single unknown ω':

$$4\Omega^2 a^2 \frac{\partial^2 \omega'}{\partial p^2} - S_p F\left(\omega'\right) = \frac{\kappa}{p}\, F\left(Q'\right)\,. \tag{10.20}$$

Finally, we use (10.8) again, to rewrite (10.20) as

$$\begin{aligned} 4\Omega^2 a^2 \frac{\partial^2}{\partial p^2}\left[W_n^{\sigma,m}(p)\;\Theta_n^{\sigma,m}(\mu)\right] - S_p F\left[W_n^{\sigma,m}(p)\;\Theta_n^{\sigma,m}(\mu)\right] \\ = \frac{\kappa}{p}\, F\left[J_n^{\sigma,m}(p)\,\Theta_n^{\sigma,m}(\mu)\right]\,. \end{aligned} \tag{10.21}$$

Here we have assumed that the $\Theta_n^{\sigma,m}(\mu)$ are linearly independent. Using the facts that $W_n^{\sigma,m}$ and $J_n^{\sigma,m}$ depend only on p, and $\Theta_n^{\sigma,m}$ depends only on μ, we can simplify and rearrange (10.21) to

$$\frac{\partial^2 W_n^{\sigma,m}/\partial p^2}{\kappa\left(J_n^{\sigma,m}/p\right)+S_p W_n^{\sigma,m}} = \frac{F\left(\Theta_n^{\sigma,m}\right)}{4\Omega^2 a^2 \Theta_n^{\sigma,m}} \,. \tag{10.22}$$

The left side of (10.22) depends only on pressure, and the right side depends only on latitude. It follows that the two sides must both be equal to a (dimensional) constant, which we take to be $-(gh_n)^{-1}$, where h_n has units of length and is called an equivalent depth. *Note that h_n has entered the problem as a separation constant.*

We can now split (10.22) into the following two equations:

$$\boxed{F\left(\Theta_n^{\sigma,m}\right)+\varepsilon_n \Theta_n^{\sigma,m}=0}\,, \tag{10.23}$$

$$\boxed{\frac{d^2 W_n^{\sigma,m}}{dp^2}+\frac{S_p}{gh_n}W_n^{\sigma,m} = -\frac{\kappa}{gh_n}\left(\frac{J_n^{\sigma,m}}{p}\right)}\,. \tag{10.24}$$

Eq. (10.23) can be called the meridional structure equation, and it is also called the Laplace Tidal Equation, or LTE. Eq. (10.24) can be called the vertical structure equation. In (10.23), we have introduced the nondimensional quantity

$$\varepsilon_n \equiv \frac{4\Omega^2 a^2}{gh_n}\,, \tag{10.25}$$

which is called Lamb's parameter (or sometimes "the terrestrial constant"). All information about the planet's radius, rotation rate, and gravity is encoded in ε_n.

In the derivation above, we have used our assumptions that the basic state is at rest, and that the basic-state temperature depends only on pressure. Separation of variables is not possible if the basic state is made more realistic, e.g., if the observed zonally averaged winds are used.

In the classical literature on the atmospheric tides, and in the current discussion, the heating amplitude $J_n^{\sigma,m}$ is treated as a known quantity, so that (10.24) contains the single unknown $W_n^{\sigma,m}$. This assumption that $J_n^{\sigma,m}$ is known makes sense only if $J_n^{\sigma,m}$ is at least approximately independent of the motion. Such an assumption is defensible (but approximate) for heating due to the absorption of solar radiation by stratospheric ozone. It would be completely nonsensical for cumulus heating.

10.2.2 Boundary conditions

The LTE is a second-order ordinary differential equation, so two boundary conditions are needed. Perhaps surprisingly, it suffices to assume that the $\Theta_n^{\sigma,m}(\mu)$ are bounded at the poles, i.e., at $\mu = \pm 1$. If $\widehat{h}$ is given, nontrivial solutions of (10.23) exist only when ν, m, and n satisfy a dispersion equation. These nontrivial solutions are called Hough functions, after Sydney Samuel Hough, who was a British applied mathematician and astronomer. The Hough functions were tabulated by Longuet-Higgins (1968). You should look that paper up; from the perspective of the 21st century, it is strange to see a journal article that contains page after page of tables of numbers.

The vertical structure equation, (10.24), is a second-order ordinary differential equation for $W_n^{\sigma,m}$, so it also needs two boundary conditions. At the top of the atmosphere we use

$$W_n^{\sigma,m} = 0 \quad \text{at} \quad p = 0\,. \tag{10.26}$$

This is exact. The exact lower boundary condition (in the absence of mountains) is $w \equiv Dz/Dt = 0$ at $p = p_S(\lambda,\varphi,t)$. This is complicated because we expect p_S to be a function of longitude, latitude, and time. As a simplifying approximation, we use the linearized lower boundary condition

$$\frac{Dz'}{Dt} \cong \left(\frac{\partial z'}{\partial t}\right)_p + \omega' \frac{\partial \overline{z}}{\partial p} = 0 \quad \text{at} \quad p = p_0\,. \tag{10.27}$$

Here p_0 is the spatially and temporally constant value of p_S in the basic state. Because

$$gz' = \phi' - \phi'_{\rm ET}(\lambda,\varphi,t)\ , \tag{10.28}$$

where $\phi'_{\rm ET}(\lambda,\varphi,t)$ is known, and using the hydrostaticity of the basic state, as expressed by

$$g\left(\frac{\partial \overline{z}}{\partial p}\right)_{p=p_0} = -\overline{\alpha_0} = -\frac{R\overline{T}_0}{p_0} \equiv -g\frac{H_0}{p_0}\,, \tag{10.29}$$

we can rewrite the linearized lower boundary condition (10.27) as

$$\frac{\partial \phi'}{\partial t} - \omega' \frac{g H_0}{p_0} \cong \frac{\partial \phi'_{\rm ET}}{\partial t} \qquad \text{at } p = p_0 \,, \tag{10.30}$$

where $H_0 \equiv RT_0/g$ has dimensions of length and is called the "scale height." Eq. (10.30) involves both ϕ' and ω'. After some additional algebra to eliminate ϕ' in (10.30), we can finally express the lower boundary condition entirely in terms of $W_n^{\sigma,m}$:

$$\boxed{\frac{dW_n^{\sigma,m}}{dp} - \frac{H_0}{h_n}\frac{W_n^{\sigma,m}}{p_0} = \frac{i\sigma}{g h_n} G_n^{\sigma,m} \text{ at } p = p_0 \,.} \tag{10.31}$$

Note that the gravitational forcing, $G_n^{\sigma,m}$, enters the problem through the lower boundary condition on the vertical structure equation. The thermal forcing enters through the vertical structure equation itself. Neither type of forcing appears directly in the LTE.

10.2.3 Solution procedure

Chapman and Lindzen (1969) remark that "there are, in general, no 'obvious' closed solutions" of the LTE. Approximate solutions cane be found using series expansions or numerical methods. Hough (1898) used an expansion in terms of the "associated Legendre functions of order n and rank m," which are given by

$$\begin{aligned} P_n^m(\mu) = \frac{(2n)!}{2^n n!\,(n-m)!}\left(1-\mu^2\right)^{m/2}\Bigg[\mu^{n-m} - \frac{(n-m)(n-m-1)}{2(2n-1)}\mu^{n-m-2} \\ + \frac{(n-m)(n-m-1)(n-m-2)(n-m-3)}{2\cdot 4(2n-1)(2n-3)}\mu^{n-m-4} - \ldots\Bigg] \,. \end{aligned} \tag{10.32}$$

For each value of n, we allow all (non-negative) values of m up to and including n. We require

$$n \geq m \,, \tag{10.33}$$

which means that the expression inside the square brackets is a polynomial in μ. If m is even, then the full expression for $P_n^m(\mu)$ is a polynomial in μ. Notice that for $\mu = \pm 1$ (i.e., at the poles) the associated Legendre functions are equal to zero, for all values of n and m. The expansion used by Hough (1898) takes the form

$$\Theta_n^{\sigma,m}(\mu) = \sum_{j=m}^{\infty} C_{j,n} P_n^m(\mu) \ . \tag{10.34}$$

Chapman and Lindzen (1969) explain how to determine the $C_{j,n}$.

10.2.4 Free oscillations

A forced oscillation is *externally* driven. As discussed above, Laplace considered both thermal and gravitational forcing. The gravitational forcing is clearly external to the atmosphere, but as discussed later only a few types of thermal forcing can only be considered even approximately external. As an example, if $J_n^{\sigma,m} \neq 0$ then (10.24) can only be satisfied if $W_n^{\sigma,m}$ is different from zero; we say that $W_n^{\sigma,m}$ is "forced" to be different from zero.

A free oscillation is one for which there is no forcing; in this context, "free" means "not forced." A free oscillation is analogous to the persistent ringing of a bell. The tone of the bell is determined by the bell's composition and shape, and is (nearly) independent of how the tone is excited. The free oscillations of the atmosphere are determined by a, Ω, g, κ, and S_p.

The LTE and its boundary conditions are satisfied by the trivial solution $\Theta_n^{\sigma,m}(\varphi) \equiv 0$. Non-trivial solutions do exist, but only for particular choices of the parameters ε_n and ν. The method of solution differs depending on whether or not forcing is included.

The equations for free oscillations can be solved using the following steps:

1. Solve the vertical structure equation. For free oscillations, the vertical structure equation (10.24) reduces to

$$\frac{d^2 W_n^{\sigma,m}}{dp^2} + \frac{S_p}{g h_n} W_n^{\sigma,m} = 0 \ . \tag{10.35}$$

   Since there is no gravitation forcing of free oscillations, the lower boundary condition (10.31) can be simplified to

$$\frac{dW_n^{\sigma,m}}{dp} - \frac{H_0}{h_n} \frac{W_n^{\sigma,m}}{p_0} = 0 \text{ at } p = p_0 \ . \tag{10.36}$$

   We also have

$$W_n^{\sigma,m} = 0 \text{ at } p = 0 \ . \tag{10.37}$$

This yields the equivalent depths, h_n, as eigenvalues. An example is given below. For some simple cases there is a single equivalent depth. In other cases there are infinitely many equivalent depths.

2. Use the h_n in (10.25) to evaluate the ε_n.

3. Solve the LTE for each value of ε_n, with the nondimensional frequencies, ν, as eigenvalues. Note that ν appears in the definition of F. The nondimensional frequencies will depend on both ε_n and the zonal wave number, m, which also appears in the definition of F. Examples are given below.

Since both the vertical structure equation and the LTE are eigenvalue problems, the solution for free oscillations is a double eigenvalue problem.

As discussed above, the system (10.35)-(10.37) has non-trivial solutions only for special values of the h_n. For $h_n \neq \widehat{h}_n$, the only solution is $W_n^{\sigma,m}(p) \equiv 0$. To find these eigenvalues and the corresponding nontrivial solutions for $W_n^{\sigma,m}(p)$, we have to specify the static stability S_p as a function of height. Different choices for S_p will give different h_n and different $W_n^{\sigma,m}$.

As a simple example, suppose that $S_p = 0$, which means that the potential temperature is uniform with height. An isentropic atmosphere cannot support internal gravity waves, because when $S_p = 0$ a vertically displaced particle does not experience a gravitational restoring force (see the discussion at the beginning of Chapter 8). A stratified (i.e., non-isentropic) atmosphere can support both external and internal gravity waves. For this isentropic case, the vertical structure equation, (10.35), reduces to

$$\frac{d^2 W_n^{\sigma,m}}{dp^2} = 0\,. \tag{10.38}$$

A solution of (10.38) that is consistent with the upper boundary condition (10.37) is

$$W_n^{\sigma,m} = A_n^{\sigma,m} p, \tag{10.39}$$

where $A_n^{\sigma,m}$ is an arbitrary constant. Use of (10.39) in the lower boundary condition (10.36) gives

$$h_n = H_0\,. \tag{10.40}$$

This is the only possible equivalent depth for free oscillations of an isentropic atmosphere.

The procedure used above can also be used to find the equivalent depth(s) associated with other stratifications. In one of the problems at the end of this chapter, you are asked to find the equivalent depth of an isothermal atmosphere.

10.2.5 Forced oscillations

Atmospheric tides are examples of forced oscillations, in which the forcing is either thermal or gravitational. A key point is that the frequency and zonal wave number of the tidal forcing are known. Within the limits of applicability of the linear equations, *the frequency and zonal wave number of the forced solution must be the same as those of the forcing*. It is of course possible for free oscillations and forced oscillations to coexist, in superposition.

The solution procedure for forced oscillations is:

1. Solve the LTE (the meridional structure equation) as an eigenvalue problem for the equivalent depths.
2. Expand the gravitational and thermal forcing in terms of Hough functions, to obtain their vertical structures $G_n^{\sigma,m}(p)$ and $J_n^{\sigma,m}(p)$, respectively.
3. For each wave type (i.e., each Hough function), solve the vertical structure equation for the response to the forcing.

For further discussion, see Lindzen (1966), Chapman and Lindzen (1969), Forbes and Garrett (1979), and Lindzen (1990).

10.2.6 The scope of Laplace's analysis

The solutions found by Laplace cover an incredibly wide variety of atmospheric phenomena, many of which were not recognized until long after Laplace's study was published. For example, Rossby waves were discovered by Carl Gustav Rossby (1939), but Rossby-Haurwitz waves, which are Rossby waves in spherical geometry, are among the solutions of Laplace's equations, which he published in 1799. The existence of Rossby waves was not recognized until Rossby found them on a midlatitude β-plane. Similarly, the several families of equatorial waves discovered by Matsuno (1966) are are among the solutions of Laplace's equations. Their existence was not recognized until Matsuno found them on an equatorial β-plane. In retrospect, Laplace's theory serves as a unifying framework that places the work of Rossby, Matsuno, and others into a broader context.

In the remainder of this chapter, we will explore various types of eddies, most of which correspond to solutions of Laplace's equations.

10.2.7 Solving the LTE for the case of no rotation

If there is no rotation, then $\nu = \sigma/(2\Omega) \to \infty$ and $\varepsilon_n \to 0$. We continue to assume that $S_p = 0$, so that (10.40) applies. For this case, we find that

$$\nu^2 F \to \frac{d}{d\mu}\left[\left(1-\mu^2\right)\frac{d}{d\mu}\right] - \frac{m^2}{1-\mu^2}\,, \tag{10.41}$$

and

$$\nu^2 \varepsilon_n \to \frac{\sigma^2 a^2}{g h_n} = \frac{\sigma^2 a^2}{g H_0}\,. \tag{10.42}$$

Then the LTE can be written as

$$\frac{d}{d\mu}\left[\left(1-\mu^2\right)\frac{d\Theta_n^{\sigma,m}}{d\mu}\right] + \left(\frac{\sigma^2 a^2}{g H_0} - \frac{m^2}{1-\mu^2}\right)\Theta_n^{\sigma,m} = 0\,. \tag{10.43}$$

It can be shown that (10.43) has solutions that are bounded at the poles if and only if

$$\frac{\sigma^2}{g H_0} = \frac{n(n+1)}{a^2}, \quad n = 1,2,3\ldots\,. \tag{10.44}$$

This is how n enters the problem. The requirement that n be a non-negative integer ensures that the solutions "fit" on the sphere. This is analogous to the requirement that m be an integer, which ensures that the longitudinal structure (sines and cosines) "fits" on a periodic latitude circle. The case $n = 0$ implies a globally uniform field, so it is trivial and uninteresting. The quantity $n(n+1)/a^2$ Eq. (10.44) is essentially the square of the total horizontal wave number. Eq. (10.44) is a dispersion equation that relates frequency to this total wave number.

Eq. (10.44) says that for a stationary solution with $\sigma = 0$, we would need $n = 0$; without rotation, there are no non-trivial stationary solutions. Note that m does not appear in the dispersion equation for this non-rotating case.

The eigenfunctions for this non-rotating case turn out to be the so-called "associated Legendre functions of order n and rank m," denoted by

$$\Theta_n^{\sigma,m} = P_n^m(\mu), \text{ for } n \geq m . \tag{10.45}$$

In other words, for the special case of no rotation, the Hough functions reduce to the associated Legrendre functions. Substitution can be used to confirm that the $P_n^m(\mu)$ really are solutions of (10.43).

By combining (10.45) with the longitudinal structure shown in (10.8), we obtain the two-dimensional horizontal structure of the waves:

$$Y_n^m(\mu,\lambda) = P_n^m(\mu)\exp(im\lambda) . \tag{10.46}$$

As discussed in Appendix H, the $Y_n^m(\mu,\lambda)$ are spherical harmonics. Here n is the total number of nodal circles, m is the zonal wave number, and $n-m$ is the number of nodes in the meridional direction, also known as the "meridional nodal number." (Try saying that five times fast.) Further discussion is given in Appendix H, which is an introduction to spherical harmonics. These solutions are external gravity waves. They are called "external" because they have no nodes in the vertical and do not propagate vertically. By rearranging (10.44), the frequencies of the external gravity waves can be written as

$$\sigma = \pm\frac{\sqrt{n(n+1)gH_0}}{a} . \tag{10.47}$$

These frequencies depend on the wave's horizontal scale through the two-dimensional index, n, but as mentioned above they are independent of m. For example, for $n = 1$, m can be either 0 or 1 (because we allow all m in the range $n \geq m \geq 0$), but both modes have the same frequency. We show below that this is not true when rotation is present, because with rotation the zonal direction, in which scale is measured by m, is physically different from the meridional direction, in which scale is measured by $n-m$.

10.2.8 With rotation: Free Oscillations of the First and Second Kinds

Now consider $\Omega \neq 0$, still for an isentropic atmosphere, and neglect all details. We will obtain the solution more directly, without using the LTE or the vertical structure equation. Define a stream function ψ and a velocity potential χ, so that

$$
\begin{aligned}
u' &= -\frac{1}{a}\left(\frac{\partial \psi}{\partial \varphi}\right)_p + \frac{1}{a\cos\varphi}\left(\frac{\partial \chi}{\partial \lambda}\right)_p, \\
v' &= \frac{1}{a\cos\varphi}\left(\frac{\partial \psi}{\partial \lambda}\right)_p + \frac{1}{a}\left(\frac{\partial \chi}{\partial \varphi}\right)_p .
\end{aligned}
\tag{10.48}
$$

The vorticity is then $\zeta_p = \mathbf{k}\cdot(\nabla_p \times \mathbf{v}_h) = \nabla_p^2\psi$, and the divergence is $\delta_p = \nabla_p \cdot \mathbf{v}_h = \nabla_p^2\chi$. The equation of horizontal motion can be differentiated to obtain

$$
\frac{\partial}{\partial t}\nabla_p^2\psi + \beta v' + f\nabla_p^2\chi = 0 \tag{10.49}
$$

(the vorticity equation) and

$$
\frac{\partial}{\partial t}\nabla_p^2\chi + \beta u' - f\nabla_p^2\psi = -g\nabla_p^2 z' \tag{10.50}
$$

(the divergence equation), where

$$
\beta \equiv \frac{1}{a}\frac{df}{d\varphi} = \frac{2\Omega\cos\varphi}{a} \tag{10.51}
$$

is the rate of change of the Coriolis parameter with latitude. We can also show (see the problems at the end of this chapter) that for the special case of an isentropic atmosphere

$$
\frac{\partial z'}{\partial t} + H_0\nabla_p^2\chi = 0, \text{ and } \psi \cong \text{constant} . \tag{10.52}
$$

Equations (10.48) through (10.52) form a closed set that can be solved for ψ, χ, and z', as well as u' and v'.

Recall that, without rotation, there are no nontrivial stationary solutions. From (10.49) and (10.52), you should be able to see that when the atmosphere is rotating, stationary motion cannot exist unless $v' = 0$. For nontrivial stationary motion with $v' = 0$, it follows from the equation of zonal motion, (10.2), that $m = 0$, i.e., the motion must be both purely zonal (because $v' = 0$) and zonally uniform. In other words, for a rotating planet, the

only stationary solutions are zonally uniform zonal currents. Observations do show very persistent zonal currents in the Earth's atmosphere, and also in the atmospheres of the "gas giant" planets of our solar system, although of course they are not truly steady.

Margules (1893) and Hough (1898) showed that the LTE has two classes of solutions for free oscillations, which they named Free Oscillations of the First and Second Class. For the case of small ε_n, which means weak rotation or large equivalent depth, approximate solutions of (10.49) and (10.50) can be obtained by expanding in spherical harmonics Longuet-Higgins (1968). These Free Oscillations of the First Class (FOFC) are "inertia-gravity waves," which are gravity waves modified by the effects of rotation.

For the case of no rotation, we have pure gravity waves, which are governed by

$$\chi \cong A_n^m P_n^m(\mu)\, e^{i(m\lambda+\sigma t)} \text{ (irrotational)}\,, \tag{10.53}$$

and the dispersion equation (10.47). Haurwitz (1937) derived a more accurate (but approximate) expression that includes the effects of rotation:

$$\sigma \cong \frac{\Omega m}{n(n+1)} \pm \sqrt{\frac{\Omega^2 m^2}{n^2(n+1)^2} + n(n+1)\frac{gH_0}{a^2}}\,. \tag{10.54}$$

This can be compared with (10.47). The additional terms in (10.54) vanish for $\Omega = 0$. For large n, (10.54) reduces to (10.47), which means that waves with small horizontal scales are hardly affected by rotation. As an example, for $n \geq 4$, the error that comes from using (10.47) instead of (10.54) is less than 1%. From (10.54) we see that eastward-propagating inertia-gravity waves have frequencies slightly different from those of westward-propagating inertia-gravity waves with the same m and n. The difference is due to rotation; after all, it is planetary rotation that gives meaning to the words "east" and "west."

The Free Oscillations of the Second Kind (FOSC) are also called Rossby-Haurwitz waves, or planetary waves. They are essentially Rossby waves in spherical geometry. They are almost nondivergent, and approximate solutions can be found by assuming $\chi =$ constant (i.e., strict nondivergence) from the beginning. For nondivergent flow, the vorticity equation (10.49) reduces to

$$\frac{\partial}{\partial t}\nabla^2\psi + \frac{2\Omega}{a^2}\frac{\partial\psi}{\partial\lambda} = 0. \tag{10.55}$$

Using the form of the Laplacian in spherical coordinates, and assuming solutions of the form

$$\psi = \widehat{\psi} e^{i(\sigma + m\lambda)} \,, \tag{10.56}$$

(10.55) can be written as

$$i\sigma \left\{ \frac{1}{a^2 \cos^2\varphi} \left[\cos\varphi \frac{d}{d\varphi} \left(\cos\varphi \frac{\partial}{\partial \varphi} \right) - m^2 \right] \right\} \widehat{\psi} + \frac{2\Omega i m}{a^2} \widehat{\psi} = 0 \,, \tag{10.57}$$

or

$$\left\{ \frac{d}{d\mu} \left[\left(1-\mu^2\right) \frac{d}{d\mu} \right] + \left(\frac{2\Omega m}{\sigma} - \frac{m^2}{1-\mu^2} \right) \right\} \widehat{\psi} = 0 \,, \tag{10.58}$$

which is very similar to (10.43). It can be shown that non-trivial solutions of (10.57) exist when $2\Omega m/\sigma = n(n+1)$, which should be compared with (10.44) and (10.54). The solutions take the form $\widehat{\psi} \sim P_n^m$, and the frequency is given by

$$\boxed{\sigma \cong \frac{2\Omega m}{n(n+1)} > 0 \,.} \tag{10.59}$$

Recall that a positive frequency implies westward propagation. These Rossby-Haurwitz waves always propagate westward (relative to the mean zonal wind), at the rate $\dfrac{-2\Omega}{n(n+1)}$; note that this is in angular propagation per unit time, rather than distance per unit time. Longer waves, with smaller values of m, propagate faster. Eq. (10.59) shows that for Rossby-Haurwitz waves, unlike pure gravity waves, σ does depend explicitly on the zonal wave number, m. This is natural, because the very existence of Rossby-Haurwitz waves depends on rotation, as can be seen from the factor of Ω in the numerator on the right-hand side of (10.59).

A direct test of the theory of non-divergent Rossby waves was made by Eliasen and Machenhauer (1965) and Deland (1965). They used spherical harmonics to analyze the 500 hPa stream function, isolating transient waves by taking the difference in 24 hours. Their results, illustrated in Fig. 10.3, show westward propagation, as expected. Table 10.1 compares the computed and observed phase speeds, in degrees of longitude per day. The

model over-predicts the westward phase speeds for low zonal wave numbers, due to the neglect of the effects of the divergence of the horizontal winds.

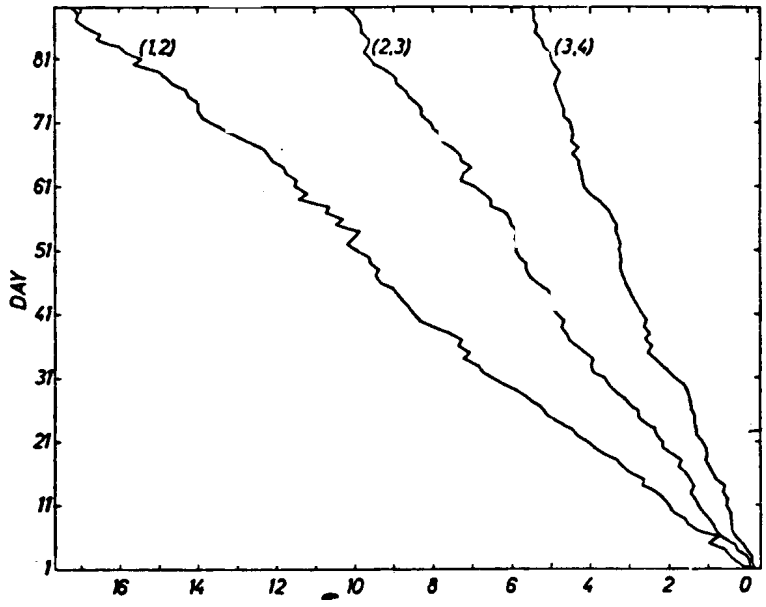


Figure 10.3: Successive daily values of the phase angle for the 24 hour tendency field, at the 500 hPa level for the components $(m,n) = (1,2),(2,3),(3,4)$, during the 90 day period beginning 1 December 1956. The ordinate is time, in days. The abscissa represents the number of westward circulations around the Earth after the first passage of the Greenwich meridian, and so it is a measure of longitudinal phase propagation. From Eliasen and Machenhauer (1965).

In the Earth's atmosphere, Rossby waves can be excited in a variety of ways, including flow over topography (discussed later in this chapter), baroclinic instability, and cumulus heating. Once excited, the waves can propagate zonally, meridionally, and vertically.

10.3 Observations of stationary and transient eddies in middle latitudes

Blackmon (1976) discussed some very basic features of the observed eddy activity in the Northern Hemisphere, as seen in the 500 hPa geopotential height. He used a ten-year record, and considered both summer and winter conditions. The data were available twice per day, at 00 Z and 12 Z.

Blackmon filtered the data in both space and time, in order to isolate particular space and time scales. He expanded the height fields into spherical harmonics Y_n^m, where superscript m denotes the zonal wave number, and subscript n denotes the "two dimensional index." As already discussed earlier in this chapter, the number of nodes in the meridional

Table 10.1: A comparison of the observed and computed westward phase speeds of selected planetary waves, from Eliasen and Machenhauer (1965). The units of the phase speed are degrees of longitude per day, towards the west relative to the mean flow. The computed phase speeds are for non-divergent Rossby-Haurwitz waves.

m,n	Observed phase speed, deg day^{-1}	Computed phase speed, deg day^{-1}
1,2	70	115
2,3	40	53
3,4	20	28
1,4	20	28
2,5	12	16
3,6	8	9

direction is $n-m \geq 0$, and $n \geq m$ is required. The largest value of n considered by Blackmon was $n = 18$. Note that $m = 18$ corresponds to a zonal wave length of 20° of longitude, which in middle latitudes corresponds to a wavelength of roughly 1000 km.

As discussed in Appendix H, the spherical harmonics form a complete orthonormal basis that can be used to represent an arbitrary function on the sphere:

$$Z(\lambda,\varphi) = \sum_{m=-M}^{M} \left(\sum_{n=|m|}^{N(m)} C_n^m Y_n^m \right). \tag{10.60}$$

Here the C_n^m are the expansion coefficients, and M is a suitably chosen positive integer. Larger values of M allow a more detailed representation of $Z(\lambda,\varphi)$. Blackmon defined three categories of spatial scales, based on the two-dimensional index:

Regime I: $0 \leq n \leq 6$, or "long waves;"

Regime II: $7 \leq n \leq 12$, or "medium-scale waves;" and

Regime III: $13 \leq n \leq 18$, or "short waves."

Because Blackmon's truncation scheme is based on the two-dimensional index, a particular wave can have nodes in the zonal direction or in the meridional direction; most have both. Note that for all three regimes $0 \leq m \leq n$. All three regimes therefore contain modes with small values of m, i.e. long zonal scales.

The expansion coefficients for each set of waves were determined for each observation time. Then they were filtered in time, using three filters:

Low pass: Admits periods in the range longer than or equal to 10 days;

Medium-pass: Admits periods in the range 2.5 days to 6 days;

High-pass: Admits periods in the range 1 day to 2 days.

Note that a period of one day represents the most rapidly fluctuating wave that can be captured by the twice-a-day data that Blackmon used.

The lower panel of Fig. 10.4 shows the time-averaged 500 hPa height averaged over nine winters. The features seen here are stationary waves. Note the two prominent troughs, one near the east coast of North America, and the other near Japan. These same features can be seen in the figures of Chapter 3. The upper panel of Fig. 10.4 shows the total root-mean-square (rms) geopotential height, for winter, without time or space filtering. There are three prominent "centers of action," in the North Pacific, the North Atlantic, and over Siberia.

Fig. 10.5 shows low-pass filtered (long period) rms winter heights, for all spatial scales (top left), for Regime I (top right), for Regime II (bottom left), and for Regime III (bottom right). It is clear that very little contribution comes from Regime III. Both Regime I and Regime II contribute significantly. The Regime II contribution shows maxima in regions where blocking commonly occurs. Blocking will be discussed later.

Fig. 10.6 is similar to Fig. 10.5, but for medium-pass eddies, i.e. those of "synoptic" periods in the range of 2.5 to 6 days. For this range of periods, most of the action comes from Regime II and Regime III; the long waves do not contribute much. Note, however, that the contour interval is smaller than that used in Fig. 10.5, so that the total amount of activity portrayed in Fig. 10.6 is less than in Fig. 10.5. The strong signal seen over the North Pacific and North Atlantic Oceans in Fig. 10.6 is associated with the "storm tracks." Winter storms often form on the western sides of the ocean basins, near the east coasts of the continents, where the horizontal temperature gradient is strong. The storms move eastward and poleward, creating the tracks seen in the figure. Further discussion is given later in this chapter.

Even less power resides in the high-pass eddies (not shown).

Table 10.2 shows that in winter the greatest power is found along a "ridge" running from lower left (small n, low frequency) to upper right (large , high frequency). By far the largest power occurs in the low frequencies and for two-dimensional index on the order of 6. The diurnal tide is also apparent near the bottom right, with $n = 1$.

The corresponding results for the summer season are omitted here for brevity. In summer, the wave amplitudes are greatly reduced, as is the strength of the mean flow, and

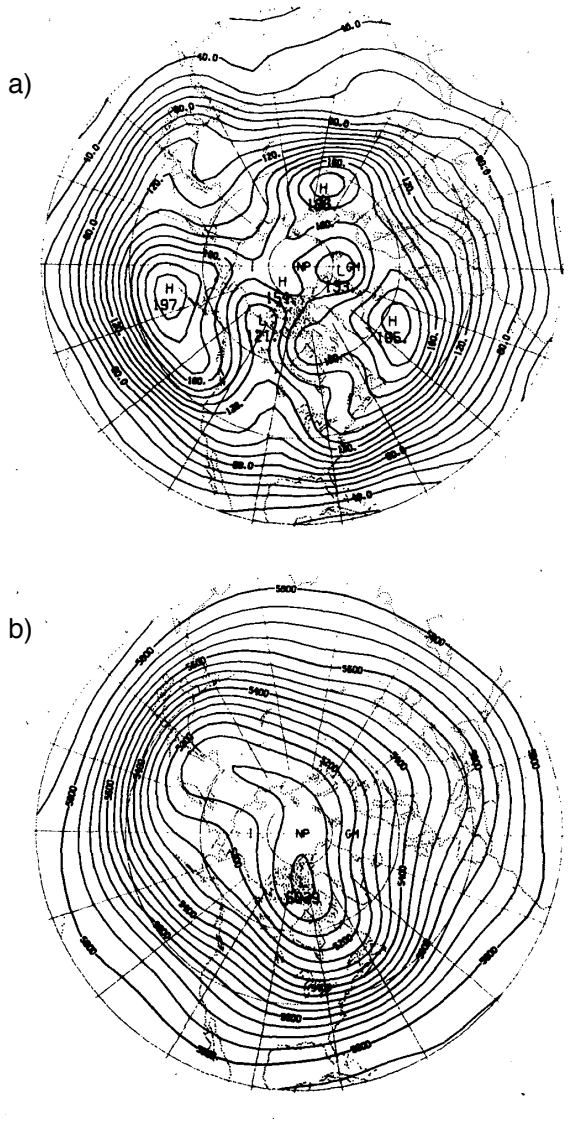


Figure 10.4: Maps of the 500 hPa geopotential height based on data for nine northern winters. The top panel shows the root-mean-square (RMS) temporal variability of the spatially unfiltered data, with a contour interval of 10 m. The lower panel shows a simple time average of the same data, with a contour interval of 50 m From Blackmon (1976).

the centers of action are shifted towards the pole. As in winter, the long waves and low frequencies dominate.

One conclusion of Blackmon's study is that most of the transient eddy energy resides at the low frequencies and long wavelengths. This has been known for a long time. For example, Wiin-Nielsen et al. (1963) analyzed the energy and momentum transport as a function of latitude and zonal wave number. More recently, Lembo et al. (2022) has used ERA5 data to show the probability distributions of meridional moist static energy transports

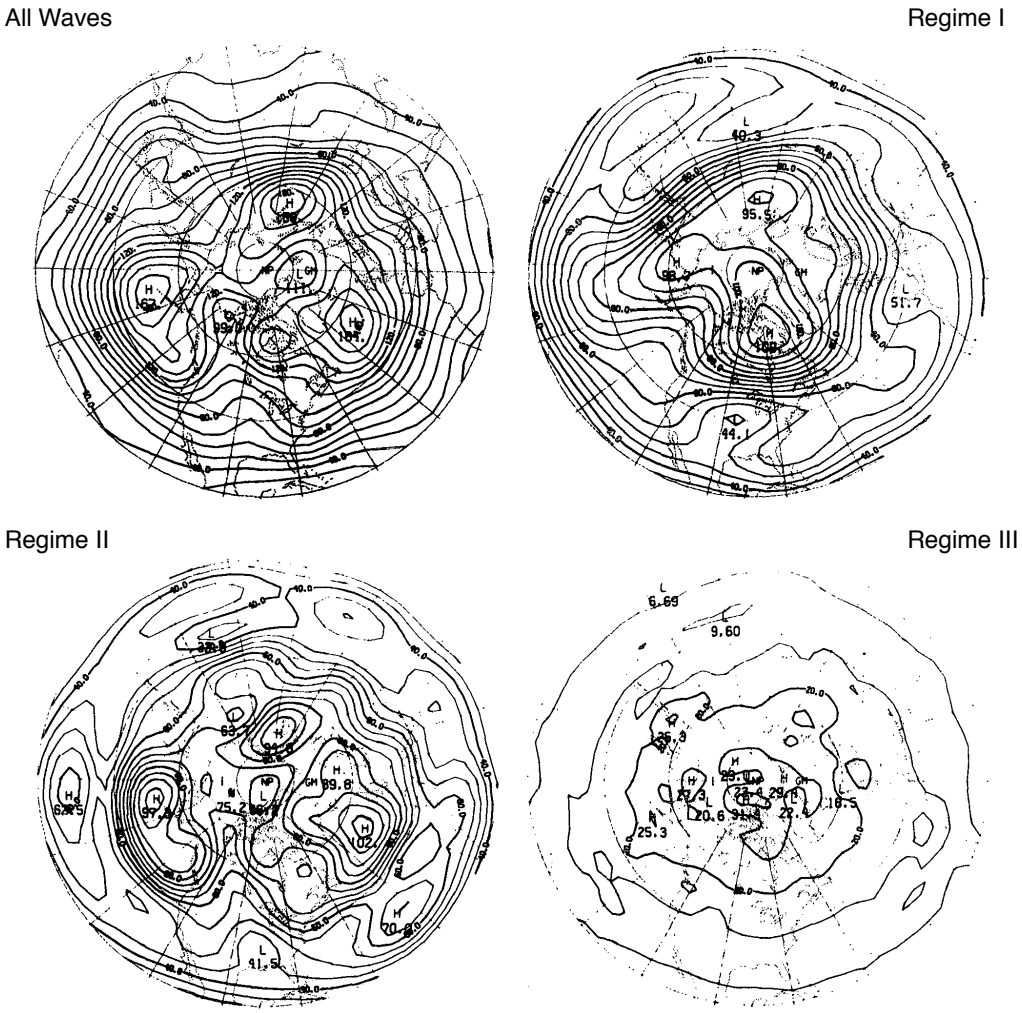


Figure 10.5: Maps of the low-pass filtered rms fields (winter): (a) all eddies, contour interval 10 m; (b) eddies in Regime I, contour interval 5 m; (c) eddies in Regime II, contour interval 5 m; (d) eddies in Regime III, contour interval 5 m. From Blackmon (1976).

by zonal wave number, for the Northern Hemisphere. Fig. 10.7 shows the results plotted as a function of latitude. The strongest contributions are from the planetary scales, with $m = 1 - 5$. Wave numbers 11 and higher contribute very little. Lembo et al. (2022) also show the corresponding results for the northern summer.

Chang et al. (2002) discuss the observed climatology of the northern hemisphere storm tracks seen in Fig. 10.6. Fig. 10.8, taken from their paper, shows that the storm tracks are associated with strong conversion from eddy available potential energy to eddy kinetic energy, i.e., the storms are baroclinic eddies.

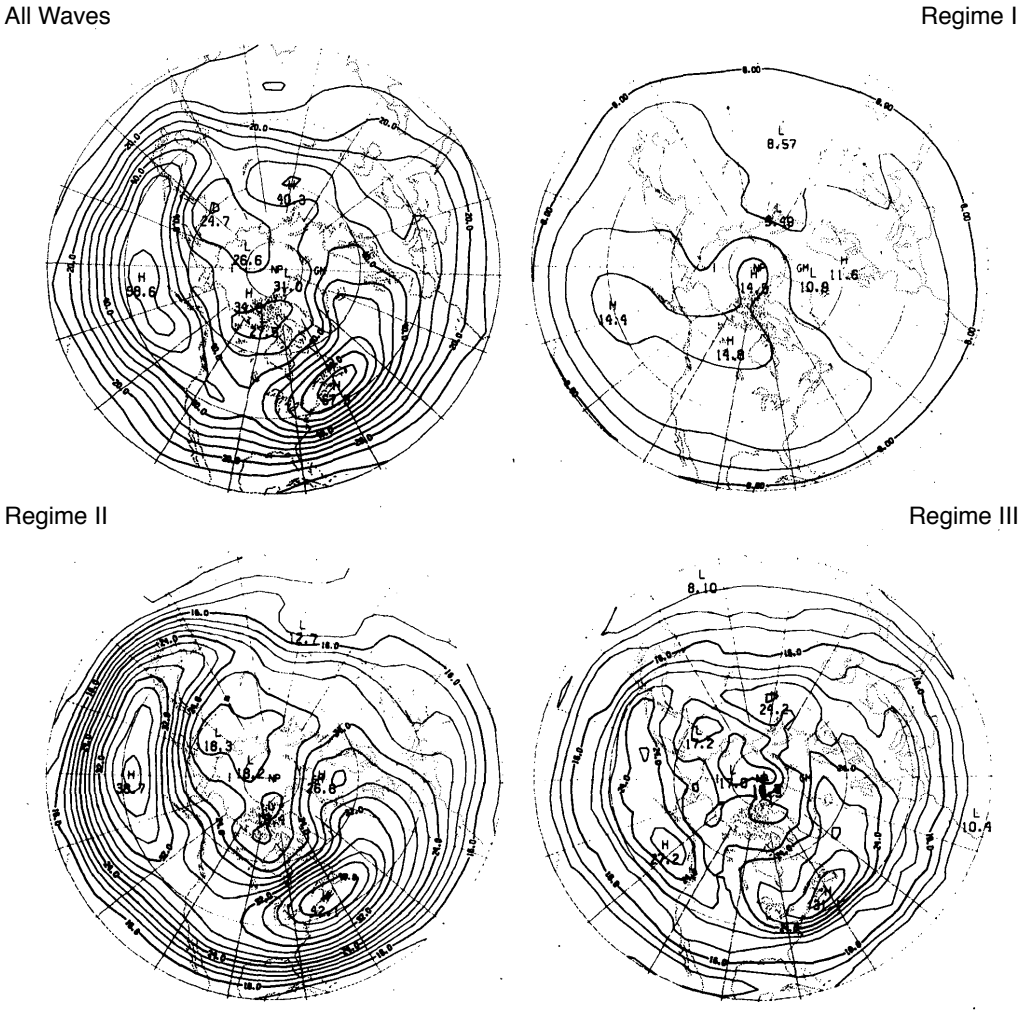


Figure 10.6: Maps of the medium-pass filtered rms fields (winter): (a) all eddies, contour interval 5 m; (b) eddies in Regime I, contour interval 2 m; (c) eddies in Regime II, contour interval 2 m; (d) eddies in Regime III, contour interval 2 m. From Blackmon (1976).

10.4 Annular modes

Thompson and Wallace (1998, 2000) and Limpasuvan and Hartmann (2000) found that variability of the winter sea-level pressure is dominated by zonally symmetric structures, which they called "annular modes." The Northern Annular Mode, or NAM, had been noticed by Lorenz (1951), and is associated with the North Atlantic Oscillation (Wallace and Gutzler, 1981; Hurrell et al., 2001, 2003; Hurrell and Deser, 2010), which strongly influences European weather. The Southern Annular Mode (SAM; Rogers and van Loon, 1982; Hartmann and Lo, 1998; Thompson et al., 2005) is also called the Antarctic Oscillation. Both the NAM and the SAM have vertically uniform (i.e., barotropic) vertical structures. They contribute about a quarter of the total temporal variance of the winds and

Table 10.2: Power as a function of wavenumber and frequency for the winter season. The units are m^2 rad^{-1} (15 days)$^{-1}$. The left-most column gives the values of the two-dimensional index, n. The second and following columns of the heading row give the frequencies, in cycles per day^{-1}.

n	0	1/15	2/15	3/15	4/15	5/15	6/15	7/15	8/15	9/15	10/15	11/15	12/15	13/15	14/15	15/15
18	51.7	49.8	45.8	39.3	37.0	26.7	21.2	17.3	13.7	10.1	7.2	5.3	4.1	3.4	3.2	3.4
17	68.7	66.5	58.6	48.1	40.1	34.0	27.9	22.1	15.8	11.0	7.6	5.4	4.3	3.1	3.2	3.3
16	98.5	97.5	82.4	69.6	56.5	45.0	34.8	26.0	18.4	12.2	7.9	5.5	4.1	3.3	3.3	3.5
15	141.6	131.0	108.1	85.9	70.3	58.6	75.6	32.3	19.7	11.6	7.3	5.1	4.1	3.3	3.3	3.4
14	210.1	195.5	158.9	122.3	99.5	79.2	53.3	32.3	18.2	10.8	6.9	4.9	3.5	3.3	3.3	3.4
13	302.8	272.5	213.2	164.7	129.0	92.6	56.1	31.6	16.8	9.4	5.7	4.4	3.2	2.9	2.9	3.3
12	502.6	435.2	311.1	225.5	159.9	99.8	53.5	26.8	13.6	8.0	5.3	3.9	3.2	3.0	3.0	3.4
11	746.3	639.7	443.3	290.6	176.5	94.4	43.8	21.1	10.6	6.0	4.1	3.6	3.0	2.8	3.0	3.4
10	1190.1	943.3	545.9	322.8	181.0	80.5	33.6	15.6	8.3	5.4	3.9	3.3	3.0	2.9	3.1	3.6
9	1815.2	1417.6	739.3	323.8	135.8	55.0	23.2	11.4	6.1	4.1	3.1	2.8	2.4	2.7	2.9	3.6
8	3305.7	2175.9	732.7	245.2	104.5	42.0	17.3	9.1	4.3	3.6	2.3	2.6	2.2	2.7	2.5	3.4
7	3732.6	2370.4	698.2	191.2	71.2	29.3	12.9	7.7	3.8	3.6	2.3	2.9	1.9	2.6	2.5	4.3
6	3731.5	2374.2	657.5	149.3	51.0	20.6	9.0	5.8	2.4	2.8	1.3	2.3	1.5	2.4	2.3	4.2
5	3563.5	2192.6	522.5	114.0	39.1	16.2	7.1	4.9	3.4	2.7	2.2	2.2	2.2	2.2	2.6	4.8
4	2553.1	1600.8	420.2	95.1	34.5	13.8	5.7	4.2	2.1	2.6	1.5	2.1	1.3	2.0	2.1	3.8
3	1448.2	992.6	329.2	79.7	27.0	9.4	4.6	3.3	2.1	1.8	1.4	1.5	1.2	1.5	5.3	4.6
2	1115.3	642.0	146.3	41.8	18.5	8.3	3.3	2.3	1.6	1.6	1.1	1.3	1.2	1.4	2.2	3.7
1	294.5	221.7	105.9	45.1	15.8	4.1	2.0	1.7	1.2	1.0	0.9	1.0	1.0	1.1	6.0	12.4
0	168.6	92.6	17.1	4.7	2.5	1.4	1.0	0.8	0.6	0.6	0.6	0.6	0.6	0.6	1.8	3.5

pressure (or geopotential height) in the middle latitudes of their respective hemispheres, on a wide range of time scales.

Recently Thompson and Woodworth (2014) and Thompson and Barnes (2014) have identified a Baroclinic Annular Mode (BAM) in the middle latitudes of the Southern Hemisphere. The BAM oscillates with a period of about four weeks. Thompson and Barnes (2014) showed that the oscillations involve a cyclic feedback loop in which a period of

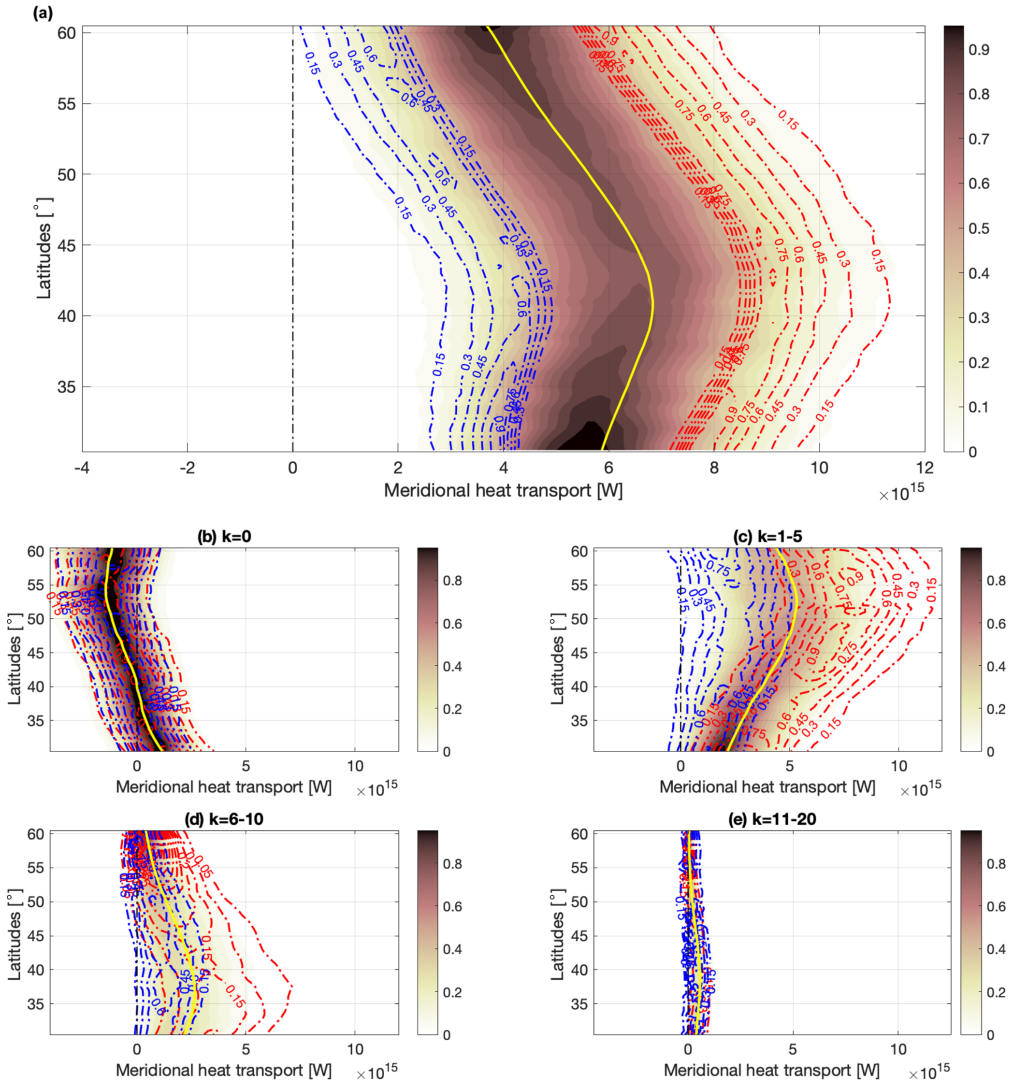


Figure 10.7: Probability density functions (PDFs) of total (filled contours) and extreme poleward (red contours) and equatorward (blue contours) DJF meridional moist static energy transports for the 1979–2012 period in ERA5. The PDFs are grouped by zonal wave number m. (a) Sum of all wavenumbers, (b) contribution from $m = 0$ (zonal mean), (c) contributions from $m = 1 - 5$ (planetary scales), (d) contributions from $m = 6 - 10$ (synoptic scales), (e) contributions from $m = 11 - 200$ (higher scales). The PDFs have been normalized by dividing by the maximum value across all latitudes. From Lembo et al. (2022).

gradually increasing meridional temperature gradient is followed by an increase in the transient eddy meridional energy flux, which then reduces the meridional temperature gradient, which in turn leads to a decrease in the transient eddy meridional energy flux.

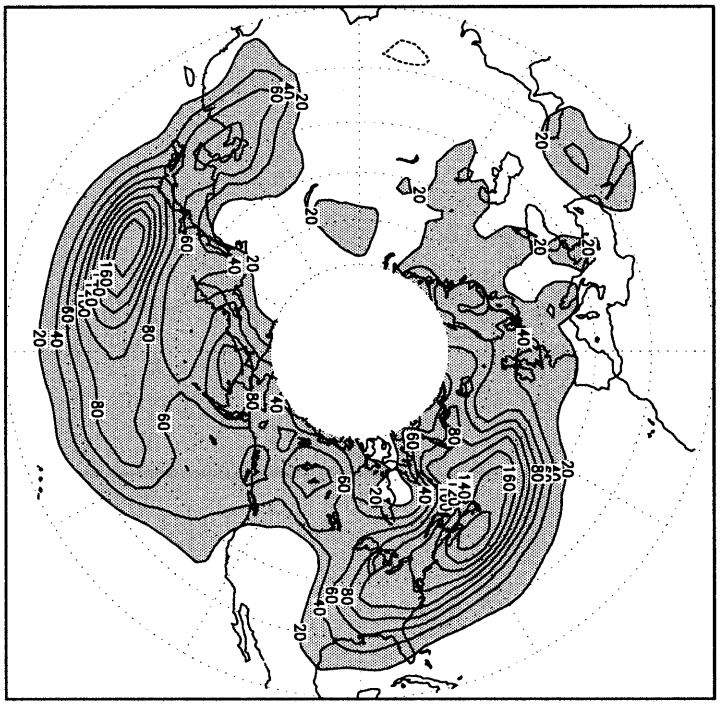


Figure 10.8: The vertically averaged rate of conversion of eddy available potential energy into eddy kinetic energy in Northern Hemisphere winter. The contour interval is 20 m^2 s^{-2} day^{-1}. From Chang et al. (2002). American Meteorological Society. Used with permission.

10.5 Propagation of planetary waves

The following discussion is based on the famous paper by Charney and Drazin (1961). Closely related work can be found in Dickinson (1968) and Matsuno (1970).

As discussed in Chapter 4, the quasi-geostrophic form of the potential vorticity equation is

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_g \cdot \nabla_p\right) Z_{\mathrm{QG}} = 0 \,, \tag{10.61}$$

where

$$Z_{\mathrm{QG}} = \frac{\nabla^2 \phi}{f_0} + \beta_0 y + \frac{\partial}{\partial p}\left(\frac{f_0}{S_p}\frac{\partial \phi}{\partial p}\right) \tag{10.62}$$

is the quasi-geostrophic pseudo-potential vorticity (QGPPV), S_p is the static stability, and f_0 and β_0 are representative values of the Coriolis parameter and its meridional derivative, respectively. We adopt the "log pressure" coordinate

$$z(p) \equiv -H_0 \ln\left(\frac{p}{p_0}\right) , \tag{10.63}$$

where T_0 is a constant reference temperature, and we define the Brunt-Väisälä frequency by

$$N^2 \equiv \frac{g}{\theta_{\rm bs}} \frac{\partial \theta_{\rm bs}}{\partial z} . \tag{10.64}$$

Here the subscript "bs" denotes the basic state. We will use

$$\psi \equiv \frac{\phi}{f_0} \tag{10.65}$$

which is called the geostrophic stream function. Recall that

$$v_g = \frac{\partial \psi}{\partial x} \quad \text{and} \quad u_g = -\frac{\partial \psi}{\partial y} . \tag{10.66}$$

Here we use Cartesian coordinates for simplicity. With the use of (10.65), (10.63), and (10.64), the QGPPV can be rewritten as

$$Z_{\rm QG} = \nabla^2 \psi + \beta_0 y + \frac{1}{\rho_{\rm bs}} \frac{\partial}{\partial z}\left(\rho_{\rm bs} \frac{f_0^2}{N^2} \frac{\partial \psi}{\partial z}\right) \tag{10.67}$$

Linearizing (10.61) about the zonal-mean state gives

$$\left(\frac{\partial}{\partial t} + \overline{u}^{\lambda} \frac{\partial}{\partial x}\right) Z^*_{\rm QG} + v^*_g \frac{\partial}{\partial y} \overline{Z_{\rm QG}}^{\lambda} = 0 . \tag{10.68}$$

We look for separable solutions of the form

$$\psi^* = \mathrm{Re}\left\{ \widehat{\psi}\,(y,\, z)\, e^{ik(x-ct)} \right\} , \tag{10.69}$$

$$Z^*_{\mathrm{QG}} = \mathrm{Re}\left\{ \widehat{Z}\,(y,z)\, e^{ik(x-ct)} \right\} , \tag{10.70}$$

where

$$\widehat{Z} = -k^2 \widehat{\psi} + \frac{\partial^2 \widehat{\psi}}{\partial y^2} + \frac{1}{\rho_{\mathrm{bs}}} \frac{\partial}{\partial z} \left(\rho_{\mathrm{bs}} \frac{f_0^2}{N^2} \frac{\partial \widehat{\psi}}{\partial z} \right) . \tag{10.71}$$

Substitution of (10.66), (10.69), and (10.70) into (10.68) gives

$$\left(\overline{u}^{\lambda} - c \right) \widehat{Z} + \widehat{\psi} \frac{\partial \overline{Z}^{\lambda}}{\partial y} = 0 , \tag{10.72}$$

Using (10.71), we can rewrite (10.72) as

$$\frac{\partial^2 \widehat{\psi}}{\partial y^2} + \frac{1}{\rho_{\mathrm{bs}}} \frac{\partial}{\partial z} \left(\rho_{\mathrm{bs}} \frac{f_0^2}{N^2} \frac{\partial \widehat{\psi}}{\partial z} \right) = - \left[\frac{1}{\left(\overline{u}^{\lambda} - c \right)} \frac{\partial \overline{Z_{\mathrm{QG}}}^{\lambda}}{\partial y} - k^2 \right] \widehat{\psi} . \tag{10.73}$$

This is a fairly general form of the quasi-geostrophic wave equation, which describes wave propagation in the latitude-height plane. We will simplify it considerably before analyzing it.

The first step is to introduce a "scaled" version of $\widehat{\psi}$. As wave energy propagates up to higher levels, it encounters decreasing values of ρ_{bs}. The energy density (energy per unit volume) scales like $\rho_{\mathrm{bs}} \left(k \widehat{\psi} \right)^2$, so if the energy density is constant with height, $\widehat{\psi}$ must increase upward in proportion to $1/\sqrt{\rho_{\mathrm{bs}}}$. Because of this effect, the equations become simpler if we introduce a scaled value of $\widehat{\psi}$:

$$\psi \equiv \frac{\sqrt{\rho_{\mathrm{bs}}}}{N} \widehat{\psi} . \tag{10.74}$$

Note that here ψ (no hat) is the scaled value; the meaning of ψ now departs from that used in (10.65). We also note that

$$\begin{aligned}
&\frac{1}{\rho_{\rm bs}}\frac{\partial}{\partial z}\left(\rho_{\rm bs}\frac{f_0^2}{N^2}\frac{\partial\widehat{\psi}}{\partial z}\right)\\
&\qquad=\frac{f_0^2}{\rho_{\rm bs}}\frac{\partial}{\partial z}\left[\frac{\sqrt{\rho_{\rm bs}}}{N}\frac{\partial}{\partial z}\left(\frac{\sqrt{\rho_{\rm bs}}}{N}\widehat{\psi}\right)-\frac{\sqrt{\rho_{\rm bs}}}{N}\widehat{\psi}\frac{\partial}{\partial z}\left(\frac{\sqrt{\rho_{\rm bs}}}{N}\right)\right]\\
&\qquad=\frac{f_0^2}{\rho_{\rm bs}}\frac{\partial}{\partial z}\left[\frac{\sqrt{\rho_{\rm bs}}}{N}\frac{\partial\widehat{\psi}}{\partial z}-\widehat{\psi}\frac{\partial}{\partial z}\left(\frac{\sqrt{\rho_{\rm bs}}}{N}\right)\right]\\
&\qquad=\frac{f_0^2}{\rho_{\rm bs}}\left[\frac{\sqrt{\rho_{\rm bs}}}{N}\frac{\partial^2\widehat{\psi}}{\partial z^2}-\widehat{\psi}\frac{\partial^2}{\partial z^2}\left(\frac{\sqrt{\rho_{\rm bs}}}{N}\right)\right].
\end{aligned}\tag{10.75}$$

Substituting from (10.74) and (10.75), we can rewrite (10.73) as

$$\boxed{\frac{\partial^2\psi}{\partial y^2}+\frac{f_0^2}{N^2}\frac{\partial^2\psi}{\partial z^2}=-\left(\frac{f_0^2}{N^2}\frac{n^2}{4H_0^2}\right)\psi\,,}\tag{10.76}$$

where

$$n^2\equiv\frac{4N^2H_0^2}{f_0^2}\left[\frac{1}{\left(\overline{u}^\lambda-c\right)}\frac{\partial\overline{Z_{\rm QG}}^\lambda}{\partial y}-k^2-\frac{f_0^2}{\sqrt{\rho_{\rm bs}}N}\frac{\partial^2}{\partial z^2}\left(\frac{\sqrt{\rho_{\rm bs}}}{N}\right)\right]\tag{10.77}$$

is called the "index of refraction." Here $H_0\equiv RT_0/g$, where T_0 is the reference temperature used in (10.63). Eq. (10.76) is the quasi-geostrophic wave equation. When $n^2>0$, the solution for ψ is oscillatory (propagating), and when $n^2<0$, the solution for ψ is "evanescent," which means that it decays exponentially away from the source of excitation.

Comparing (10.76) - (10.77) with (10.73), we see that the left-hand side of (10.76) has become simpler, but the expression for the index of refraction has become more complicated. So it goes. Using some idealizations, we can drastically simplify (10.77) without altering its basic meaning. First, consider the special case of an isothermal atmosphere (i.w., T constant with height). This is not unrealistic for the lower stratosphere. It can be shown that for an isothermal atmosphere $N^2=\frac{g^2}{c_pT_0}=$ constant and $\rho_{\rm bs}\sim e^{-z/H_0}$, so that (10.77) reduces to

$$n^2 \cong \frac{4N^2H_0^2}{f_0^2}\left[\frac{1}{\left(\overline{u}^\lambda - c\right)}\frac{\partial\overline{Z_{\rm QG}}^\lambda}{\partial y} - k^2\right] - 1\ . \tag{10.78}$$

Inspection of (10.78) shows that $\overline{u}^\lambda - c > 0$ is necessary for $n^2 > 0$, i.e., to have propagating solutions. Now we simplify further by assuming stationary waves, for which the phase speed, c, is zero. As discussed later in this chapter, stationary waves can be excited by flow over mountains, or by land-sea contrast. For the special case of stationary waves, (10.78) reduces to

$$n^2 = \frac{4N^2H_0^2}{f_0^2}\left(\frac{1}{\overline{u}^\lambda}\frac{\partial\overline{Z_{\rm QG}}^\lambda}{\partial y} - k^2\right) - 1\ . \tag{10.79}$$

To simplify even further, note from (10.67) that

$$\frac{\partial\overline{Z_{\rm QG}}^\lambda}{\partial y} = \beta_0 - \frac{\partial^2\overline{u}^\lambda}{\partial y^2} - \frac{1}{\rho_{\rm bs}}\frac{\partial}{\partial z}\left(\rho_{\rm bs}\frac{f_0^2}{N^2}\frac{\partial\overline{u}^\lambda}{\partial z}\right)\ . \tag{10.80}$$

When the meridional and vertical shears of $\overline{u}^\lambda$ are not too strong,

$$\frac{\partial\overline{Z_{\rm QG}}^\lambda}{\partial y} \cong \beta_0 \geq 0\ . \tag{10.81}$$

Using (10.81) in (10.79), we finally obtain

$$n^2 \cong \frac{4N^2H_0^2}{f_0^2}\left(\frac{\beta_0}{\overline{u}^\lambda} - k^2\right) - 1\ . \tag{10.82}$$

From (10.82), we can draw the following conclusions:

- To have propagation ($n^2 > 0$), we need $\beta_0/\overline{u}^\lambda > 0$. Because $\beta_0 > 0$, $\overline{u}^\lambda$ must be positive (westerly). The interpretation is straightforward: Stationary Rossby waves cannot exist in easterlies, because they propagate westward relative to the air, so that easterlies cannot hold them in place. Recall that the summer hemisphere stratosphere is dominated by easterlies, while the winter hemisphere stratosphere is dominated by

westerlies. Note, however, that large positive $\overline{u}^{\lambda}$ also makes $n^2 < 0$. Stationary waves cannot propagate through very strong westerlies, which would sweep them downstream. Fig. 10.9, from Charney and Drazin (1961), shows the vertical distribution of n^2 for summer and winter, averaged over the Northern Hemisphere middle latitudes, for stationary waves of three different wavelengths.

- Even when $\beta_0/\overline{u}^{\lambda} > 0$, waves with sufficiently large k (i.e., sufficiently short zonal wavelength) cannot propagate. Short waves are, therefore, "trapped" near their excitation levels. Since $\overline{u}^{\lambda}$ has a maximum near the tropopause in middle latitudes, many short waves are trapped in the troposphere, even in winter. Only longer waves can propagate to great heights. This suggests that long waves will dominate in the stratosphere and mesosphere even more than they do in the troposphere.

- A level where $\overline{u}^{\lambda} = 0$ is called a "critical level" for stationary waves. Eq. (10.79) shows that, at a critical level, $n^2 \to \infty$. Suppose that $\overline{u}^{\lambda} > 0$ below a critical level, and $\overline{u}^{\lambda} < 0$ above. Then, for waves excited at the lower boundary (e.g., by flow over topography), upward propagation is blocked at the critical level, and no wave activity will be seen above the critical level. Meridionally propagating waves can also be blocked at a critical latitude, where $\overline{u}^{\lambda} = 0$. As discussed in Chapter 3, such critical latitudes are often found in the subtropics. In general, we can speak of "critical lines" in the latitude-height plane, along which $\overline{u}^{\lambda} = 0$. If we allowed $c \neq 0$, we would find that the critical surfaces are those for which $\overline{u}^{\lambda} - c = 0$.

Fig. 10.10 provides evidence that the theory is correct. It shows the geopotential height fields at 500 hPa, 100 hPa, and 10 hPa, for Northern Hemisphere summer and winter. In winter, planetary waves clearly propagate upward to the 10 hPa level, while in summer they do not. Note that the apparent horizontal scale of the dominant eddies increases upward in winter. This is consistent with the theory, which predicts that the upward propagation of the shorter modes is blocked at lower levels while longer modes can continue to propagate upward to great heights.

Matsuno (1970) used the observed winds for the Northern Hemisphere winter to compute $\partial \overline{Z_{\mathrm{QG}}}^{\lambda}/\partial y$, the index of refraction, and the energy flow in the latitude-height plane for zonal wave number 1. His results are shown in Fig. 10.11. The upward-propagating waves are directed equatorward by the spatially varying index of refraction.

10.6 Theory of orographically forced stationary waves

Stationary waves are forced by mechanical and/or thermal effects that are anchored to the Earth's surface. Mountain ranges can produce waves either by blocking the flow or by acting as elevated heat sources. Geographically fixed thermal forcing is also associated with land-sea contrasts, sea surface temperature gradients, and the like.

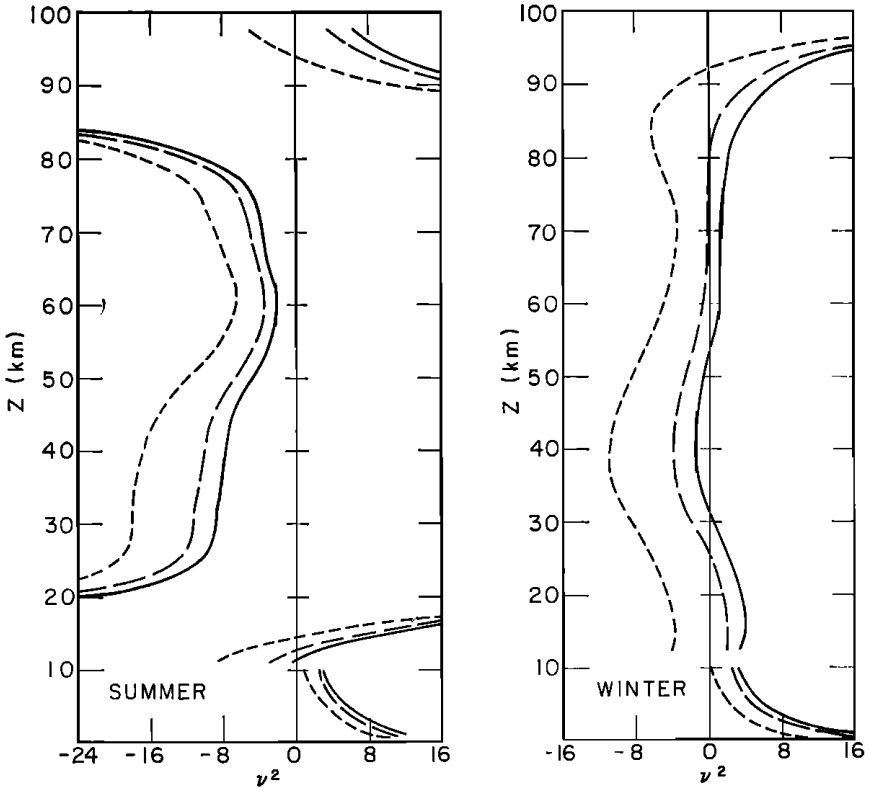


Figure 10.9: The square of the index of refraction for summer and winter, averaged between 30° and 60°N, for waves of different wavelengths, L. The short-dashed lines correspond to $L = 6,000$ km, the long-dashed lines correspond to $L = 10,000$ km, and the solid lines correspond to $L = 14,000$ km. From Charney and Drazin (1961).

Held (1983) summarized the work of Charney and Eliassen (1949), who studied the effects of orographic barriers on stationary waves in middle latitudes. The starting point is consideration of the conservation of potential vorticity in shallow water equations introduced in Chapter 4, and also using the quasi-geostrophic approximation discussed in the same chapter. The governing equation is

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right) Z_{\mathrm{SWQG}} = -\frac{r\zeta}{h}, \tag{10.83}$$

where

$$Z_{\mathrm{SWQG}} \equiv \frac{\zeta + f}{h}. \tag{10.84}$$

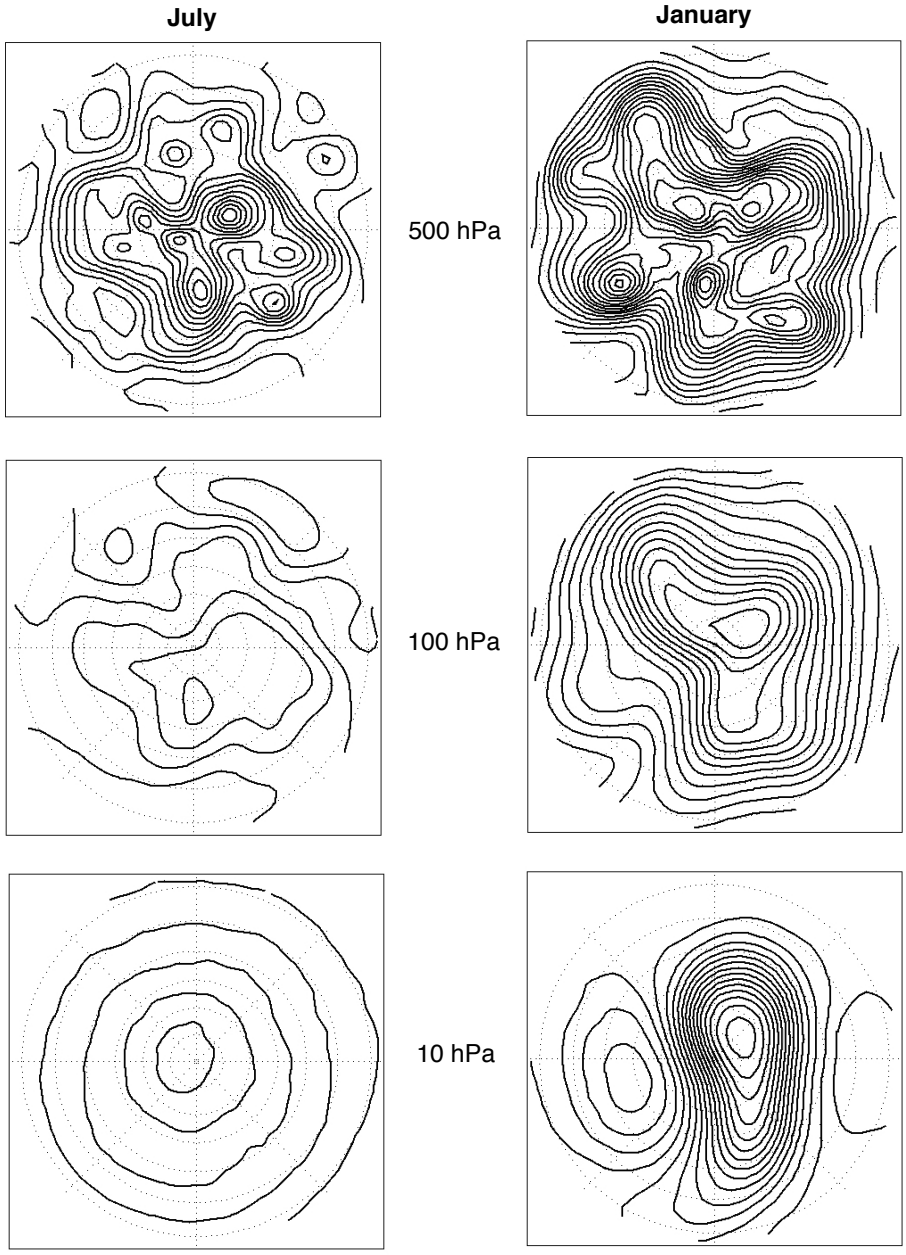


Figure 10.10: These Northern Hemisphere data were collected during the International Geophysical Year. Geopotential heights for July 15, 1958 are shown on the left, and those for January 15, 1959 are shown on the right. The levels plotted are 500 hPa, 100 hPa, and 10 hPa. From Charney (1973).

Here h is the depth of the water. The geostrophic wind is $\mathbf{v}_g \equiv \frac{g}{f_0}\mathbf{k} \times \nabla h_{\text{fs}}$, where $h_{\text{fs}} \equiv h_{\text{T}} + h$ is the height of the free surface, h_{T} is the height of the topography, and r is a nonnegative Rayleigh friction coefficient. Quasigeostrophy allows us to write

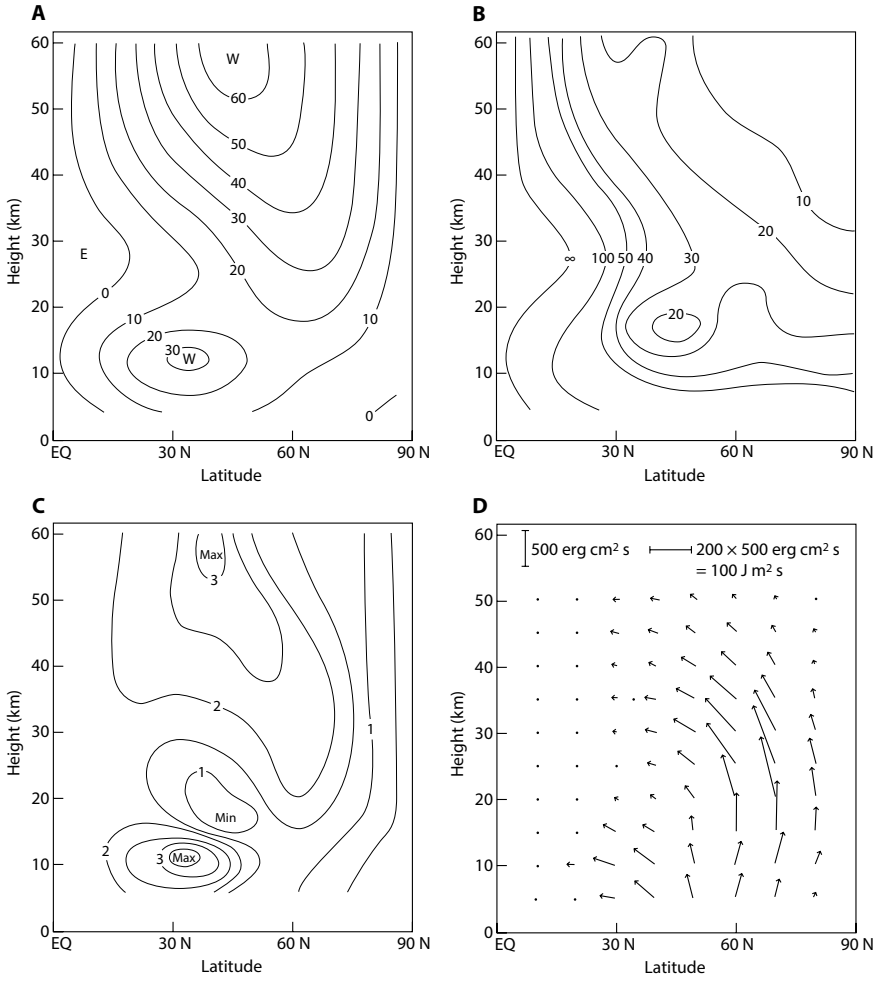


Figure 10.11: A) An idealized distribution of $\overline{u}$ (in m s^{-1}) for the Northern Hemisphere winter. B) The latitudinal gradient of the potential vorticity, $\partial \overline{Z_{\mathrm{QG}}}^{\lambda}/\partial y$, expressed as a multiple of the Earth's rotation rate. C) The refractive index square n^2, for the $k = 0$ wave. D) The computed wave energy flow in the meridional plane associated with zonal wave number 1. From Matsuno (1970).

$$\zeta = \frac{g}{f_0}\nabla^2 h_{\mathrm{fs}} \, . \tag{10.85}$$

We linearize (10.83) about a stationary, geostrophically balanced, zonally averaged state. We assume that $\overline{u}^{\lambda}$ is independent of latitude and time, and that $\overline{h_{\mathrm{T}}}^{\lambda} = 0$. The basic-state zonal wind satisfies

$$\overline{u}^{\lambda} = \frac{g}{f_0}\frac{\partial \overline{h}^{\lambda}}{\partial y}. \tag{10.86}$$

Here we use Cartesian coordinates for simplicity. Then the PV gradient of the zonally averaged flow is given by

$$\overline{h}^{\lambda}\frac{\partial}{\partial y}\overline{Z_{\rm SWQG}}^{\lambda} = \beta + \frac{f_0^2\overline{u}^{\lambda}}{g\overline{h}^{\lambda}} \tag{10.87}$$

In linearizing about the basic state given by (10.86) and (10.87), we assume not only that the perturbation vorticity and height are small, but also that $h_{\rm T}$ is small compared to $\overline{h}^{\lambda}$, so that we can neglect any products of $h_{\rm T}$ with a perturbation quantity. We use

$$h^*_{\rm fs} = h_{\rm T} + h^* \,. \tag{10.88}$$

The linearized version of (10.83) is then

$$\left(\frac{\partial}{\partial t} + \overline{u}^{\lambda}\frac{\partial}{\partial x}\right) Z^*_{\rm SWQG} + v^*\frac{\partial \overline{Z_{\rm SWQG}}^{\lambda}}{\partial y} = -\frac{r\zeta^*}{\overline{h}^{\lambda}} \,. \tag{10.89}$$

As usual with QG models, all perturbation quantities can be expressed in terms of $h^*_{\rm fs}$, as follows:

$$v^* = \frac{g}{f_0}\frac{\partial h^*_{\rm fs}}{\partial x} \,, \tag{10.90}$$

$$\zeta^* = \frac{g}{f_0}\nabla^2 h^*_{\rm fs} \,, \tag{10.91}$$

$$\begin{aligned}\overline{h}^{\lambda} Z^*_{\mathrm{SWQG}} &= \zeta^* - f_0 \frac{h^*}{\overline{h}^{\lambda}} \\ &= \frac{g}{f_0}\nabla^2 h^*_{\mathrm{fs}} - f_0 \frac{h^*}{\overline{h}^{\lambda}} \ .\end{aligned} \tag{10.92}$$

By substitution from (10.87) - (10.88) and (10.90)- (10.92), we can rewrite (10.89) as

$$\left(\frac{\partial}{\partial t} + \overline{u}^{\lambda}\frac{\partial}{\partial x}\right)\left[\frac{g}{f_0}\nabla^2 h^*_{\mathrm{fs}} - f_0\left(\frac{h_{\mathrm{T}} + h^*}{\overline{h}^{\lambda}}\right)\right] + \frac{g}{f_0}\frac{\partial h^*_{\mathrm{fs}}}{\partial x}\left(\beta + \frac{f_0^2 \overline{u}^{\lambda}}{g\overline{h}^{\lambda}}\right) = -r\frac{g}{f_0}\nabla^2 h^*_{\mathrm{fs}} \ , \tag{10.93}$$

or, after rearranging,

$$\frac{\partial}{\partial t}\left(\nabla^2 h^*_{\mathrm{fs}} - d^{-2} h^*_{\mathrm{fs}}\right) + \overline{u}^{\lambda}\frac{\partial}{\partial x}\left(\nabla^2 h^*_{\mathrm{fs}}\right) + \beta\frac{\partial h^*_{\mathrm{fs}}}{\partial x} + r\nabla^2 h^*_{\mathrm{fs}} = -\overline{u}^{\lambda} d^{-2}\frac{\partial h_{\mathrm{T}}}{\partial x} \ , \tag{10.94}$$

where d is the *Rossby radius of deformation*, which is defined by

$$d \equiv \frac{\sqrt{g\overline{h}^{\lambda}}}{f_0} \ . \tag{10.95}$$

The radius of deformation can be interpreted as the distance that a gravity wave propagates before it feels the effects of rotation. For scales smaller than the radius of deformation, the effects of rotation are weak (or even negligible), while for scales larger than the radius of deformation the effects of rotation are strong. The "topographic forcing" term of (10.94), involving $\partial h_{\mathrm{T}}/\partial x$, has been placed on the right-hand side of (10.94), to set it apart. Eq. (10.94) can describe waves that are "forced" by topography, which enters mathematically through the inhomogeneous term on the right-hand side. In the presence of such forcing, a non-zero h^* is demanded by (10.94). Homogeneous "free wave" solutions can also exist, but they are damped by the friction term, as discussed below.

Assume that the perturbations have the form

$$h_{\rm fs}^* = {\rm Re}\left\{\widehat{h}_{\rm fs}\exp\left[i\left(kx+ly-\sigma t\right)\right]\right\}\ , \tag{10.96}$$

where $\widehat{h}$ is a constant, and also that the topography is described by

$$h_{\rm T} = {\rm Re}\left\{\widehat{h}_{\rm T}\exp\left[i\left(kx+ly\right)\right]\right\}\ . \tag{10.97}$$

In this linear model, the wave numbers of the topography and the wave numbers of $h_{\rm fs}^*$ must be the same. We assume that k and l are non-negative, which means that the sign of σ determines the direction of propagation; a positive value of σ corresponds to eastward propagation. Substitution of (10.96) and (10.97) into (10.94) gives:

$$\left[-\sigma\left(K^2+d^{-2}\right)+k\left(\overline{u}^\lambda K^2-\beta\right)-irK^2\right]\widehat{h}_{\rm fs}\ \exp\left(-i\sigma t\right) = \overline{u}^\lambda d^{-2}k\widehat{h}_{\rm T}\ . \tag{10.98}$$

Here K is the total wave number, defined by

$$K^2 \equiv k^2+l^2\ . \tag{10.99}$$

We see directly from (10.98) that the amplitude of the waves, as measured by $\widehat{h}$, is proportional to the amplitude of the forcing, as measured by $\widehat{h}_{\rm T}$. The proportionality factor is rather complicated, however.

First consider the special case in which a free wave exists in the absence of topographic forcing. Then (10.98) reduces to a dispersion formula, which can be written as

$$-\sigma\left(K^2+d^{-2}\right)+k\left(\overline{u}^\lambda K^2-\beta\right)-irK^2 = 0\ , \tag{10.100}$$

or

$$\sigma = \frac{k\left(\overline{u}^\lambda K^2-\beta\right)-irK^2}{K^2+d^{-2}}\ . \tag{10.101}$$

Eq. (10.101) is the dispersion equation for a damped, free Rossby wave in a balanced mean flow. The wave dies out after a finite time, because there is no forcing to sustain it against the frictional damping.

To see how the friction leads to damping, write

$$\sigma = \sigma_0 - i\left(\tau_f\right)^{-1} , \tag{10.102}$$

where

$$\sigma_0 \equiv \frac{k\left(\overline{u}^{\lambda} K^2 - \beta\right)}{K^2 + d^{-2}} , \tag{10.103}$$

and

$$\left(\tau_f\right)^{-1} \equiv \frac{rK^2}{K^2 + d^{-2}} . \tag{10.104}$$

Then (10.96) can be rewritten as

$$h^*_{\rm fs}(x,\, y,\, t) = e^{-t/\tau_f}\, {\rm Re}\left\{\widehat{h}_{\rm fs} \exp\left[i\left(kx + ly - \sigma_0 t\right)\right]\right\} . \tag{10.105}$$

This shows that the wave-like solution, with frequency σ_0, decays with e-folding time τ_f. In other words, the free waves are killed off by friction, as mentioned earlier.

A stationary wave is one for which ${\rm Re}(\sigma) = 0$. Under what conditions can a free Rossby wave be stationary? Eq. (10.101) shows that a stationary free wave is possible only for $\overline{u} > 0$. The reason is that a Rossby wave propagates toward the west relative to the mean flow, so in order to hold the wave stationary relative to the Earth's surface the mean flow must be from the west. In such a case, the total wave number of the stationary wave is

$$K^2 = \beta / \overline{u}^{\lambda} \equiv K_S^2 . \tag{10.106}$$

Here the subscript K_S stands for "stationary."

With this preparation, we return now to the topographically forced case, and assume a stationary, neutral wave, i.e., $c = 0$. For the case of no friction ($r = 0$), we find from (10.96) that

$$\widehat{h}_{\rm fs} = \frac{\widehat{h}_{\rm T}}{d^2\left(K^2 - K_S^2\right)} \ . \tag{10.107}$$

Since the ratio of $\widehat{h}_{\rm fs}$ to $\widehat{h}_{\rm T}$ is a real number, they are "in phase" with each other. According to (10.107), in the absence of friction the forced wave has "infinite amplitude" for $K^2 = K_S^2$. This is the phenomenon of resonance, familiar from introductory physics. The infinity can be avoided by turning on the friction, i.e., allowing r to be positive. With friction, (10.107) is replaced by

$$\widehat{h}_{\rm fs} = \frac{\widehat{h}_{\rm T}}{d^2\left(K^2 - K_S^2 - i\frac{rK^2}{k\overline{u}^\lambda}\right)} \ . \tag{10.108}$$

Eq. (10.108) shows that friction shifts the phase of $\widehat{h}_{\rm fs}$, relative to $\widehat{h}_{\rm T}$. Also, the denominator on the right-hand side of (10.108) cannot become zero, so long as $r > 0$. For a stationary wave, (10.108) simplifies to

$$\widehat{h}_{\rm fs} = \widehat{h}_{\rm T}\left[\frac{ik\left(\overline{u}^\lambda\right)^2}{d^2 r\beta}\right] \ . \tag{10.109}$$

Here we have used (10.106). The factor of i in the numerator of (10.109) shifts the phase of $\widehat{h}_{\rm fs}$ by a quarter cycle (i.e., $\pi/2$) relative to $\widehat{h}_{\rm T}$. Fig. 10.12 shows how the squared amplitude of the steady-state wave response varies as the zonal wind speed changes, for several different values of r. Resonance occurs for a wind speed close to 20 m s^{-1}. Sufficiently strong friction leads to smooth behavior near resonance.

As seen in Fig. 10.13, the Charney-Eliassen model, despite its extreme simplicity, can explain reasonably well the observed zonal structure of the 500 hPa height field at 45° N in January. This strongly suggests that the observed midlatitude stationary eddies in winter are forced primarily by topography. A similar conclusion was reached by Manabe and Terpstra (1974) through numerical experiments with a global circulation model.

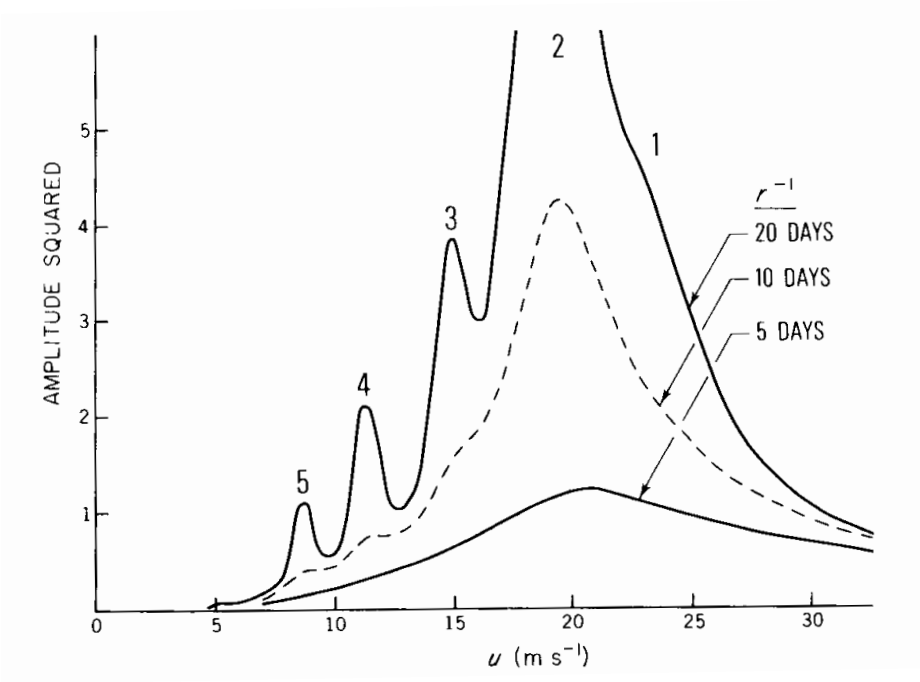


Figure 10.12: The mean-square height response, in the Charney-Eliassen model, as a function of the zonal wind speed, for different values of the Rayleigh friction coefficient. The units are 10^4 m^2. The integers written above the curves indicate the values of the zonal wind speed for which particular zonal wavenumbers resonate. From Held (1983). Copyright 1983 by Academic Press Ltd.

Fig. 10.14 shows the observed time-latitude sections of the zonal and meridional contributions to the stationary wave kinetic energy, i.e. $\overline{u^{*2}}^{\lambda}$ and $\overline{v^{*2}}^{\lambda}$, respectively. In winter, the zonal component is strongest in middle latitudes, and we can think of this as corresponding to the orographically forced waves analyzed above, although of course there is also a thermally forced component. In summer, there is a subtropical maximum of stationary eddy kinetic energy associated with the monsoons, which are discussed later in this chapter.

A more detailed discussion of orographically and thermally forced stationary Rossby waves is given by Held et al. (2002).

10.7 Tropical waves

In his dissertation research, Taroh Matsuno (1966, Fig. 10.15) studied the linearized shallow water equations (which were discussed at the end of Chapter 4) applied to the equatorial β-plane, i.e., a region centered on the Equator that is meridionally narrow enough so that β can be treated as a constant. When Matsuno began this work, his motivation was to investigate to what extent near-equatorial motions are geostrophic. In the process, however,

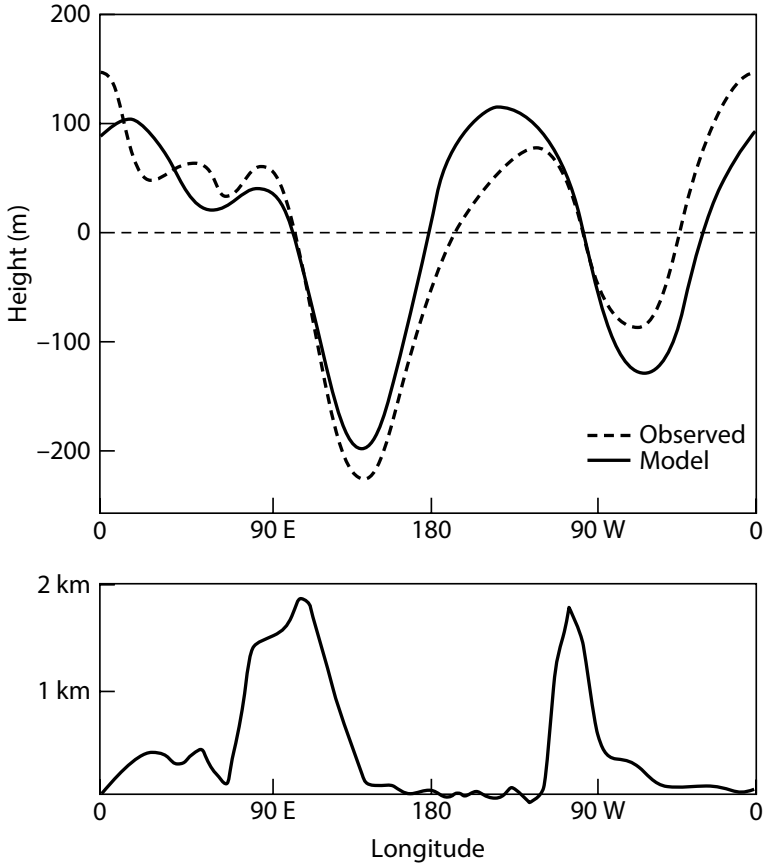


Figure 10.13: Upper panel: The height response as a function of longitude in the Charney-Eliassen model (solid line) and the observed climatological 500 hPa eddy heights at 45°N in January (dashed line). The lower panel shows the topography used. From Held (1983). Copyright 1983 by Academic Press Ltd.

he discovered two new classes of tropical waves, which soon after were detected in the observations. These same waves are actually among the solutions found by Laplace, but this was not recognized until later (Lindzen, 1967). The model created by Matsuno turns out to be relevant to a wide variety of phenomena, including monsoons, the Madden-Julian oscillation, the Quasi-Biennial Oscillation, and El Niño. This wide applicability is amazing, in view of the model's extreme simplicity.

The shallow-water equations on the equatorial β-plane, linearized about a state of rest, are:

$$\frac{\partial u}{\partial t} - fv + g\frac{\partial h}{\partial x} = 0\,,$$

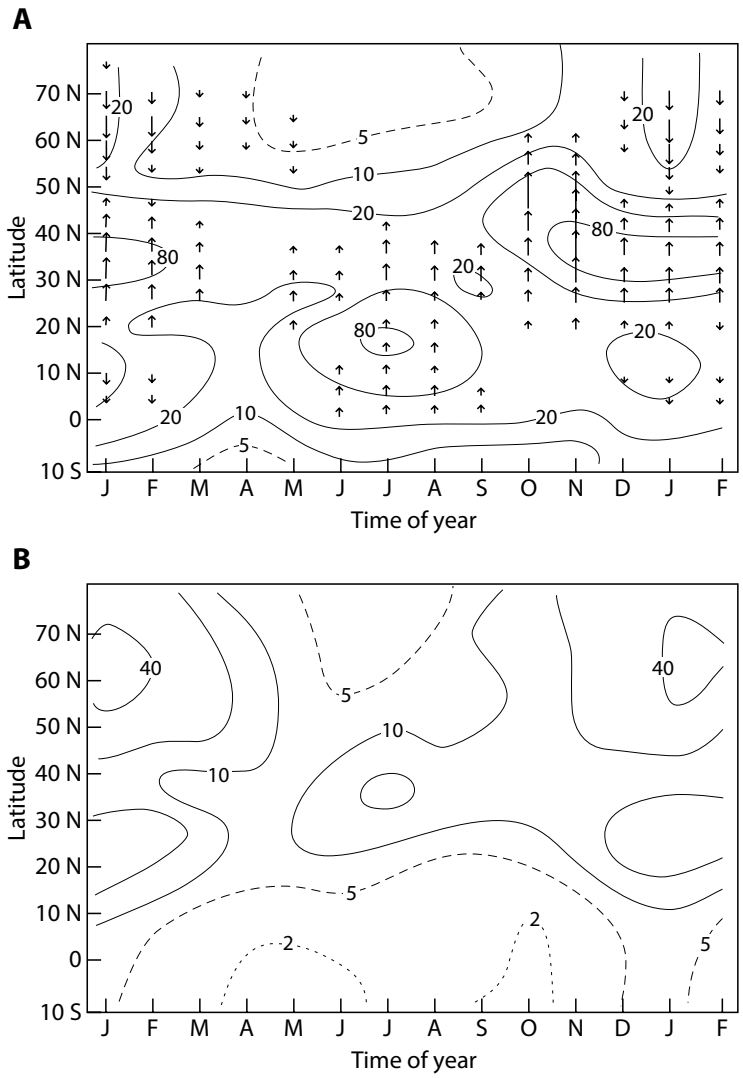


Figure 10.14: Time-latitude sections of Northern Hemisphere climatological mean stationary wave kinetic energy at the 200 hPa level, based on data of Oort and Rasmusson (1971). (A) Zonal wind component; (B) meridional wind component, in units of $m^2\ s^{-2}$. Arrows in panel A denote the direction and relative magnitude of the meridional flux of zonal momentum by the stationary waves. From Wallace (1983). Copyright 1983 by Academic Press Ltd.

$$\frac{\partial v}{\partial t} + fu + g\frac{\partial h}{\partial y} = 0\,,$$

$$\frac{\partial h}{\partial t} + H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0\,. \tag{10.110}$$



Figure 10.15: Professor Taroh Matsuno, who made major contributions to several branches of atmospheric science, including tropical wave dynamics and stratospheric sudden warmings. Photo kindly provided by Prof. Matsuno.

See Chapter 4 for an explanation of the shallow water equations. We omit the effects of "bottom topopgraphy" in this discussion. In (10.110),

$$\frac{\partial h}{\partial t} + H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0. \,, \tag{10.111}$$

where y is distance in the meridional direction, measured from $y = 0$ at the Equator (i.e., $y = a\varphi$), and β is approximated by its Equatorial value, which is $2\Omega/a$.

In Eq. (10.95), we defined the radius of deformation. This definition fails at the equator, where $f = 0$, but we can use the idea in (10.111) to define an alternative "equatorial radius of deformation, " $d_{\rm eq}$, as follows. Let

$$d_{\rm eq} \equiv \frac{c}{\beta d_{\rm eq}} \,, \tag{10.112}$$

where $c \equiv \sqrt{gH}$ is the speed of a pure gravity wave. It follows that

$$d_{\rm eq} = \sqrt{c/\beta}\ . \tag{10.113}$$

As a corresonding time scale, we can take

$$\begin{aligned} T_{\rm eq} \equiv & d_{\rm eq}/c \\ = & \left(\sqrt{c/\beta}\right)/c \\ = & 1/\sqrt{c\beta}\ . \end{aligned} \tag{10.114}$$

With these length and time scales, the velocity scale is simply c. For a representative phase speed of $c = 10$ m s^{-1}, we find that $d_{\rm eq} \cong 1000$ km and $T_{\rm eq} \cong 1$ day. Non-dimensionalizing the governing equations using $d_{\rm eq}$ and $T_{\rm eq}$, we obtain

$$\begin{aligned} \frac{\partial u}{\partial t} - yv + \frac{\partial \phi}{\partial x} = 0\ , \\ \frac{\partial v}{\partial t} + yu + \frac{\partial \phi}{\partial y} = 0\ , \\ \frac{\partial \phi}{\partial t} + \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \end{aligned} \tag{10.115}$$

Here ϕ is the non-dimensional form of gh.

As a side comment, we note that these equations can actually apply to a model with vertical structure (e.g., McCreary, 1981; Fulton and Schubert, 1985), and so are more readily applicable to the real atmosphere than one might guess. As a very simple example, consider a two-level model, governed by

$$\frac{\partial \mathbf{v}_1}{\partial t} + f\mathbf{k} \times \mathbf{v}_1 + \nabla \phi_1 = 0,$$

$$\frac{\partial \mathbf{v}_3}{\partial t} + f\mathbf{k} \times \mathbf{v}_3 + \nabla \phi_3 = 0,$$

$$\frac{\partial}{\partial t}(\phi_3 - \phi_1) + (S_p \Delta p)\,\omega_2 = 0. \tag{10.116}$$

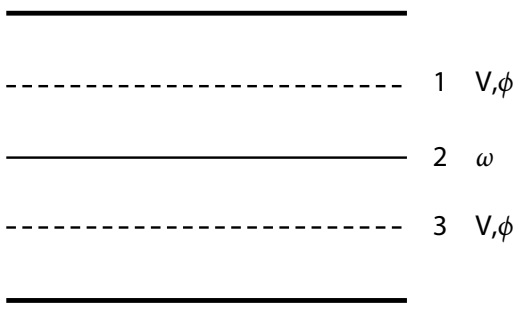


Figure 10.16: Schematic representing the two-level model represented by (10.7) - (10.116).

As shown in Fig. 10.16, subscript 1 denotes the upper level and subscript 3 denotes the lower level. The vertical velocity is defined in between, at level 2. We use $\Delta p \equiv p_3 - p_1$ to denote the pressure thickness between the two layers, and S is the static stability of the basic state. Let

$$\mathbf{v}_d \equiv \mathbf{v}_3 - \mathbf{v}_1 \tag{10.117}$$

be the vertical shear (actually, difference) of the horizontal wind between the two layers, and

$$\phi_d \equiv \phi_3 - \phi_1 \tag{10.118}$$

be the thickness between the two layers, respectively. Then (10.116) implies that

$$\frac{\partial \mathbf{v}_d}{\partial t} + f\mathbf{k} \times \mathbf{v}_d + \nabla \phi_d = 0, \tag{10.119}$$

and

$$\frac{\partial \phi_d}{\partial t} + \frac{S_p \Delta p^2}{2} \nabla \cdot \mathbf{v}_d = 0, \tag{10.120}$$

which are identical to the shallow water equations, and we can identify

$$c_i \equiv \Delta p \sqrt{\frac{S_p}{2}} \tag{10.121}$$

as the phase-speed of the internal gravity waves that derive their restoring force from the static stability.

We return now to our discussion of (10.115). Assume solutions of the form

$$\begin{aligned} u &= \widehat{u}(y) \exp i(kx + \sigma t), \\ v &= \widehat{v}(y) \exp i(kx + \sigma t), \\ \phi &= \widehat{\phi}(y) \exp i(kx + \sigma t) \cdot \end{aligned} \tag{10.122}$$

Following Matsuno, we adopt the convention that $k \geq 0$. This implies that $\sigma > 0$ for westward propagation, and $\sigma < 0$ for eastward propagation. Substitution into (10.115) gives

$$\begin{aligned} i\sigma\widehat{u} - y\widehat{v} + ik\widehat{\phi} &= 0, \\ i\sigma\widehat{v} + y\widehat{u} + \frac{d\widehat{\phi}}{dy} &= 0, \\ i\sigma\widehat{\phi} + ik\widehat{u} + \frac{d\widehat{v}}{dy} &= 0\,. \end{aligned} \tag{10.123}$$

We can solve the first of these equations for $\widehat{u}$ in terms of $\widehat{v}$ and $\widehat{\phi}$, and the result can be used to eliminate $\widehat{u}$ in the other two equations. Then the system (10.123) can be rewritten as

$$\begin{aligned} \sigma\widehat{u}+k\widehat{\phi} &= -iy\widehat{v}, \\ \left(ky-\sigma\frac{d}{dy}\right)\widehat{\phi} &= i\left(\sigma^2-y^2\right)\widehat{v}, \\ \left(\sigma^2-k^2\right)\widehat{\phi} &= i\left(ky+\sigma\frac{d}{dy}\right)\widehat{v} \quad . \end{aligned} \tag{10.124}$$

Here we have arranged the terms involving $\widehat{v}$ on the right-hand side of each equation.

Before considering the general case, we discuss a special solution, called the *equatorial Kelvin wave*, for which the meridional wind is identically zero. Setting $\widehat{v}=0$ in (10.124) gives

$$\begin{aligned} k\widehat{\phi}+\sigma\widehat{u} &= 0\,, \\ ky\widehat{\phi}-\sigma\frac{d\widehat{\phi}}{dy} &= 0\,, \\ \left(\sigma^2-k^2\right)\widehat{\phi} &= 0\,. \end{aligned} \tag{10.125}$$

In order to have non-trivial solutions, we need

$$\sigma = \pm k \tag{10.126}$$

where σ and k are non-zero. With the use and inclusion of (10.126), the system (10.125) is simplified to

$$\begin{aligned} \widehat{\phi}\pm\widehat{u} &= 0\,, \\ \left(-y\pm\frac{d}{dy}\right)\widehat{\phi} &= 0\,, \\ \sigma &= \pm k\,. \end{aligned} \tag{10.127}$$

The second equation in (10.127) determines the meridional structure of ϕ. Its solution is

$$\widehat{\phi} = e^{\pm y^2/2} \tag{10.128}$$

If we choose the plus sign, we get solutions that grow exponentially away from the Equator, which is unacceptable, especially since the Equatorial β-plane approximation is only useful near the Equator. We therefore choose the minus sign, which gives a "bell-shaped" solution that has a maximum on the Equator and decays strongly away from the Equator. The solution for the equatorial Kelvin wave can now be written as

$$\begin{aligned} \widehat{u} &= e^{-y^2/2} \,, \\ \widehat{\phi} &= e^{-y^2/2} \,, \\ \sigma &= -k \,. \end{aligned} \tag{10.129}$$

Note that $\sigma = -k < 0$ implies eastward propagation. An equatorial Kelvin wave always propagates towards the east. The bell-shaped solutions for $\widehat{u}$ and $\widehat{\phi}$ imply that Kelvin waves are symmetrical across the Equator.

Returning now to the general case, we can eliminate $\widehat{\phi}$ between the second and third equations of (10.124) to obtain an equation that governs $\widehat{v}$:

$$ky\left(ky\widehat{v} + \sigma\frac{d\widehat{v}}{dy}\right) - \sigma\frac{d}{dy}\left(ky\widehat{v} + \sigma\frac{d\widehat{v}}{dy}\right) = \left(\sigma^2 - k^2\right)\left(\sigma^2 - y^2\right)\widehat{v}, \tag{10.130}$$

which can be simplified to

$$\boxed{\frac{d^2\widehat{v}}{dy^2} + \left(\sigma^2 - k^2 + \frac{k}{\sigma} - y^2\right)\widehat{v} = 0.} \tag{10.131}$$

The substitution used to eliminate $\widehat{\phi}$ between the second and third equations of (10.124) is only valid for $\sigma^2 - k^2 \neq 0$, so (10.131) does not apply to the Kelvin wave, for which $\sigma^2 - k^2 = 0$.

We expect the solutions of (10.131) to have oscillatory behavior for $\sigma^2 - k^2 + \dfrac{k}{\sigma} - y^2 > 0$, and exponential behavior for $\sigma^2 - k^2 + \dfrac{k}{\sigma} - y^2 < 0$. Because of the $-y^2$ term, exponential behavior is guaranteed to emerge sufficiently far from the Equator, and, as with the Kelvin wave, we want this to be exponential decay rather than exponential growth. Therefore, as boundary conditions, we use

$$\widehat{v} \to 0 \text{ as } y \to \pm\infty. \tag{10.132}$$

It can be shown that nontrivial solutions satisfying these boundary conditions exist when

$$\boxed{\sigma^2 - k^2 + \frac{k}{\sigma} = 2n+1 \text{ for } n = 0,\ 1,\ 2\ldots.} \tag{10.133}$$

This is a dispersion equation. The expression on the right-hand side of (10.133) generates all positive odd integers, so that Eq. (10.133) is equivalent to the statement that $\sigma^2 - k^2 + \dfrac{k}{\sigma}$ is a positive odd integer. You can confirm by substitution that the solutions of (10.131) are

$$\widehat{v}(y) = Ce^{-\frac{1}{2}y^2} H_n(y), \tag{10.134}$$

where C is an arbitrary real constant and $H_n(y)$ is the nth Hermite polynomial, which is given by

$$H_n(y) \equiv (-1)^n e^{y^2} \frac{d^n}{dy^n}\left(e^{-y^2}\right) \tag{10.135}$$

(see Appendix I on Hermite Polynomials). Just as for the equatorial Kelvin wave, the factor $e^{-\frac{1}{2}y^2}$ in (10.134) ensures that these modes decay rapidly away from the Equator. For realistic values of the dimensional parameters, the dimensional e-folding distance turns out to be about 1000 km.

Because the solutions (10.134) proportional to Hermite polynomials, and because the different Hermite polynomials are distinguished by their values of the parameter n, we can think of the meridional shapes of the Equatorial waves as being determined by the value of n. Solutions with the same value of n will have the same meridional shape. Although (10.133) does not apply to the Kelvin wave and is only valid for non-negative values of n, it turns out that $\sigma = -k$ (the dispersion relation for the Kelvin wave) is a solution of (10.133) if we set $n = -1$. For this reason, the Kelvin wave can be called the $n = -1$ solution of Matsuno's model.

The dispersion equation, (10.133), is cubic in σ, which means that there are three σs for each $(k,\ n)$ pair. Two of these correspond to inertia-gravity waves. They can be approximated by

$$\sigma_{1,2} \cong \pm\sqrt{k^2+2n+1} \tag{10.136}$$

The third root corresponds to an equatorial Rossby wave. It can be approximated by

$$\sigma_3 \cong \frac{k}{\sqrt{k^2+2n+1}} \tag{10.137}$$

Notice that the frequency of the Rossby wave is smaller (in absolute value) than that of the Kelvin wave. For a given value of k, the eastward phase speed of the Kelvin wave is greater than the westward phase speed of the Rossby wave.

For the special case $n = 0$, the dispersion equation (10.133) can be factored, giving

$$(\sigma - k)\left(\sigma^2 + k\sigma - 1\right) = 0\,, \tag{10.138}$$

Matsuno showed that for $n = 0$ the three roots can be interpreted as follows:

$$\text{Eastward gravity wave:} \quad \sigma_1 = -\frac{k}{2} - \sqrt{\left(\frac{k}{2}\right)^2 + 1}\,, \tag{10.139}$$

$$\text{Westward gravity wave:} \quad \sigma_2 = \begin{cases} -\dfrac{k}{2} + \sqrt{\left(\dfrac{k}{2}\right)^2 + 1} & \text{for} \quad k \le \dfrac{1}{\sqrt{2}} \\ k & \text{for} \quad k \ge \dfrac{1}{\sqrt{2}}\,, \end{cases} \tag{10.140}$$

$$\text{Rossby wave:} \quad \sigma_3 = \begin{cases} k & \text{for} \quad k \le \dfrac{1}{\sqrt{2}} \\ -\dfrac{k}{2} + \sqrt{\left(\dfrac{k}{2}\right)^2 + 1} & \text{for} \quad k \ge \dfrac{1}{\sqrt{2}}\,. \end{cases} \tag{10.141}$$

Notice that (with $n = 0$) the westward gravity wave and the Rossby wave are not really distinct. They coincide for $k = 1/\sqrt{2}$. For reasons that were discussed earlier, the root $\sigma = k$ has to be thrown out in (10.140) and (10.141). Matsuno concluded, therefore, that for $n = 0$ there are actually only two waves: an eastward-moving gravity wave, and a "mixed Rossby-gravity wave," which is also known as the Yanai wave, after Michio Yanai [1]. The Yanai wave behaves like a gravity wave for $k < 1/\sqrt{2}$, and like a Rossby wave for $k > 1/\sqrt{2}$. The dispersion relation for the Yanai wave is

$$\sigma = -\frac{k}{2} + \sqrt{\left(\frac{k}{2}\right)^2 + 1}\,, \tag{10.142}$$

for all values of k. Because $H_0(y) = 1$, as can be seen from (10.135), Eq. (10.134) reduces to

$$\widehat{v}(y) = Ce^{-\frac{1}{2}y^2} \tag{10.143}$$

for the Yanai wave. This shows that in Yanai waves the meridional velocity has the same sign on both sides of the Equator and is a maximum on the Equator. A Yanai wave consists of a chain of vortices propagating westward along the Equator. As discussed later, the Yanai wave bears some resemblance to the so-called "tropical Easterly waves," which are found off the Equator near the latitude of the ITCZ, and which sometimes transition into tropical cyclones.

The dispersion relations associated with the various-solutions of Matsuno's model are plotted in Fig. 10.17. Recall that positive values of the frequency correspond to westward propagating waves, and negative values to eastward propagating waves. The dashed line proceeding downwards towards the right from the origin represents the Kelvin wave. The solid curves arcing upward from the origin represent Rossby waves, with positive values of n. The dashed curves in the upper part of the diagram correspond to westward-propagating inertia-gravity waves, and the solid curves in the lower part of the diagram correspond to eastward propagating inertia-gravity waves. The westward propagating wave represented by the curve that is partly solid and partly dashed is the Yanai wave. The dashed portion of this curve, plotted for $k < 1/\sqrt{2}$, represents those wave numbers for which the Yanai wave behaves like a westward propagating gravity wave. The solid portion of the curve, for $k > 1/\sqrt{2}$, represents those wave numbers for which the Yanai wave behaves like a Rossby wave.

[1] Matsuno and Yanai shared an office when they were graduate students.

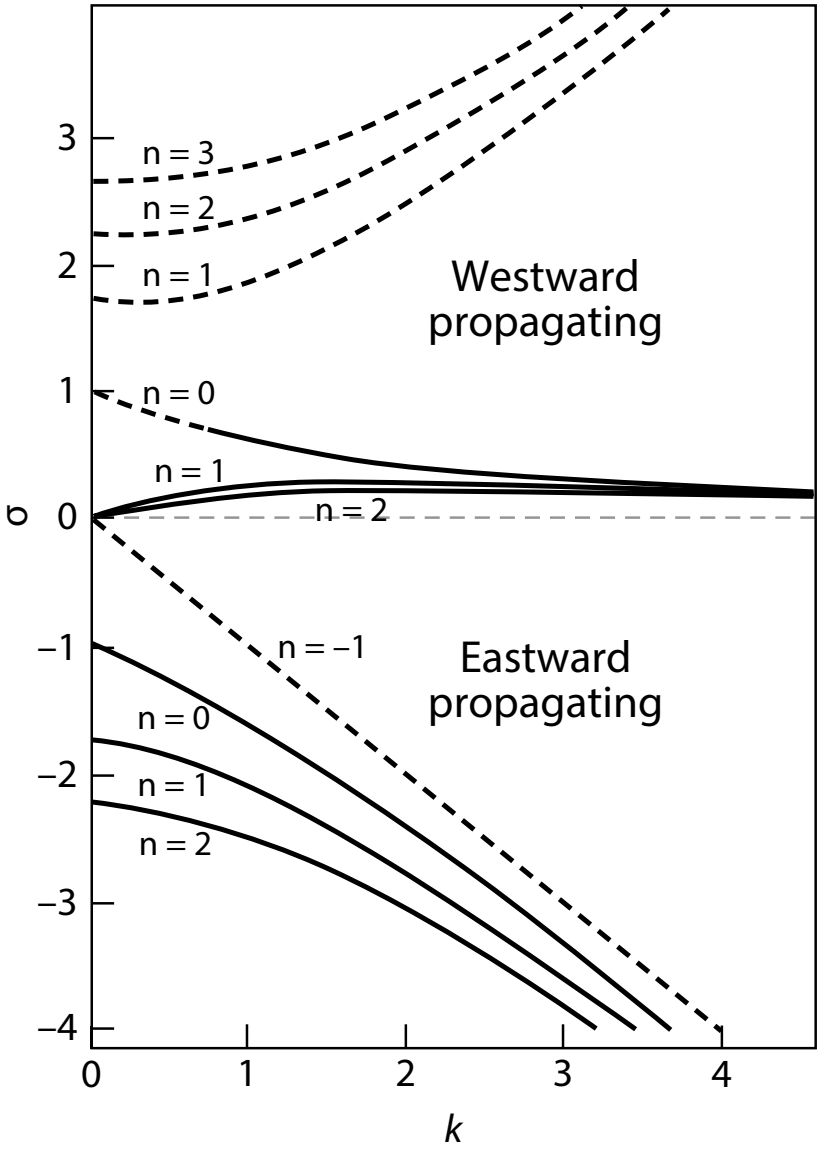


Figure 10.17: Frequencies as functions of zonal wavenumber. Positive frequencies correspond to westward propagation. The eastward-propagating modes labeled $n = 0,\ n = 1,$ and $n = 2$ (solid lines) are eastward-propagating inertia-gravity waves. The westward-propagating modes labeled $n = 3,\ n = 2,$ and $n = 1$ (dashed lines) are westward-propagating inertia-gravity waves. The westward-propagating modes labeled $n = 1$ and $n = 2$ (solid lines) are Rossby waves. The eastward-propagating mode labeled $n = -1$ (straight dashed line) is the Kelvin wave. The westward-propagating wave labeled $n = 0$ is the mixed Rossby-gravity or Yanai wave. It behaves like a gravity wave and is denoted by a dashed line for $k < 1/\sqrt{2}$, and it behaves like a Rossby wave and is denoted by a solid line for $k > 1/\sqrt{2}$. From Matsuno (1966).

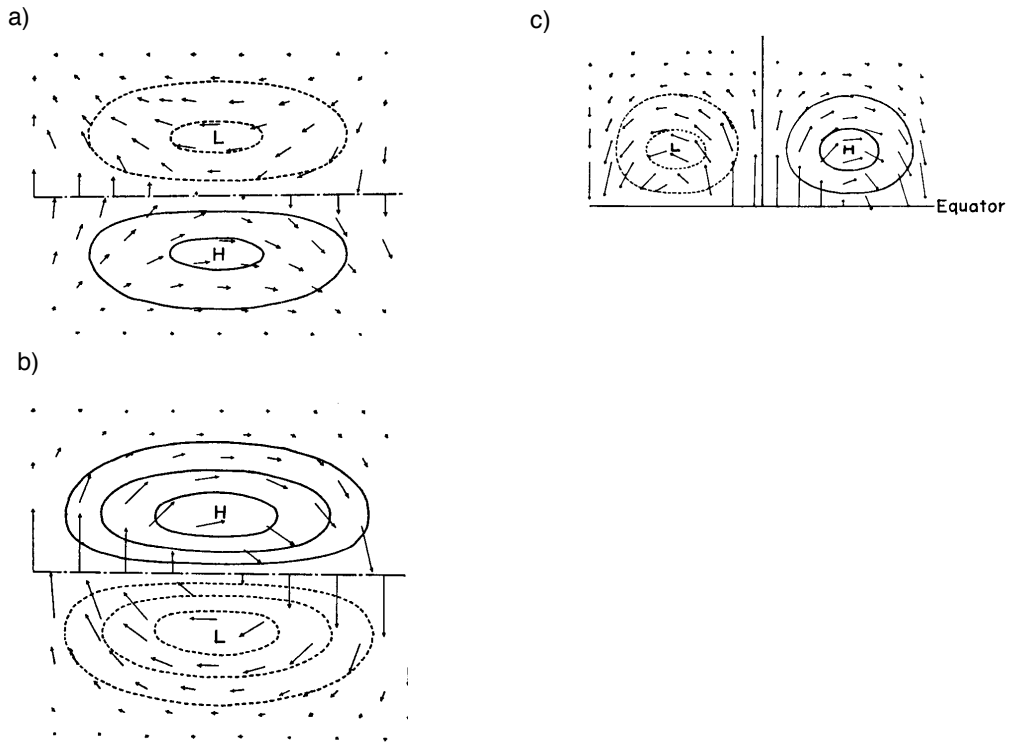


Figure 10.18: Pressure and velocity distributions of solutions for $n = 0$ and $k = 0.5$. (A) Eastward-moving inertia-gravity wave. (B) Westward-moving Yanai wave, which for this value of k behaves like an inertia-gravity wave. (C) The structure of the Yanai wave for $n = 0$ and $k = 1$, in which case the Yanai wave acts like a Rossby wave. For each mode, v is a maximum on the equator and does not pass through zero anywhere. This is characteristic of $n = 0$. From Matsuno (1966).

For $n = 0$, the eastward-moving inertia-gravity wave and westward-moving Yanai wave have the structures shown in panels a) and b) of Fig. 10.18. For a pure gravity wave we expect the winds to be perpendicular to the isobars. When rotation is dominant, the winds are parallel to the isobars. The waves shown look like pure gravity waves near the Equator. For $n = 0$ and $k = 1$, the Yanai wave takes on the characteristics of a Rossby wave, as shown in the lower panel.

Solutions for are shown on the left side of Fig. 10.20. The corresponding results for $n = 2$ are shown on the right side of the figure. Recall that the subscript n denotes the solution whose meridional structure is described by the nth Hermite polynomial. As can be seen in the figure, higher values of n correspond to more nodes in the meridional direction.

The very simple structure of the Kelvin wave is shown in Fig. 10.20. The velocity vectors are purely zonal, as expected, and the zonal wind is in phase with the pressure, as in a gravity wave.

The equatorially trapped waves found by Matsuno correspond to solutions of Laplace's

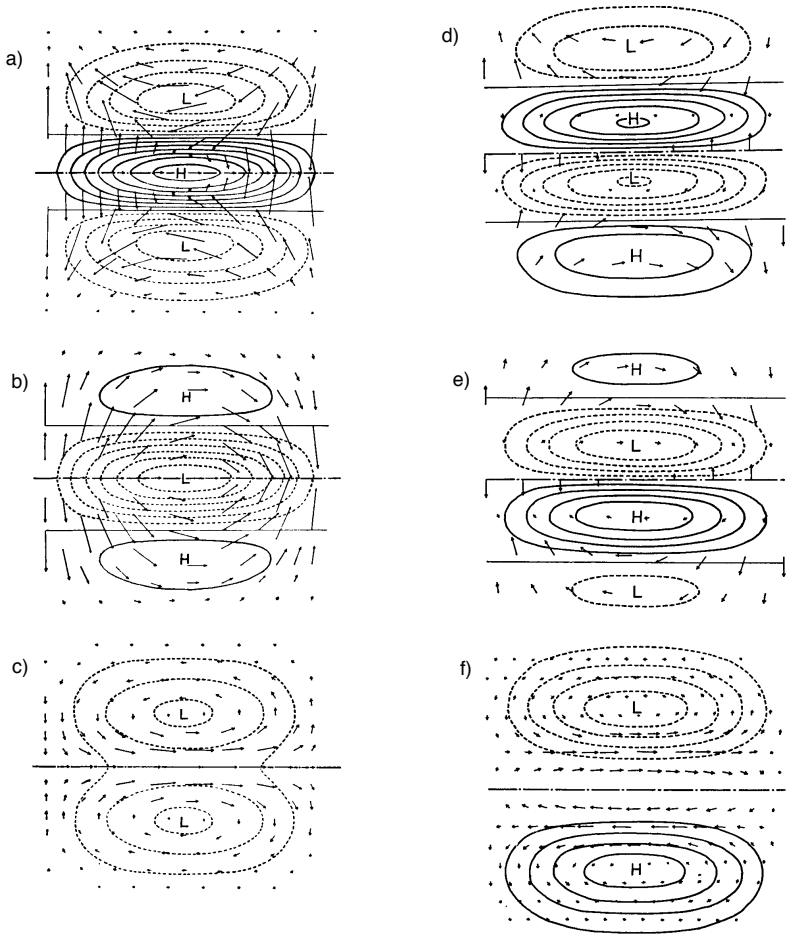


Figure 10.19: Left side: Pressure and velocity distributions of solutions for $n = 1$. a) Eastward-propagating inertia-gravity wave. For each mode, $u = 0$ on the equator, as expected for $n = 1$. b) Westward-propagating inertia-gravity wave. c) Rossby wave. Right side: Corresponding results for $n = 2$. For each mode, u is symmetric across the equator, as expected for $n = 2$. From Matsuno (1966).

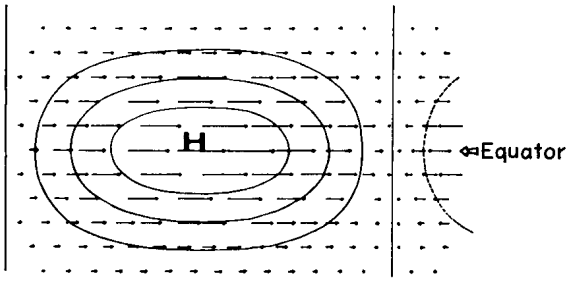


Figure 10.20: Pressure and velocity distributions for $n = -1$ and $k = 0.5$. This is the Kelvin wave. From Matsuno (1966).

equations, if allowances are made for the effects of stratification and sphericity. A vertically continuous (i.e., non-shallow-water) version of Matsuno's analysis was presented by Lindzen (1967).

Matsuno's theoretical discoveries were confirmed by observations in remarkably short order. Maruyama and Yanai (1967) observed the Yanai wave not long after Matsuno had predicted its existence, and Wallace and Kousky (1968) found the Kelvin wave soon thereafter.

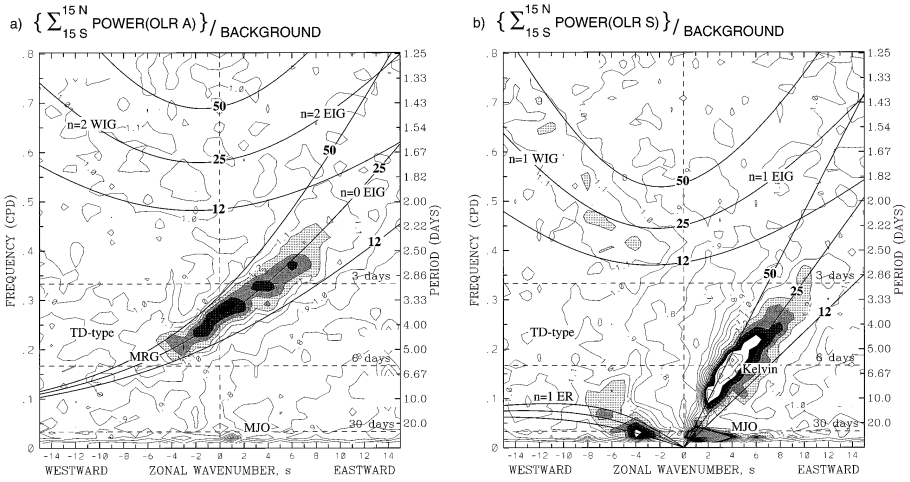


Figure 10.21: The variability of the outgoing longwave radiation (OLR) as a function of frequency and zonal wavenumber for modes that are symmetric across the equator (right panel) and antisymmetric (left panel). Eastward propagation is associated with positive wavenumbers. The boxes indicated with dashed lines select particular wave types. These dispersion diagrams are analogous to those used by Matsuno, as shown in Fig. 10.17, but the conventions are different. With Matsuno's convention the zonal wavenumber is nonnegative, and the sign of the frequency determines the direction of zonal propagation, while in the figures shown here the frequency is nonnegative, and the sign of the zonal wavenumber determines the direction of zonal propagation. From Wheeler and Kiladis (1999). © American Meteorological Society. Used with permission.

Tropical waves continue to be the subject of many observational, theoretical, and numerical studies (Kiladis et al., 2009). In a particularly famous study, Wheeler and Kiladis (1999) examined the space-time variability of the tropical outgoing long wave radiation. In Fig. 10.21, the data have been separated into modes that are symmetric across the Equator (right panel), such as the Kelvin wave, and modes that are anti-symmetric across the Equator (left panel), such as the mixed-Rossby-gravity wave. Through the use of additional filtering procedures motivated by Matsuno's results, Wheeler and Kiladis were able to show the longitudinal propagation of various types of equatorially trapped disturbances (see Figs. 10.22 and 10.23).

Matsuno's model describes free waves that can exist without any forcing, although as discussed in the next section he also considered stationary forced solutions. Both dry and moist Equatorial waves are observed, and the latter are often described as "convectively

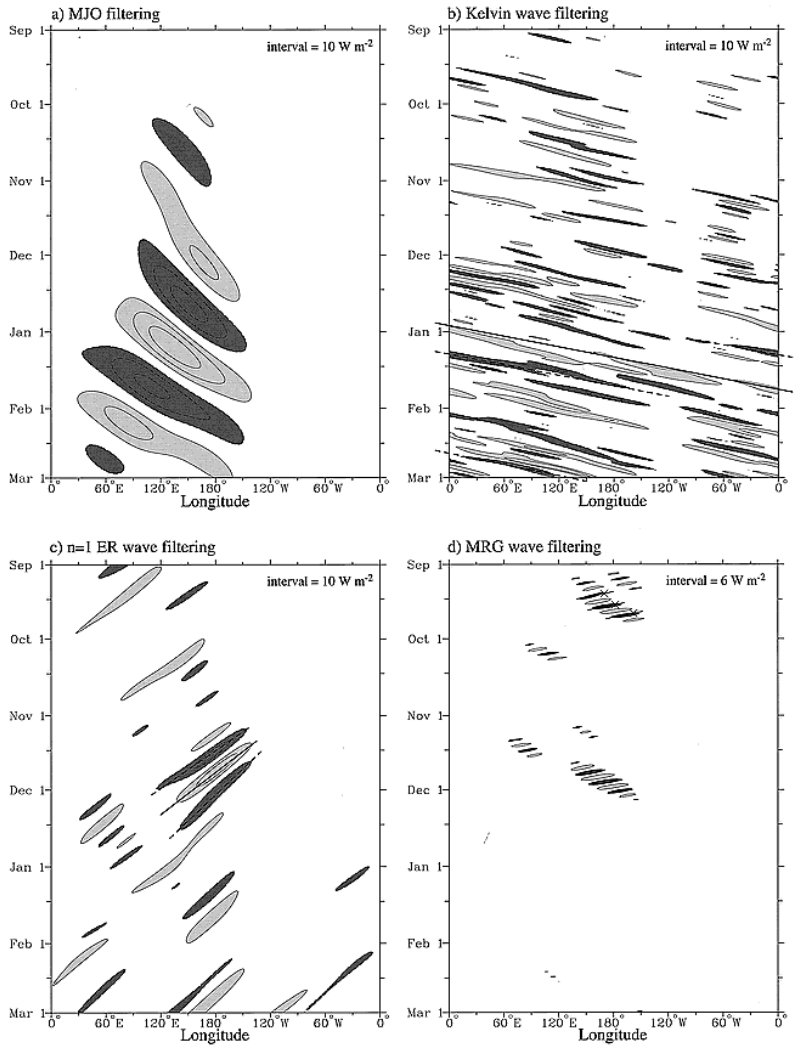


Figure 10.22: Longitudinal propagation of the Madden-Julian oscillation (MJO), Kelvin waves, equatorial Rossby (ER) waves, and mixed-Rossby-gravity (MRG) waves, as seen in the OLR. The zero contour has been omitted. The various modes are selected by including only the contributions from wavenumbers and frequencies that fall within the corresponding boxes in Fig. 10.21. This is what is meant by filtering. From Wheeler and Kiladis (1999). © American Meteorological Society. Used with permission.

coupled." This rather vague phrase can mean various things. As the waves propagate across the tropics, they inevitably influence the ambient convection that is supported by surface evaporation and radiative cooling. For example the waves can promote or suppress convective activity in regions of wave-induced upward or downward motion. The convection also influences the waves, of course. For example, convective heating in regions of wave-induced upward motion reduces the cooling that would otherwise occur there, so that the waves feel an "effective static stability" that is somewhat weaker than the actual static stability, and this reduces their the phase speed. Interactions of this type do represent

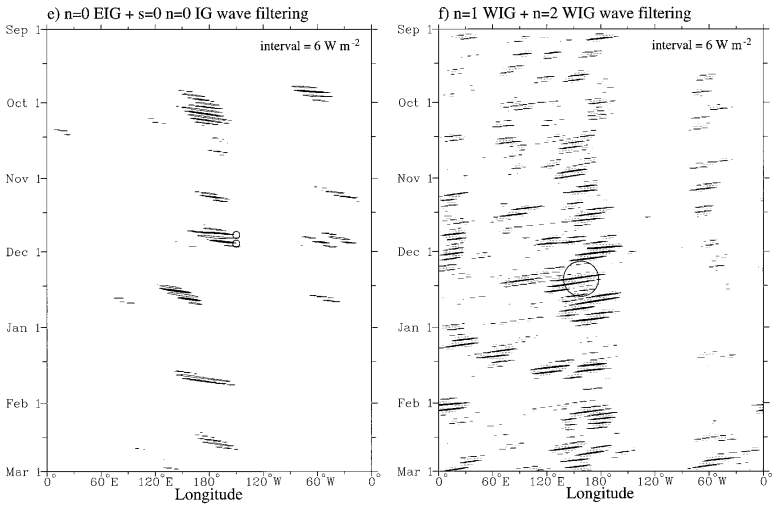


Figure 10.23: Longitudinal propagation of eastwardand westward-propagating inertia-gravity waves, as seen in the OLR. The zero contour has been omitted. As in Fig. 10.22, the various modes are selected by including only the contributions from wave numbers and frequencies that fall in the corresponding boxes in Fig. 10.21. From Wheeler and Kiladis (1999). © American Meteorological Society. Used with permission.

coupling of the waves with the convection, but it is incidental coupling in the sense that the free-wave solutions of Matsuno's model can exist without the convection (although the real-world waves are naturally modified by convection), and the convection can exist without the waves (although it is modified by the waves). Convective excitation of the waves can be quasi-random, like a cat walking on the keys of a piano.

We can also imagine a more essential coupling of convection with tropical eddies, in which the very existence of the eddies depends on the convective heating. Tropical cyclones and the Madden-Julian oscillation are examples of eddies that cannot exist without convection. They are discussed later in this chapter.

10.8 The response of the tropical atmosphere to stationary heat sources and sinks

The discussion above is all about "free waves." Forced solutions to Matsuno's model are also very relevant to observations of monsoons and the Madden-Julian Oscillation, both of which are discussed later in this chapter. Fig. 10.24, which appears near the end of the paper by Matsuno (1966), shows the stationary circulation driven by a mass source and sink on the Equator. Think of this figure in terms of the low-level flow. The mass sink can be interpreted as a region of rising motion, where the air is converging at low levels, e.g., in the western equatorial Pacific. The mass source can be interpreted as a region of sinking motion, where the air is diverging at low levels, e.g., in the eastern equatorial Pacific.

(Unfortunately the mass sink is plotted on the east side and the mass source is plotted on the west side, but this does not really matter because the solution is periodic in the zonal direction anyway.) The model predicts strong westerlies converging (from the west, of course) at low levels into the region of rising motion, and low-level easterlies converging on the east side of the region of low-level convergence. The easterlies can be interpreted as the trades, and as the lower branch of the Walker circulation, which is discussed later in this chapter. The westerlies can be interpreted as a "monsoon-like" westerly inflow to a region of heating. Monsoons are also discussed later in this chapter.

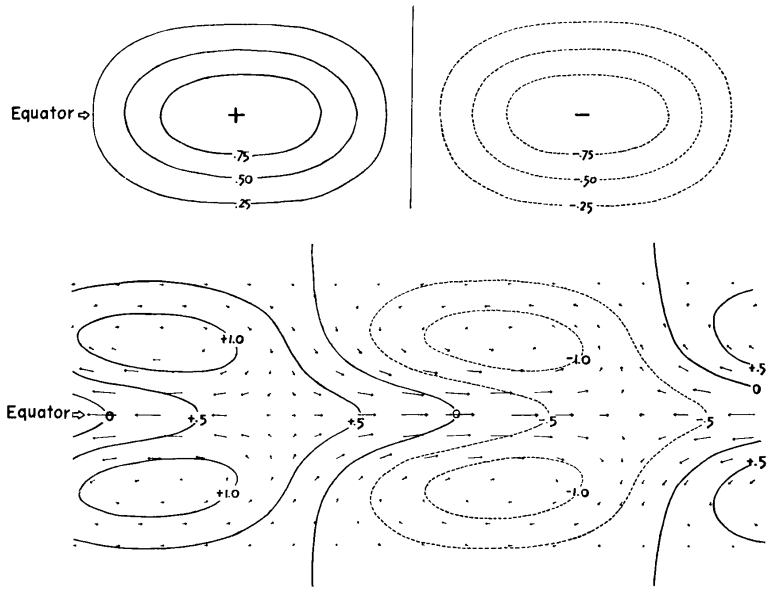


Figure 10.24: Stationary circulation pattern (lower panel) forced by the mass source and sink shown in the upper panel. From Matsuno (1966).

Webster (1972) and Gill (1980) followed Matsuno's lead by developing simple analytic models of the response of a resting tropical atmosphere to heat sources and sinks. Gill examined the atmospheric response to a relatively small-scale heating source that is centered on the equator. If the atmosphere is abruptly heated at some initial time, Kelvin waves propagate rapidly eastward and generate easterly trade winds to the east of the heating. Thus the easterly trade winds in the Pacific could result from Kelvin waves produced by convective heating over Indonesia. Similarly, equatorial Rossby waves propagate westward and generate westerlies to the west of the heating. Because the fastest Rossby wave travels at only one-third the speed of the Kelvin wave, the effects of the Rossby waves would be expected to reach only one-third as far those as those of the Kelvin wave. Gill interpreted the westerlies over the Indian ocean as a response to Rossby waves generated by convective heating over Indonesia.

Gill (1980) studied what amounts to a steady-state version of Matsuno's model, and introduced forcing in the form of mass sources and sinks, along with very simple damping. Corresponding to (10.115), we have

$$\begin{aligned} \varepsilon u - yv + \frac{\partial \phi}{\partial x} &= 0 \,, \\ yu + \frac{\partial \phi}{\partial y} &= 0 \,, \\ \varepsilon \phi + \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= -Q \,. \end{aligned} \tag{10.144}$$

and, as a purely diagnostic relation,

$$w = \varepsilon \phi + Q \tag{10.145}$$

Here u and v should be interpreted as the lower-tropospheric wind components. In (10.144) and (10.145), ε^{-1} is a dissipation time scale, and Q is a "heating rate" that must be specified. The variables ϕ, w, and Q are defined in the middle troposphere. Gill included damping in the form of Rayleigh friction and Newtonian cooling, and for simplicity assumed that their time scales, given by ε^{-1}, are equal. Rayleigh friction is a simple parameterization of friction in which the velocity is divided by a frictional time scale. The friction term is neglected in the meridional momentum equation of (10.144); see Gill (1980) for an explanation.

Gill focused primarily on cases for which the heating is symmetric or anti-symmetric about the equator. The solution for symmetric heating resembles a Walker circulation, with lower-tropospheric inflow into the heating region and upper-tropospheric outflow. The Walker Circulation is discussed in detail later in this Chapter. The surface easterlies cover a larger area than the surface westerlies because the phase speed of the eastward-propagating Kelvin wave is three times faster than that of the westward-moving Rossby wave.

By using (10.144) to form a vorticity equation for the case of no damping, and then substituting from the continuity equation, Gill found that

$$v = yQ. \tag{10.146}$$

This simple equation is closely related to what is sometimes called "Sverdrup balance," in which the "meridional advection of the Coriolis parameter," i.e., the so-called β-term of the

vorticity equation, is balanced by the divergence term, which is represented by the heating rate on the right-hand side of (10.146). According to (10.146), v changes sign across the equator in a region where Q has a single sign. In a region of cooling ($Q < 0$), the flow is towards the Equator on both sides, and in a region of heating it is away from the Equator on both sides.

For $Q > 0$, (10.146) implies poleward motion in the lower layer and equatorward motion in the upper layer. This suggests that in regions of heating, e.g. the western Pacific, the Walker circulation produces a north-south circulation that opposes the Hadley circulation. Geisler (1981) found the same result. For $Q < 0$, the low-level motion is Equatorward; this is what we see in the subtropical highs, e.g., in the eastern Pacific.

The solution for anti-symmetric heating consists of a mixed Rossby-gravity wave and a Rossby wave. There is no Kelvin-wave response because the Kelvin wave is intrinsically symmetric across the Equator. Long mixed Rossby-gravity waves do not propagate, and so the response of this wave type is largely confined to the region of heating. Due to the westward propagation of Rossby waves, no response is generated to the east of the forcing region. To the west, the region of westerly flow into the heating region is limited because the Rossby modes travel slowly, and so are dissipated before they can propagate far to the west.

Gill interpreted the asymmetric case as a simulation of the Hadley circulation, and the symmetric case as a simulation of the Walker circulation, which is discussed later in this chapter.

For heating centered on the Equator, as in Fig. 10.25, Gill found strong westerlies on the west side, and strong easterlies on the east side, combining to give strong zonal convergence on the heating. The westerlies can be interpreted as the time-averaged response to westward -propagating Rossby waves excited by the heating, and the easterlies can be interpreted as the time-averaged response to eastward-propagating Kelvin waves excited by the heating. This implies an east-west overturning circulation, shown the bottom panel of Fig. 10.25, and a surface pressure field with a minimum pressure slightly to the west of the heating, depicted by the contours in the middle panel of the figure. Later in this chapter we will discuss the Walker Circulation, which resembles the bottom panel in Fig. 10.25.

When the heating is antisymmetric across the Equator, as in Fig. 10.26, the model produces something like a Hadley circulation, with a low-level cyclonic circulation on the side with positive heating, and a low-level anticyclone on the other side. When the symmetric and antisymmetric heatings are combined, as in Fig. 10.27, the model produces a circulation that looks remarkably similar to that of the Asian summer monsoon, as discussed in the next subsection.

Although Gill demonstrated that heating of limited extent generates tropical waves that produce broad wind and pressure fields resembling the observations, his results must be viewed with caution due to several limitations. First, the model was linearized about a

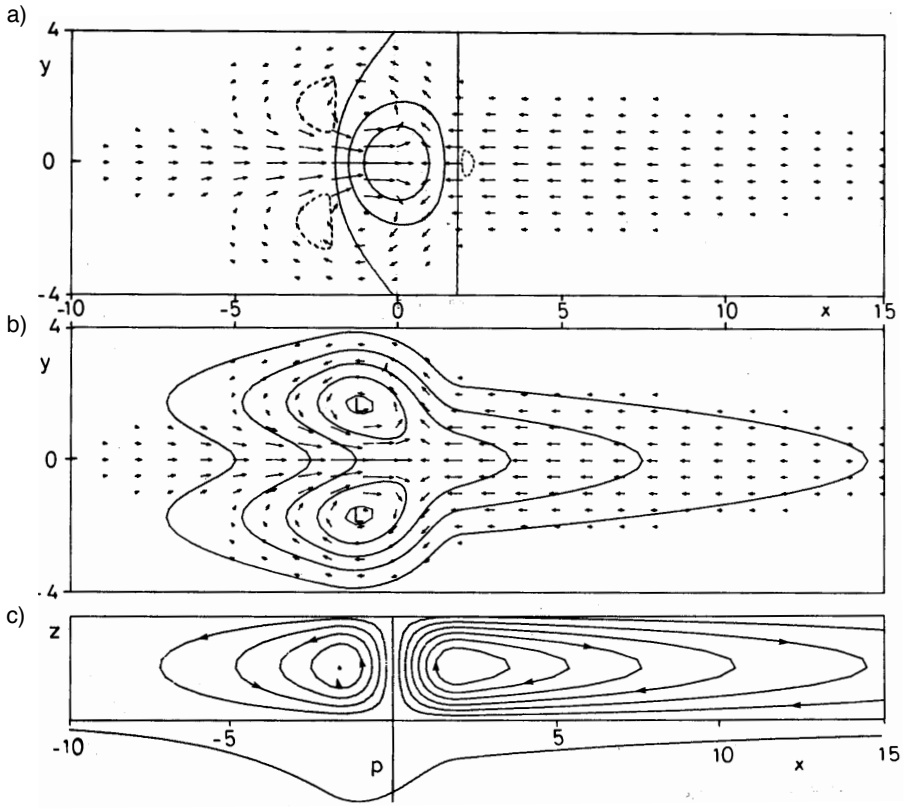


Figure 10.25: Solution of Gill's model for the case of heating symmetric about the equator. Panel A shows the heating field and the low-level wind field. Panel B shows the perturbation pressure field, which features low pressure along the equator generally, with twin cyclones slightly off the equator. Panel C shows the implied vertical motion and the zonal variation of the pressure along the equator. (Panel C is very famous.) From Gill (1980). © Quarterly Journal of Royal Meteorological Society.

specified basic state, and so it cannot explain the basic state. Second, results were generated for a specified rather than predicted heating of the tropical troposphere, and so ocean-atmosphere interactions and feedbacks involving moist convection were excluded. Third, the model includes neither a moisture budget nor cloud radiative effects.

10.9 Monsoons

Monsoons occur in many parts of the world. In the older literature, a monsoon is defined in terms of a dramatic seasonal reversal of the low-level prevailing winds (Lighthill and Pearce, 1981). From a more modern perspective, a monsoon can be viewed as a thermally forced stationary planetary wave associated with land-sea contrast. Comprehensive overviews are given by Webster et al. (1998) and Chang et al. (2011).

The most spectacular monsoon on Earth is the one associated with Asia, Earth's largest

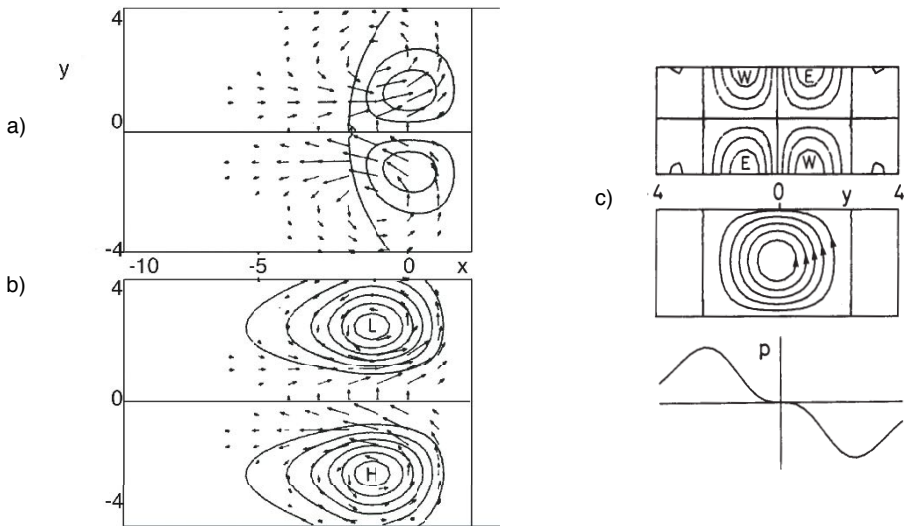


Figure 10.26: The response to antisymmetric heating. On the left side, panel A shows contours of the midlevel vertical velocity superimposed on the horizontal wind vectors for the lower layer. Panel B shows contours of the perturbation surface pressure, again with the lower-layer horizontal wind field superimposed. The right-hand panels show the zonally integrated solution corresponding to the results in the left-hand panels. The upper right-hand panel shows the latitude-height distributions of the zonal velocity and the streamfunction of the mean meridional circulation, as well as the meridional profile of the surface pressure. From Gill (1980). © Quarterly Journal of Royal Meteorological Society.

continent. The Asian monsoon can be defined in a number of ways. In the dramatic case of the Asian monsoon, the winds near the surface reverse from the northeast in winter to the southwest in summer, as seen in Fig. 10.28. The 15 m s^{-1} low-level southwest wind that crosses the Equator and flows from the coast of east Africa to the shores of India is known as the Somali jet. It is one of the strongest low-level jets in the world. There are other seasonal changes of wind in the world, but none have the geographical scope or the socioeconomic impact of the Asian monsoon.

The basic trigger for the monsoon is a contrast between the temperature of the Asian continent and that of the surrounding ocean. The thermal anomaly in the middle troposphere is enhanced by the spectacular topography of the Tibetan Plateau (Fig. 10.29), which extends upward to about the 500 hPa level. Much of the "surface" heating associated with the summer monsoon actually occurs in the middle troposphere, because it is located on the Tibetan Plateau. The Tibetan Plateau towers above the surrounding land surface, with average elevations over the central Plateau of over 3000 m. The observed JJA mean 500 hPa temperature for the monsoon region is shown in Fig. 10.30. An "island" of warm air is centered over the Tibetan plateau. The warm air is in contact with the mountainous terrain, which is heated by the sun.

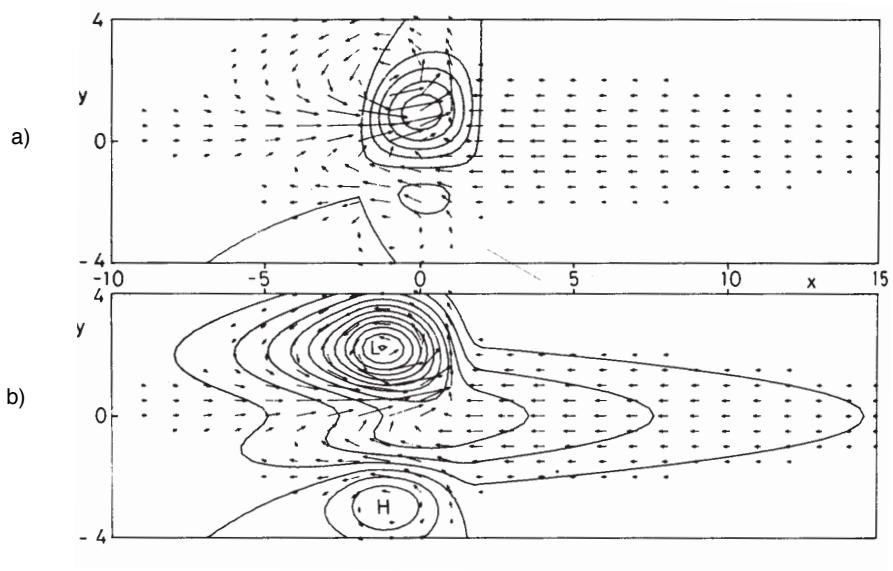


Figure 10.27: The response of Gill's model to a combination of symmetric and antisymmetric heating. Panel A shows the heating, and panel B shows the surface pressure. The low-level winds are shown in both panels. From Gill (1980). © Quarterly Journal of Royal Meteorological Society.

Naturally, the lowest surface air temperatures in the monsoon region occur at the highest elevations. There is a large area of low (less than 280 K) surface air temperature on the Tibetan Plateau, and a large area of high (greater than 305 K) surface air temperatures on the Arabian Peninsula. After the seasonal snow on the Plateau has melted in late spring and early summer, the surface and the air above it are heated to a temperature higher than that of the surrounding atmosphere. Rising motion balances this heating, and this forces convergence in the lower and middle troposphere, and compensating divergence aloft. As the seasonal heating builds, a trough forms over southern India in late May, and subsequently moves north and west. The monsoon trough is one of the most prominent signatures of the Asian monsoon. It can extend across the Indian Ocean, the Indian subcontinent, and the Arabian Sea, from Bangladesh to the Arabian Peninsula. The sea level pressure field is dominated by the monsoon trough, and an area of high pressure over the Tibetan Plateau.

Because of the agricultural importance of monsoon precipitation, and the vast populations living in Asia, the onset of the Asian summer monsoon is one of the most anticipated events in the world. The onset is can be defined as the beginning of consistent rainfall of the monsoon season. Fig. 10.31 shows typical dates of onset. Cloudiness and precipitation usually begin to increase at the southern tip of India in late May. Nearly the entire subcontinent is receiving rainfall by the end of June. The onset of the summer monsoon brings cooler surface temperatures to India and other areas that receive monsoonal precipitation, due to the increase in clouds as well as the increase in soil moisture that accompanies the precipitation.

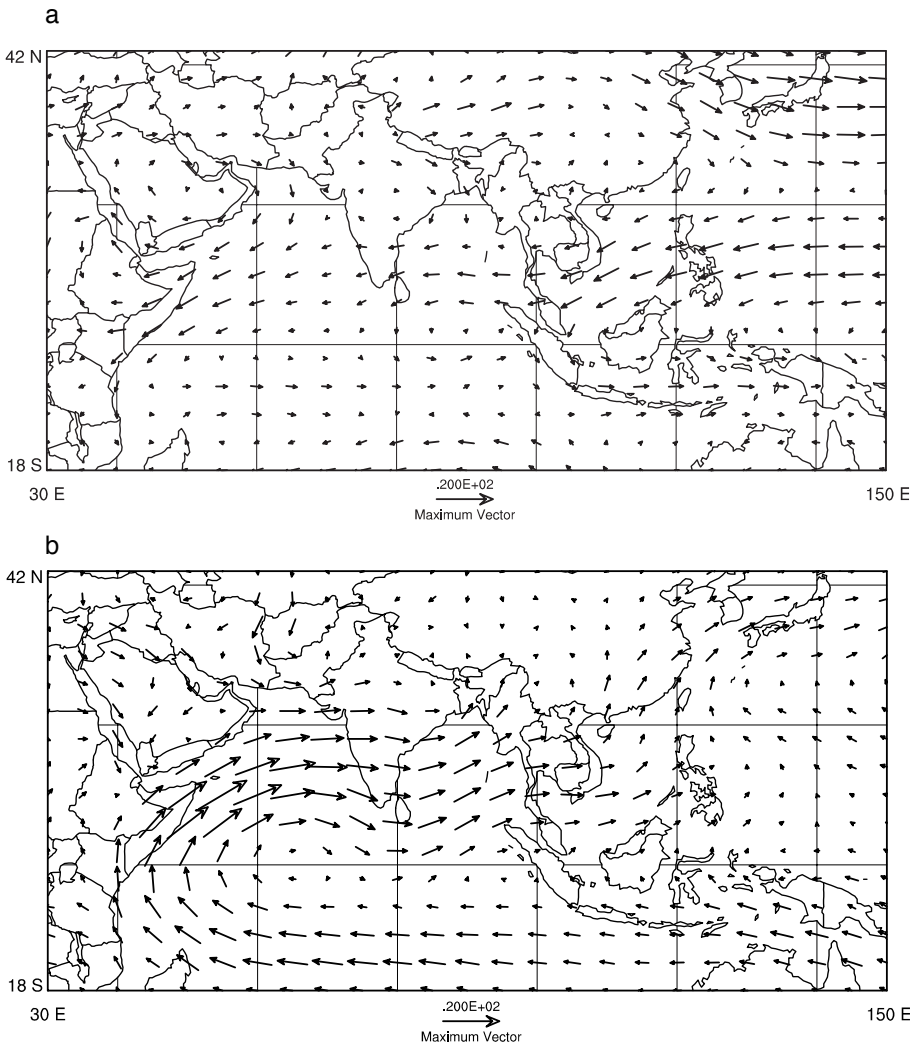


Figure 10.28: Observed 850 hPa wind vectors for (A) January and (B) July.

The monsoon trough moves northward episodically, as shown in Fig. 10.32. In some cases, the progression of the monsoon trough is expedited by the passage of a tropical or extratropical cyclone to the north (Shukla and Mooley, 1987). The Ganges valley receives copious amounts of rain from monsoon depressions that form in the Bay of Bengal and propagate northward and westward. An example is shown in Fig. 10.32a. Many areas within the monsoon region also receive significant amounts of precipitation from tropical cyclones. Fig. 10.32b shows the daily precipitation totals at an average of several stations on the southwest coast of India, and gives a sense as to the observed intraseasonal variability of monsoon precipitation. Much of the precipitation of the Asian monsoon is forced by southwest winds flowing over the western shores of India and Southeast Asia, as well as the foothills of the Tibetan Plateau.

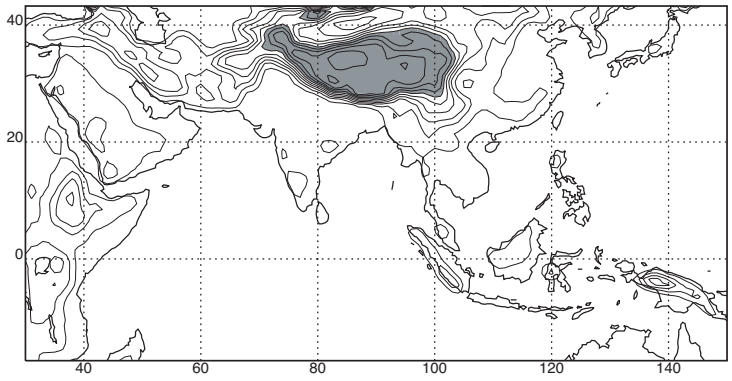


Figure 10.29: Average elevation of the monsoon region. Before plotting, the data were averaged to $1^\circ \times 1^\circ$, and then 9-point smoothed. Terrain over 3000 m high is shaded.

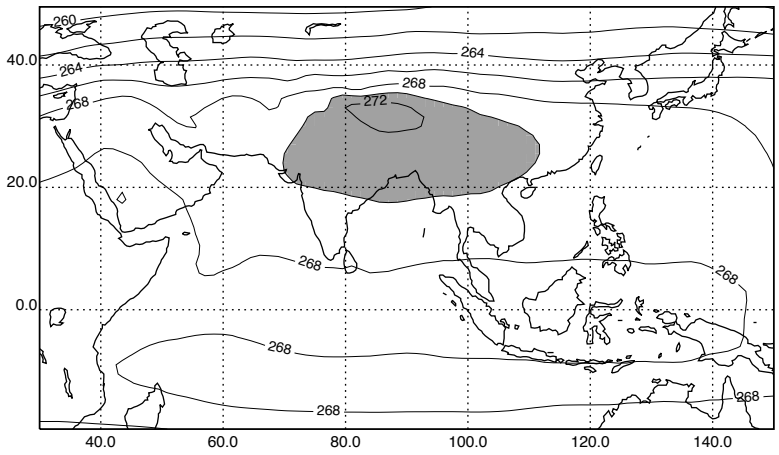


Figure 10.30: Observed JJA climatological 500 hPa temperatures. The contour interval is 2 K; values greater than 270 K are shaded.

Fig. 10.33 shows summer-seasonal means of the precipitation across the monsoon region. There are two major precipitation maxima, one west of the southwest coast of India, and the other west of southern Myanmar. Both of these areas receive strong onshore flow, showing that the ocean is the source of the moisture. Minima occur near Sri Lanka and the east coast of Vietnam. These areas appear to be in orographic rain shadows. The northern

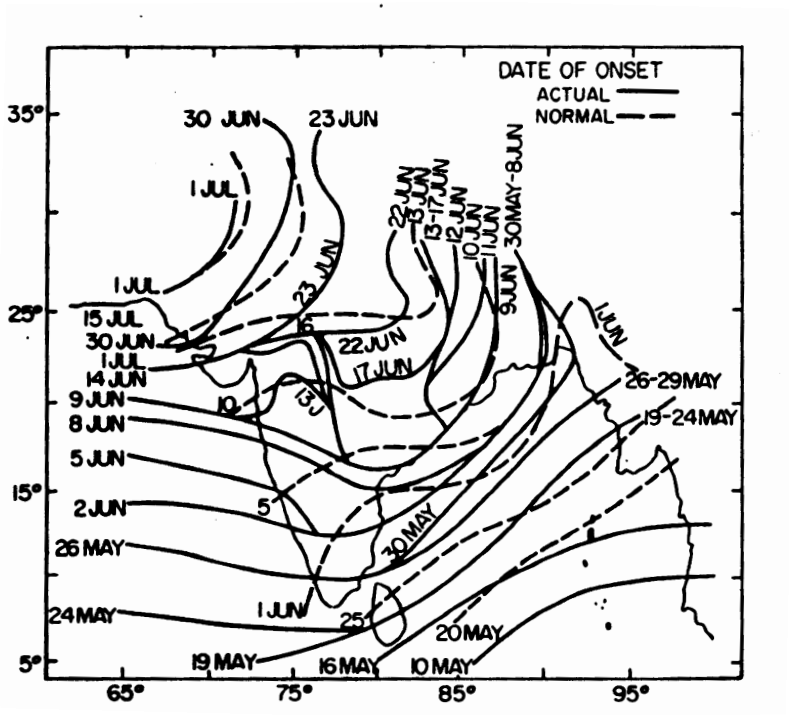


Figure 10.31: Dates of onset of the Asian monsoon near India in 1988 (actual) and mean (normal). From Krishnamurti et al. (1989). Used with kind permission from Springer Science and Business Media.

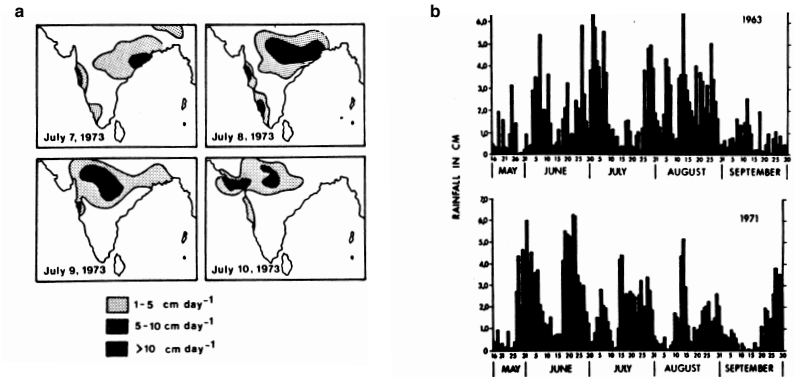


Figure 10.32: (A) An example of the progress of a monsoon depression across India. Many such depressions occur throughout the summer monsoon. (B) Observed daily rainfall along the southwest coast of India for the summer monsoon seasons of 1963 and 1971. From Webster (1987).

and western parts of the monsoon region are quite dry, receiving less than 2 mm day-1 of rain. The observed 850 hPa wind analysis shows that these areas do not receive much moisture from the Indian Ocean during JJA. Only a small portion of the monsoon region receives more than 20 mm day-1 of precipitation.

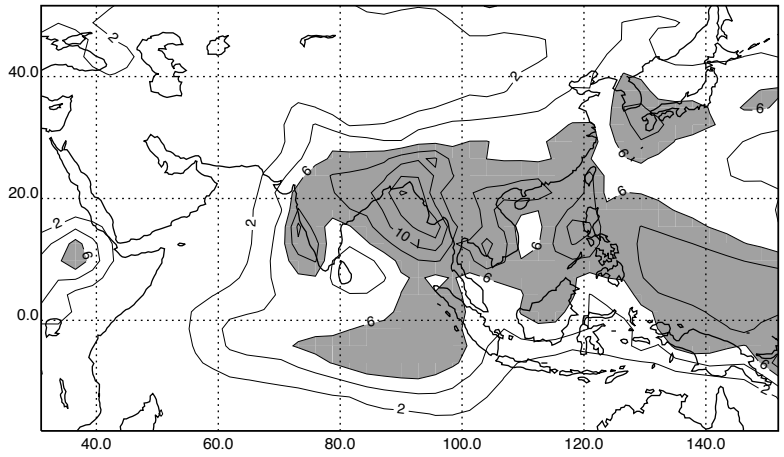


Figure 10.33: The observed JJA climatological precipitation from the Global Precipitation Climatology Project. The contour interval is 2 mm day^{-1}; values greater than 6 mm day^{-1} are shaded.

The precipitation in much of the monsoon region varies on many time scales, including the diurnal cycle (Johnson, 2011). The individual disturbances that cause the precipitation associated with the Asian monsoon last only a few days at any single location. However, there are also prominent variations known as "breaks," which occur at periods of approximately 10-20 days and 40-50 days (Webster, 1987). The 10-20 day variation is related to the periodic northward propagation of the Intertropical Convergence Zone (ITCZ), which begins near the Equator, and progresses to the foothills of the Tibetan Plateau in roughly 15 days, as seen in Fig. 10.34. The ITCZ usually reforms in the south after it has progressed to the foothills of the Tibetan Plateau, but occasionally it stays near the Plateau for an "extended break" period of 40-50 days. These extended break periods have been linked (Webster, 1987) in this chapter.

The heating of the middle troposphere by the Tibetan plateau induces convergence in the middle and lower troposphere (Yanai et al., 1992), largely through the Somali jet. To balance the convergence at the lower levels of the atmosphere, there must be large-scale rising motion, and divergence aloft. The observed JJA mean 500 hPa vertical velocity for the monsoon region is shown in Fig. 10.35. The strongest areas of rising motion are over southwestern India and the Bay of Bengal. These areas both receive copious amounts of precipitation in JJA. Two of the strongest areas of sinking motion are over the eastern Mediterranean Sea and northern China. Both areas are quite dry in JJA, receiving less than of rain. The area-mean vertical velocity is upward, at the rate of about -10 hPa day-1.

The upper-level divergence is associated with a broad, strong anticyclonic circulation over the Plateau at 200 hPa, as shown in Fig. 10.36. The strong, upper-level easterlies are consistent with the thermal wind relationship, because the lower and middle tropospheric temperatures actually increase towards the north in the Northern Hemisphere (Yanai et al., 1992; Murakami et al., 1986; Yanai and Li, 1994). There is a slight northerly component

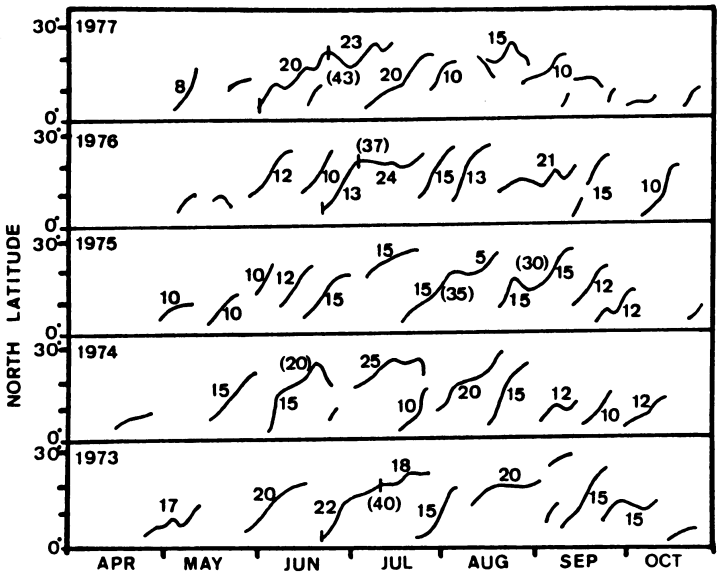


Figure 10.34: Mean latitudinal position of the monsoon trough in the Indian Ocean for the summers of 1973-1977, as obtained from the maximum cloudiness zone and the 700 hPa trough. Numbers refer to longevity of a particular cloudiness zone, with extended break periods indicated by parentheses. From Webster (1987), Webster (1983), and Sikka and Gadgil (1980). Copyright 1988 Royal Meteorological Society.

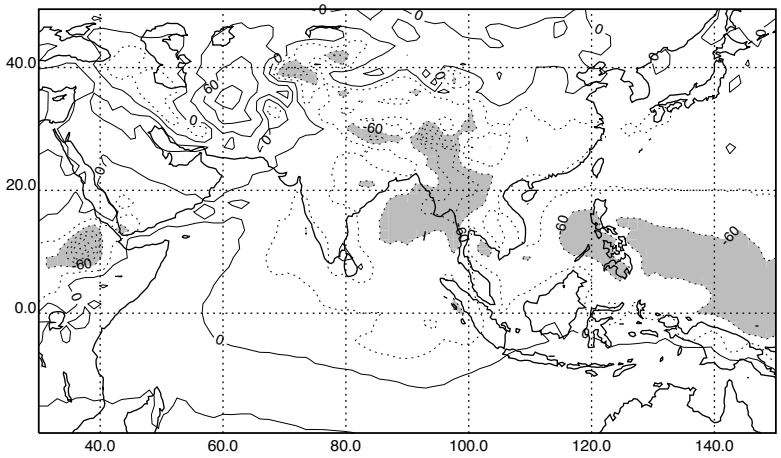


Figure 10.35: Observed JJA climatological 500 hPa vertical velocity over the Asian monsoon region. Contour interval is 20 hPa day^{-1}. Areas with vertical velocities more negative than -60 hPa day^{-1} are shaded.

to the 200-hPa winds at the Equator, especially on the eastern side of the region. The winds shift abruptly from easterlies to westerlies at the northern and southern fringes of

the monsoon region, with westerlies of up to at about 35 °N. Fig. 10.37 shows a latitude-pressure cross-section of the observed zonal wind at 77.5° E.

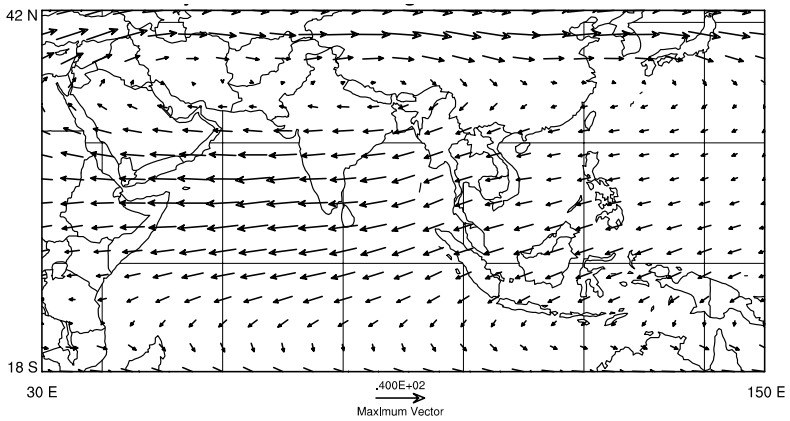


Figure 10.36: JJA climatological 200 hPa winds. The scale vector is 50 m s^{-1}.

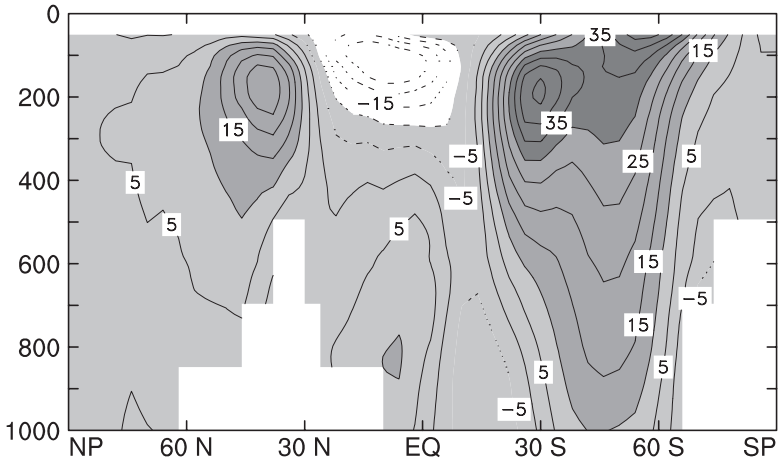


Figure 10.37: Latitude-pressure plot of the northern summer-season climatological zonal winds at 77.5° E. The contour interval is 5 m s^{-1}. Areas of high terrain are blocked out.

These observations show that the monsoon can be interpreted as a locally enhanced Hadley circulation (Webster, 1987), with a meridional low-level branch that moves towards a warm, convectively active ascending branch, and an upper-level return flow that feeds a relatively cool descending branch. In fact, in the northern summer the Hadley circulation is more or less contained within the longitudes of the monsoon region. The monsoon is thus a direct circulation, which converts potential energy to kinetic energy.

Observations show strong interannual variability of the Asian monsoon. Charney and Shukla (1981) argued on the basis of modeling studies that the such variations are due to changes in the "boundary conditions," such as the sea-surface temperature (SST). They

speculated that if these boundary conditions could be specified without error, then seasonal predictions of quantities such as monthly-mean monsoon precipitation would be possible. Later work has strongly supported this idea. Dry Asian summer monsoons have been statistically correlated to anomalously warm SSTs in the eastern Pacific (El Niño), and wet monsoons have been correlated to anomalously cold SSTs in the eastern Pacific (Philander, 1990). The interannual variability of monsoons has also been linked to variations in Tibetan snow cover (Yanai and Li, 1994; Barnett et al., 1989).

10.10 The Madden-Julian Oscillation

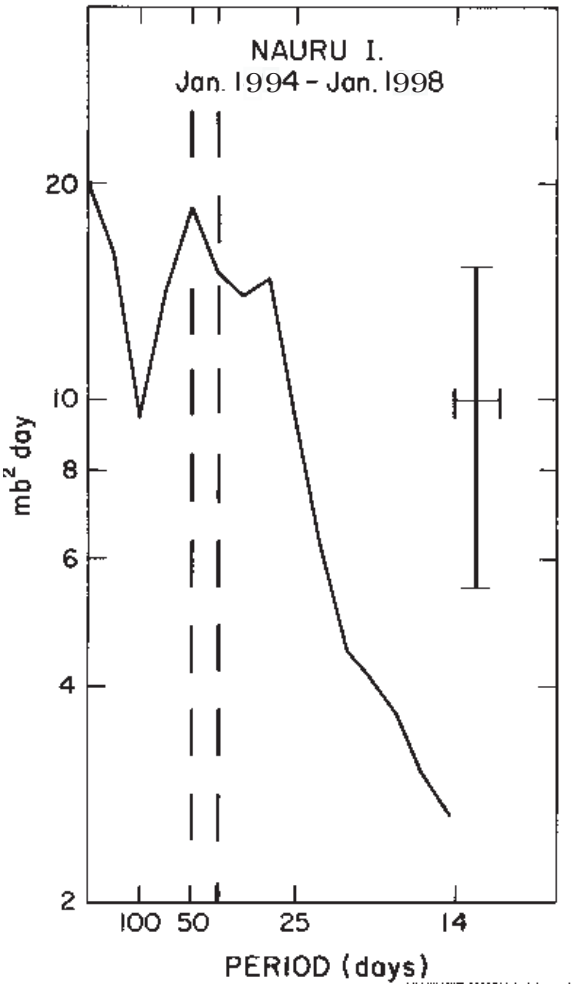


Figure 10.38: Variance spectrum for station pressures at Nauru Island, 0.4° S, 161.0° E. Ordinate (variance/ frequency) is logarithmic, and the abscissa (frequency) is linear. The 40-50 day range is indicated by the dashed vertical lines. Prior 95% confidence limits and the bandwidth of the analysis (0.008 day^{-1}) are indicated by the cross. From Madden and Julian (1994), based on Madden and Julian (1972) © American Meteorological Society. Used with permission.

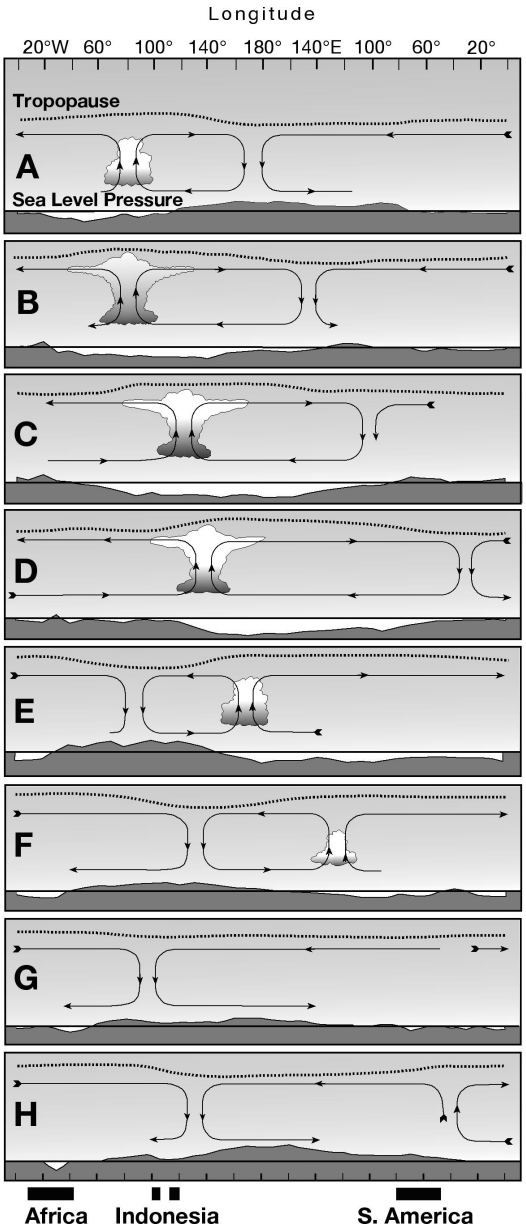


Figure 10.39: Schematic depiction of the time and space (zonal plan) variations of the disturbance associated with the Madden-Julian oscillation. Dates are indicated symbolically by the letters at the left of each chart and correspond to dates associated with the oscillation in Canton's station pressure. The letter A refers to the time of low pressure at Canton, and E refers to the time of high pressure there. The other letters represent intermediate times. The mean pressure disturbance is plotted at the bottom of each chart, with negative anomalies shaded. The circulation cells are based on the mean zonal wind disturbance. Regions of enhanced large-scale convection are indicated schematically by the cumulus and cumulonimbus clouds. The relative tropopause height is indicated at the top of each chart. From Madden and Julian (1994), based on Madden and Julian (1972). © American Meteorological Society. Used with permission.

The MJO (the Madden-Julian Oscillation) can be defined as a broad region of humid air and vigorous precipitation that maintains itself as it drifts slowly eastward across the tropical Indian and Western Pacific Oceans (Madden and Julian, 1971, 1972; Xie et al., 2018; Li et al., 2018). It reoccurs with a period in the range 30 - 60 days (Fig. 10.38). The MJO appears most clearly over the Indian and western Pacific Oceans, and involves many meteorological variables, including the zonal wind, surface pressure, temperature, and humidity. Observations show corresponding oscillations of various measures of cumulus activity (Murakami et al., 1986), including precipitation (Hartmann and Gross, 1988) and outgoing longwave radiation (e.g., Weickmann and Khalsa, 1990). A schematic depiction of the MJO is provided in Fig. 10.39. The observed phase propagation is shown in Fig. 10.40. The MJO does not produce strong oscillations of convection over the Amazon basin or the Congo basin, even though an MJO signal is sometimes observed in the winds at those longitudes.

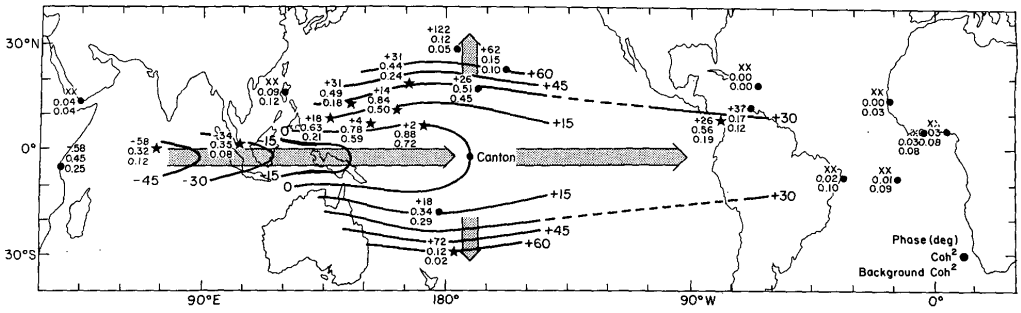


Figure 10.40: Mean phase angles (in degrees), coherence squares, and background coherence squares for approximately the 36-50 day range of cross spectra between surface pressures at all stations and those at Canton. The plotting model is given in the lower right-hand corner. Positive phase angle means Canton time series leads. Stars indicate stations where coherence squares exceed a smooth background at the 95% level. Mean coherence squares at Shemya (52.8°N, 174.1°E) and Campbell Island (52.6°S, 169.2 °E; not shown) are 0.08 and 0.02, respectively. Both are below their average background coherence squares. Values at Dar es Salaam (0.8°S, 39.3°E) are from a cross spectrum with Nauru. The arrows indicate propagation direction. From Madden and Julian (1994), adapted from Madden and Julian (1972). © American Meteorological Society. Used with permission.

The MJO propagates eastward at about 5 m s^{-1} from the Indian Ocean to the dateline, at which point it decouples from the convection and increases its phase speed to about 12 m s^{-1}. Global circumnavigation can sometimes be seen in the upper-tropospheric winds, but is harder to detect in other fields closer to the surface. Embedded within the broad convectively active region are fluctuations on smaller space and time scales (Nakazawa, 1988, Fig. 10.41). In the convectively active phase, strong surface westerlies and high surface latent heat fluxes are observed.

As would be expected from the theoretical work of Matsuno (1966) and Gill (1980), the precipitation maximum of the MJO is accompanied by low-level winds that trace out

twin cyclones on the west side of the precipitation maximum, and a zonally broader patch of easterly winds on the east side. The zonal wind field thus converges at low levels near the precipitation maximum, and diverges aloft.

The MJO has been the subject of intense research because of its intraseasonal time scale, which suggests that intraseasonal weather anomalies may be predictable; and also because of their apparent relationships with the Indian summer monsoon Yasunari (1979); Krishnamurti and Subrahmanyam (1982), the likelihood of tropical Pacific storms (Shaw, 1978), and the initiation of El Niño events (Lau and Chan, 1986). Since tropical convection can force Rossby waves that propagate into the extratropics, the MJO can also influence the weather in middle latitudes (e.g., Rueda, 1991).

Atmospheric global circulation models (GCMs) have difficulty in simulating the MJO. In addition, there is no consensus on the basic physical mechanisms that give rise to the MJO, although progress is being made. The problem of understanding the MJO can be separated into several linked parts. First of all, the steady motion generated by a moving heat source on the Equatorial beta plane is well described by the model of Matsuno (1966) and Gill (1980), hereafter called the "MG model," and has been studied by Salby and Hendon (1994) and Schubert and Masarik (2006), among others. The possible relevance of the MG model to the MJO has been recognized for decades (e.g., Chao, 1987). The low-level wind pattern associated with the MJO closely resembles that shown in Fig. 10.24, for the case of heating centered on the Equator. Given such a moving heat source, a numerical model should be able to simulate a wind field similar to that of the observed MJO.

On the other hand, the theory of Matsuno (1966) fails to produce free modes resembling the MJO, despite its success in predicting the other observed Equatorial waves (with the possible exception of the Easterly waves mentioned earlier (Kiladis et al., 2009). This failure suggests that the MJO depends fundamentally on processes that were not included in Matsuno's model. It is now believed that moist processes are essential to the MJO (e.g., Raymond, 2001; Grabowski and Moncrieff, 2004; Bony and Emanuel, 2005), and the term "moisture mode," which was coined by Fuchs and Raymond (2007), is widely used to describe the MJO (Sugiyama, 2009). The dry MG model cannot describe moisture modes.

Variations of longwave radiative heating and surface evaporation are believed be important for the MJO (e.g., Raymond, 2001; Bony and Emanuel, 2005; Andersen and Kuang, 2012; Arnold et al., 2015). GCMs should be able to simulate them, at least qualitatively, so they are probably not the missing ingredients that prevent many conventional GCMs from simulating the MJO.

The eastward propagation of the MJO is favored by moistening of the atmosphere on its east side, and drying on its west side. Moisture advection is a key process in the MJO (Maloney, 2009; Maloney et al., 2010; Andersen and Kuang, 2012; Pritchard and Bretherton, 2014). Both meridional and vertical advection favor strong drying to the west of the humid, strongly precipitating core of the MJO; under such conditions, positive water vapor anomalies must move eastward or be destroyed. In modeling terms, the drying is due

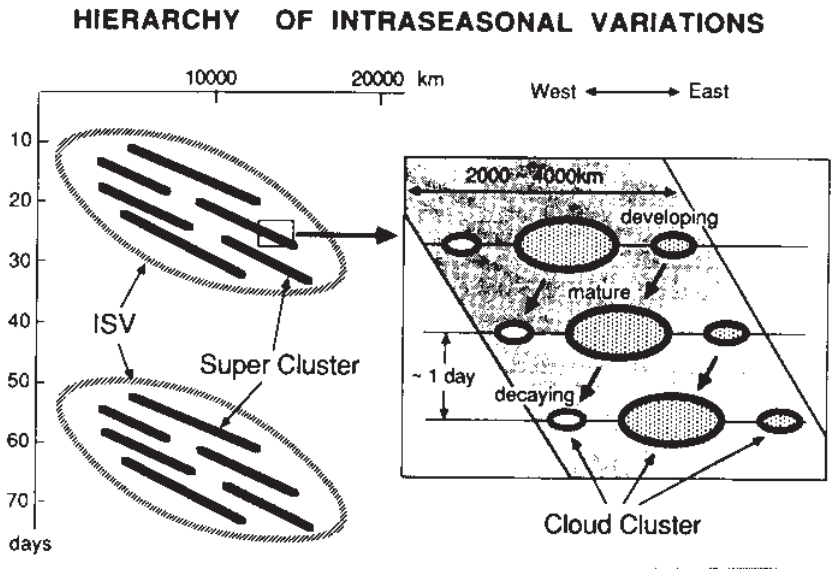


Figure 10.41: Schematic describing the details of the large-scale eastward-propagating cloud complexes (slanting ellipses marked ISV [intraseasonal variability] on the left-hand side). Slanting heavy lines represent super cloud clusters (SCC) within the larger complexes or ISV. The right-hand side illustrates the fine structure of the SCC with smaller westward-moving cloud clusters that develop, grow to maturity, and decay in a few days. From Nakazawa (1988).

to resolved-scale advection, rather than parameterized processes. The precipitation rate is observed to decrease as advection dries the air on the west side of the MJO. Perhaps surprisingly, not all models simulate this (Holloway et al., 2013). In an analysis of numerical experiments with an aquaplanet version of the SP-CAM, Andersen and Kuang (2012) found that the eastward drift of the MJO speeds up if the subtropical reservoir of dry air is brought closer to the Equator.

WISHE (wind-induced surface heat exchange) theories of the MJO focus on the moistening on the east side (Emanuel, 1987a; Neelin and Yu, 1994; Yu and Neelin, 1994). Neelin and Yu (1994) and Emanuel (1987a) hypothesized a feedback in which the maintenance and eastward propagation of convection depend on the existence of a "basic state" easterly flow. The perturbation easterlies on the east side of the MJO reinforce the basic state easterlies, favoring strong evaporation on the east side. Westward zonal advection can carry the evaporatively moistened air towards the convective disturbance (Sobel and Maloney, 2012, 2013). The observed specific humidity fluctuations are strongest near the 700 hPa level, however (Sherwood, 1999; Holloway and Neelin, 2009). The MJO needs a process that lifts the water vapor from the surface and moistens the air near the 700 hPa level. The upward transport of moisture on the east side of the MJO is due to processes that are parameterized in GCMs. Some GCMs that fail to simulate the MJO are unable to moisten the air in regions of strong precipitation (Thayer-Calder and Randall, 2009; Kim et al., 2009; Landu and Maloney, 2011; Mapes and Bacmeister, 2012; Kim et al., 2014; Hung et al.,

2013).

As discussed by Thayer-Calder and Randall (2009), an essential requirement for successful simulation of the MJO is the ability to produce realistically deep layers of high relative humidity in regions of strong precipitation. Chikira (2014) presents an insightful analysis of the processes that lead to mid-tropospheric humidity changes in tropical convective systems. Recall from Chapter 8 that the dry static energy and water vapor are governed by

$$\rho \frac{\partial \overline{s}}{\partial t} = -\rho \overline{\mathbf{v}} \cdot \nabla \overline{s} - \widetilde{M} \frac{\partial \overline{s}}{\partial z} + \overline{Q_R} + \rho L \widetilde{C} + D\left(s_c - \overline{s}\right), \tag{10.147}$$

$$\rho \frac{\partial \overline{q_v}}{\partial t} = -\rho \overline{\mathbf{v}} \cdot \nabla \overline{q_v} - \widetilde{M} \frac{\partial \overline{q_v}}{\partial z} - \rho \widetilde{C} + D\left[\left(q_v\right)_c - \overline{q_v}\right] , \tag{10.148}$$

where $\widetilde{M}$ is the environmental mass flux (positive upward), $\widetilde{C}$ is the net environmental condensation rate, and D is the rate at which mass is detrained from the convective updrafts. As explained in Chapter 3, we invoke the "weak temperature gradient approximation" of Charney (1963) to neglect the time-tendency and horizontal advection terms of (10.147). Then (10.147) reduces to

$$0 \cong -\widetilde{M} \frac{\partial \overline{s}}{\partial z} + \overline{Q_R} + \rho L \tilde{C} + D\left(s_c - \overline{s}\right) , \tag{10.149}$$

which can be rearranged to

$$\widetilde{M} \cong \left[\overline{Q_R} + \rho L \tilde{C} + D\left(s_c - \overline{s}\right)\right] \left(\frac{\partial \overline{s}}{\partial z}\right)^{-1} . \tag{10.150}$$

According to (10.150), the environmental mass flux is whatever it takes to maintain a constant dry static energy. Assuming that $\partial \overline{s} / \partial z > 0$, the condition for the environmental mass flux to be upward is $\overline{Q_R} + \rho L \tilde{C} + D\left(s_c - \overline{s}\right) > 0$. This is relevant because an upward environmental mass flux will moisten the environment.

We can use (10.150) to eliminate $\widetilde{M}$ in (10.148). The resulting equation does not contain an explicit vertical advection term. We obtain

$$\rho \frac{\partial \overline{q_v}}{\partial t} = -\rho \overline{\mathbf{v}} \cdot \nabla \overline{q_v} + \frac{\alpha}{L} \left[\overline{Q_R} + \rho L \tilde{C} + D \left(s_c - \overline{s} \right) \right] - \rho \widetilde{C} + D \left[\left(q_v \right)_c - \overline{q_v} \right], \tag{10.151}$$

where

$$\alpha \equiv -L \left(\frac{\partial \overline{q_v}}{\partial z} \right) \left(\frac{\partial \overline{s}}{\partial z} \right)^{-1}, \tag{10.152}$$

so that

$$1 - \alpha = \left(\frac{\partial \overline{h}}{\partial z} \right) \left(\frac{\partial \overline{s}}{\partial z} \right)^{-1}. \tag{10.153}$$

From (10.152), we expect $\alpha > 0$. From (10.153), we see that that throughout most of the tropics α should be greater than one in the lower troposphere, and close to zero in the upper troposphere. If the environment becomes saturated with a moist adiabatic lapse rate, then $\alpha = 1$.

By combining terms, we can rewrite (10.151) as

$$\rho \frac{\partial \overline{q_v}}{\partial t} = -\rho \overline{\mathbf{v}} \cdot \nabla \overline{q_v} + \frac{\alpha}{L} overline Q_R - (1 - \alpha) \rho \widetilde{C} + D \left\{ \left[\left(q_v \right)_c - \overline{q_v} \right] + \frac{\alpha}{L} \left(s_c - \overline{s} \right) \right\}. \tag{10.154}$$

If detrainment occurs at levels where $s_c \cong \overline{s}$, then (10.154) can be approximated by

$$\boxed{\rho \frac{\partial \overline{q_v}}{\partial t} \cong -\rho \overline{\mathbf{v}} \cdot \nabla \overline{q_v} + \frac{\alpha}{L} \overline{Q_R} - (1 - \alpha) \rho \widetilde{C} + D \left[\left(q_v \right)_c - \overline{q_v} \right].} \tag{10.155}$$

Inspection of (10.155) shows that:

1. Radiative cooling dries, and radiative heating moistens. If water vapor and radiative heating are positively correlated, then radiation may be able to drive an instability, as suggested by Raymond (2001).

2. For $\alpha > 1$, which is expected in the lower troposphere, evaporation into the environment (i.e., $\widetilde{C} < 0$) dries the air. This counterintuitive result means that the evaporation of detrained cloud water and stratiform precipitation in the lower troposphere can actually have a drying effect. A process that reduces α to values less than one can change this.

3. Condensation in the environment (i.e., $\widetilde{C} > 0$) leads to a net moistening when $\alpha > 1$. However, if the environment is saturated with a moist adiabatic lapse rate, then $\alpha \cong 1$ is expected, in which case environmental condensation has no effect on the water vapor.

4. For $0 < \alpha < 1$, the evaporation into the environment of detrained cloud water and falling rain can drive an increase in moisture variance, consistent with the "stratiform instability" of Mapes (2000) and Kuang (2008). It appears that the best chance for instability comes with $0 < \alpha < 1$.

10.11 Tropical Easterly Waves

Tropical Easterly Waves (TEWs) are westward-moving disturbances with synoptic-scale horizontal structures typically ranging from 1000 to 4000 kilometers and periods of 3 to 6 days (Takayabu and Nitta, 1993; Reed and Recker, 1971). They often form in the easterly flow along the equatorial side of a subtropical high. They are associated with convective disturbances. TEWs are of great interest because they sometimes transform into tropical depressions (TDs) and tropical cyclones (TCs), especially in the Atlantic and northeastern Pacific basins. Matsuno (1966) does not explicitly describe TEWs, but his work provides a foundation for understanding them. MRG waves ($n = 0$) are particularly relevant because, like TEWs, they propagate westward and exhibit an antisymmetric pressure pattern relative to the equator, with alternating cyclonic and anticyclonic circulations.

Takayabu and Nitta (1993) highlighted a notable connection between equatorial MRG waves and off-equatorial TD disturbances. They proposed that TD disturbances may emerge due to the shortening of the wavelengths of MRGs, influenced by variations in the background zonal wind. Sobel and Bretherton (1999) described the transition from MRGs to TD-type disturbances as a complex process that involves both a reduction in MRG wave wavelength and the breakdown of its structure. Aiyyer and Molinari (2003) later argued that MRGs can evolve into TEWs when they interact with the the Madden-Julian Oscillation, which is described in the next section.

10.12 Tropical cyclones

Emanuel et al. (2005) gives an excellent introduction to tropical cyclones. TCs are characterized by cyclonic low-level circulations surrounding a warm core, with an anticyclonic circulation aloft. Fig. 10.42 shows where TCs form, and how they move. As is well known,

TCs form over warm water, often within the ITCZ (Schubert and Masarik, 2006). They are "heat engines" that extract thermodynamic energy from the warm ocean and transport it upward, fueling their own circulations in the process (Emanuel, 1987b). TCs are home to very deep and vigorous cumulus convection, which is concentrated in the eye wall and, to a lesser extent, in the spiral bands. Fig. 10.42 shows that TC formation is suppressed near the equator, because the Coriolis parameter is too small there, and perhaps also because the SST is usually cooler along the equator than in slightly higher latitudes. TCs are also inhibited by vertical wind shear, which prevents the formation of the deep convection needed for cyclogenesis.

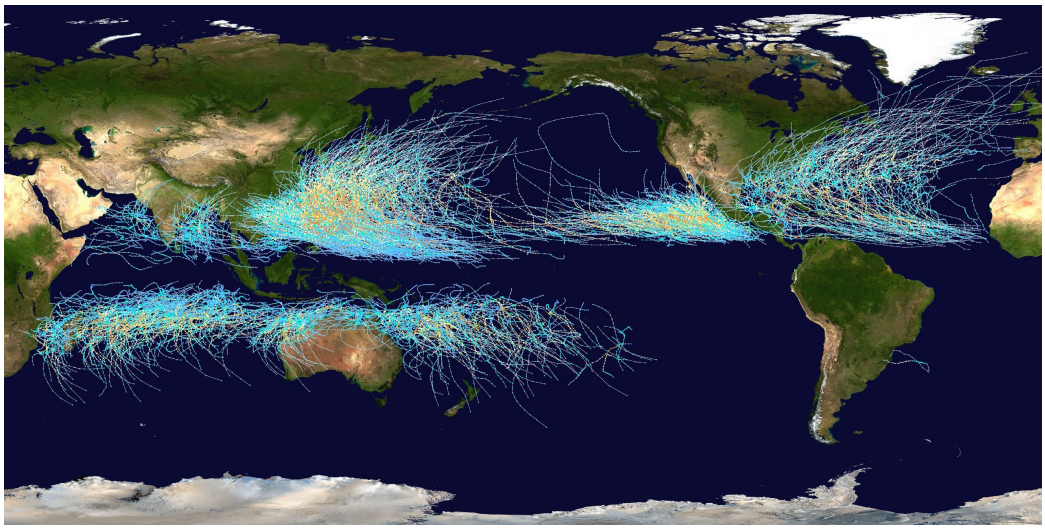


Figure 10.42: Where on Earth do cyclones go? Usually known as hurricanes when in the Atlantic Ocean and typhoons when in the Pacific, the featured map shows the path of all major storms from 1985 through 2005. The map shows graphically that cyclones usually occur over water, which makes sense since evaporating warm water gives them energy. The map also shows that cyclones never cross – and rarely approach – the Earth's equator, since the Coriolis effect goes to zero there, and cyclones need the Coriolis force to circulate. The Coriolis force also causes cyclone paths to arc away from the equator. Although long-term trends remain a topic of research, evidence indicates that hurricanes have become, on the average, more powerful in the North Atlantic over the past 30 years, and their power is projected to keep increasing. From `https://apod.nasa.gov/apod/image/2109/StormPaths_NHC_8000.jpg`.

Unlike baroclinic eddies, TCs do not have an intrinsic life cycle; they can persist as long as they move over warm water (Khairoutdinov and Emanuel, 2013; Chavas and Reed, 2019). Recent work has shown that TC-like circulations can even occur over heated land (Cronin and Chavas, 2019).

Fig. 10.42 also shows that TCs move westward and poleward. Their movement is influenced by trade winds, and also by "β-drift," which arises from the interaction between the cyclonic circulation of a storm and the variation of Earth's Coriolis force with latitude. A storm advects stronger absolute vorticity equatorward on its west side, and weaker absolute vorticity poleward on its east side. This gives rise to a cyclonic gyre equatorward and

westward of the storm center, and an anticyclonic gyre poleward and eastward of the storm center (9Fig. 10.43). The combined action of these two gyres causes the storm to drift poleward and westward relative to the mean flow, with a speed of about 1–3 m s^{-1}. Larger and more intense tropical cyclones experience stronger β drift. The poleward component of the β drift decreases with increasing latitude, while the westward component changes little with latitude. If a storm enters the mid-latitude westerlies, it will be carried toward the east and undergo a transformation to an extratropical storm.

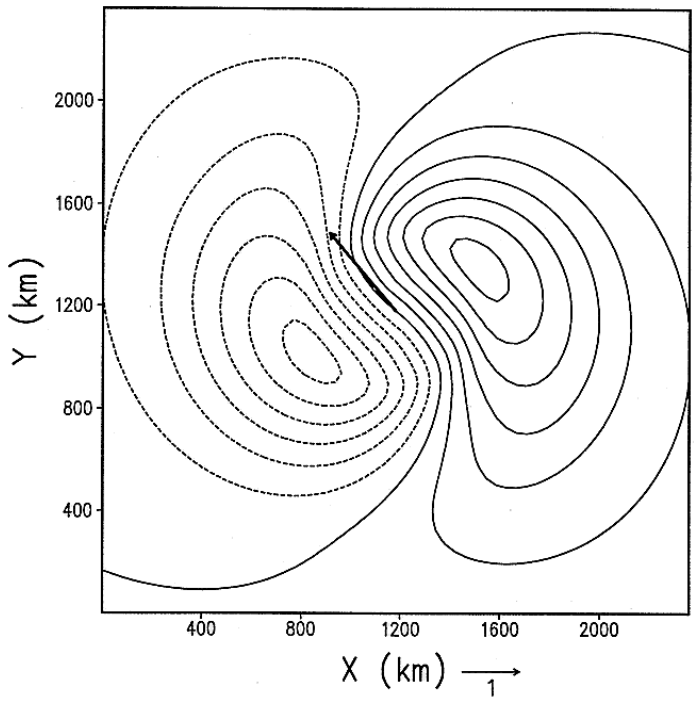


Figure 10.43: Diagram illustrating the streamfunction pattern associated with the "beta gyres" that result from the advection of planetary vorticity by a tropical cyclone. The TC the gives rise to these beta gures would be in the center of the figure. The arrow shows the direction (i.e., northwest) towards which the beta gyres will advect the TC. From Wang et al. (1997).

On average, approximately 80 to 90 named TCs occur globally each year, with over half intensifying to hurricane-force winds (119 km h^{-1}). TC activity peaks globally during late northern summer. September is the most active month worldwide, while May is the least active. For the North Atlantic, the season runs from June 1 to November 30, peaking around September 10. For the Eastern Pacific, activity spans May 15 to November 30, similar to the Atlantic but slightly broader. For the Western Pacific, cyclones occur year-round, with a minimum in February-March and a peak in early September. For the North Indian Ocean, activity is concentrated from April to December, with peaks in May and November. For the Southern Hemisphere, tropical cyclones can occur at any time of year, but activity peaks between mid-February and early March.

Fig 10.42 shows that there are fewer tropical cyclones in the Southern Hemisphere

compared to the Northern Hemisphere. The South Atlantic lacks precursor disturbances, such as TEWs, which serve as triggers for TC development in other basins. TCs are extremely rare in the South Atlantic basin due to several unfavorable factors. Potential formation zones are closer to the equator in the South Atlantic, where the Coriolis parameter is weaker, making it harder for storms to organize. The South Atlantic also experiences frequent strong vertical wind shear, which disrupts the development and organization of TCs. SSTs are generally cooler in the South Atlantic, relative to the North Atlantic due to the influence of the cold Malvinas Current, which blocks warm water from moving poleward. Warm ocean waters are crucial for cyclone development and maintenance. Throughout the Southern Hemisphere, cool ocean currents and proximity to Antarctica mean that warm waters do not extend as far south as warm waters extend north in the Northern Hemisphere, restricting how far south cyclones can travel before losing strength. This is apparent in Fig. 10.42.

As discussed in Chapter 4, vorticity tends to move with the air. There is a tendency for tropical cyclones to carry air with them as they move. Through this mechanism, they transfer moisture transfer from the ocean to land and from tropical regions to subtropical and extratropical areas. They can generate very large precipitation totals over land.

TCs also transport energy vertically through several mechanisms, including convective processes, gravity waves, and ocean-atmosphere interactions. They pump heat into the ocean's interior, which can be carried poleward by ocean currents Emanuel (2001); Li and Sriver (2018). The deep convective towers in their eyewalls and rainbands transport energy to the upper atmosphere. TCs also excite vigorous gravity waves that transport energy and momentum vertically.

10.13 Segue

This chapter describes both observations and theories of some of the many eddies motions that fill the atmosphere. We began with Laplace's theory of the free and forced oscillations of a thin spherical atmosphere. There are many solutions, notably including inertia-gravity waves and Rossby waves. The vertical propagation of Rossby waves was discussed, using the QG framework. We then turned to Matsuno's theory of equatorially trapped waves, which was based on the shallow water equations but can be extended to a stratified atmosphere. Laplace's equations also describe the waves that Matsuno found, but with differences due to sphericity.

Most of the eddies that we have discussed are "free modes," in the sense that they are not responses to external forcing. The topographically forced Rossby waves and the monsoon are major exceptions.

The effects of the eddies on the zonally averaged flow are the subject of the next Chapter.

10.14 Problems

1. Show that the static stability increases strongly upward in an isothermal atmosphere.

2. Prove that an isothermal atmosphere with temperature T has only one equivalent depth for free oscillations, given by

$$\widehat{h}_{\rm fs} = (c_p/c_v)\,H, \tag{10.156}$$

   where $c_p/c_v \cong 1.4$, and $H \equiv RT/g$.

3. Show that for a resting isentropic basic state with no heating and no gravitational forcing, perturbations satisfy

$$\frac{\partial z'}{\partial t} + H_0 \nabla_p^2 \chi = 0, \tag{10.157}$$

   for all p, where $H_0 \equiv RT_0/g$ and subscript zero denotes a surface value. Eq. (10.157) looks very much like the continuity equation for shallow water.

4. Show that the FOSC with the nondivergent approximation must satisfy

$$\nabla_p \cdot (f \nabla_p \psi) = g \nabla_p^2 z'. \tag{10.158}$$

   This implies that the oscillations are not in exact geostrophic balance.

5. Show that when there is a basic zonal current from the west, with a constant angular velocity $\dot{\lambda}$, the apparent "phase speed" relative to the Earth's surface of the free oscillation of the second class is given by

$$-\frac{\sigma}{m} = \dot{\lambda} - \frac{2\left(\Omega + \dot{\lambda}\right)}{n(n+1)}\,. \tag{10.159}$$

6. Solberg (1936) showed that the Laplace Tidal Equations have a special solution for which the period of oscillation is 1/2 day and the zonal wave number is zero.

(a) Show that the solution is given by

$$\Theta_n = A\sin\left(\sqrt{\varepsilon_n}\mu\right) + B\cos\left(\sqrt{\varepsilon_n}\mu\right) \ , \tag{10.160}$$

where ε_n is given by (10.25).

(b) Show that the two conditions

$$\varepsilon_n = (n\pi/2)^2 , \quad \text{and} \quad A/B = \tan(n\pi/2) \quad \text{for} \quad n = 1,2,\ldots \tag{10.161}$$

are required to ensure that the winds remain finite at the poles.

7. Prove by substitution that the associated Legendre functions are solutions of (10.43).

8. Matsuno derived the "meridional structure equation" that governs equatorially trapped waves (except for Kelvin waves):

$$\frac{d^2\widehat{v}}{dy^2} + \left(\sigma^2 - k^2 + \frac{k}{\sigma} - y^2\right)\widehat{v} = 0 \, . \tag{10.162}$$

The solution is

$$\widehat{v}(y) = Ce^{-\frac{1}{2}y^2} H_n(y) \ . \tag{10.163}$$

Prove by substitution that the dispersion equation is

$$\sigma^2 - k^2 + \frac{k}{\sigma} = 2n+1 \text{ for } n = 0,\ 1,\ 2\ldots \ . \tag{10.164}$$

9. Consider linear quasi-geostrophic frictionally damped Rossby waves produced by shallow-water flow over periodic mountains, as in the model of Charney and Eliassen. For simplicity, set the meridional wave number l to zero. Assume that the shallow water has a uniform density ρ_0. Find an expression for the "form drag" that the mountains exert on the mean flow, for the case of stationary resonant waves (i.e., $K^2 = K_S^2$).

10. As discussed in the text, Matsuno (1966) discovered equatorial Kelvin waves, which propagate eastward and have no meridional velocity component. It is also possible to have oceanic "coastal" Kelvin waves, such that the currents are everywhere parallel to the coast line. A coastal Kelvin wave propagates parallel to the coast.

    Consider a coastal Kelvin wave in the Northern Hemisphere. The wave is propagating along a north-south oriented "west coast" with water to the west and a continent to the east. Derive the relevant wave equation, and adopt the convention that the meridional wave number l is non-negative. Does the wave propagate north or south? Explain your answer. Include a sketch that shows the spatial structures of the height and current fields.

Chapter 11

Interactions and non-interactions of waves with the mean flow

11.1 Interactions and non-interactions of gravity waves with the mean flow

We have seen how eddies can affect the mean flow, through eddy flux divergences and energy conversions. Despite the presence of such terms in the equations, however, it turns out that under surprisingly general conditions the eddies have no net effect on the mean flow. Several related theorems demonstrate this. They are called, reasonably enough, "*non-interaction theorems*." The earliest such ideas were published by Eliassen and Palm (1961), and the following discussion of this section is based on that paper. The same material is also discussed in more detail, and in somewhat more general form, in Chapter 8 of the book by Lindzen (1990).

Consider the equation of zonal motion in the simplified form

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + w\frac{\partial u}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial x} \,. \tag{11.1}$$

Here we have omitted rotation, sphericity, friction, and meridional motions. We will apply Eq. (11.1) to small-scale gravity waves that are forced by a mean flow over topography. We define a zonally uniform but vertically varying basic state, and eddies, by

$$\begin{aligned} u =& \overline{u}^{\lambda}(z) + u^* \,, \\ w =& w^* \,, \\ p =& \overline{p}^{\lambda}(z) + p^* \,, \\ \rho =& \overline{\rho}^{\lambda}(z) + \rho^* \,. \end{aligned} \tag{11.2}$$

Here we are using our notation for zonal means and departures from zonal means, although in reality we are only considering an isolated mountain or mountain range. The starred quantities (the eddies) are small-amplitude wave-like perturbations with zero means. The mean flow is expected to respond to the momentum flux according to

$$\overline{\rho}^{\lambda} \frac{\partial \overline{u}^{\lambda}}{\partial t} \sim -\frac{\partial}{\partial z}\left(\overline{\rho}^{\lambda} \overline{w^* u^*}^{\lambda}\right) \,. \tag{11.3}$$

Our goal is to understand what determines the wave momentum flux divergence, $\partial\left(\overline{\rho}^{\lambda}\overline{w^*u^*}^{\lambda}\right)/\partial z$.

Substituting (11.2) into (11.1) and linearizing, we obtain

$$\overline{\rho}^{\lambda} \frac{\partial u*}{\partial t} = -\left(\overline{\rho}^{\lambda}\overline{u}^{\lambda}\frac{\partial u*}{\partial x} + \overline{\rho}^{\lambda} w^* \frac{\partial \overline{u}^{\lambda}}{\partial z} + \frac{\partial p*}{\partial x}\right) \,. \tag{11.4}$$

We assume that the perturbations are steady, so that

$$\frac{\partial u^*}{\partial t} = 0 \,. \tag{11.5}$$

This implies both that the waves are neutral, i.e., neither amplifying or decaying, and also that they are stationary, i.e., their phase speed is zero. The latter assumption is reasonable, e.g., for mountain waves. Then (11.4) reduces to

$$\begin{aligned} 0 =& \overline{\rho}^{\lambda}\overline{u}^{\lambda}\frac{\partial u^*}{\partial x} + \overline{\rho}^{\lambda} w^* \frac{\partial \overline{u}^{\lambda}}{\partial z} + \frac{\partial p^*}{\partial x} \\ =& \frac{\partial}{\partial x}\left(\overline{\rho}^{\lambda}\overline{u}^{\lambda}u^* + p^*\right) + \overline{\rho}^{\lambda} w^* \frac{\partial \overline{u}^{\lambda}}{\partial z} \,. \end{aligned} \tag{11.6}$$

We will apply this form of the steady-state equation of motion to the eddies.

Next, multiply (11.6) by $\left(\overline{\rho}^{\lambda}\overline{u}^{\lambda}u^{*}+p^{*}\right)$, to obtain

$$0=\frac{1}{2}\frac{\partial}{\partial x}\left(\overline{\rho}^{\lambda}\overline{u}^{\lambda}u^{*}+p^{*}\right)^{2}+\left(\overline{\rho}^{\lambda}\right)^{2}\overline{u}^{\lambda}\frac{\partial\overline{u}^{\lambda}}{\partial z}w^{*}u^{*}+\overline{\rho}^{\lambda}\frac{\partial\overline{u}^{\lambda}}{\partial z}w^{*}p^{*}\,. \tag{11.7}$$

Averaging (11.7) with respect to x, from far upstream of the mountain to far downstream of the mountain, gives

$$\frac{\partial\overline{u}^{\lambda}}{\partial z}\left(\overline{\rho}^{\lambda}\overline{u}^{\lambda}\overline{w^{*}u^{*}}^{\lambda}+\overline{w^{*}p^{*}}^{\lambda}\right)=0\,, \tag{11.8}$$

which can be simplified to

$$\boxed{\overline{\rho}^{\lambda}\overline{u}^{\lambda}\overline{w^{*}u^{*}}^{\lambda}+\overline{w^{*}p^{*}}^{\lambda}=0\,.} \tag{11.9}$$

provided that $\partial\overline{u}^{\lambda}/\partial z\neq 0$. Eq. (11.9) shows that the wave momentum flux, $\overline{\rho}^{\lambda}\overline{w^{*}u^{*}}^{\lambda}$, and the wave energy flux, $\overline{w^{*}p^{*}}^{\lambda}$, are closely related. At a "critical" level, where $\overline{u}^{\lambda}=0$, the wave energy flux must go to zero; the only other possibility is that our assumptions, e.g., a steady state with no friction, do not apply at the critical level. For a wave forced by flow over a mountain, the energy flux is, of course, upward, but (11.9) shows that the upward energy flux is blocked at the critical level. *This means that the wave does not exist above the critical level.* A more detailed analysis (Booker and Bretherton, 1967; Bretherton, 1969) shows that the degree of wave blocking at a critical level actually depends on the Richardson number, which measures the relative importance of buoyancy and shear.

Eq. (11.9) also shows that a wave with an upward energy flux will produce a downward momentum flux in westerlies and an upward momentum flux in easterlies. In either case, the wave exerts a drag on the mean flow aloft.

Let e_E be the total eddy energy per unit mass associated with the wave (i.e., the sum of the eddy kinetic, eddy internal, and eddy potential energies). It can be shown that e_E satisfies

$$\frac{\partial}{\partial x}\left(\overline{\rho}^{\lambda} e_E \overline{u}^{\lambda} + p^* u^*\right) + \frac{\partial}{\partial z}\left(w^* p^*\right) = -\overline{\rho}^{\lambda} u^* w^* \frac{\partial \overline{u}^{\lambda}}{\partial z} \,. \tag{11.10}$$

Averaging again, from far upstream to far downstream, gives

$$\boxed{\frac{\partial}{\partial z}\left(\overline{w^* p^*}^{\lambda}\right) = -\overline{\rho}^{\lambda} \overline{w^* u^*}^{\lambda} \frac{\partial \overline{u}^{\lambda}}{\partial z} \,.} \tag{11.11}$$

The right-hand side of (11.10) is a "gradient production" term, of the type first discussed in Chapter 7. It represents conversion between the kinetic energy of the mean state and the total eddy energy, e_E. Eq. (11.11) simply says that the production term on the right-hand side is balanced by the transport term on the left-hand side. An energy source is balanced by an energy flux divergence, and an energy sink is balanced by an energy flux convergence.

By combining (11.9) and (11.11) we can show that

$$\boxed{\overline{u}^{\lambda} \frac{\partial}{\partial z}\left(\overline{\rho}^{\lambda} \overline{w^* u^*}^{\lambda}\right) = 0 \,.} \tag{11.12}$$

Eq. (11.12) tells us that *when $\overline{u}^{\lambda} \neq 0$, the wave momentum flux $\overline{\rho}^{\lambda} \overline{w^* u^*}^{\lambda}$ is independent of height*. From (11.3), this implies that *the wave momentum flux has no effect on $\overline{u}$, except at the critical level where $\overline{u} = 0$*. Eq. (11.12) is a non-inteaction theorem for vertically propagating gravity waves. The wave momentum flux is absorbed at the critical level. From (11.3), it follows that $\overline{u}^{\lambda}$ will tend to change with time at the critical level, which means that it will become different from zero. The critical level will move.

If we allowed the phase speed, c, to be non-zero, we would find $\overline{u}^{\lambda} - c$ everywhere in place of $\overline{u}^{\lambda}$. The momentum would be absorbed at the level where $\overline{u}^{\lambda} - c = 0$.

Since (11.12) tells us that $\overline{\rho}^{\lambda} \overline{w^* u^*}^{\lambda}$ is independent of height (where $\overline{u}^{\lambda} \neq 0$), we see from (11.9) that the wave energy flux is just proportional to $\overline{u}^{\lambda}$. Alternatively, we can combine (11.9) and (11.12) to write

$$\overline{w^* p^{*\lambda}} / \overline{u}^{\lambda} = \text{constant} \,. \tag{11.13}$$

The quantity $\overline{w^* p^{*\lambda}} / \overline{u}^{\lambda}$ is called the "wave action." Eq. (11.9) can be read as "wave action + wave momentum flux = zero."

The work of Eliassen and Palm, reviewed above, was published in 1961. The importance of their ideas for the global circulation began to be appreciated about 25 years later. Since the mid-1980s, there has been a lot of interest in the effects of gravity wave momentum fluxes on the general circulation; because the waves act to decelerate the mean flow, these interactions are referred to as "gravity-wave drag" (Palmer et al., 1986; McFarlane, 1987). At first, most of the discussion was about gravity waves forced by flow over topography, but later gravity waves forced by convective storms and fronts were also recognized as important (e.g., Fovell et al., 1992; Beres et al., 2005; Richter et al., 2010).

Fig. 11.1a shows the deceleration of the zonally averaged zonal wind induced by gravity-wave drag in a general circulation model, as reported by McFarlane (1987). Here the gravity wave drag has been parameterized based on the assumption that the waves are produced by flow over mountains. The plot shows the "tendency" of the zonally averaged zonal wind due to this orographic gravity-wave drag, for northern-winter conditions. The very strong response of the zonally averaged zonal wind is shown in Fig. 11.1b. In order for thermal wind balance to be maintained, there have to be corresponding changes in the zonally averaged temperature; these are shown in Fig. 11.1c. The polar troposphere has warmed dramatically, so as to be consistent with the weaker westerly jet. The responses to gravity-wave drag shown in Figs. 11.1a and b make the model results more realistic than before, suggesting that gravity-wave drag is an important process in nature.

11.2 Rossby waves

11.2.1 The effects of a mean zonal flow

In Chapter 10, we derived the dispersion equation for Rossby waves for the case of no mean flow. Now consider what happens when a mean zonal flow is present.

The westward propagation of Rossby-Haurwitz waves is due to the Earth's sphericity, which is associated with variations of the Coriolis parameter with latitude. To see this, consider the vorticity equation in the simplified form

$$\frac{\partial \zeta}{\partial t} = -\beta v \tag{11.14}$$

where β was defined in (10.51). This is a statement of absolute vorticity conservation. Consider a chain of vortices along a latitude circle, as shown in Fig. 11.2. Where $v > 0$, to the west of the place where $\zeta < 0$, we have $\beta v > 0$, so we get $\partial \zeta / \partial t < 0$. Similarly, where $v < 0$, to the west of the place where $\zeta > 0$, we have $\beta v < 0$, so we get $\partial \zeta / \partial t > 0$. The implication is that Rossby waves propagate westward relative to the mean flow. All other things being equal, the waves propagate westward more quickly near the Equator, where β

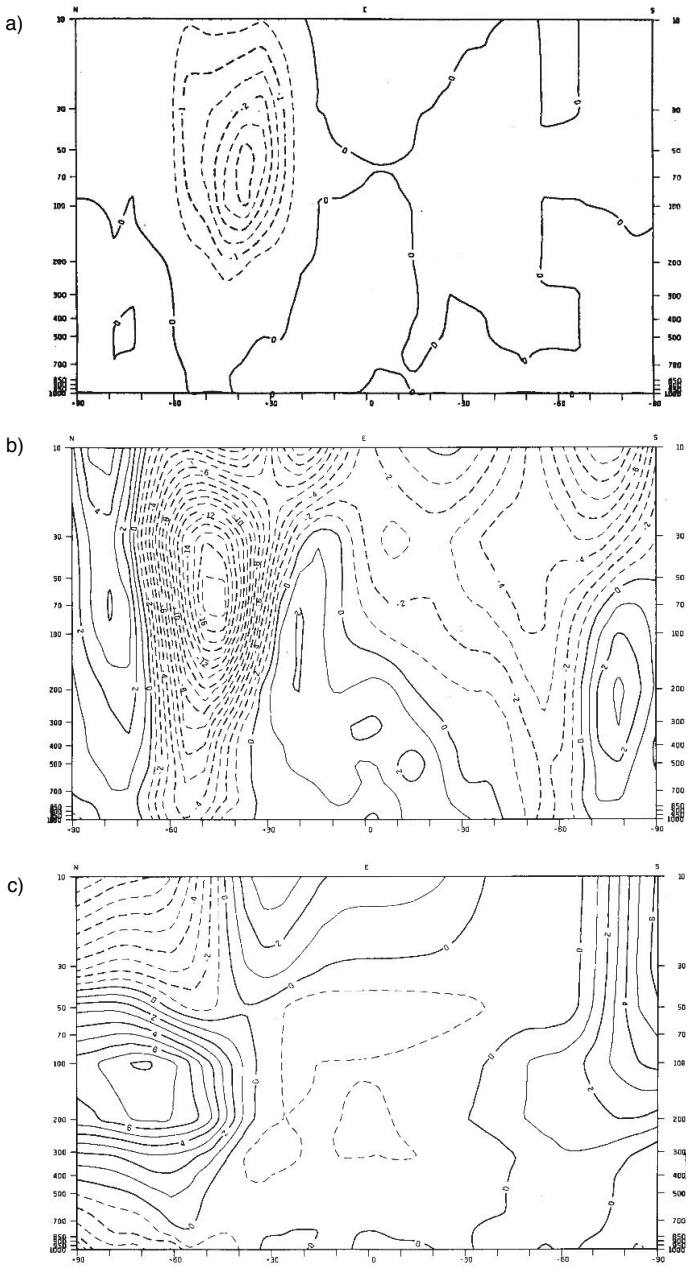


Figure 11.1: a) The deceleration of the zonally averaged zonal flow, induced by orographically forced gravity waves, as simulated with a general circulation model. The units are m s^{-1} day^{-1}. b) The actual change in the zonally averaged wind caused by the introduction of gravity wave drag in a general circulation model, as inferred by comparison with a control run. The units are m s^{-1}. c) The actual change in the zonally averaged temperature caused by the introduction of gravity wave drag in a general circulation model, as inferred by comparison with a control run. The units are K. From McFarlane (1987).

is large, than at higher latitudes. Rossby wave propagation on the sphere is discussed by

Grose and Hoskins (1979).

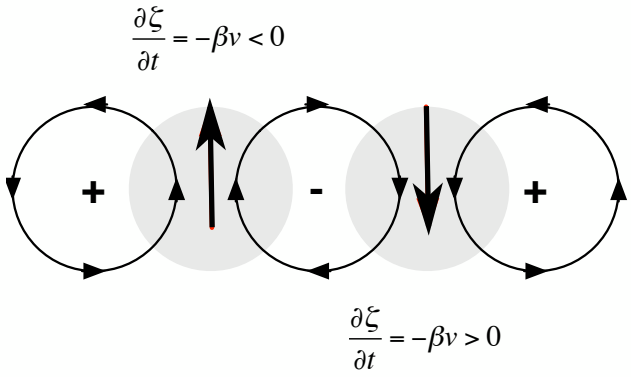


Figure 11.2: A chain of vortices along a latitude circle, illustrating the westward propagation of Rossby waves. See the text for an explanation.

For simplicity, we consider purely rotational horizontal motions in Cartesian coordinates.

The linearized vorticity equation with a constant zonal current is

$$\frac{\partial \zeta^*}{\partial t} + \overline{u}^{\lambda} \frac{\partial \zeta^*}{\partial x} + \beta v^* = 0 \, . \tag{11.15}$$

For purely rotational flow, we can represent the solution in terms of a stream function, ψ:

$$v^* = \frac{\partial \psi^*}{\partial x}, \quad u^* = -\frac{\partial \psi^*}{\partial y}, \quad \zeta^* = \nabla^2 \psi^* \, . \tag{11.16}$$

We look for solutions of the form

$$\psi^* = a \cos\left(kx + ly - \sigma t\right) \, , \tag{11.17}$$

where A is an arbitrary constant "amplitude" with dimensions of $\mathrm{m}^2\ \mathrm{s}^{-1}$, and the zonal wave number k, meridional wave number l, and frequency σ are constant in both space and time. Here we treat the frequency as non-negative, so that the signs of k and l determine the directions of zonal and meridional phase propagation, respectively. Substitution of (11.16) and (11.17) into (11.15) gives

$$-\left(-\sigma+\overline{u}^{\lambda}k\right)\left(k^2+l^2\right)+\beta k=0\,, \tag{11.18}$$

or

$$\sigma-\overline{u}^{\lambda}k=\frac{-\beta k}{k^2+l^2}\,. \tag{11.19}$$

The zonal phase speed, c, satisfies

$$\sigma=kc\,, \tag{11.20}$$

so (11.19) can also be written as

$$\boxed{\overline{u}^{\lambda}-c=\frac{\beta}{k^2+l^2}>0\,.} \tag{11.21}$$

For $\overline{u}^{\lambda}=0$, Eq. (11.21) tells us that $c<0$, which means westward phase propagation, as discussed in Chapter 10. More generally, the phase propagation must be westward relative to the mean zonal flow. Rossby wave phase lines can only travel towards the east when they are advected by a westerly mean flow that is faster than the (westward) phase speed of the wave. A "stationary" Rossby wave, with $c<0$, cannot exist in an easterly flow.

The special case $c-\overline{u}^{\lambda}=0$ occurs at critical latitudes. Eq. (11.21) suggests that the total wave number becomes very large near a critical latitude, but this is only a suggestion (rather than a true deduction) because, as mentioned above, we have assumed in writing (11.17) that the zonal and meridional wave numbers are spatially (and temporally) constant.

Eq. (11.21) also tells us that long Rossby waves (with small total wave numbers) propagate faster than short Rossby waves.

11.2.2 Angular momentum transports by Rossby waves

An expression for the meridional flux of zonal momentum by the waves can be derived by using (11.17) in (11.16) to write

$$v^*=-Ak\sin\left(kx+ly-\sigma t\right)\text{ and }u^*=Al\sin\left(kx+ly-\sigma t\right)\,, \tag{11.22}$$

so that

$$u^* v^* = -A^2 kl \sin^2 (kx + ly - \sigma t) \ . \tag{11.23}$$

Averaging over longitude (which is equivalent to averaging over one zonal wavelength), we find that

$$\begin{aligned} \overline{u^* v^*}^\lambda &= -\frac{A^2 l}{2\pi} \int_0^{2\pi} \sin^2 (kx + ly - \sigma t) \, d(kx) \\ &= -\frac{A^2 l}{2\pi} \left[\frac{kx}{2} - \frac{\sin(2kx)}{4} \right]_0^{2\pi} \\ &= -\frac{A^2 kl}{2} \ . \end{aligned} \tag{11.24}$$

The product kl determines the "tilt" of the wave. For $kl > 0$, the wave is said to be positively tilted, and for $kl < 0$ it is negatively tilted. According to (11.24), a positively tilted wave transports zonal momentum towards the south, and a negatively tilted wave transports zonal momentum towards the north.

11.2.3 Energy transports by Rossby waves

The zonal component of the group velocity is given by

$$\begin{aligned} c_{gx} &\equiv \frac{\partial \sigma}{\partial k} \\ &= \overline{u}^\lambda + \beta \left[\frac{k^2 - l^2}{(k^2 + l^2)^2} \right] \ . \end{aligned} \tag{11.25}$$

The sign of the zonal group velocity depends on the ratio of k and l. For zonally long waves ($k/l < 1$), the group velocity is typically westward. For zonally short waves ($k/l > 1$), the group velocity is typically eastward. Stationary and most synoptic-scale Rossby waves tend to have eastward zonal group velocities.

The meridional component of the group velocity is

$$\begin{aligned} c_{gy} \equiv & \frac{\partial \sigma}{\partial l} \\ = & \frac{2\beta kl}{\left(k^2+l^2\right)^2} \, . \end{aligned} \tag{11.26}$$

The wave energy travels meridionally with the speed c_{gy}. Comparing (11.24) and (11.26), we see that both involve the product kl. We can write

$$\boxed{\frac{\overline{u^* v^*}^\lambda}{c_{gy}} = -\frac{A^2\left(k^2+l^2\right)^2}{4\beta} < 0 \, .} \tag{11.27}$$

This shows that the sign of the meridional flux of zonal momentum is always opposite to the sign of the meridional flux of wave energy. If wave energy is going equatorward, then the momentum flux is poleward, and vice versa.

It follows that a latitude belt that is a source of Rossby-wave energy will tend to acquire westerly momentum. The wave momentum flux is into the region of westerlies, so momentum is transported "up the gradient." Recall from Chapter 7 that when the momentum transport is up-gradient, the gradient production term of the eddy kinetic energy equation converts the kinetic energy of the eddies into the kinetic energy of the mean flow. As discussed already in Chapter 7, this is observed, and it can only happen if the eddies have a source of kinetic energy other than the kinetic energy of the mean flow. That source can be baroclinic conversion from eddy available potential energy, as discussed in Chapter 10. Baroclinic eddies in the midlatitude storm tracks excite Rossby waves. The Rossby wave energy propagates away from its source in the storm tracks (primarily towards the Equator), and westerly momentum is transported into the storm tracks, where the westerlies are already strong. Hence the expression "eddy-driven jet."

Comparison of (11.21) and (11.26) shows that

$$c_{gy} = \frac{2kl\left(\overline{u}^\lambda - c\right)}{k^2+l^2} \, . \tag{11.28}$$

This shows that $c_{gy} = 0$, which means that the wave energy flux goes to zero, at a critical latitude where $\overline{u}^\lambda - c = 0$. The wave energy can propagate to a critical latitude, but not beyond it. Since the wave momentum flux travels in the direction opposite to the wave energy flux, momentum is removed from the vicinity of the critical latitude. As a result, the winds near the critical latitude tend to become easterly.

11.3 More about the interaction of planetary waves with the mean flow

11.3.1 How the eddies affect the zonally averaged quasigeostrophic PV

Recall from Chapter 4 that, within the framework of QG theory, the quasi-geostrophic pseudo-potential vorticity (QGPPV) controls the dynamics. We should therefore ask how the eddies affect the QGPPV. Many of the ideas discussed in this section originated with Charney and Drazin (1961), Charney and Stern (1962), and Dickinson (1969).

From the QGPPV equation discussed in Chapters 4 and 10, we can show that the zonally averaged QGPPV is governed by

$$\frac{\partial}{\partial t}\overline{Z_{\rm QG}}^{\lambda} = -\frac{\partial}{\partial y}\left(\overline{v_{\rm g}^{*}Z_{\rm QG}^{*}}^{\lambda}\right). \tag{11.29}$$

This means that, apart from the effects of heating and friction, the zonally averaged QGPPV changes only due to the convergence of the meridional eddy flux of QGPPV. We can show (see the problems at the end of this chapter) that this flux is related to the meridional eddy fluxes of both momentum and temperature, in the following interesting way:

$$\boxed{\begin{aligned}\overline{v_{\rm g}^{*}Z_{\rm QG}^{*}}^{\lambda} &= \overline{v_{\rm g}^{*}\zeta_{\rm g}^{*}}^{\lambda} - \frac{\partial}{\partial p}\left(\frac{Rf_0}{pS_p}\overline{v_{\rm g}^{*}T^{*}}^{\lambda}\right)\\ &= -\frac{\partial}{\partial y}\left(\overline{u_{\rm g}^{*}v_{\rm g}^{*}}^{\lambda}\right) - \frac{\partial}{\partial p}\left(\frac{Rf_0}{pS_p}\overline{v_{\rm g}^{*}T^{*}}^{\lambda}\right).\end{aligned}} \tag{11.30}$$

This shows that the meridional eddy flux of $Z_{\rm QG}$ can be written as a flux convergence. Specifically, the meridional eddy flux of potential vorticity is related to the convergence of the meridional eddy flux of zonal momentum, and to the rate of change with height of the meridional eddy sensible heat flux. When we form the convergence of the eddy potential vorticity flux, i.e., $-\partial\left(\overline{v_{\rm g}^{*}Z_{\rm QG}^{*}}^{\lambda}\right)/\partial y$, (11.30) will give us the term $-\partial^2\left(\overline{u_{\rm g}^{*}v_{\rm g}^{*}}^{\lambda}\right)/\partial y^2$. This affects the meridional shear of $\overline{u}^{\lambda}$. We will also get a term proportional to $-\partial^2\left(\overline{v_{\rm g}^{*}T^{*}}^{\lambda}\right)/\partial y\,\partial p$. This affects the static stability.

The flux whose convergence appears on the right-hand side of (11.30) is the quasi-geostrophic form of the so-called Eliassen-Palm flux vector, which we denote by **QGEPF** $\equiv \left(0,\ \mathrm{QGEPF}_{\varphi}, \mathrm{QGEPF}_{p}\right)$, where

$$\mathrm{QGEPF}_{\varphi} = -\overline{u_{\rm g}^{*}v_{\rm g}^{*}}^{\lambda}, \tag{11.31}$$

$$\mathrm{QGEPF}_p = -\frac{Rf_0}{pS_p}\,\overline{v_{\mathrm{g}}^* T^*}^{\lambda}\ . \tag{11.32}$$

When the QG pseudo-potential vorticity flux vanishes, the Eliassen-Palm flux is non-divergent. Later in this chapter we will see that these ideas can be generalized considerably.

11.3.2 Wave fluxes of potential temperature, wave energy, and potential vorticity

As discussed in Chapter 4, the QG form of the thermodynamic energy equation is

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_{\mathrm{g}} \cdot \nabla\right) \frac{\partial \phi}{\partial p} + S_p \omega = 0\ . \tag{11.33}$$

Eq. (11.33) can be written as

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_{\mathrm{g}} \cdot \nabla\right) \psi_z + \frac{N^2}{f_0} w = 0\ , \tag{11.34}$$

where z is the "log-p" coordinate defined in Chapter 10, $\psi \equiv \phi / f_0$ is the geostrophic stream function, $\psi_z \equiv \partial\psi/\partial z$, and w is defined by $-\omega/(\rho g)$. Linearization of (11.34) about a zonally averaged zonal flow gives

$$\left(\frac{\partial}{\partial t} + \overline{u_{\mathrm{g}}}^{\lambda} \frac{\partial}{\partial x}\right) \psi_z^* - v_{\mathrm{g}}^* \frac{\partial \overline{u_{\mathrm{g}}}^{\lambda}}{\partial z} + \frac{N^2}{f_0} w^* = 0\ . \tag{11.35}$$

Here we have used the thermal wind equation.

11.3.2.1 Baroclinically amplifying waves

Multiplication of (11.35) by ψ_z^* gives a form of the "*temperature variance equation*:"

$$\left(\frac{\partial}{\partial t} + \overline{u_{\mathrm{g}}}^{\lambda} \frac{\partial}{\partial x}\right) \left[\frac{1}{2}\left(\psi_z^*\right)^2\right] - v_{\mathrm{g}}^* \psi_z^* \frac{\partial \overline{u_{\mathrm{g}}}^{\lambda}}{\partial z} + \frac{N^2}{f_0} w^* \psi_z^* = 0\ . \tag{11.36}$$

Note the two gradient production terms (refer to Chapter 7). The triple-moment and dissipation terms are not included here.

Take the zonal mean of (11.36), so that the $\overline{u_g}^\lambda \partial/\partial x$ term drops out. Rearrange the result to isolate the meridional energy flux on the left-hand side:

$$\left(\overline{v^* \psi_z^*}^\lambda\right) \frac{\partial \overline{u}_g{}^\lambda}{\partial z} = \frac{\partial}{\partial t}\left[\frac{1}{2}\overline{\left(\psi_z^*\right)^2}^\lambda\right] + \frac{N^2}{f_0}\overline{w^* \psi_z^*}^\lambda \ . \tag{11.37}$$

Note that $\overline{w^* \psi_z^*}^\lambda / f_0 > 0$ implies an upward temperature flux in either hemisphere, and similarly $\overline{v^* \psi_z^*}^\lambda > 0$ implies a poleward temperature flux in either hemisphere.

Recall from Chapter 7 that baroclinic eddies derive their kinetic energy by conversion from the potential energy of the mean state, by means of an upward temperature flux that lowers the mean state's center of mass. For a baroclinically amplifying wave, $\frac{\partial}{\partial t}\left[\frac{1}{2}\overline{\left(\psi_z^*\right)^2}^\lambda\right] > 0$ and $\overline{w^* \psi_z^*}^\lambda / f_0 > 0$. (How do we know?) Eq. (11.37) therefore implies that *a baroclinically amplifying wave produces a poleward temperature flux (in either hemisphere) when $\partial \overline{u_g}^\lambda / \partial z > 0$, i.e., when the temperature decreases towards the pole*. Such a temperature flux is downgradient, so the gradient-production term is positive, which means that temperature variance tends to increase with time.

11.3.2.2 Upward propagating neutral waves

Next, consider a wave of the form $e^{ik(x-ct)}$, for which $\partial/\partial t = -c\partial/\partial x$. We assume that c is real, which means that the wave is neutral, i.e., its amplitude is not changing with time. Multiply (11.35) by ψ^* and take the zonal mean, to obtain

$$\left(\overline{u_g}^\lambda - c\right) \overline{v_g^* \psi_z^*}^\lambda = \frac{N^2}{f_0}\overline{w^* \psi^*}^\lambda \ . \tag{11.38}$$

Note that $\overline{w^* \psi^*}^\lambda / f_0 > 0$ means an upward propagation of wave energy in either hemisphere. Recall also that $\overline{u_g}^\lambda - c > 0$ is needed in order for the wave to propagate at all. It follows that *an upward-propagating neutral wave transports energy poleward*. Such a wave might be excited, for example, by flow over topography.

In summary, a poleward heat flux is produced by either a baroclinically amplifying wave with $\partial \overline{u_g}^\lambda / \partial z > 0$ or a neutral wave that propagates upward.

11.3.2.3 A non-interaction theorem for the quasigeostrophic system

Now recall from Chapter 10 that the QG eddy PV equation is

$$\left(\frac{\partial}{\partial t}+\overline{u_{\rm g}}^{\lambda}\frac{\partial}{\partial x}\right)Z^{*}_{\rm QG}+v^{*}_{\rm g}\frac{\partial}{\partial y}\overline{Z_{\rm QG}}^{\lambda}=0\,. \tag{11.39}$$

Application of (11.39) to a neutral wave gives

$$\left(\overline{u_{\rm g}}^{\lambda}-c\right)\frac{\partial Z^{*}_{\rm QG}}{\partial x}+v^{*}_{\rm g}\frac{\partial}{\partial y}\overline{Z_{\rm QG}}^{\lambda}=0\,. \tag{11.40}$$

Multiply (11.40) by $Z^{*}_{\rm QG}$ and take the zonal mean to show that

$$\boxed{\overline{v^{*}_{\rm g}Z^{*}_{\rm QG}}^{\lambda}=0 \quad \text{except where} \quad \overline{u_{\rm g}}^{\lambda}=c\,.} \tag{11.41}$$

This means that the eddy QGPPV flux vanishes except at a critical latitude. It follows from (11.29) that neutral waves do not affect $\overline{Z_{\rm QG}}^{\lambda}$ except at a critical latitude. *This is a non-interaction theorem for quasi-geostrophic planetary waves*, analogous to the non-interaction theorem for gravity waves obtained by Eliassen and Palm (1961). Eq. (11.41) implies that the QG Eliassen-Palm flux is non-divergent except at a critical latitude.

11.3.2.4 Acceleration of the mean flow by upward propagating neutral waves

Referring back to (11.30), we see that $\overline{v^{*}Z^{*}_{\rm QG}}^{\lambda}=0$ means

$$-\frac{\partial}{\partial y}\left(\overline{u^{*}_{\rm g}v^{*}_{\rm g}}^{\lambda}\right)-\frac{\partial}{\partial p}\left(\frac{Rf_0}{pS_p}\overline{v^{*}_{\rm g}T^{*}}^{\lambda}\right)=0\,. \tag{11.42}$$

This means that the quasi-geostrophic Eliasssen-Palm flux is nondivergent.

Vertically integrate (11.42) through the depth of the atmosphere to obtain

$$-\int_{0}^{p_S}\frac{\partial}{\partial y}\left(\overline{u^{*}_{\rm g}v^{*}_{\rm g}}^{\lambda}\right)\,\frac{dp}{g}=f_0^2\frac{\rho}{N^2}\left(\overline{v^{*}_{\rm g}\psi^{*}_{z}}^{\lambda}\right)_{S}\,, \tag{11.43}$$

which applies to neutral waves. The left-hand side of (11.43) represents the vertically integrated convergence of meridional momentum flux, and the right-hand side represents the near-surface value of the eddy meridional energy flux. Recall our earlier conclusion that an upward-propagating neutral wave produces a poleward heat flux, i.e., $\overline{v_g^* \psi_z^*}^\lambda > 0$. It follows from (11.43) that

$$\boxed{-\int_0^{p_S} \frac{\partial}{\partial y}\left(\overline{u_g^* v_g^*}^\lambda\right)\frac{dp}{g} > 0 \quad \text{for an upward - propagating neutral wave .}} \tag{11.44}$$

11.3.2.5 Wave momentum exchange with the Earth's surface

This means that the vertically integrated meridional momentum flux convergence tends to *strengthen* the vertically integrated zonal-mean zonal wind. This is consistent with the discussion of Rossby wave momentum transport given earlier in this chapter, and it is also consistent with the observed conversion from eddy kinetic energy to zonal kinetic energy, discussed in Chapter 10. On the other hand, the waves also transport temperature poleward, which tends to reduce the meridional temperature gradient and so (as implied by thermal-wind balance) tend to *weaken* the jet. The momentum flux and heat flux thus have opposing effects on the mean flow. If these two opposing effects were to cancel, then the eddies would have no net effect on the mean flow.

An upward propagating neutral wave in westerly shear tends to produce a downward momentum flux at the Earth's surface. To see why, consider the angular momentum equation, in Cartesian coordinates for simplicity:

$$\frac{\partial}{\partial t}(\rho M) + \frac{\partial}{\partial x}(\rho u M) + \frac{\partial}{\partial y}(\rho v M) + \frac{\partial}{\partial z}(\rho w M) = -\frac{\partial p}{\partial \lambda} . \tag{11.45}$$

Here we have used height as the vertical coordinate, and we have assumed no friction above the boundary layer. Taking the zonal mean of (11.45) gives

$$\frac{\partial}{\partial t}\left(\overline{\rho M}^\lambda\right) + \frac{\partial}{\partial y}\left(\overline{\rho v^* M^*}^\lambda\right) + \frac{\partial}{\partial z}\left(\overline{\rho w^* M^*}^\lambda\right) = 0 . \tag{11.46}$$

Here advection of $\overline{M}^\lambda$ by $\overline{v}^\lambda$ and $\overline{w}^\lambda$ has been neglected; this can be justified for the midlatitudes. To the extent that $\overline{v}^\lambda$ is geostrophic, it vanishes anyway. Now assume $\partial \overline{\rho M}^\lambda / \partial t = 0$, consistent with $\partial \overline{Z}^\lambda / \partial t = 0$. This leads to

$$\frac{\partial}{\partial y}\left(\overline{\rho v^* M^*}^\lambda\right) = -\frac{\partial}{\partial z}\left(\rho\overline{w^* M^*}^\lambda\right) \ . \tag{11.47}$$

Integrating (11.47) vertically with respect to mass, and using (11.44), we conclude that

$$\int_0^\infty \frac{\partial}{\partial z}\left(\rho\overline{w^* M^*}^\lambda\right)\, dz > 0 \, . \tag{11.48}$$

We know that $\rho\overline{w^* M^*}^\lambda$ must vanish at great height, so (11.48) implies that

$$\boxed{\left(\rho\overline{w^* M^*}^\lambda\right)_S < 0 \quad \text{for an upward-propagating neutral wave} \ .} \tag{11.49}$$

This shows that, in the presence of an upward propagating neutral wave, surface friction and/or mountain torque must carry angular momentum into the Earth's surface. An alternative interpretation is that, in a belt of westerlies where (11.37) is satisfied, surface friction and/or mountain torque will produce an upward-propagating planetary wave which, as discussed above, will transport energy poleward.

Compare (11.44) and (11.49). The meridional momentum flux accelerates the westerlies, while the vertical momentum flux decelerates them. Angular momentum is carried poleward into the elevated jet, and then downward into the Earth's surface.

11.3.3 Momentum transport by upward-propagating neutral waves

As shown in Eq. (6.38), in isentropic coordinates the upward flux of zonal momentum due to a wave takes the form

$$\text{upward wave flux of zonal momentum} = -\left(\overline{p^* \frac{\partial z^*}{\partial \lambda}}^\lambda\right)_\theta \ . \tag{11.50}$$

From Eq. (4.133), which is the mechanical energy equation in isentropic coordinates, we see that

$$\text{upward flux of wave energy} = -\overline{z^* \left(\frac{\partial p^*}{\partial t}\right)_\theta}^\lambda \ . \tag{11.51}$$

For a neutral wave propagating zonally with phase velocity c, we have $\partial/\partial t = -(c/a\cos\varphi)\,\partial/\partial\lambda$. This allows us to make a connection between (11.50) and (11.51):

$$\begin{aligned} -\overline{z^*\left(\frac{\partial p^*}{\partial t}\right)_\theta}^{\lambda} &= \frac{c}{a\cos\varphi}\overline{z^*\left(\frac{\partial p^*}{\partial \lambda}\right)_\theta}^{\lambda} \\ &= -\frac{c}{a\cos\varphi}\overline{p^*\left(\frac{\partial z^*}{\partial \lambda}\right)_\theta}^{\lambda} \,. \end{aligned} \tag{11.52}$$

We conclude that

$$\text{upward flux of wave energy } = -\frac{c}{a\cos\varphi}\overline{p^*\left(\frac{\partial z^*}{\partial \lambda}\right)_\theta}^{\lambda} \,. \tag{11.53}$$

Comparing (11.53) with (11.50), we see that for a neutral wave

$$\text{upward wave energy flux } = \frac{c}{a\cos\varphi} \times \text{ upward wave flux of zonal momentum} \tag{11.54}$$

This shows that *for eastward-propagating neutral waves (with $c > 0$) the momentum flux and the energy flux have the same sign, while for westward-propagating neutral waves (with $c < 0$) the momentum and energy fluxes have opposite signs. These conclusions apply to all wave types.*

11.4 The Transformed Eulerian Mean (TEM) system

Earlier in this chapter we discussed non-interaction theorems for pure gravity waves and for quasi-geostrophic eddies on a β-plane. Charney and Drazin (1961), Andrews and Mcintyre (1976), Boyd (1976), and Andrews and McIntyre (1978) discovered that non-interaction theorems can be derived for more general flows.

The zonally averaged equations in spherical coordinates can be written as

$$\begin{aligned}&\frac{\partial \overline{M}^{\lambda}}{\partial t}+\overline{v}^{\lambda}\frac{\partial \overline{M}^{\lambda}}{\partial \varphi}+\overline{w}^{\lambda}\frac{\partial \overline{M}^{\lambda}}{\partial z}-\overline{F_x}^{\lambda}a\cos\varphi\\&=\frac{-1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^*M^*}^{\lambda}\cos\varphi\right)-\frac{1}{\overline{\rho}^{\lambda}}\frac{\partial}{\partial z}\left(\overline{\rho}^{\lambda}\overline{w^*M^*}^{\lambda}\right),\end{aligned}\tag{11.55}$$

$$\begin{aligned}&\frac{\partial \overline{v}^{\lambda}}{\partial t}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial \overline{v}^{\lambda}}{\partial \varphi}+\overline{w}^{\lambda}\frac{\partial \overline{v}^{\lambda}}{\partial z}+\overline{u}^{\lambda}\left(f+\frac{\overline{u}^{\lambda}\tan\varphi}{a}\right)+\frac{1}{a}\frac{\partial \overline{\phi}^{\lambda}}{\partial \varphi}-\overline{F_y}^{\lambda}\\&=\frac{-1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\overline{(v^*)^2}^{\lambda}\cos\varphi\right]-\frac{1}{\overline{\rho}^{\lambda}}\frac{\partial}{\partial z}\left(\overline{\rho}^{\lambda}\overline{v^*w^*}^{\lambda}\right)-\frac{\overline{(u^*)^2}^{\lambda}\tan\varphi}{a},\end{aligned}\tag{11.56}$$

$$\begin{aligned}&\frac{\partial \overline{\theta}^{\lambda}}{\partial t}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial \overline{\theta}^{\lambda}}{\partial \varphi}+\overline{w}^{\lambda}\frac{\partial \overline{\theta}^{\lambda}}{\partial z}-\overline{Q}^{\lambda}\\&=\frac{-1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^*\theta^*}^{\lambda}\cos\varphi\right)-\frac{1}{\overline{\rho}^{\lambda}}\frac{\partial}{\partial z}\left(\overline{\rho}^{\lambda}\overline{w^*\theta^*}^{\lambda}\right),\end{aligned}\tag{11.57}$$

$$\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\rho}^{\lambda}\overline{v}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial z}\left(\overline{\rho}^{\lambda}\overline{w}^{\lambda}\right)=0\,,\tag{11.58}$$

$$\frac{\partial \overline{\phi}^{\lambda}}{\partial z}-\frac{R\overline{\theta}^{\lambda}}{H}e^{-\kappa z/H}=0\,.\tag{11.59}$$

Here $z\equiv -H\log\left(p/p_0\right)$ is the log-pressure vertical coordinate, and $w\equiv Dz/Dt$. The scale height H is RT_0/g, where T_0 is a constant. In (11.57), Q represents a heating process. In the above equations, $\overline{\rho}^{\lambda}\left(z\right)\equiv\rho_0 e^{-z/H}$, where ρ_0 is the density at $z=0$. We have assumed for simplicity that the temperature is uniform with height, but this assumption is not necessary.

We assume that the meridional momentum equation, (11.56), can be approximated by gradient wind balance, i.e.

$$\overline{u}^{\lambda}\left(f+\overline{u}^{\lambda}\frac{\tan\varphi}{a}\right)+\frac{1}{a}\frac{\partial\overline{\phi}^{\lambda}}{\partial\varphi}\cong 0\,. \tag{11.60}$$

The smallness of $\overline{v}^{\lambda}$ outside the tropics, discussed in Chapter 3, implies that the first three terms on the left-hand side of (11.56) are small. Eq. (11.60) can then be obtained by neglecting friction and the various eddy terms on the right-hand side of (11.56). Note, however, that (11.60) breaks down at the equator, where f and $\tan\varphi$ are zero.

We now *define* a "*residual circulation*" by

$$\overline{\rho}^{\lambda}V\equiv\overline{\rho}^{\lambda}\overline{v}^{\lambda}-\frac{\partial}{\partial z}\left[\left(\overline{\rho}^{\lambda}\frac{\partial z}{\partial\overline{\theta}^{\lambda}}\right)\overline{v^{*}\theta^{*}}^{\lambda}\right]\,, \tag{11.61}$$

$$\overline{\rho}^{\lambda}W\equiv\overline{\rho}^{\lambda}\overline{w}^{\lambda}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\left(\overline{\rho}^{\lambda}\frac{\partial z}{\partial\overline{\theta}^{\lambda}}\right)\overline{v^{*}\theta^{*}}^{\lambda}\cos\varphi\right]\,. \tag{11.62}$$

Here $\partial z/\partial\overline{\theta}\equiv\left(\partial\overline{\theta}/\partial z\right)^{-1}$. Recall that $\rho\,\partial z/\partial\theta$ is the isentropic pseudodensity. The eddy contributions to V and W are both proportional to the eddy meridional flux of potential temperature, multiplied by the isentropic pseudodensity averaged along surfaces of constant z. In the absence of eddies, $V=\overline{v}^{\lambda}$ (which, again, is expected to be small outside the tropics), and $W=\overline{w}^{\lambda}$. The eddy terms are of course largest in middle latitudes, while the "Eulerian" terms, involving $\overline{v}^{\lambda}$ and $\overline{w}^{\lambda}$, are dominant in the tropics. Note that the residual circulation is independent of longitude. Fig. 11.3 shows the stream function of the observed residual circulation from 10°S to 70°N, as analyzed by Edmon Jr. et al. (1980). The residual circulation for winter looks like a giant Hadley Cell, extending from the tropics to the poles. As discussed by Juckes et al. (1994) and later in this chapter, it is not a coincidence that the residual circulation looks similar to the mean meridional circulation as seen in isentropic coordinates (see Figure 6.3). For summer, we see the northern edge of a southern-hemisphere Hadley Cell that has its rising branch in the Northern Hemisphere. Clearly, the residual circulation is partially driven by tropical heating. Notice that the stream function lines intersect the lower boundary, especially in higher latitudes where the eddy terms are dominant. This happens because the TEM system does not force the vertically averaged residual circulation to vanish. It is a weakness of the TEM system.

Use of (11.61) and (11.62) to eliminate $\overline{v}$ and $\overline{w}$ in (11.58) shows that V and W satisfy a continuity equation with the same form as (11.58):

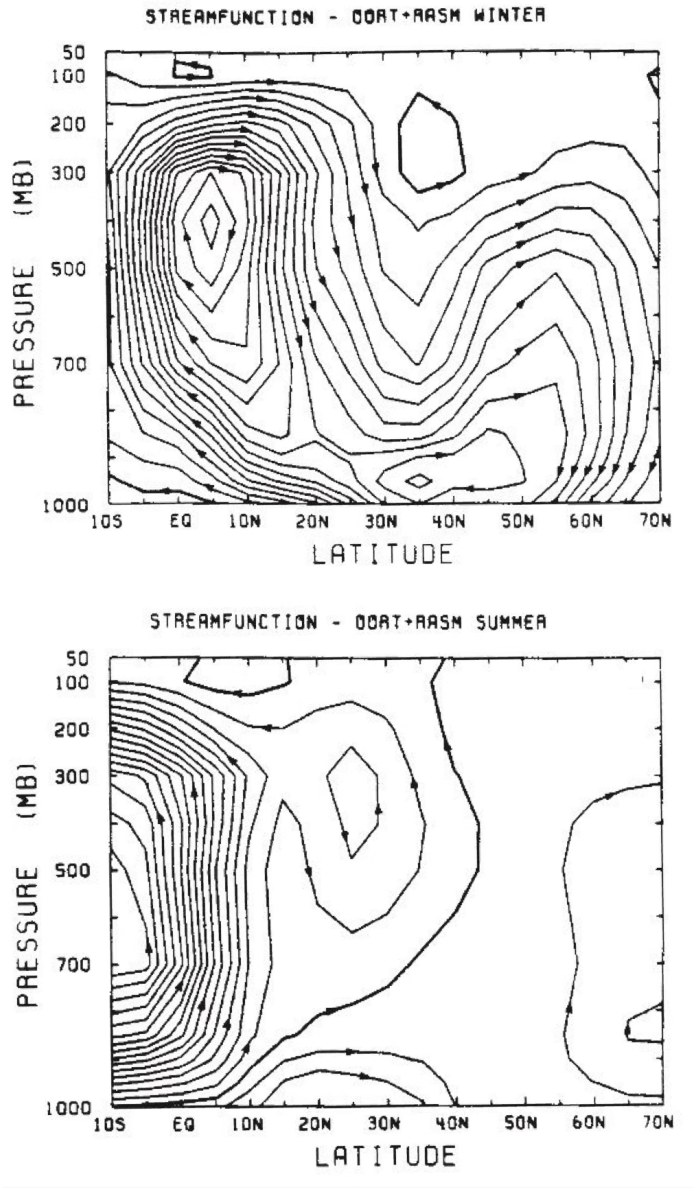


Figure 11.3: Top panel: The stream function of the winter residual circulation. The contour interval is 7.5 x 10^{16} m^2 s Pa. Bottom panel: Same as the top panel but for summer, and with a much smaller contour interval of 1 x 10^{15} m^3. From Edmon Jr. et al. (1980).

$$\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\rho}^{\lambda}V\cos\varphi\right)+\frac{\partial}{\partial z}\left(\overline{\rho}^{\lambda}W\right)=0\,. \tag{11.63}$$

Similarly, use of (11.61) and (11.62) to eliminate $\overline{v}$ and $\overline{w}$ in (11.55) and (11.57) leads to:

$$\frac{\partial \overline{M}^{\lambda}}{\partial t}+\frac{V}{a}\frac{\partial \overline{M}^{\lambda}}{\partial \varphi}+W\frac{\partial \overline{M}^{\lambda}}{\partial z}-a\cos\varphi\overline{F_x}^{\lambda}=\frac{1}{\overline{\rho}^{\lambda}}\nabla\cdot\mathbf{EPF}\,, \tag{11.64}$$

and

$$\begin{aligned}&\frac{\partial \overline{\theta}^{\lambda}}{\partial t}+\frac{V}{a}\frac{\partial \overline{\theta}^{\lambda}}{\partial \varphi}+W\frac{\partial \overline{\theta}^{\lambda}}{\partial z}-\overline{Q}^{\lambda}\\ &=-\frac{1}{\overline{\rho}^{\lambda}}\frac{\partial}{\partial z}\left[\left(\overline{\rho}^{\lambda}\frac{\partial z}{\partial \overline{\theta}^{\lambda}}\right)\left(\overline{v^*\theta^*}^{\lambda}\frac{1}{a}\frac{\partial \overline{\theta}^{\lambda}}{\partial \varphi}+\overline{w^*\theta^*}^{\lambda}\frac{\partial \overline{\theta}^{\lambda}}{\partial z}\right)\right]\,,\end{aligned} \tag{11.65}$$

respectively, where

$$\mathbf{EPF}\equiv\left[0,(\mathrm{EPF})_{\varphi},(\mathrm{EPF})_{z}\right] \tag{11.66}$$

is the "Eliassen-Palm flux," whose components are

$$(\mathrm{EPF})_{\varphi}\equiv\left(\overline{\rho}^{\lambda}\frac{\partial z}{\partial \overline{\theta}^{\lambda}}\right)\left(\frac{\partial \overline{M}^{\lambda}}{\partial z}\cdot\overline{v^*\theta^*}^{\lambda}-\overline{v^*M^*}^{\lambda}\frac{\partial \overline{\theta}^{\lambda}}{\partial z}\right)\,, \tag{11.67}$$

and

$$(\mathrm{EPF})_{z}\equiv\left(\overline{\rho}^{\lambda}\frac{\partial z}{\partial \overline{\theta}^{\lambda}}\right)\left(-\frac{1}{a}\frac{\partial \overline{M}^{\lambda}}{\partial \varphi}\overline{v^*\theta^*}^{\lambda}-\overline{w^*M^*}^{\lambda}\frac{\partial \overline{\theta}^{\lambda}}{\partial z}\right)\,. \tag{11.68}$$

In (11.67), the $-\overline{v^*M^*}^{\lambda}$ term is usually dominant, and in (11.68) the $\overline{v^*\theta^*}^{\lambda}$ term is usually dominant. When the **EPF** vector points upward, the meridional flux of potential temperature is in control. When it points in the meridional direction, the meridional flux of zonal momentum is in control. From (11.64) we see that a positive Eliassen-Palm flux divergence tends to increase $\overline{M}^{\lambda}$. Compare (11.67) and (11.68) with the QG forms (11.31) and (11.32). Eqs. (11.64) - (11.65) are called the "transformed Eulerian mean" (TEM) equations.

You should recognize that the eddy terms inside the vertical derivative on the right-hand side of (11.65) are proportional to the gradient-production terms of the equation for the eddy potential temperature variance, which was discussed in Chapter 7.

The derivation above appears to be nothing more than an algebraic shuffle. We wrote down (11.61) and (11.62) without any explanation or motivation. What is the point of all this? The point is that for steady small-amplitude eddies with no friction or heating it can be shown that

$$\nabla \cdot (\mathbf{EPF}) = 0 \,. \tag{11.69}$$

It turns out that the eddy forcing term on the right-hand side of (11.65) is zero under the same conditions, i.e.

$$\overline{\rho}^{\lambda} \overline{v^* \theta^*}^{\lambda} \frac{1}{a} \frac{\partial \overline{\theta}^{\lambda}}{\partial \varphi} + \overline{\rho}^{\lambda} \overline{w^* \theta^*}^{\lambda} \frac{\partial \overline{\theta}^{\lambda}}{\partial z} = 0 \,. \tag{11.70}$$

This means that the gradient-production terms of the potential temperature variance equation vanish.

For the case of steady, small-amplitude eddies, in the absence of friction and heating, the TEM system reduces to

$$\begin{aligned}
\frac{\partial \overline{M}^{\lambda}}{\partial t} + \frac{V}{a} \frac{\partial \overline{M}^{\lambda}}{\partial \varphi} + W \frac{\partial \overline{M}^{\lambda}}{\partial z} &= 0, \\
\overline{u}^{\lambda} \left(f + \overline{u}^{\lambda} \frac{\tan \varphi}{a} \right) + \frac{1}{a} \frac{\partial \overline{\phi}^{\lambda}}{\partial \varphi} &= 0, \\
\frac{\partial \overline{\theta}^{\lambda}}{\partial t} + \frac{V}{a} \frac{\partial \overline{\theta}^{\lambda}}{\partial \varphi} + W \frac{\partial \overline{\theta}^{\lambda}}{\partial z} &= 0, \\
\frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} \left(\overline{\rho}^{\lambda} V \cos \varphi \right) + \frac{\partial}{\partial z} \left(\overline{\rho}^{\lambda} W \right) &= 0, \\
\frac{\partial \overline{\phi}^{\lambda}}{\partial z} - \frac{R \overline{\theta}^{\lambda}}{H} e^{-\kappa z / H} &= 0 \,.
\end{aligned} \tag{11.71}$$

This system has the following steady solution:

$$\begin{aligned} \frac{\partial \overline{M}^{\lambda}}{\partial t} &= 0, \\ V = 0 \quad \text{and} \quad W &= 0, \\ \frac{\partial \overline{\theta}^{\lambda}}{\partial t} &= 0\,, \end{aligned} \tag{11.72}$$

where $\overline{M}^{\lambda}$ is in gradient-wind balance, and $\overline{\theta}^{\lambda}$ is specified from the past history or radiative - convective equilibrium. From the definitions of V and W, we can solve for the mean meridional circulation implied by $V = 0$ and $W = 0$:

$$\overline{\rho}^{\lambda}\overline{v}^{\lambda} = \frac{\partial}{\partial z}\left(\overline{\rho}^{\lambda}\frac{\overline{v^*\theta^*}^{\lambda}}{\partial\overline{\theta}^{\lambda}/\partial z}\right)\,, \tag{11.73}$$

$$\overline{w}^{\lambda} = -\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\cos\varphi\frac{\overline{v^*\theta^*}^{\lambda}}{\partial\overline{\theta}^{\lambda}/\partial z}\right)\,. \tag{11.74}$$

As an aid for the interpretation of these results, suppose we had a solution with no eddies at all. "No eddies" certainly qualify as "steady small-amplitude eddies," so the above argument applies, and from (11.73) and (11.74) we conclude that the MMC vanishes in the absence of eddies (and heating).

Now add steady small-amplitude eddies, so that $\nabla\cdot(\mathbf{EPF})$ continues to be zero. Then exactly the same $\overline{M}^{\lambda}$ and $\overline{\theta}^{\lambda}$ will satisfy the equations! Of course, $\overline{v}^{\lambda}$ and $\overline{w}^{\lambda}$ will be different, i.e., the MMC will be different, because it will have to be just what it takes to ensure that $V = W = 0$, i.e., to satisfy (11.60) and (11.61). This MMC is eddy-driven. The system produces the MMC in order to prevent the eddies from disrupting the gradient-wind balance. Perhaps a better way to say this is that the processes that act to maintain thermal wind balance (i.e., geostrophic and hydrostatic adjustment) accomplish this feat by using the wave-induced MMC as a tool.

The interpretation of this amazing result is that if you try to modify $\overline{M}^{\lambda}$ and $\overline{\theta}^{\lambda}$ by applying steady, linear eddy forcing such that $\nabla\cdot(\mathbf{EPF}) = 0$, you will be disappointed! All that will happen is that the MMC will change, in such a way that V and W continue to be zero. In effect, the eddies induce an MMC whose effects exactly cancel the *direct* effects of the eddies on $\overline{M}^{\lambda}$ and $\overline{\theta}^{\lambda}$.

When the eddies are unsteady, the residual circulation is different from zero, and $\overline{M}^{\lambda}$ and $\overline{\theta}^{\lambda}$ are modified by the combined effects of the eddies and the eddy-induced MMC. The effects of the eddies and the MMC still tend to cancel, but the cancellation is incomplete.

Edmon Jr. et al. (1980) used the QG form of the non-interaction theorem to analyze the data of Oort and Rasmusson (1971). Fig. 11.4 shows the Eliassen-Palm flux (arrows) and its divergence (contours). First consider the winter results, shown in the upper panels. In midlatitudes, near the surface, the EP flux vectors point strongly upward, indicating an intense poleward potential temperature flux. Near the tropopause, the arrows curve toward the tropics and become nearly horizontal. This indicates a strong poleward eddy momentum flux. Keep in mind that $\nabla \cdot (\mathbf{EPF}) > 0$ means $\partial \overline{M}^{\lambda} / \partial t > 0$, i.e., a positive EPF divergence favors westerly acceleration. The negative divergence (i.e., convergence) near 200 mb at about 30° N indicates that the net effect of the eddies is to decelerate the jet. In fact, the westerlies are being decelerated throughout middle latitudes, except near the surface. Note that this EPF convergence results mainly from the upward decrease of the upward component of the flux, i.e., it is mainly due to the meridional potential temperature flux. The results for summer (panel c) are quite similar, except that the action is generally weaker, and shifted poleward. Further analysis (Edmon Jr. et al., 1980) shows that the contributions of the transient eddies dominate over those of the stationary eddies, in both winter and summer.

Fig. 11.4 shows seasonal means. In the remainder of this chapter, we discuss three fairly dramatic examples of time-dependent interactions between the eddies and the zonally averaged circulation, and one example of a seasonal-mean interaction.

The TEM framework, as presented above and originally formulated by Andrews and Mcintyre (1976), is most rigorously justified in the limit of small-amplitude eddies. Eddy perturbations to the mean flow are assumed to be small compared to the mean flow, and it follows that the residual circulation induced by the eddies is, in effect, assumed to be weak. In practice, however, the TEM framework is often applied in contexts where eddies are not small-amplitude. In these cases, the TEM is used as an approximation. It does not apply to highly turbulent regimes in which eddy amplitudes can be comparable to or larger than the mean flow. Recent work Pauluis et al. (2011) has led to statistical generalizations of the TEM that relax the small-amplitude assumption and allow for finite-amplitude eddy effects.

11.5 The Eliassen-Palm theorem in isentropic coordinates

The following discussion is based on the work of Andrews (1983), Tung (1986), and Andrews et al. (1987).

The tendency term of the angular momentum equation can be broken into two parts:

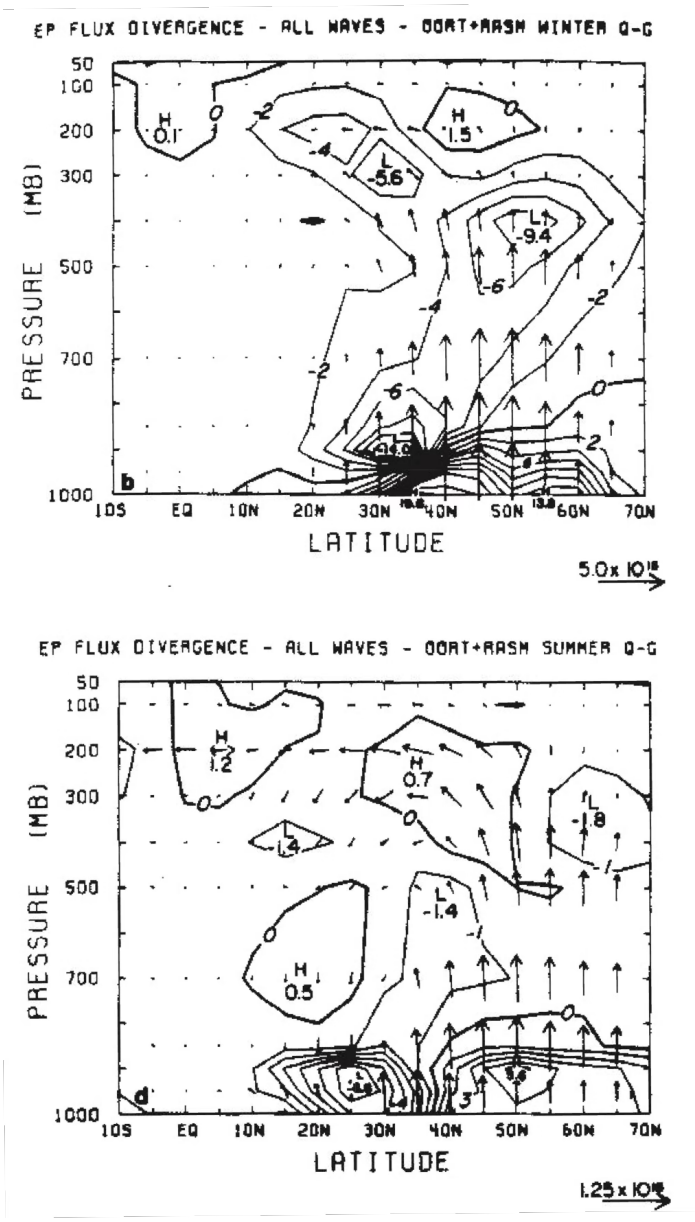


Figure 11.4: Top panel: Contribution of all eddies to the seasonally averaged Eliassen-Palm fluxes for winter. The contour interval is 2 x 10^{15} m^3. The horizontal arrow scale in units of m^3 is indicated at bottom right. Bottom panel: Same as the top panel, but for summer. From Edmon Jr. et al. (1980).

$$\frac{\partial}{\partial t}\left(\overline{\rho_\theta M}^\lambda\right) = \frac{\partial}{\partial t}\left(\overline{\rho_\theta}^\lambda \overline{M}^\lambda\right) + \frac{\partial}{\partial t}\left(\overline{\rho_\theta {}^* M^*}^\lambda\right) . \tag{11.75}$$

Using (11.75) in (6.38), we obtain

$$\begin{aligned}
&\frac{\partial}{\partial t}\left(\overline{\rho_\theta}^\lambda \overline{M}^\lambda\right) + \frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\rho_\theta v M}^\lambda \cos\varphi\right) \\
&= -\frac{\partial}{\partial t}\left(\overline{\rho_\theta^* M^*}^\lambda\right) + \frac{\partial}{\partial\theta}\left(\overline{p^*\frac{\partial z^*}{\partial\lambda}}^\lambda - \overline{\rho_\theta M\dot{\theta}}^\lambda + \overline{F_u}^\lambda a\cos\varphi\right) .
\end{aligned} \tag{11.76}$$

Here the "eddy part" of the time-rate-of-change term has been moved to the right-hand-side of the equals sign, for reasons that will be discussed below. We want an "advective form" of (11.76), so we subtract $\overline{M}^\lambda$ times the zonally averaged continuity equation to obtain

$$\begin{aligned}
&\overline{\rho_\theta}^\lambda \frac{\partial \overline{M}^\lambda}{\partial t} + \frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\rho_\theta v M}^\lambda \cos\varphi\right) - \frac{\overline{M}^\lambda}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{\rho_\theta v}^\lambda \cos\varphi\right) \\
&= -\frac{\partial}{\partial t}\left(\overline{\rho_\theta{}^* M^*}^\lambda\right) + \overline{M}^\lambda \frac{\partial}{\partial\theta}\left(\overline{\rho_\theta\dot{\theta}}^\lambda\right) + \frac{\partial}{\partial\theta}\left(\overline{p^*\frac{\partial z^*}{\partial\lambda}}^\lambda - \overline{\rho_\theta M\dot{\theta}}^\lambda + \overline{F_u}^\lambda a\cos\varphi\right) .
\end{aligned} \tag{11.77}$$

We cannot yet combine the meridional and vertical derivative terms of (11.77) to obtain an advective form.

In order to obtain the desired advective form (a few steps further on), we introduce a mass-weighted zonal mean, defined by

$$\widehat{A} \equiv \frac{\overline{\rho_\theta A}^\lambda}{\overline{\rho_\theta}^\lambda} , \tag{11.78}$$

where A is an arbitrary scalar. Note that with this definition the "hat" quantities are independent of longitude. Using (11.78), we can write

$$\begin{aligned}
\rho_\theta A &= \overline{\rho_\theta A}^\lambda + (\rho_\theta A)^* \\
&= \overline{\rho_\theta}^\lambda \widehat{A} + (\rho_\theta A)^* ,
\end{aligned} \tag{11.79}$$

and

$$\begin{aligned}\overline{\rho_\theta AB}^\lambda &= \overline{\rho_\theta A}^\lambda \overline{B}^\lambda + \overline{(\rho_\theta A)^* B^*}^\lambda \\ &= \overline{\rho_\theta}^\lambda \widehat{A}\overline{B}^\lambda + \overline{(\rho_\theta A)^* B^*}^\lambda \,,\end{aligned} \tag{11.80}$$

where B is a second arbitrary variable. As special cases of (11.80), we can write

$$\begin{aligned}\overline{\rho_\theta vB}^\lambda &= \overline{\rho_\theta v}^\lambda \overline{B}^\lambda + \overline{(\rho_\theta v)^* B^*}^\lambda \\ &= \overline{\rho_\theta}^\lambda \widehat{v}\overline{B}^\lambda + \overline{(\rho_\theta v)^* B^*}^\lambda \,,\end{aligned} \tag{11.81}$$

and

$$\begin{aligned}\overline{\rho_\theta \dot{\theta} B}^\lambda =& \overline{\rho_\theta \dot{\theta}}^\lambda \overline{B}^\lambda + \overline{\left(\rho_\theta \dot{\theta}\right)^* B^*}^\lambda \\ =& \overline{\rho_\theta}^\lambda \widehat{\dot{\theta}}\overline{B}^\lambda + \overline{\left(\rho_\theta \dot{\theta}\right)^* B^*}^\lambda \,.\end{aligned} \tag{11.82}$$

Using (11.81) and (11.82), we can rewrite the zonally averaged continuity and angular momentum equations as

$$\frac{\partial \overline{\rho_\theta}^\lambda}{\partial t} + \frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{\rho_\theta}^\lambda \widehat{v}\cos\varphi\right) = -\frac{\partial}{\partial \theta}\left(\overline{\rho_\theta}^\lambda \widehat{\dot{\theta}}\right) \,, \tag{11.83}$$

and

$$\begin{aligned}&\overline{\rho_\theta}^\lambda \frac{\partial \overline{M}^\lambda}{\partial t} + \frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left\{\left[\overline{\rho_\theta}^\lambda \widehat{v}\overline{M}^\lambda + \overline{(\rho_\theta v)^* M^*}^\lambda\right]\cos\varphi\right\} - \frac{\overline{M}^\lambda}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{\rho_\theta}^\lambda \widehat{v}\cos\varphi\right) \\ &= -\frac{\partial}{\partial t}\left(\overline{\rho_\theta^* M^*}^\lambda\right) + \overline{M}^\lambda \frac{\partial}{\partial \theta}\left(\overline{\rho_\theta}^\lambda \widehat{\dot{\theta}}\right) + \frac{\partial}{\partial \theta}\left[\overline{p^*\frac{\partial z^*}{\partial \lambda}}^\lambda - \overline{\rho_\theta}^\lambda \widehat{\dot{\theta}}\overline{M}^\lambda - \overline{\left(\rho_\theta \dot{\theta}\right)^* M^*}^\lambda - \overline{F_u}^\lambda a\cos\varphi\right] \,,\end{aligned} \tag{11.84}$$

respectively. The meridional and vertical derivatives in (11.84) can now be combined to obtain the desired advective form. We also simplify and rearrange, to obtain

$$\overline{\rho_\theta}^\lambda \left(\frac{\partial \overline{M}^\lambda}{\partial t} + \frac{\widehat{v}}{a}\frac{\partial \overline{M}^\lambda}{\partial \varphi} + \widehat{\dot{\theta}}\frac{\partial \overline{M}^\lambda}{\partial \theta} \right) = -\frac{\partial}{\partial t}\left(\overline{\rho_\theta^* M^*}^\lambda \right) \\ -\frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left[\overline{(\rho_\theta v)^* M^*}^\lambda \cos\varphi \right] + \frac{\partial}{\partial \theta}\left[\overline{p^* \frac{\partial z^*}{\partial \lambda}}^\lambda - \overline{\left(\rho_\theta \dot{\theta}\right)^* M^*}^\lambda - \overline{F_u}^\lambda a\cos\varphi \right] . \tag{11.85}$$

Here all of the eddy terms (and friction) have been collected on the right-hand side, and the remaining terms have been collected on the left-hand side.

The terms on the right-hand side of (11.85) have the form of a flux divergence. This motivates us to define the isentropic Eliassen-Palm flux vector as

$$\mathbf{IEPF} \equiv \left(0,\, \mathrm{IEPF}_\varphi,\, \mathrm{IEPF}_\theta\right) , \tag{11.86}$$

where

$$\mathrm{IEPF}_\varphi \equiv -\overline{(\rho_\theta v)^* M^*}^\lambda , \tag{11.87}$$

and

$$\mathrm{IEPF}_\theta \equiv \overline{p^* \frac{\partial z^*}{\partial \lambda}}^\lambda - \overline{\left(\rho_\theta \dot{\theta}\right)^* M^*}^\lambda . \tag{11.88}$$

The meridional component of the **IEPF** is simply *minus* the eddy meridional angular momentum flux. The vertical component is *minus* the "total" vertical eddy angular momentum flux due to the combination of isentropic form drag and the vertical mass flux associated with heating. With these definitions, (11.85) can be written as

$$\overline{\rho_\theta}^\lambda \left(\frac{\partial \overline{M}^\lambda}{\partial t} + \frac{\widehat{v}}{a}\frac{\partial \overline{M}^\lambda}{\partial \varphi} + \widehat{\dot{\theta}}\frac{\partial \overline{M}^\lambda}{\partial \theta} \right) = -\frac{\partial}{\partial t}\left(\overline{\rho_\theta {}^* M^*}^\lambda \right) + \nabla \cdot \mathbf{IEPF} - a\cos\varphi \frac{\partial \overline{F_u}^\lambda}{\partial \theta} , \tag{11.89}$$

where the divergence of the isentropic Eliassen-Palm flux is given by

$$\nabla \cdot \mathbf{IEPF} = \frac{-1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\overline{(\rho_\theta v)^* M^*}^\lambda \cos\varphi\right] + \frac{\partial}{\partial\theta}\left[\overline{p^*\frac{\partial z^*}{\partial\lambda}}^\lambda - \overline{\left(\rho_\theta \dot{\theta}\right)^* M^*}^\lambda\right] . \quad (11.90)$$

Eq. (11.89) is essentially the same as Eq. (3.9.7a) of Andrews et al. (1987).

As already noted, the residual circulation that is defined with the TEM approach looks the same as the zonally averaged circulation as seen in isentropic coordinates, which is shown in Figure 6.3. With the isentropic system, there is no need to define a "residual" circulation, because *the true zonally averaged circulation as seen in isentropic coordinates is the residual circulation*. It vanishes for a steady state (or time average) with no heating, even when friction is present.

Consider a steady state (or time average) with no heating. Then the continuity equation (11.83) reduces to

$$\frac{\partial}{\partial\varphi}\left(\overline{\rho_\theta}^\lambda \widehat{v}\cos\varphi\right) = 0 \text{ for steady flow without heating .} \quad (11.91)$$

Since $\cos\varphi = 0$ at both poles, we conclude that

$$\overline{\rho_\theta v}^\lambda = 0 \text{ for all } \varphi \text{ for steady flow without heating ,} \quad (11.92)$$

from which it follows that

$$\widehat{v} = 0 \text{ for all } \varphi \text{ for steady flow without heating .} \quad (11.93)$$

In other words, when we see $\widehat{v} \neq 0$ in a time average, it is due to diabatic processes. This was pointed out already in Chapter 6.

Eq. (11.93) tells us that for steady flow with no heating the meridional advection term of (11.89) vanishes. Since we are talking about steady flow, the two tendency terms of (11.89) are also zero. It follows that:

$$\nabla_\theta \cdot \mathbf{IEPF} = 0 \text{ for steady flow without heating or friction .} \quad (11.94)$$

Another way of saying this is that, for steady flow in the absence of heating, the zonally averaged meridional transport of angular momentum is due only to the eddies, and is balanced by the form drag on isentropic surfaces. This beautifully simple result is pretty nearly exact. It is a statement of the Eliassen-Palm theorem.

In Chapter 4, we derived the potential vorticity equation in the form

$$\frac{\partial}{\partial t}\left(\rho_\theta Z\right) + \nabla_\theta \cdot \left[\rho_\theta \mathbf{v} Z - \mathbf{k} \times \left(\dot{\theta}\frac{\partial \mathbf{v}}{\partial \theta} + \frac{1}{\rho_\theta}\frac{\partial \mathbf{F_v}}{\partial \theta}\right)\right] = 0\,. \tag{11.95}$$

This result was derived by Haynes and McIntyre (1987; their Eq. 3.4). Applying zonal and time averages to (11.95), we find that

$$\frac{\partial}{\partial \varphi}\left\{\left[\overline{\rho_\theta v Z}^{\lambda,t} - \left(\overline{\dot{\theta}\frac{\partial u}{\partial \theta}}^{\lambda,t} + \frac{1}{\rho_\theta}\frac{partial\overline{F_u}^{\lambda,t}}{\partial \theta}\right)\right] cos\varphi\right\} = 0\,. \tag{11.96}$$

Eq. (11.96) says that quantity in curly brackets is independent of latitude. It is zero at both poles, because of the factor of $\cos\varphi$. It must therefore be zero at every latitude, which means that

$$\overline{\rho_\theta v Z}^{\lambda,t} - \left(\overline{\dot{\theta}\frac{\partial u}{\partial \theta}}^{\lambda,t} + \frac{1}{\rho_\theta}\frac{\partial \overline{F_u}^{\lambda,t}}{\partial \theta}\right) = 0 \quad \text{for all} \quad \varphi\,. \tag{11.97}$$

We can further conclude from (11.97) that

$$\text{for steady flow,} \quad \overline{\rho_\theta v Z}^{\lambda,t} = 0 \text{ for all } \varphi \text{ except where there is heating or friction}\,. \tag{11.98}$$

This says that the conditions for the eddy PV flux to vanish are the same as the conditions for the isentropic Eliassen-Palm flux to be non-divergent. Compare this with the very similar (but less exact) result (11.41), which was derived for the quasi-geostrophic case.

The Eliassen-Palm theorem is simpler, more general, and easier to interpret when expressed using isentropic coordinates. The TEM system has been compared with the isentropic system by Held and Schneider (1999), Ferrari and Plumb (2003), Plumb and Ferrari

(2005), Egger and Hoinka (2008), Juckes (2001), Pauluis et al. (2011), Noda (2010), and Hardiman et al. (2010), among others. The isentropic system provides a more direct physical interpretation of mass transport, but the TEM is better suited for use with reanalysis data or numerical model output, because they are typically provided in pressure or height coordinates. The TEM framework is conceptually and diagnostically powerful, but the isentropic system is simpler and more exact.

11.6 Segue

As discussed in Chapter 10, atmospheric eddies exist on a wide range of space and time scales. In this chapter, we discussed the interactions of the eddies with the zonally averaged flow. The eddies interact with each other too. Chapter 13 considers the global circulation as a kind of large-scale turbulence, with a broad range of scale interactions.

11.7 Problems

1. (a) Prove that

$$\overline{v_g^* Z_{QG}^*}^{\lambda} = -\frac{\partial}{\partial y}\left(\overline{u_g^* v_g^*}^{\lambda}\right) - \frac{\partial}{\partial p}\left(\frac{R f_0}{p S_p}\overline{v_g^* T^*}^{\lambda}\right) . \tag{11.99}$$

   (b) Prove that for nondivergent horizontal motion on the sphere the meridional flux of vorticity and the divergence of the meridional flux of angular momentum are related by

$$\overline{v^* \zeta^*}^{\lambda} \cos\varphi = -\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^* M^*}^{\lambda}\cos\varphi\right) . \tag{11.100}$$

2. Solve the one-dimensional steady-state version of the Charney-Devore model, as given by

$$\left(\frac{\upsilon}{K_m}\right)A + \left(U - \frac{\beta}{K_m^2}\right)B = 0 , \tag{11.101}$$

$$\left(U - \frac{\beta}{K_m^2}\right)A + \left(\frac{\upsilon}{K_m}\right)B + \left(\frac{f_0 h_T}{K_m^2 H}\right)U = 0 , \tag{11.102}$$

$$\left(\frac{f_0 h_T K_m}{4H}\right) B - \upsilon\,(U - U^*) = 0\,. \tag{11.103}$$

In order to do this you will have to choose appropriate values for the various parameters of the model. Explain your choices.

3. Show that

$$\psi_z^* = \frac{g}{f_0}\frac{T^*}{T_S}\,. \tag{11.104}$$

Chapter 12

Examples of wave-mean flow interactions

12.1 Blocking

Blocking (Berggren, 1949; Rex, 1950 a, b), is a low-frequency, mid-latitude phenomenon characterized by a nearly stationary anticyclone that persists for at least several days and sometimes up to several weeks. The anticyclone splits the westerly jet and steers cyclonic disturbances around itself, mainly on the poleward side. The flow in the vicinity of a blocking high is strongly meridional, and is an example of what the older literature calls a "low-index" regime (Rossby, 1939); a high-index regime is strongly zonal. Blocking highs tend to fluctuate in intensity, weakening briefly and then re-intensifying. They remain nearly stationary in an average sense, although they may wander around a little over their lifetimes.

Blocks tend to occur preferentially in the Northern Hemisphere, especially in the eastern North Atlantic and the eastern North Pacific in winter, and northern Asia in summer. Figure 12.1 shows the frequency of Northern Hemisphere blocking for summer and winter. It is clear that blocking is more frequent in winter. Blocks frequently occur near the eastern edges of the North Pacific and North Atlantic storm tracks. Southern Hemisphere blocking is less frequent, but does occur, most commonly near New Zealand.

The formation of a block can excite upward-propagating Rossby waves. Wintertime blocking events sometimes precede sudden stratospheric warmings Quiroz (1986); Martius et al. (2009). Figure 12.2 shows examples of a variety of blocking events in the North Atlantic region.

Blocks strongly influence weather patterns for one or more weeks at a time, sometimes in association with heat waves and/or persistent droughts (e.g., Green, 1977). If the formation and dissipation of blocks can be predicted, then it may be possible to predict some weather patterns weeks in advance. Until recently, weather-prediction models were not very successful in forecasting blocks, and the same models produced blocks less often than observed when run in climate simulation mode. Within the last few years, this situation has

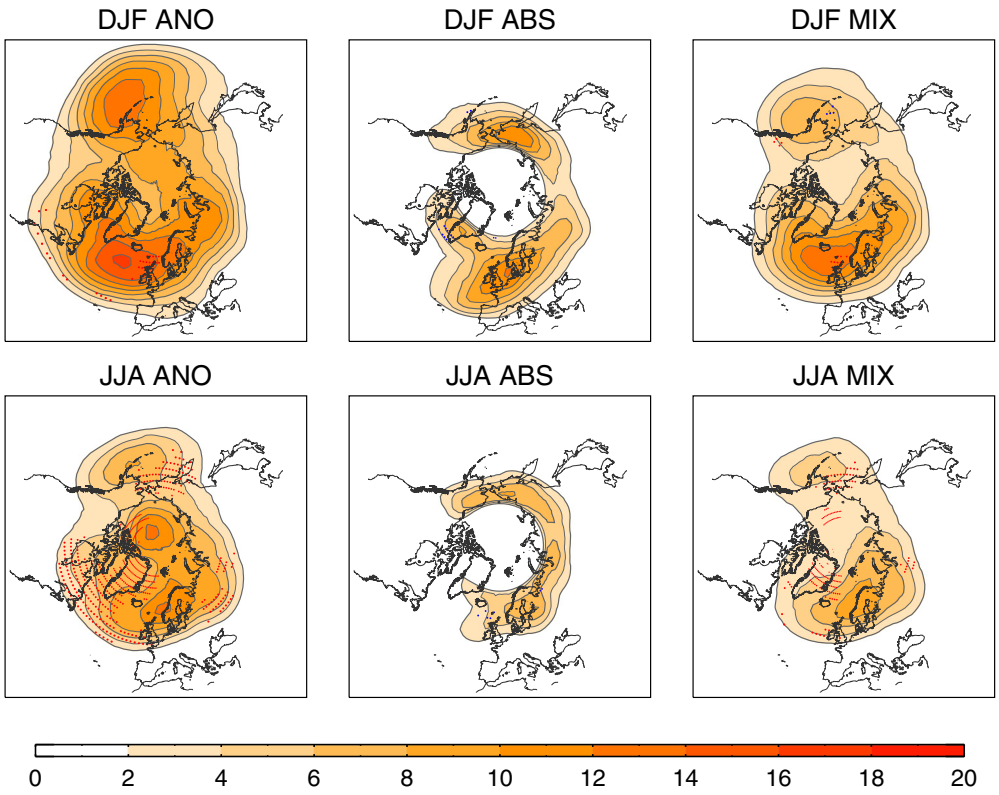


Figure 12.1: Blocking climatology. Winter (top) and summer (bottom) blocking frequency (percentage of days in the season, with 2% corresponding approximately to 2 blocked days per season) for 1958–2012. "ANO," "ABS", and "MIX" refer to three alternative methods for defining blocking events. The red and blue dots indicate regions with significant increasing/ decreasing trends in at least two of the three reanalyses. From Woollings et al. (2018).

improved dramatically, apparently as a result of increased model resolution.

We still lack a simple and widely accepted dynamical theory of blocking. What causes the formation of a blocking anticyclone? Why are blocks so persistent? What causes a block to break down?

As discussed by Colucci (1985), blocks sometimes begin when particularly intense cyclones advect air with low potential vorticity (PV) from the subtropics (where low PV is the norm) into middle latitudes (where low PV is an anomaly). An example of this is shown in Fig. 12.3, which is taken from (Shutts, 1986). At the same time the air with low PV is advected poleward, a mass of high PV air is advected into position on the equatorward side of the low-PV anomaly. The block thus has a dipole PV structure. Near the longitude of the block, the PV decreases poleward, whereas it normally increases poleward. The blocking anticyclone could be called a "cut-off high," while the cyclone is a "cut-off low." Fig. 12.4 shows the corresponding sea level pressure and 500 hPa height fields. The dipole structure

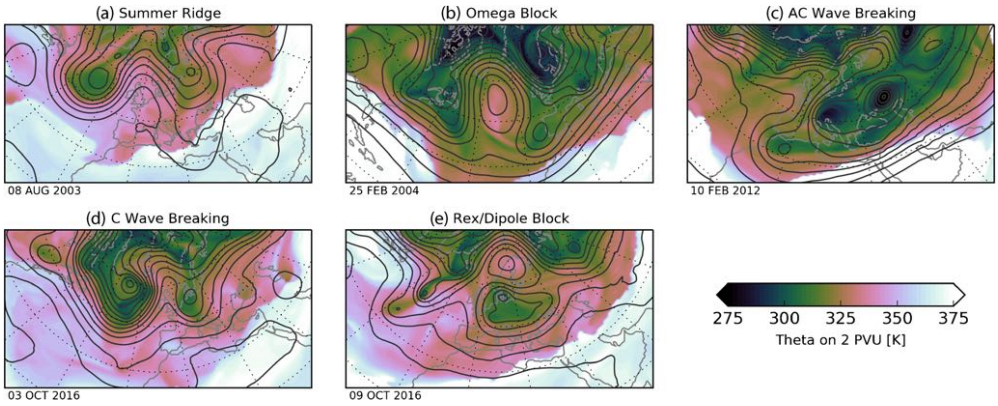


Figure 12.2: Examples of North Atlantic blocks. Snapshots of potential temperature on the dynamical tropopause, where the potential vorticity is equal to 2 PVU (color shading), and contours of the 500 hPa geopotential height (with a contour spacing of 60 m) for the dates indicated. Each example has been given a descriptive name that is commonly used for similar blocks. From Woollings et al. (2018).

is clearly evident in the latter. Note that the westerlies have split into a strong branch to the north of the high, and a weaker branch to the south of the low.

How is it possible for blocking highs to persist as well defined, isolated "objects," even while they are embedded in the turmoil of the midlatitude winter circulation? The destruction of the low seems natural; the persistence of the high demands an explanation. Hoskins et al. (1985) speculated that lows are disrupted by the convection that often accompanies them, while highs can survive because they suppress convection. Observations suggest that the smaller-scale transient eddies that are steered around a blocking high actually help to maintain the high (e.g., Hansen and Chen, 1982; Hoskins et al., 1983; Egger et al., 1986; Shutts, 1986; Dole, 1986; Mullen, 1987; Nakamura et al., 1997; Chang et al., 2002; Luo and Chen, 2006; Ren et al., 2009). Blocks may be examples of up-scale energy transport, which, as discussed in Chapter 13, is expected on theoretical grounds in two-dimensional turbulence.

The nearly stationary character of blocks suggests that they are somehow anchored to fixed geographic features such as topography, although Hu et al. (2008) reported a simulation of blocks with an "aqua-planet" model that uses zonally symmetric boundary conditions. A highly simplified theory that involves topographic anchoring was proposed by Charney and DeVore (1979, hereafter CDV), who suggested that in the presence of topography the large-scale circulation can adopt either of two equilibrium states; one of these corresponds to blocking, and the other to a more zonal flow. The basic idea is that for some configurations of the mean flow there can be stationary waves that are resonantly forced by the topography. The waves then feed back on the mean flow. Under certain conditions, the resonant waves alter the mean flow in a way that favors the formation of resonant waves, i.e., the waves create a mean flow that is favorable for their own continuing existence. On

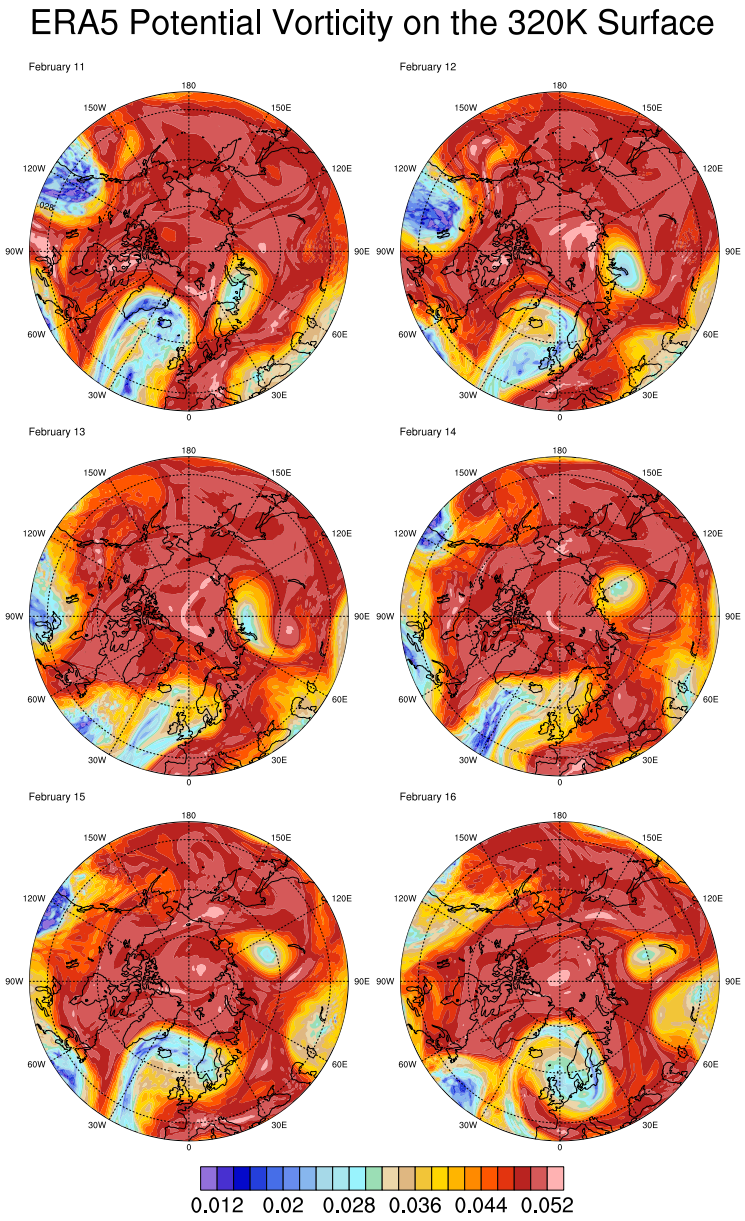


Figure 12.3: Contours of the fourth root of the Ertel potential vorticity on the 320 K isentropic surface, which slopes from about 200 hPa near the pole to about 600 hPa in the tropics. The panels are for six successive days, spanning February 11-16, 1983. The contour interval is 0.001 in SI units. The bold contour is 0.005. Based on Shutts (1986).

the other hand, if the waves are not present, the resulting mean flow is not favorable for the formation of the waves. Hence the system can exist in either of two possible configura-

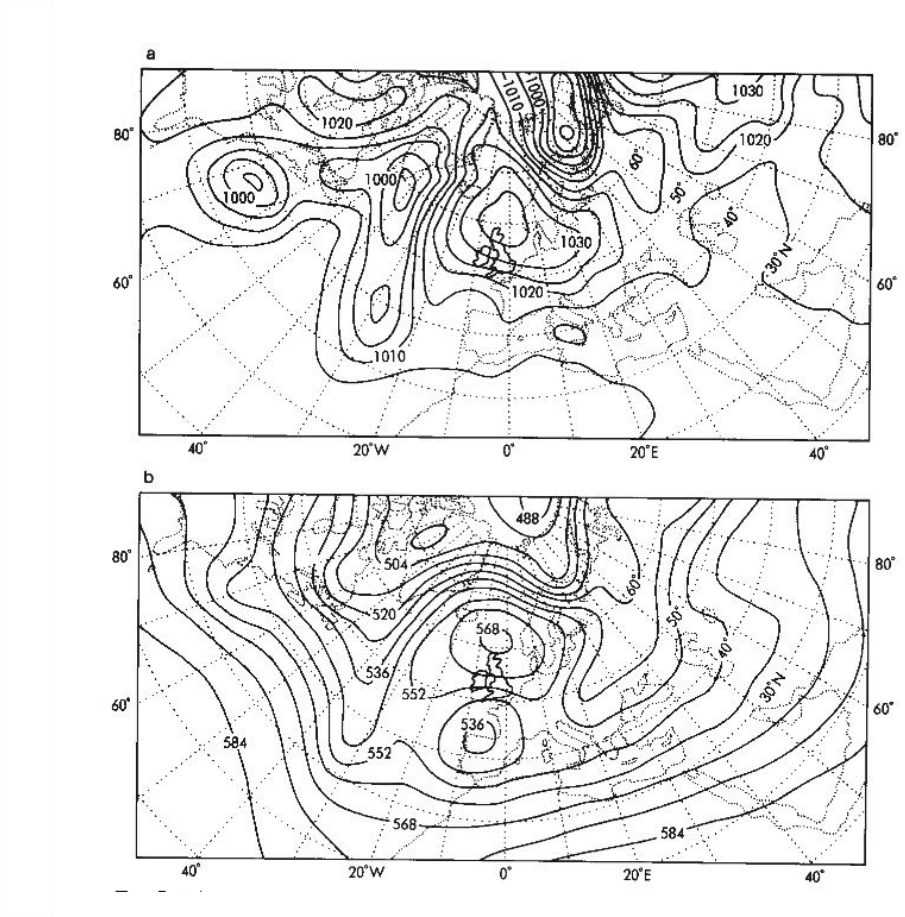


Figure 12.4: (a) Mean sea level pressure field for 15 February 1983, 12Z. Contour interval: 5 hPa. (b) Height of the 500-hPa surface for 15 February, 12Z. Contour interval: 8 dam. From Shutts (1986).

tions. These are referred to as "multiple equilibria," and the possible existence of multiple equilibria has been a key reason for continuing interest in the CDV theory. Because the theory involves the interactions of the waves with the mean flow, it is inherently nonlinear.

A simplified version of the CDV theory is as follows. The topography is assumed to vary sinusoidally in longitude, i.e.,

$$h(x) \equiv h_T \cos(K_m x) \ . \tag{12.1}$$

The wind field is described by a stream function, ψ, which is assumed to be of the form

$$\psi(x,y,t) = -U(t)\,y + A(t)\cos(K_m x) + B(t)\sin(K_m x) \ . \tag{12.2}$$

Think of this as a highly truncated functional expansion. See Fig. 12.5. Recall that the stream function is defined by the relations $u = -\partial\psi/\partial y$ and $v = \partial\psi/\partial x$; using these formulae, we see that the zonal component of the flow described by (12.2) is simply $U(t)$, i.e., it depends only on time. The "A" part of the wave is in phase with the topography, in the sense that for $A > 0$ the maximum of the stream function (corresponding to ridging behavior) occurs over the mountain. The "B" part of the wave represents a trough downstream of the mountain. The wave-like meridional component of the flow varies with both x and t:

$$v(x,\, t) = K_m \left[-A(t)\sin(K_m x) + B(t)\cos(K_m x)\right] . \tag{12.3}$$

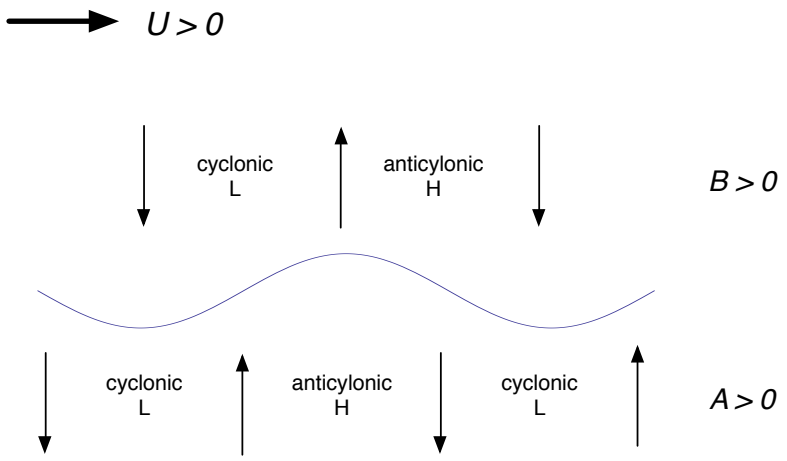


Figure 12.5: The relationships between the and components of the meridional wind and the topography in the model of Charney and Devore. The topography is indicated by the wavy line in the center. The arrows indicate the direction of the v-wind component, with "up" corresponding to "southerly." The arrows above the topography are for the -component of the wave, and the arrows below are for the A-component, assuming that A and B are positive. The indicated regions are cyclonic or anticyclonic for the case of $f > 0$ (the Northern Hemisphere).

By substituting (12.2) into a nonlinear vorticity equation describing a Rossby wave flow with friction, CDV showed that the wave motion satisfies

$$\frac{1}{K_m}\frac{dA}{dt} + \left(\frac{\upsilon}{K_m}\right)A + \left(U - \frac{\beta}{K_m^2}\right)B = 0 , \tag{12.4}$$

$$\frac{1}{K_m}\frac{dB}{dt} - \left(U - \frac{\beta}{K_m^2}\right)A + \left(\frac{\upsilon}{K_m}\right)B + \left(\frac{f_0 h_T}{K_m^2 H}\right)U = 0\,, \tag{12.5}$$

while the mean zonal flow obeys

$$\frac{dU}{dt} = \left(\frac{f_0 h_T K_m}{4H}\right)B - \upsilon\,(U - U^*)\ . \tag{12.6}$$

In (12.4) - (12.5), the terms involving υ represent friction. In (12.6), the "B" term represents the orographic form drag (i.e., mountain torque) that the topography exerts on the mean flow when the wave has its trough over the mountain, and the U^* term represents a "momentum forcing" that maintains the mean flow against friction. Note that (12.4) and (12.5) "blow up" if the wave number of the topography, K_m, is equal to zero. This simply means that there is no topographically forced wave in the absence of topography. Also note that (12.4) and (12.5) are nonlinear, because they involve the products of A and B with U. This nonlinearity represents the wave-mean flow interactions.

CDV considered equilibrium (steady) solutions of (12.4) - (12.6). These equilibria can be found by setting the time-rate of change terms to zero, solving the resulting linear system (12.4) - (12.5) for A and B as functions of U, and selecting the appropriate value of U by requiring that the steady-state version of (12.6) is also satisfied. Fig. 12.6 shows an example, in which the straight line represents the (B,U) pairs that satisfy (12.6), while the peaked line represents the (B,U) pairs that satisfy (12.4) - (12.5). There are three equilibria, but it can be shown that the middle one is unstable. The stable equilibrium with large U has a small wave amplitude, and the stable equilibrium with small U has a large wave amplitude. This is understandable because the wave term of (12.6) represents a drag on U, due to mountain torque. The solution with large wave amplitude and a weak zonal flow is interpreted as representing blocking. CDV suggested that if the model is subjected to stochastic forcing, representing the random fluctuations of the weather, perhaps manifested through fluctuations of U^*, then the system can undergo occasional transitions back and forth between the neighborhood of the "blocked" equilibrium and the neighborhood of the "unblocked" equilibrium.

Charney and Straus (1980) extended the CDV theory to the baroclinic case, and it has been studied by many other authors. The CDV theory has been heavily criticized (e.g., Tung and Rosenthal, 1985), in part because its extreme idealizations limit the possible behaviors of the model, suggesting that the small number of discrete equilibria (i.e., two) is an artifact. Nevertheless the theory continues to be cited frequently as an important background concept for the interpretation of blocking, and also in research on the possible existence of multiple discrete weather "regimes" (e.g., O'Kane et al., 2013).

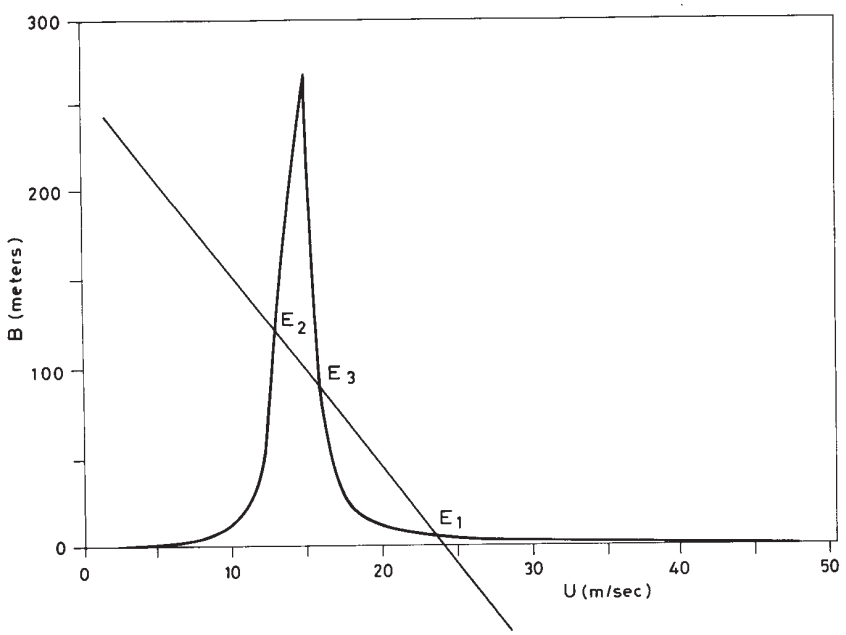


Figure 12.6: From Speranza (1986). Equilibrium solutions of (12.4) - (12.5) , found by setting the time-rate of change terms to zero, solving the resulting linear system for A and B as functions of U, and selecting the appropriate value of U by requiring that (12.6) is also satisfied. The straight line represents the (B,U) pairs that satisfy (12.6), while the peaked line represents the (B,U) pairs that satisfy (12.4)-(12.5). Equilibrium solutions occur where the two curves intersect. The figure shows that there are three solutions, but it turns out that the middle one is unstable.

McWilliams (1980) suggested that "modons" could be considered as idealized models of blocks. Modons are exact solutions of the nonlinear vorticity equation (Flierl, 1978). They have dipole structures, in which a high (a negative vorticity center) is paired with a neighboring low (a positive vorticity center). Modons must have finite amplitudes in order to exist; there is no such thing as a "linear" modon. An interesting and block-like property of modons is that they are resistant to disruption by perturbations. In addition, a modon "translates" relative to the mean flow, and it is possible to set up a modon that is stationary relative to the Earth, in the presence of background westerlies. The conditions for this are special, however, and it is not clear why such special conditions should occur in real cases.

Nakamura and Huang (2018) proposed a novel theory of blocking, based on an analogy between intense wave activity in the westerlies and heavy traffic on a highway. When traffic is light, an increase in the number of cars leads to an increase in the rate of traffic flow (cars per unit time past a fixed location), but if the number of cars continues to increase a traffic jam develops and the rate of traffic flow actually decreases. In short, the traffic flow is a non-montonic function of the number of cars.

Nakamura and Huang (2018) defined a vertically averaged measure of local tropospheric wave activity called $\langle A \rangle$. In their analogy, $\langle A \rangle$ corresponds to the number of cars.

Figure 12.7, which is taken from their paper, shows that $\langle A \rangle$ is negatively correlated with the strength of the westerlies in regions where blocks commonly occur. Nakamura and Huang (2018) also defined a measure of the zonal flux of wave activity, which is analogous to the rate of traffic flow. Figure 12.8 shows that the flux of wave activity increases with $\langle A \rangle$ when $\langle A \rangle$ is small, but decreases with further increases of $\langle A \rangle$ when $\langle A \rangle$ becomes large. Motivated by the traffic-jam analogy, Nakamura and Huang (2018) proposed a simple but ad hoc nonlinear differential equation that can qualitatively simulate the life cycles of blocking events.

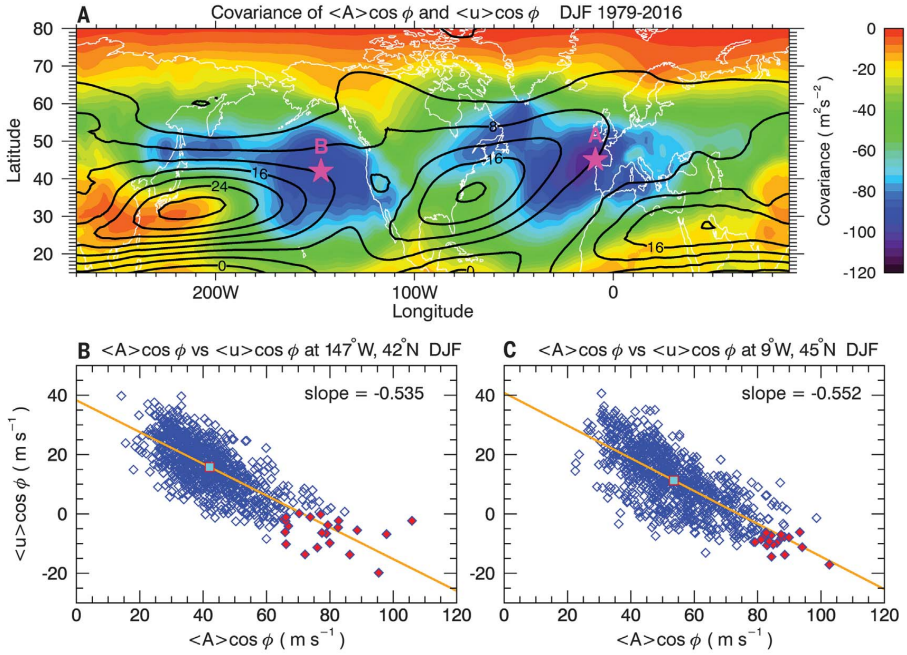


Figure 12.7: This figure shows that the wave activity $\langle A \rangle$ and the strength of the westerlies are negatively correlated. In panel (A), the color shading depicts the covariance between $\langle A \cos\varphi \rangle$ and the vertically averaged westerlies $\langle u \cos\varphi \rangle$ in the boreal winter, and the contours show the DJF climatology of $\langle u \cos\varphi \rangle$ expressed in meters per second. The locations of peak covariance are labeled A (9°W, 45°N) and B (147°W, 42°N). Data are from the ERA-Interim reanalysis for 1979 to 2016. (B) Scatter diagram of 4-day average $\langle A \cos\varphi \rangle$ versus 4-day average $\langle u \cos\varphi \rangle$ for the same period at location B. The cyan box represents the DJF climatological mean. The red diamonds represent periods in which $\langle A \cos\varphi \rangle$ and $\langle u \cos\varphi \rangle$ are simultaneously in the top and bottom fifth percentiles, respectively. The orange line is a least-squares fit whose slope is indicated. (C) Same as (B) but for location A. From Nakamura and Huang (2018).

To close this section, we list questions about blocking that have not yet been fully resolved:

- What causes the formation of a blocking anticyclone?
- What determines the preferred geographical locations of blocking activity?
- How are blocking highs maintained against the noisy background flow?

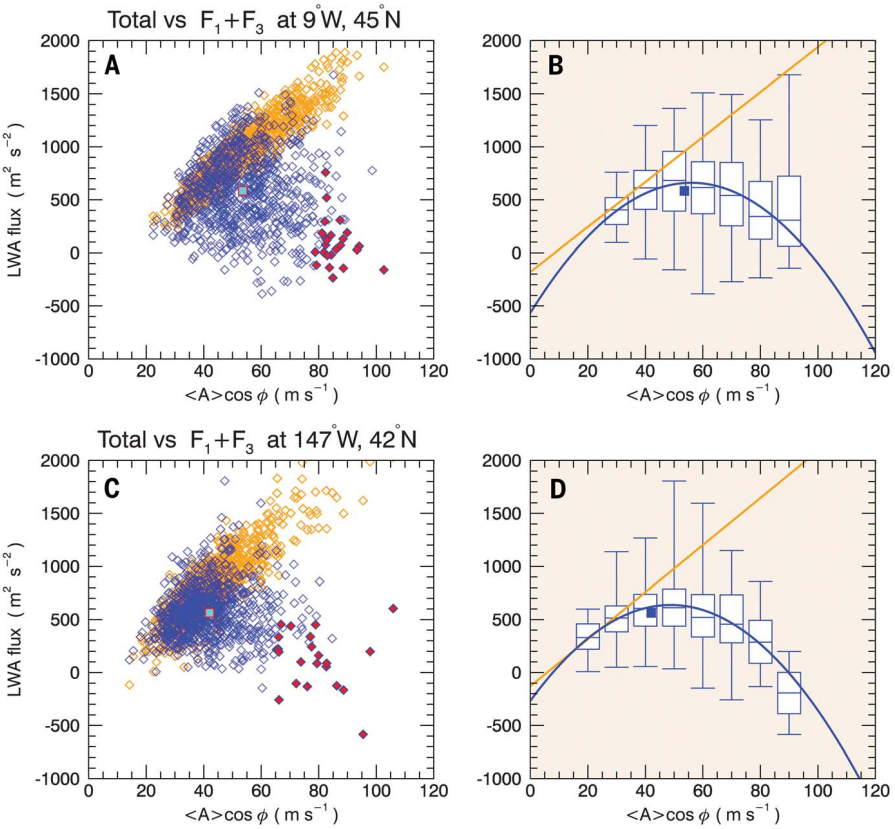


Figure 12.8: The wave activity and its zonal flux at the jet exit. Relationships between $\langle A\cos\varphi\rangle$ (abscissa) and the flux of wave actvity (ordinate). (A and B) 9°W, 45°N (location A). (C and D) 147°W, 42°N (location B). Panels (A) and (C) show four-day averages of two different measures of the wave activity flux, represented by the orange and blue dots. They are plotted against the 4-day average of $\langle A\cos\varphi\rangle$, sampled based on data for DJF of the years 1979 to 2016. The red diamonds correspond to those in panels B and C of 12.7. Panels (B) and (D) show diagrams based on panels (A) and (C) using a least-squares fit (orange), a chi-square fit with a parabolic curve (blue), and quartile plots. The quartile plots show maximum, minimum, and median values, together with the upper and lower quartiles of the wave activity flux, for 10-m s^{-1} intervals of $\langle A\cos\varphi\rangle$. The cyan squares in (A) and (C) and blue squares in (B) and (D) indicate the DJF climatological mean values. From Nakamura and Huang (2018).

- Why are blocks nearly stationary even though they are embedded in strong westerly currents?
- What causes the breakdown of a block?
- Why do we observe persistent, quasi-stationary anticyclones but not persistent, quasi-stationary cyclones?

The discussion given above shows that we have at least partial answers to some of these questions.

12.2 Sudden stratospheric warmings

Sudden stratospheric warmings (SSWs) were discovered by Richard Scherhag (Scherhag, 1952, 1960). Early reviews were published by Schoeberl (1978) and Holton (1980), and a more recent review has been published by Baldwin et al. (2021). In an SSW, polar stratospheric temperatures increase by tens of Kelvins in just a few days. As would be expected from thermal wind balance, the usually strong polar stratospheric westerlies (Chapter 3) dramatically weaken and sometimes even give way to easterlies. As shown in Fig. 12.9, SSWs occur in two different ways. Sometimes the polar vortex simply shifts away from the pole, but in other cases it splits into two parts (Charlton and Polvani, 2007). The polar vortex is re-established when a positive vorticity anomaly moves back into position over the pole.

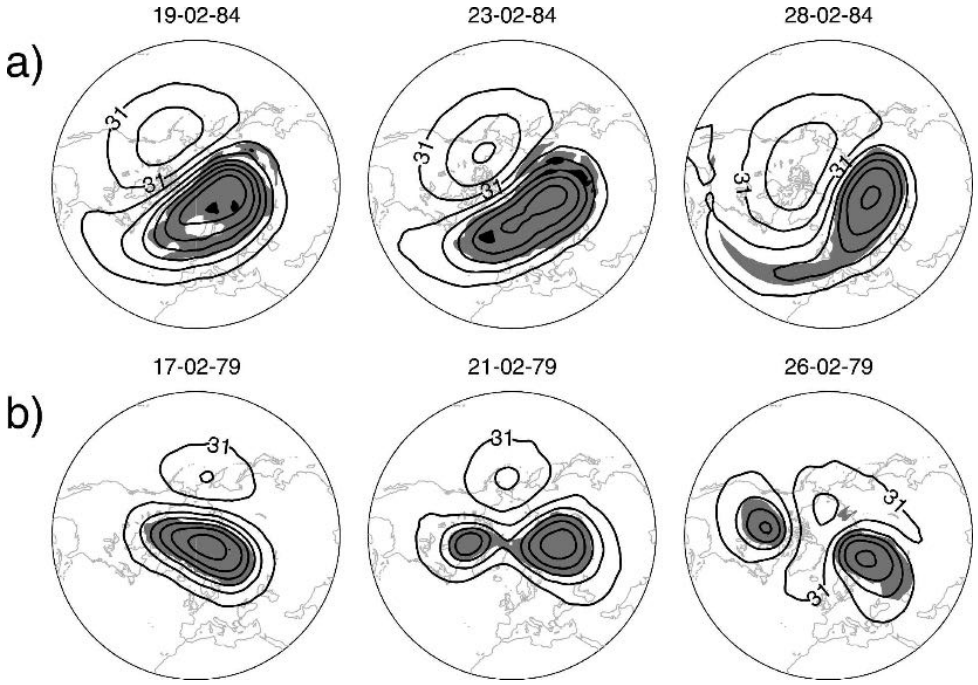


Figure 12.9: Polar stereographic plot of geopotential height (contours) on the 10-hPa pressure surface. Contour interval is 0.4 km, and shading shows potential vorticity greater than 4.0 PVUs. Recall from Chapter 4 that a PVU is 10^{-6} m^2 K s^{-1}. (a) A vortex displacement type warming that occurred in February 1984. (b) A vortex splitting type warming that occurred in February 1979. From Charlton and Polvani (2007). © American Meteorological Society. Used with permission.

Major SSWs occur a bit more often than once every other year in the Northern Hemisphere, usually in January or February. They occur much less often in the Southern Hemisphere, although there was a spectacular south polar event in September of 2002 (e.g., Newman and Nash, 2005; Simmons et al., 2005).

As first proposed by Matsuno (1970, 1971) and Matsuno and Nakamura (1979), SSWs are triggered by the rapid growth of quasi-stationary tropospheric planetary waves, which propagate upward into the stratosphere. Recall from Chapter 10 that planetary waves can propagate from the troposphere into the stratosphere in winter but not in summer. The rapid intensification of wave activity is often associated with the formation of a blocking high in the troposphere (e.g., Barriopedro and Calvo, 2014).

SSWs are far from "steady," so non-interaction theorems do not apply. Fig. 12.10 shows the life cycle of a composite SSW, constructed from the observations by averaging many SSWs together. The polar stratosphere is observed to warm up, while the lower latitudes cool off. As the polar warming progresses, the stratospheric meridional temperature gradient actually reverses, causing the usually strong westerly polar vortex to give way to polar easterlies! When easterlies form above, further wave propagation is blocked. The deceleration of the polar westerlies is observed to coincide with a strong convergence of the Eliassen-Palm vectors. The vectors are nearly vertical in the stratosphere, indicating a dominance of the meridional energy flux. As the SSW matures, the warm anomaly and the easterlies migrate downward towards the troposphere. It has recently been recognized that SSWs are often followed by persistent changes in tropospheric weather (Baldwin and Dunkerton, 2001; Thompson et al., 2005).

12.3 The quasi-biennial oscillation

As discussed by Baldwin et al. (2001), the zonal winds of the tropical stratosphere undergo an amazing "Quasi-Biennial Oscillation" (QBO), with an average period of about 28 months. A "ring" of air in the tropical stratosphere, extending over all longitudes, is observed to reverse the direction of its zonal motion roughly every 14 months, like a giant, zonally oriented equatorial Ferris Wheel, as shown in Fig. 12.11. The reversal propagates downward through the stratosphere, and disappears near the tropopause.

The QBO was discovered by Richard Reed in about 1960 (see the early review by Reed, 1966). At first, the evidence was not sufficient to establish the existence of a true oscillation in a statistically significant fashion, but additional decades of data have made it clear that a quasi-periodic oscillation really does exist. The oscillation is strong; the winds shift from about 20 m s^{-1} westerly to 20 m s^{-1} easterly, and back again. They are observed to propagate down from the middle stratosphere to near the tropopause. Corresponding oscillations are seen in other stratospheric fields, and much more weakly in the troposphere.

Lindzen and Holton (1968) and Holton and Lindzen (1972) proposed that the oscillation is due to the interactions of upward-propagating Kelvin and Yanai waves with the mean flow. More recently, it has been suggested that eastward- and westward-propagating gravity waves play an important role, and the Kelvin and Yanai waves are now being deemphasized. The various waves are observed to produce energy propagation upward into the stratosphere. The source of wave energy must, therefore, be in the troposphere, and is

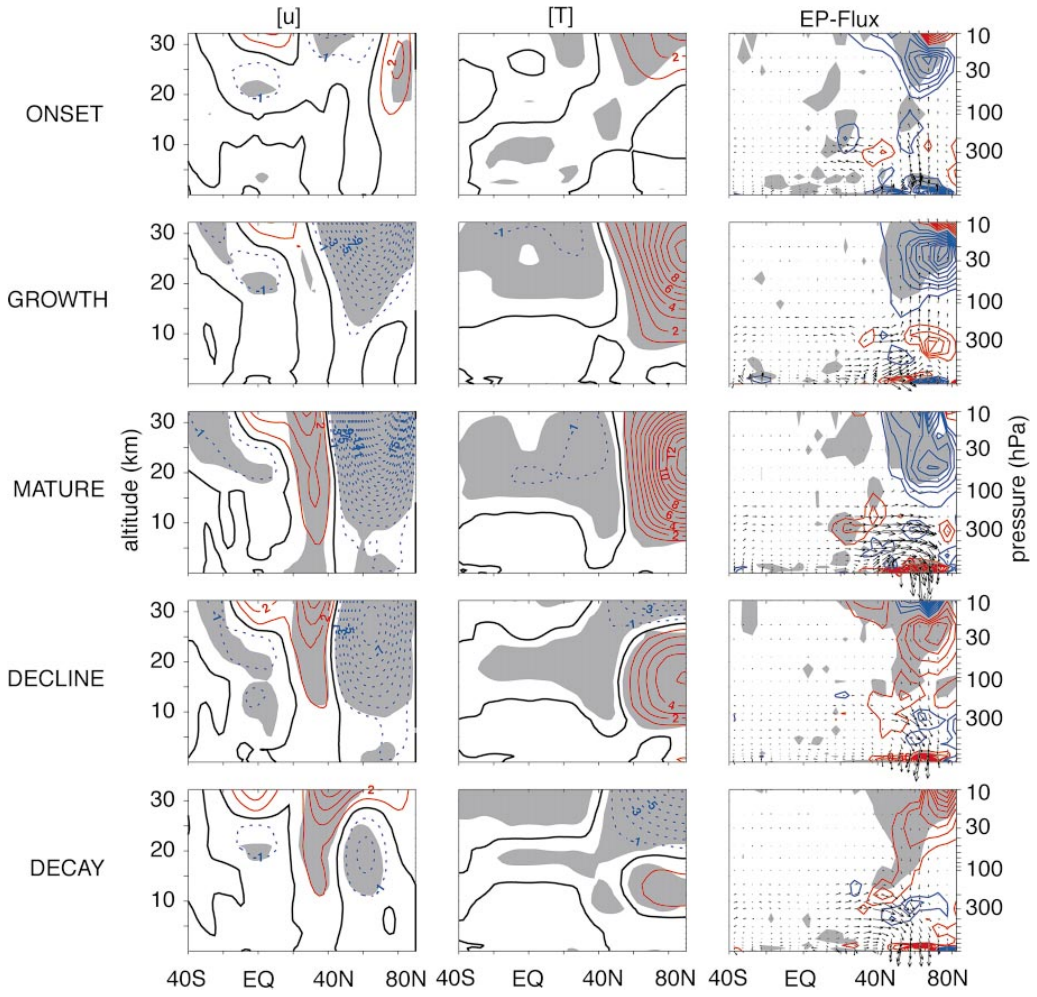


Figure 12.10: (left) Anomalous zonal-mean zonal wind, (middle) zonal-mean temperature, and (right) EP flux with its divergence during the composite life cycle of SSWs. Negative contours are given as blue dashes. Zero contours are given as bold solid lines. The wind (temperature) contour interval is 1 m s^{-1} (1 K). The EP flux divergence (divided by $\rho a \cos\varphi$) is contoured in the right column at every 0.25 m s^{-1} day^{-1}. For all fields, positive values are shown in red and negative values in blue. The vector lengths in the right column are referenced with respect to the top figure in the column. Gray shading indicates areas with a 95% confidence level. Adapted from Limpasuvan et al. (2004). © American Meteorological Society. Used with permission.

believed to be associated with latent heating. Each type of wave can be blocked by a critical level where $\overline{u}^{\lambda} - c = 0$. Eastward-propagating waves, such as Kelvin waves and eastward-propagating inertia-gravity (EIG) waves, can have critical levels in westerlies. Westward-propagating waves, such as Rossby waves, Yanai waves and westward-propagating inertia-gravity (WIG) waves, can have critical levels in easterlies.

Recall from (11.54) that for eastward-propagating neutral waves (with $c > 0$) the momentum flux and the energy flux have the same sign, while for westward-propagating neu-

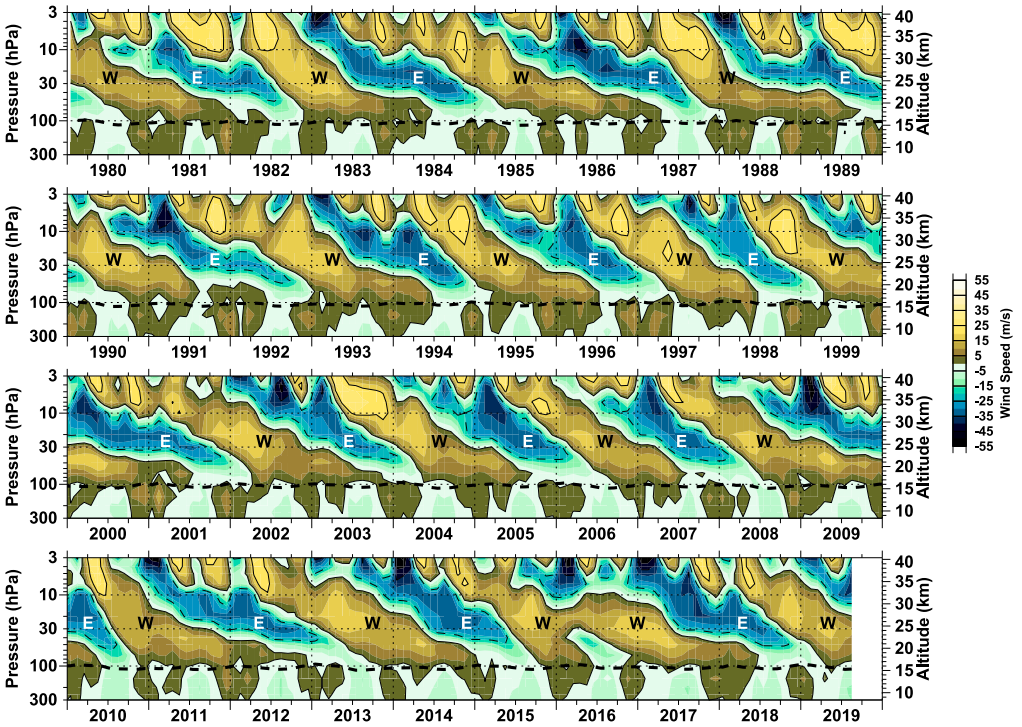


Figure 12.11: Height–time section of the monthly averaged, zonally averaged zonal wind observed at the equator showing the QBO. Data are from the NASA Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). The dashed line shows the mean altitude of the tropopause. From Baldwin et al. (2019).

tral waves (with $c < 0$) the momentum and energy fluxes have opposite signs. In particular, Kelvin waves that produce an upward energy flux transport westerly momentum upward from the level where they are generated, and deposit it aloft, producing a westerly acceleration when they encounter a critical level, in a layer of westerlies. This causes the westerlies to descend with time, as observed in the QBO. Eastward-propagating inertia-gravity waves will produce a similar effect.

Westward-propagating Yanai waves produce an upward energy flux transport easterly momentum from the level where they are generated, and deposit it aloft. They produce an easterly acceleration when they encounter a critical level, in a layer of easterlies. Westward-propagating inertia-gravity waves will produce a similar effect, and recent work suggests that they are in fact important for the QBO.

In the tropical atmosphere, eastward- and westward-propagating waves coexist. The eastward propagating waves, such as the Kelvin wave, can happily propagate through easterlies, but will cause the westerlies to descend when the waves encounter a critical level. The westward propagating waves, such as the Yanai waves, can happily propagate through westerlies, but will cause the easterlies to descend when the waves encounter a critical level. See Fig. 12.12. This is how the QBO works.

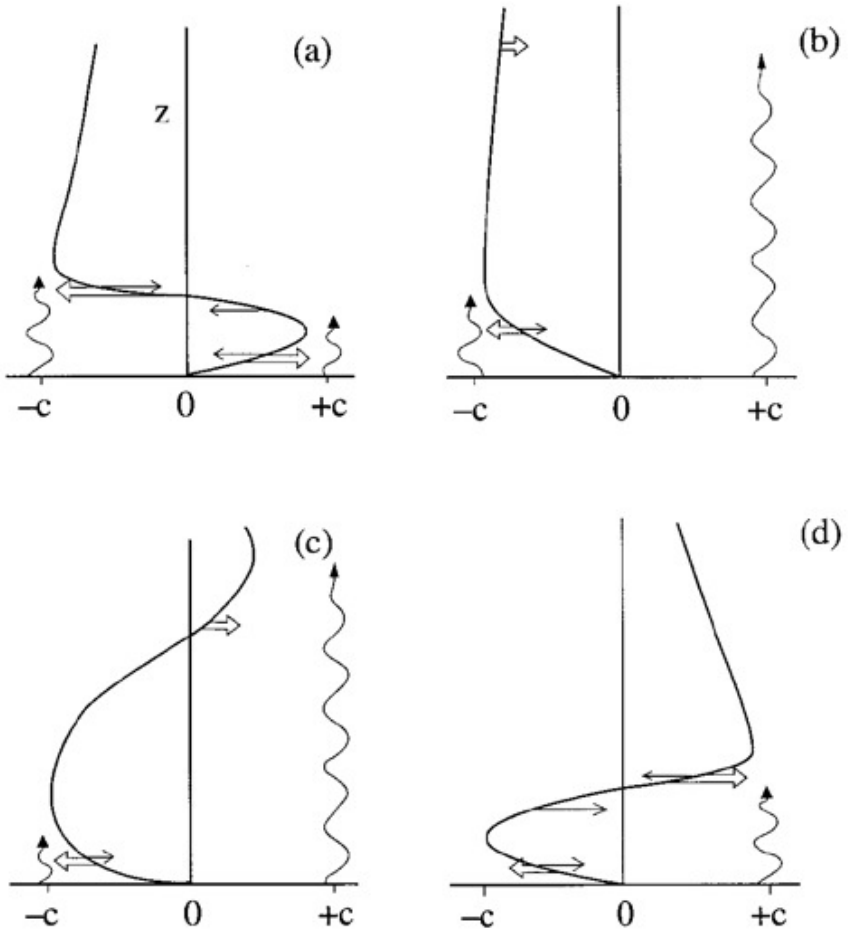


Figure 12.12: In each panel, the vertical axis is height, and the horizontal axis is wind speed and/or phase speed. The curves show the zonal wind as a function of height. The labels +c and -c denote eastward and westward phase speeds, respectively. From Plumb (1984).

You may find this explanation hard to believe. That was a common reaction when Lindzen and Holton (1968) and Holton and Lindzen (1972) first proposed it. Most such doubts were extinguished when Plumb et al. (1978) produced a remarkable laboratory simulation of the QBO in an annular tank full of stratified salt water. In their experiment, "eastward" and "westward" propagating internal gravity waves are artificially excited by an oscillating diaphragm at the bottom of the tank. At a given level, the direction of the mean flow reverses periodically with time, and these reversals propagate downward. The oscillations are caused by wave-mean flow interactions. Videos can be found here. A "direct" numerical simulation (no parameterizations) of the laboratory experiment was reported by Wedi and Smolarkiewicz (2006).

For many years, GCMs failed to simulate the QBO. Cariolle et al. (1993) produced a somewhat promising simulation, and Takahashi and Shiobara (1995) simuilated a QBO in

a simplified GCM. Finally, M. Takahashi (1996) produced reasonably successful simulations of the QBO using a full GCM that included both the troposphere and stratosphere. During 1997, the ECMWF model began for the first time to produce a QBO when run in climate-simulation mode, i.e., without data assimilation. This improvement in the model's performance was associated with an increase in the vertical resolution. It appears that high vertical resolution and possibly also weak damping are needed for a successful numerical simulation of the QBO.

There is some evidence for phenomena similar to the QBO in the atmospheres of Jupiter (Orton et al., 1991, 1994; Leovy et al., 1991; Friedson, 1999; Flasar et al., 2004) and Saturn (Fouchet et al., 2008).

12.4 The semi-annual oscillation

The semi-annual oscillation (SAO) is a periodic phenomenon observed in the equatorial middle atmosphere, characterized by alternating easterly and westerly zonal winds. The period of the SAO is approximately six months. It was first identified and studied in the middle atmosphere by researchers such as Reed (1962) and later expanded upon by Garcia et al. (1997). The SAO is a strong mode of wind and temperature variability in the tropical middle atmosphere, , particularly between the stratopause ($\tilde{5}$0 km altitude) and the upper mesosphere ($\tilde{9}$0 km altitude) – mostly above the QBO. The SAO features a westerly phase during equinoxes and an easterly phase during solstices. The westerly phase is primarily driven by Kelvin waves and gravity waves that carry westerly momentum upward. The easterly phase is associated with cross-equatorial advection of angular momentum from the summer hemisphere to the winter hemisphere, influenced by the Brewer–Dobson circulation (see below). The oscillating zonal winds induce secondary meridional circulations that maintain thermal wind balance.

Further discussion is given by Smith et al. (2017), Shangguan and Wang (2022), and Butchart (2022).

12.5 The Brewer-Dobson Circulation

The Brewer-Dobson circulation involves a slow zonally averaged poleward drift of the air in the winter stratosphere. It was discovered by Newell (1963), and is discussed in the review by Butchart (2014). It is our final, rather sedate example of an interaction of the eddies with the zonally averaged flow.

Fig. 12.13 shows the isentropic stream function for the mass circulation, plotted from the surface up to 800 K, which is in the upper stratosphere. The figure shows rising motion in the Equatorial lower stratosphere, where radiative heating leads to an upward increase of the potential temperature. The upward-moving current diverges toward the winter pole,

where it converges to feed slow subsidence. Radiative cooling causes a downward decrease of the potential temperature there.

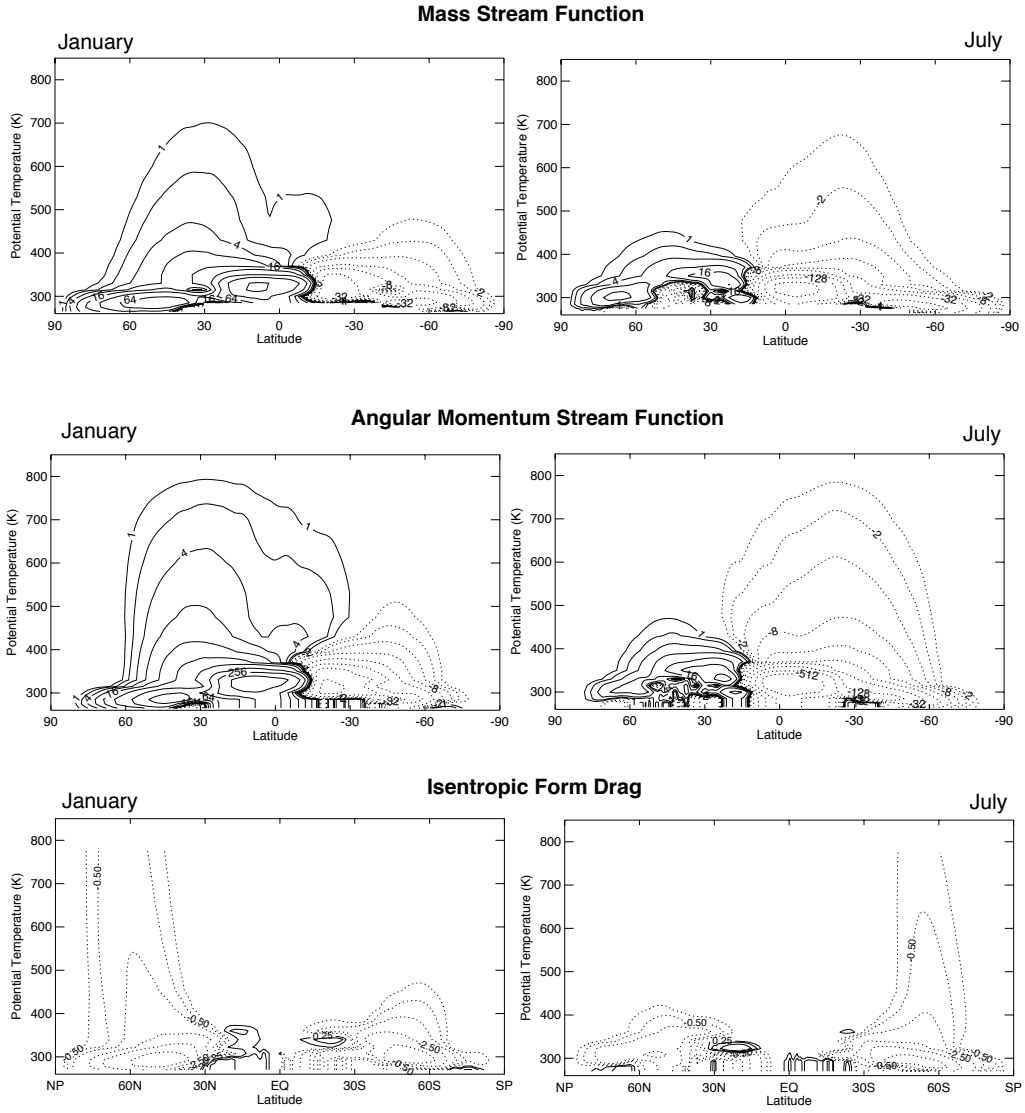


Figure 12.13: Top two panels: The isentropic mass stream function, plotted up to the 800 K level, for January and July. The contours are logarithmically spaced, in units of 10^9 kg s^{-1}. Center panels: The stream function of the total angular momentum transport. The contours are logarithmically spaced, in units of Hadleys. Bottom panels: The downward momentum flux due to isentropic form drag. The contour interval is 10^5 Nt m^{-1}.

The middle two panels of the figure show the stream function of the isentropic angular momentum transport. In the stratosphere, the contours of the angular momentum stream function resemble those of the mass stream function. Angular momentum is carried poleward to about 60° of latitude in the winter hemisphere, where the meridional flux sharply

converges. The bottom two panels of the figure show that this convergence is compensated for by a strong downward flow of angular momentum due to isentropic form drag.

In effect, we have told the story backward. The real starting point is the downward momentum flux due to isentropic form drag in middle and high latitudes of the winter hemisphere stratosphere. The tropospheric continuation of this downward momentum transport was already discussed in Chapter 6. The meridional mass flow of the Brewer-Dobson circulation is "induced" by the wave drag; it is required so that poleward angular momentum transport by the zonally averaged circulation can compensate for the downward angular momentum transport by the waves.

12.6 Segue

12.7 Problems

Chapter 13

The global circulation as turbulence

13.1 The nature of turbulence

Turbulence is not easy to define. A turbulent patch of fluid contains many interacting vortices of different scales, created through a chain of inflectional instabilities. To understand what this means, consider the dynamics of small-scale vortices in three dimensions. The relevant momentum equation is the *Boussinesq* form, which can be written as

$$\frac{\partial \mathbf{V}}{\partial t} = \mathbf{V} \times \boldsymbol{\omega} - \nabla K - \nabla \left(\frac{\delta p}{\rho_0} \right) + g \frac{\delta \theta}{\theta_0} \mathbf{k} + \nu \nabla^2 \mathbf{V} \, . \tag{13.1}$$

Here $\mathbf{V}$ is the three-dimensional wind vector; $\boldsymbol{\omega}$ is the three-dimensional vorticity vector; $K \equiv \mathbf{V} \cdot \mathbf{V}/2$ is the kinetic energy per unit mass; ρ_0, p_0, and θ_0 are reference values of the density and potential temperature, respectively; δp is the departure of the pressure from a reference profile that depends only on height; $\delta \theta \equiv \theta - \theta_0$; $\mathbf{k}$ is a unit vector pointing upwards; and ν is the (molecular) kinematic viscosity. The term $g(\delta \theta / \theta_0)\mathbf{k}$ represents the effects of buoyancy. In writing (13.1), we have used the Boussinesq approximation, and we have neglected the effects of rotation. The corresponding Boussinesq form of the continuity equation is

$$\nabla \cdot \mathbf{V} = 0 \, . \tag{13.2}$$

Taking (13.1), and using (13.2), we obtain a diagnostic equation for δp:

$$\nabla^2 \left(\frac{\delta p}{\rho_0} \right) = \nabla \cdot \left[\mathbf{V} \times \boldsymbol{\omega} - \nabla K + g \frac{\delta \theta}{\theta_0} \mathbf{k} + \nu \nabla^2 \mathbf{V} \right] \, . \tag{13.3}$$

This shows that δp is entirely determined by the wind and temperature fields. It plays only a passive role. For this reason, it is useful to eliminate δp by taking $\nabla\times$ (13.1), and using (13.2). The result is

$$\frac{\partial \boldsymbol{\omega}}{\partial t} = -\left(\mathbf{V}\cdot\nabla\right)\boldsymbol{\omega} + \left(\boldsymbol{\omega}\cdot\nabla\right)\mathbf{V} + \nabla\times\left(g\frac{\delta\theta}{\theta_0}\mathbf{k}\right) + \nu\nabla^2\boldsymbol{\omega}\ . \tag{13.4}$$

This is the three-dimensional vector vorticity equation. The buoyancy term of (13.4) acts on the part of the vorticity vector that lies in the horizontal plane.

Let Ω be a vorticity scale, and let L be the diameter of the largest vortices. When the Reynolds number $varOmegaL^2/\nu$ is sufficiently large, the flow described by (13.4) becomes turbulent, due to inflectional instability. The mechanism is illustrated in Fig. 13.1. The upper panel of the figure shows a "balanced" vortex sheet that is considered to extend to infinity in both directions. The sheet is said to be balanced because each vortex is advected up by its neighbor to the left, and down by its neighbor to the right, so that no net vertical motion occurs and the vortices do not move. Suppose, however, that one vortex is perturbed upward,as in the middle panel. It will be carried to the left by the combined effects of the unperturbed vortices. After being displaced to the left, it experiences a net upward advection, away from the sheet. This means that the initial upward perturbation has been amplified, which implies that the balanced sheet is unstable. The tilted lines in the bottom panel show a new shear zone, which is also unstable. After some time, the flow becomes highly disordered. This is the mechanism that leads to turbulence. It can work in either two or three dimensions.

The simplified discussion of inflectional instability given above refers only to vorticity advection, i.e., to the $-\left(\mathbf{V}\cdot\nabla\right)\boldsymbol{\omega}$ term of (13.4), which is of course nonlinear. Vorticity advection is the essential process that gives rise to inflectional instability, and it is the essential process that gives rise to turbulence.

The buoyancy term of (13.4) can make a strong contribution to vorticity production, and so you may wonder whether it represents a second mechanism for the production of turbulence. The answer is "not really." The buoyancy term of (13.4) is linear; as discussed later, turbulence itself is intrinsically nonlinear. Buoyancy acts as an indirect source of turbulence, but not as a direct source. Buoyancy generates highly organized coherent structures that contain regions of strong shear, e.g., the vortex rings associated with thermals. The shear in these buoyancy-driven structures sets the stage for inflectional instability. In other words, buoyancy creates conditions in which turbulence can develop through inflectional instability, but buoyancy itself does not actually produce the turbulence.

The $\left(\boldsymbol{\omega}\cdot\nabla\right)\mathbf{V}$ term of (13.4), which comes from the advection term of (13.1), was

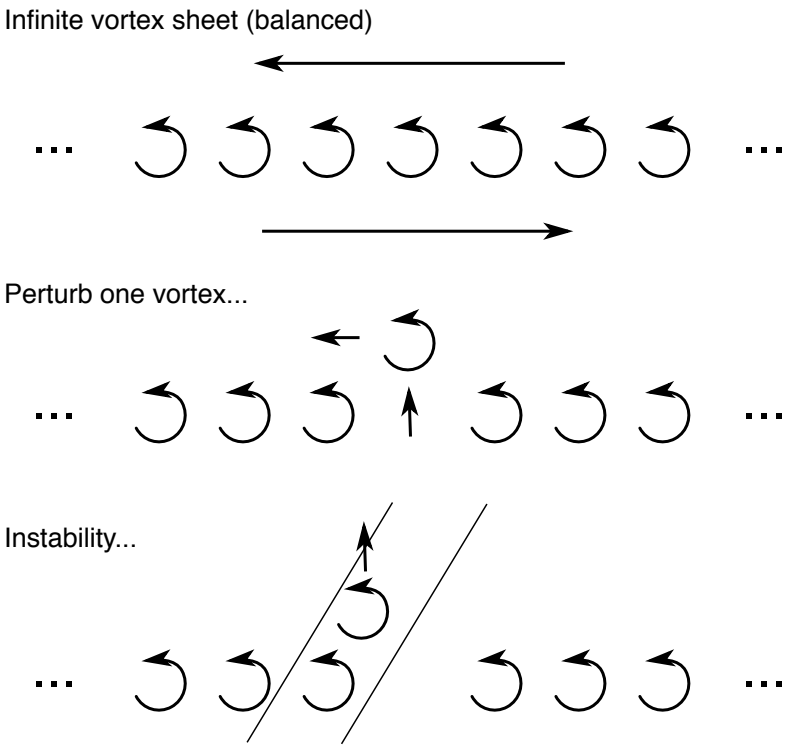


Figure 13.1: The basic mechanism of inflectional instability. If a balanced vortex sheet is perturbed, the perturbation is amplified by vorticity advection.

discussed in Chapter 4. It represents both stretching and twisting, neither of which can happen in two-dimensional flows. It is not essential for the generation of turbulence. It does, however, very powerfully *influence* three-dimensional turbulence. To see why, it is usesful to consider the enstrophy, which is the square of the magnitude of the vorticity vector (or one half of that, if you prefer). Integrated over a volume, the enstrophy is a measure of the strength of the vorticity, just as the kinetic energy is a measure of the strength of the wind. To derive an equation for the time change of the enstrophy, we form the dot product of (13.4) with the vorticity vector, :

$$\begin{aligned}\frac{\partial}{\partial t}\left(\frac{|\boldsymbol{\omega}|^2}{2}\right) = &-(\mathbf{V}\cdot\nabla)\left(\frac{|\boldsymbol{\omega}|^2}{2}\right)+|\boldsymbol{\omega}|^2\frac{\partial V_s}{\partial s}+\boldsymbol{\omega}\cdot\left[\nabla\times\left(g\frac{\delta\theta}{\theta_0}\mathbf{k}\right)\right] \\ &+\nu\left\{\nabla\cdot[\boldsymbol{\omega}\cdot(\nabla\boldsymbol{\omega})]-[(\nabla\boldsymbol{\omega})\cdot\nabla]\cdot\boldsymbol{\omega}\right\}\,.\end{aligned} \tag{13.5}$$

Here we have exposed the effects of vortex stretching on the enstrophy, using

$$\boldsymbol{\omega}\cdot[(\boldsymbol{\omega}\cdot\nabla)\,\mathbf{V}] = |\boldsymbol{\omega}|^2\frac{\partial V_s}{\partial s}\,, \tag{13.6}$$

which you should prove as an exercise. The tilting term contributes nothing in (13.5), because the direction of the tilting term is always perpendicular to the direction of $\boldsymbol{\omega}$. By analogy with our analysis of the friction terms of the kinetic energy equation in Chapter 4, we have separated the effects of viscosity on the enstrophy into two parts:

$$\begin{aligned}\boldsymbol{\omega}\cdot\left[\nu\nabla^2\boldsymbol{\omega}\right] &= \nu\boldsymbol{\omega}\cdot\left[\nabla\cdot(\nabla\boldsymbol{\omega})\right] \\ &= \nu\left\{\nabla\cdot\left[\boldsymbol{\omega}\cdot(\nabla\boldsymbol{\omega})\right]-\left[(\nabla\boldsymbol{\omega})\cdot\nabla\right]\cdot\boldsymbol{\omega}\right\}\,.\end{aligned} \tag{13.7}$$

The term $-\nu\left[(\nabla\boldsymbol{\omega})\cdot\nabla\right]\cdot\boldsymbol{\omega}$ represents enstrophy dissipation, and is always an enstrophy sink.

Now use the continuity equation to rewrite (13.5) in flux form, drop the buoyancy term, and combine the two flux-divergence terms:

$$\frac{\partial}{\partial t}\left(\frac{1}{2}|\boldsymbol{\omega}|^2\right)=\nabla\cdot\left\{-\mathbf{V}\frac{1}{2}|\boldsymbol{\omega}|^2+\nu\left[\boldsymbol{\omega}\cdot(\nabla\boldsymbol{\omega})\right]\right\}+|\boldsymbol{\omega}|^2\frac{\partial V_s}{\partial s}-\nu\left[(\nabla\boldsymbol{\omega})\cdot\nabla\right]\cdot\boldsymbol{\omega}\,. \tag{13.8}$$

In a steady state or time average, the left-hand side of (13.8) drops out, and when we spatially average over the whole domain the flux divergence terms on the right-hand side vanish as well, so that we are left with

$$\int\limits_M\left[|\boldsymbol{\omega}|^2\frac{\partial v_s}{\partial s}-\nu\left[(\nabla\boldsymbol{\omega})\cdot\nabla\right]\cdot\boldsymbol{\omega}\right]dM=0\,. \tag{13.9}$$

Because the viscous term of (13.9) is a sink of enstrophy, we conclude that, on the average, the stretching term has to be a source of enstrophy, i.e.,

$$\int\limits_M|\boldsymbol{\omega}|^2\frac{\partial V_s}{\partial s}dM>0\,. \tag{13.10}$$

Similar arguments were used in Chapter 4, in connection with the energy and entropy budgets of the global atmosphere.

Recall that the rate of production of vorticity by vortex stretching is $|\boldsymbol{\omega}|\,\partial V_s/\partial s$. Eq. (13.10) implies that in a statistically steady turbulence there is a tendency for $\partial V_s/\partial s$ to be positive where $|\boldsymbol{\omega}|$ is larger than average, and for $\partial V_s/\partial s$ to be negative where $|\boldsymbol{\omega}|$ is smaller than average. The rich get richer – strong vortices are stretched and made even stronger.

13.2 The kinetic energy cascade

We conclude that the net effect of $(\boldsymbol{\omega}\cdot\nabla)\,\mathbf{V}$ is to increase $|\boldsymbol{\omega}|$ in an average sense. The advective terms of the kinetic energy equation do not change the domain-averaged kinetic energy, however. A process that increases the enstrophy without changing the kinetic energy tends to shift the spectrum of kinetic energy towards shorter scales. To see why, note that the ratio of the kinetic energy to the enstrophy has the units of a length squared:

$$\frac{\frac{1}{2}|\mathbf{V}|^2}{|\boldsymbol{\omega}|^2} \sim L^2\,. \tag{13.11}$$

An interpretation is that L is the diameter of the most energetic vortices. Circulations with lots of vorticity per unit kinetic energy have their energy concentrated in relatively small vortices.

Momentum advection increases the enstrophy through vortex stretching, but it does not change the kinetic energy. It therefore tends to decrease L. This systematic migration of the kinetic energy from larger to smaller scales under inertial processes is called a kinetic energy "cascade." The terminology evokes a waterfall; see Figure 13.2, in which a single stream falls off a cliff and repeatedly splits on the way down, as it strikes rocks or other obstacles. The kinetic energy cascade is like a "flow" of kinetic energy from larger scales to smaller scales. Because viscosity acts most effectively on small scales, vortex stretching promotes kinetic energy dissipation. If viscosity could somehow be eliminated, kinetic energy would accumulate on small scales, leading to a very noisy wind field. Viscosity prevents this by dissipating the noise.

Later we will use the term "anti-cascade" to describe the very interesting and important processes that can transfer energy from small scales to larger scales.

13.3 Dimensional analysis of the kinetic energy spectrum

We now turn to an analysis of the distribution of kinetic energy with scale, i.e., the kinetic energy "spectrum." Consider a three-dimensional flow. Let δ be the rate of kinetic energy dissipation per unit mass, and ν the kinematic viscosity. These quantities have physical dimensions (see Appendix B) as follows:



Figure 13.2: A cascading waterfall.

$$\delta \sim L^2 T^{-3},\ \nu \sim L^2 T^{-1}\ , \tag{13.12}$$

Kolmogorov (1941) hypothesized that the statistics of locally homogeneous and isotropic three-dimensional turbulence are determined by δ and ν; according to his hypothesis, if δ and ν are known, no other information is needed to determine the statistics. Note that δ is a property of the flow, while ν is a property of the fluid.

Kolmogorov further hypothesized that there exists an "inertial subrange" – a range of scales – within which turbulence energy is neither generated nor dissipated, but just "passes through," like a little town on a big highway, where nobody stops except perhaps to buy

gas. By definition, the "viscous subrange" consists of scales for which both δ and ν are important. Together, the viscous subrange and the inertial subrange make up the homogeneous isotropic component of the turbulence. The smallest scale in the inertial subrange, which is the same as the largest scale in the viscous subrange, can be estimated by forming a length from δ and ν:

$$L_{\mathrm{K}} = \left(\frac{\nu^3}{\delta} \right)^{\frac{1}{4}} , \tag{13.13}$$

This is called the Kolmogorov microscale. According to (13.13), if δ increases for a given ν (which means for a given kind of fluid), then L_{K} has to become smaller, but weakly. For the Earth's atmosphere, a typical value of L_{K} is in the range 10^{-2} 10^{-3} m.

The inertial subrange consists of homogeneous isotropic eddies that are larger than L_{K}. Kolmogorov hypothesized that for the scales of motion that lie within the inertial subrange, the turbulence statistics are determined, as functions of wave number, by only one dimensional parameter that characterizes the particular flow in question: the dissipation rate, δ. Note that δ represents dissipation that occurs on scales outside of (shorter than those of) the inertial subrange.

In the boundary layer of the Earth's atmosphere, where the turbulent eddies have scales on the order of a few kilometers or less, kinetic energy is typically generated at lower wave numbers (on the scale of the largest turbulent eddies), and migrates to higher wave numbers by way of the inertial subrange, and is finally dissipated on the smallest scales. This is illustrated in Fig. 13.3. We want to find $K(k)$, called the "spectrum of kinetic energy," in the inertial subrange. This is the squared modulus of the Fourier transform of the velocity. It has units of energy per unit mass per unit wave number. Dimensionally,

$$K(k) \sim L^3 T^{-2} . \tag{13.14}$$

Because the turbulence statistics in the inertial subrange depend only on δ, $K(k)$ can depend only on δ and k. Assuming that

$$K(k) = \alpha \delta^m k^n , \tag{13.15}$$

where $\alpha \cong 1.5$ is an empirically determined non-dimensional constant, we find that $m = 2/3$, and $n = -5/3$. In other words,

$$K(k) = 1.5\delta^{2/3}k^{-5/3} \,. \tag{13.16}$$

As discussed for example by Lesieur (2008), Eq. (13.16) is remarkably well supported by measurements of small-scale turbulence in the laboratory, the atmosphere, and the oceans.

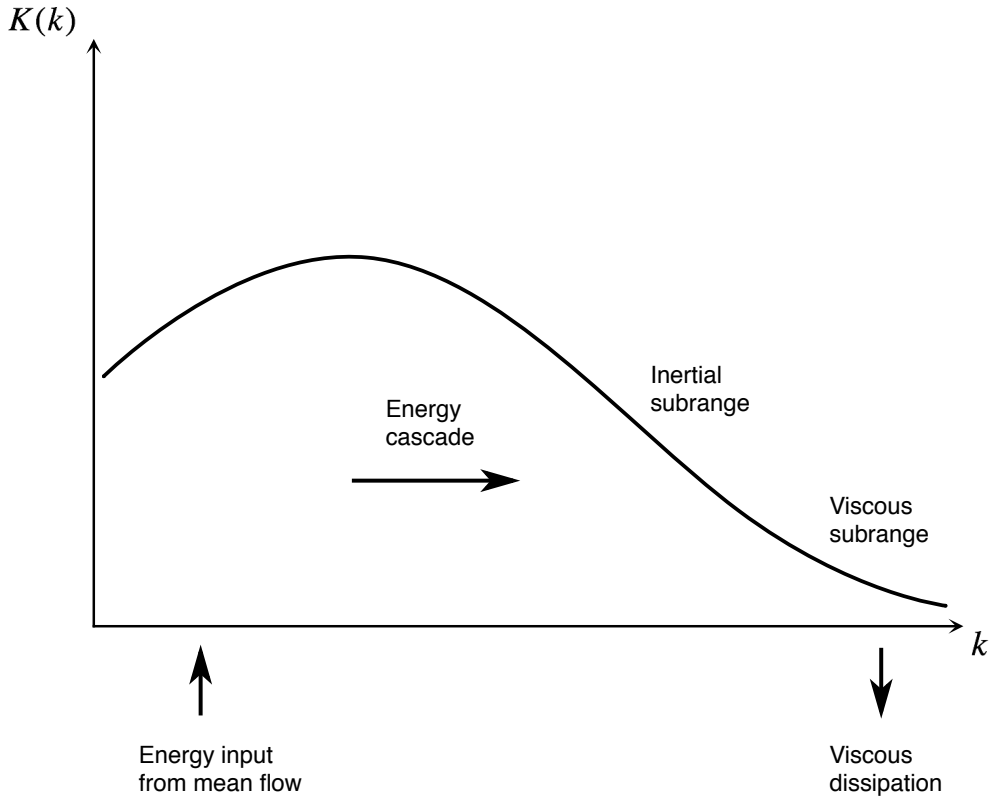


Figure 13.3: Sketch illustrating, in a highly idealized way, the flow of kinetic energy through wavenumber space, from a low-wave number region with an energy source, to a high-wave number region in which dissipation removes kinetic energy. This diagram applies to the case of three-dimensional turbulence, which we find, for example, in the boundary layer.

Nonlinearity is essential to turbulence because vorticity advection allows interactions among vortices with a variety of scales.

13.4 Nonlinearity and scale interactions

Scale interactions are intrinsically nonlinear, i.e., they can only arise from the nonlinear terms of the equations, such as the stretching term. To see how this works from a mathematical perspective, consider the following simple example. Suppose that we have two functions of a spatial coordinate x given by

$$A(x) = \hat{A}e^{ikx} \text{ and } B(x) = \hat{B}e^{ilx} . \tag{13.17}$$

respectively. Here the wave numbers of the functions A and B are denoted by k and l, respectively. If we combine A and B linearly, e.g., if we form

$$\alpha A + \beta B , \tag{13.18}$$

where α and β are spatially constant coefficients, then no "new" wave numbers are generated; k and l continue to be the only wave numbers present. In contrast, if we multiply A and B together, which is a nonlinear operation, then we generate the new wave number, $k+l$:

$$AB = \hat{A}\hat{B}e^{i(k+l)x} , \tag{13.19}$$

Other nonlinear operations such as division, exponentiation, etc., will also generate new wave numbers.

"Nonlinear" is a mathematical term. When a physical process is described as nonlinear, the physical implication is that it involves scale interactions and may introduce new scales. In atmospheric science, there are two main types of nonlinearity:

- *advection*, in which a wind vector multiplies the gradient of an advected quantity, and
- *rectification*, in which a fluctuating input leads to a nonzero output only when the input is positive. Precipitation is an example of a process for which rectification is relevant, because precipitation tends to occur when the large-scale vertical motion is upward but typically does not occur when the large-scale vertical motion is downward.

13.5 Two-dimensional turbulence

An "inertial process" is one that includes only momentum advection and the effects of the Earth's rotation. In two dimensions, the stretching/twisting term of (13.4) is zero, because $\partial/\partial z$ is zero. It follows that vorticity and enstrophy are both conserved under inertial processes in two-dimensional turbulence, and as a result the length scale L is also

conserved. The implication is that *kinetic energy does not cascade in frictionless two-dimensional flows*; in fact, it actually shifts towards larger scales.

When the effects of viscosity are included in two-dimensional turbulence, they act most effectively on the smallest scales, where the enstrophy is concentrated. Friction therefore removes or "dissipates" enstrophy quite effectively at the small-scale end of the spectrum. In order for enstrophy dissipation to continue, additional small-scale enstrophy must be supplied by nonlinear transfer of enstrophy from larger scales to the smaller scales. It is as if the enstrophy dissipation "pulls" enstrophy from larger to smaller scales, through the nonlinear terms. This is an enstrophy cascade. Note that enstrophy dissipation will also tend to increase L, over and above the increase that comes through inertial processes. The scale of the most energetic vortices will increase with time. This is a kinetic energy "anti-cascade." Further discussion is given below.

We conclude that in three-dimensional flow both kinetic energy and enstrophy cascade and are dissipated, while in a two-dimensional flow enstrophy cascades and is dissipated, but kinetic energy is nearly conserved and migrates towards larger scales.

The exchanges of energy and enstrophy among scales in two-dimensional turbulence were studied by Fjortoft (1953), who obtained some very fundamental and famous results, which can be summarized in a simplified way as follows. Consider three equally spaced wave numbers λ_1, λ_2, and λ_3, as shown in Fig. 13.4. By "equally spaced" we mean that

$$\begin{aligned} \lambda_2 - \lambda_1 =& \lambda_3 - \lambda_2 \\ \equiv& \Delta\lambda \ . \end{aligned} \tag{13.20}$$

The enstrophy, E, is given by

$$E = E_1 + E_2 + E_3 \ , \tag{13.21}$$

and the kinetic energy is

$$K = K_1 + K_2 + K_3 \ . \tag{13.22}$$

It can be shown that

$$E_n = {\lambda_n}^2 K_n \ . \tag{13.23}$$

where λ is a wave number, and the subscript n denotes a particular Fourier component.

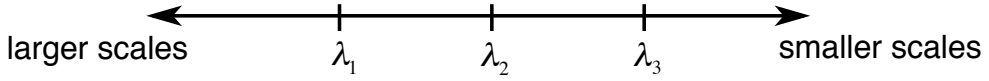


Figure 13.4: Diagram used in the explanation of Fjortoft's (1953) analysis of the exchanges of energy and enstrophy among differing scales in two-dimensional motion.

Consider an inertial process such that kinetic energy and enstrophy are redistributed in a two-dimensional frictionless flow, i.e.

$$K_n \to K_n + \delta K_n \,, \tag{13.24}$$

$$E_n \to E_n + \delta E_n \,. \tag{13.25}$$

Because kinetic energy and enstrophy are both conserved under two-dimensional inertial processes, we know that

$$\sum \delta K_n = 0 \,, \tag{13.26}$$

$$\sum \delta E_n = 0 \,. \tag{13.27}$$

From (13.26) we see that

$$\delta K_1 + \delta K_3 = -\delta K_2 \,. \tag{13.28}$$

Note from (13.23) that

$$\delta E_n = \lambda_n^2 \delta K_n \, . \tag{13.29}$$

From (13.27) and (13.29), we obtain

$$\begin{aligned} \lambda_1^2 \delta K_1 + \lambda_3^2 \delta K_3 &= -\lambda_2^2 \delta K_2 \\ &= \lambda_2^2 \left(\delta K_1 + \delta K_3 \right) \, . \end{aligned} \tag{13.30}$$

Collecting terms in (13.30), we find that

$$\frac{\delta K_3}{\delta K_1} = \frac{\lambda_2^2 - \lambda_1^2}{\lambda_3^2 - \lambda_2^2} \, . \tag{13.31}$$

Using (13.20), we can simplify (13.31) to

$$\frac{\delta K_3}{\delta K_1} = \frac{\lambda_2 + \lambda_1}{\lambda_3 + \lambda_2} < 1 \, . \tag{13.32}$$

This is the first result.

Eq. (13.32) shows that the energy transferred to higher wave numbers (δK_3) is less than the energy transferred to lower wave numbers (δK_1). This conclusion rests on both (13.26) and (13.27), i.e., on both kinetic energy conservation and enstrophy conservation. The implication is that kinetic energy actually "migrates" from higher wave numbers to lower wave numbers, i.e.,from smaller scales to larger scales. As mentioned earlier, this is sometimes called an "anti-cascade."

We now perform a similar analysis for the enstrophy. As a first step, we note from (13.29) and (13.32) that

$$\begin{aligned} \frac{\delta E_3}{\delta E_1} &= \frac{\lambda_3^2}{\lambda_1^2} \left(\frac{\lambda_2 + \lambda_1}{\lambda_3 + \lambda_2} \right) \\ &= \frac{(\lambda_2 + \Delta\lambda)^2}{(\lambda_2 - \Delta\lambda)^2} \frac{\left(\lambda_2 - \frac{1}{2}\Delta\lambda\right)}{\left(\lambda_2 + \frac{1}{2}\Delta\lambda\right)} \, . \end{aligned} \tag{13.33}$$

To show that $\delta E_3/\delta E_1$ is greater than one, we demonstrate that it can be written as $a \cdot b \cdot c$, where a, b, and c are each greater than one. We can choose:

$$a = \frac{\lambda_2 + \Delta\lambda}{\lambda_2 + \frac{1}{2}\Delta\lambda} > 1, \quad b = \frac{\lambda_2 - \frac{1}{2}\Delta\lambda}{\lambda_2 - \Delta\lambda} > 1, \quad c = \frac{\lambda_2 + \Delta\lambda}{\lambda_2 - \Delta\lambda} > 1 \, . \tag{13.34}$$

The conclusion is that enstrophy does cascade to higher wave numbers in two-dimensional turbulence. In the presence of viscosity, such a cascade ultimately leads to enstrophy dissipation.

13.6 Quasi-two-dimensional turbulence

Large-scale motions are quasi-two-dimensional, so it is reasonable to suspect that kinetic energy is transferred upscale on large scales. Consider the isentropic potential vorticity equation in the form (4.188), which is repeated here for convenience:

$$\left(\frac{\partial}{\partial t}\right)_\theta (\rho_\theta \text{ IPV}) + \nabla_\theta \cdot \left[\rho_\theta \mathbf{v} \text{ IPV} - \mathbf{e}_r \times \left(\dot{\theta}\frac{\partial \mathbf{v}}{\partial \theta} + \frac{1}{\rho_\theta}\frac{\partial \mathbf{F}_\mathbf{v}}{\partial \theta}\right)\right] = 0 \, . \tag{13.35}$$

Combining this with the continuity equation in isentropic coordinates, Eq. (C.24), we obtain an advective form of the PV equation:

$$\rho_\theta \left(\frac{\partial \text{ IPV}}{\partial t} + \mathbf{v} \cdot \nabla_\theta \text{ IPV}\right) = \text{ IPV}\frac{\partial \left(\rho_\theta \dot{\theta}\right)}{\partial \theta} - \nabla_\theta \cdot \left[\mathbf{e}_r \times \left(\dot{\theta}\frac{\partial \mathbf{v}}{\partial \theta} + \frac{1}{\rho_\theta}\frac{\partial \mathbf{F}_\mathbf{v}}{\partial \theta}\right)\right] \tag{13.36}$$

According to (13.36), in the absence of heating and friction the potential vorticity is conserved following a particle that moves along an isentropic surface. When heating is present, the first term on the right-hand side of (13.36) can cause the IPV to exponentially increase or decrease. For example, in a region where latent heating increases upward to a maximum in the middle troposphere, and then decreases upward from the middle troposphere to the tropopause, a positive potential vorticity anomaly is generated in the lower troposphere and and a negative anomaly is generated in the upper troposphere. This happens, for example, in the eyewall of a tropical cyclone.

Multiplying (13.36) by IPV gives

$$\rho_\theta \left[\frac{\partial}{\partial t} \left(\frac{\text{IPV}^2}{2} \right) + \mathbf{v} \cdot \nabla_\theta \left(\frac{\text{IPV}^2}{2} \right) \right] = \text{IPV}^2 \frac{\partial \left(\rho_\theta \dot{\theta} \right)}{\partial \theta} - \text{IPV} \nabla_\theta \cdot \left[\mathbf{e}_r \times \left(\dot{\theta} \frac{\partial \mathbf{v}}{\partial \theta} + \frac{1}{\rho_\theta} \frac{\partial \mathbf{F}_\mathbf{v}}{\partial \theta} \right) \right] . \tag{13.37}$$

The quantity IPV^2 is called the "potential enstrophy." Eq. (13.37) says that in the absence of heating and friction the potential enstrophy does not change following the motion.

Using (C.24), we can convert (13.37) back to flux form:

$$\frac{\partial}{\partial t} \left(\rho_\theta \frac{\text{IPV}^2}{2} \right) + \nabla_\theta \cdot \left(\rho_\theta \mathbf{v} \frac{\text{IPV}^2}{2} \right) = - \text{IPV} \nabla_\theta \cdot \left[\mathbf{e}_r \times \left(\dot{\theta} \frac{\partial \mathbf{v}}{\partial \theta} + \frac{1}{\rho_\theta} \frac{\partial \mathbf{F}_\mathbf{v}}{\partial \theta} \right) \right] , \tag{13.38}$$

Eq. (13.38) implies that

$$\overline{\rho_\theta \, \text{IPV}^2}^\theta = \text{ constant in the absence of heating and friction } . \tag{13.39}$$

Here $\overline{(\,)}^\theta$ denotes an average over an isentropic surface.

Consider the very special case of a nondivergent flow (i.e., nondivergent on an isentropic surface) with no heating. Then (C.24) reduces to

$$\frac{\partial \rho_\theta}{\partial t} + \mathbf{v} \cdot \nabla_\theta \rho_\theta = 0 , \tag{13.40}$$

which says that ρ_θ is conserved following the (isentropic) motion. Since

$$\rho_\theta \, \text{IPV}^2 = \frac{\eta^2}{\rho_\theta} , \tag{13.41}$$

where η is the absolute vorticity, (13.38) reduces to

$$\overline{\eta^2}^\theta = \text{ constant for nondivergent flow with no heating or friction } . \tag{13.42}$$

It follows that, in this limiting case,

$$L^2 \equiv \frac{\overline{K}^{\theta}}{\overline{\eta^2}^{\theta}} = \text{constant} . \tag{13.43}$$

This is closely analogous to the conclusion that we reached earlier for the case of two-dimensional turbulence.

For a more general divergent flow, we can write

$$\overline{\eta^2}^{\theta} < \overline{\left\{ \frac{(\rho_\theta)_{\max}}{\rho_\theta} \right\} \eta^2}^{\theta} , \tag{13.44}$$

where $(\rho_\theta)_{\max}$ is an appropriately chosen constant upper bound on ρ_θ. We can consider $(\rho_\theta)_{\max}$ to be a constant. Then (13.44) can be written as

$$\overline{\eta^2}^{\theta} < (\rho_\theta)_{\max} \overline{\left(\frac{\eta^2}{\rho_\theta} \right)}^{\theta} = (\rho_\theta)_{\max} \overline{(\rho_\theta q^2)}^{\theta} = \text{constant} , \tag{13.45}$$

According to (13.45), $\overline{\eta^2}^{\theta}$ has an upper bound. Then from (13.43) we see that L has a lower bound. This demonstrates that even for divergent motion the kinetic energy cannot cascade into arbitrarily small scales; it tends to stay in the larger scales. This important result is a partial explanation for the "smoothness" of the global circulation.

Charney (1971) showed that even though real atmospheric motions are not two-dimensional, the constraint of quasi-geostrophy causes large-scale geostrophic turbulence to behave much like idealized two-dimensional turbulence, so that potential enstrophy cascades to smaller scales and is dissipated while kinetic energy is conserved (is not dissipated) and anti-cascades to larger scales.

When kinetic energy is conserved while enstrophy decreases due to dissipation, the result is an increase in L. (Our earlier conclusion that L remains constant was based on the assumption that both kinetic energy and enstrophy are invariant.) Because the total amount of "room" available to the eddies is fixed (the planet is not getting any bigger), the only way that L can increase is if the "number of eddies" decreases; for this reason, *eddies tend to merge in two-dimensional or geostrophic turbulence*.

In an idealized numerical simulation, McWilliams et al. (1994) showed that through vortex mergers pure two-dimensional turbulence gradually organizes itself into a single large cyclone-anticyclone pair. Their spectacular results are shown in Fig. 13.5.

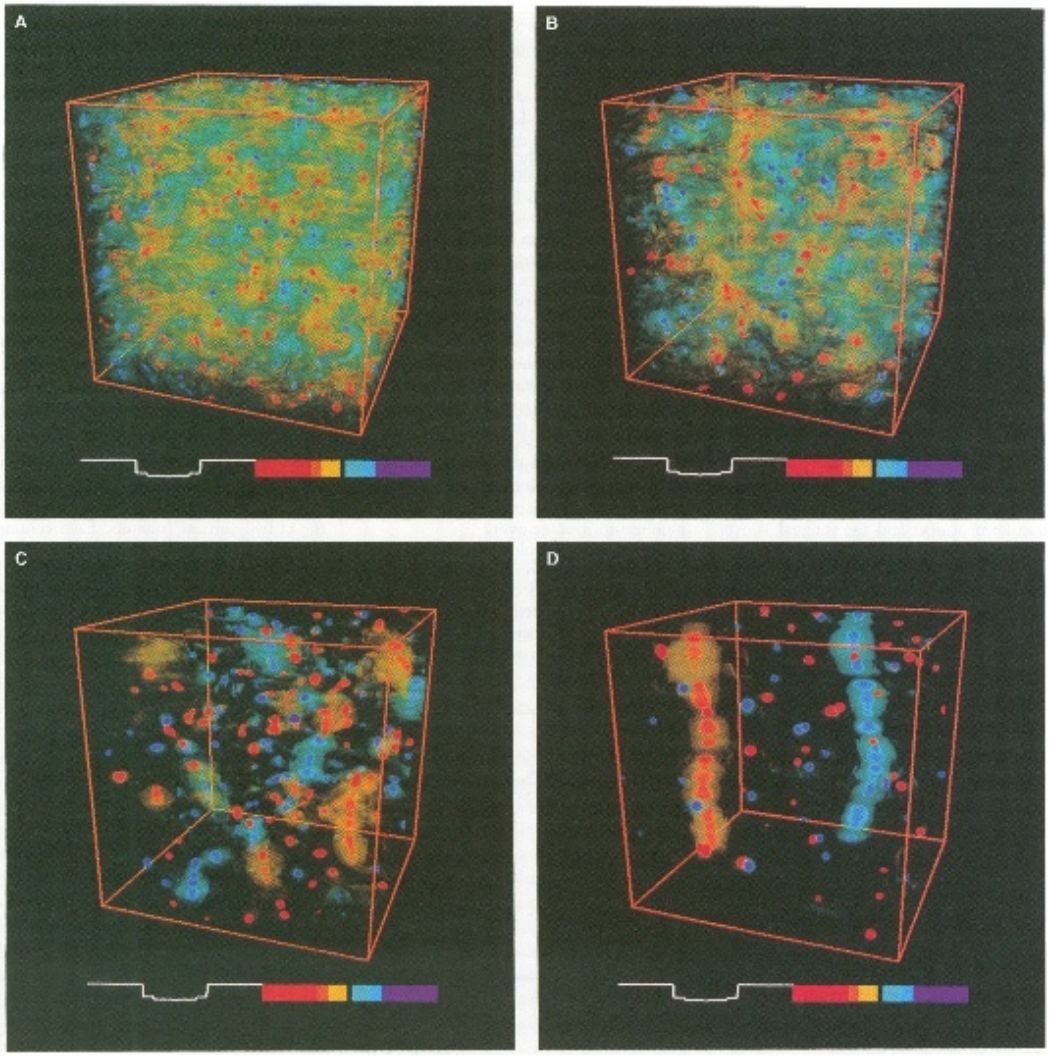


Figure 13.5: A time sequence of the vorticity distribution in an idealized numerical simulation of quasi-two-dimensional turbulence, by McWilliams et al. (1994). Initially, the vorticity is highly disorganized. Through a process of enstrophy dissipation and vortex mergers, the circulation gradually organizes itself into a single pair of vortices - one cyclonic and one anti-cyclonic.

It has been hypothesized that the anticyclonic Great Red Spot on Jupiter, and similar phenomena observed elsewhere on Jupiter and the other giant gaseous planets of the outer Solar System, may be the end products of a kinetic energy anti-cascade in quasi-two-dimensional turbulence.

13.7 The anti-cascade of two-dimensional turbulence

As mentioned earlier, kinetic energy does not cascade in two-dimensional turbulence, but enstrophy does cascade. Suppose that in two-dimensional turbulence there exists an inertial subrange, presumably on rather large scales, in which enstrophy is neither generated nor dissipated, and that within this inertial subrange the turbulence statistics are determined by the rate of enstrophy dissipation. Using methods similar to those used to derive (13.16), we can show that the kinetic energy spectrum follows k^{-3}. This conclusion is supported by numerical simulations of two-dimensional turbulence. Note that k^{-3} decreases more steeply with increasing k than does $k^{-5/3}$. This means that there is less kinetic energy at high wave numbers in two-dimensional turbulence than in three-dimensional turbulence, consistent with our earlier conclusion that kinetic energy dissipation is weak in two-dimensional turbulence and strong in three-dimensional turbulence.

An analysis very similar to that used to derive (13.16) can be used to determine the kinetic energy spectrum for scales longer than those at which kinetic energy is generated. We assume that in an inertial subrange upscale from the kinetic energy source, the kinetic energy spectrum depends only on the strength of the source and the wave number. The conclusion is that upscale from the energy source, the kinetic energy spectrum follows $k^{-5/3}$. This is true for either two-dimensional or three-dimensional turbulence.

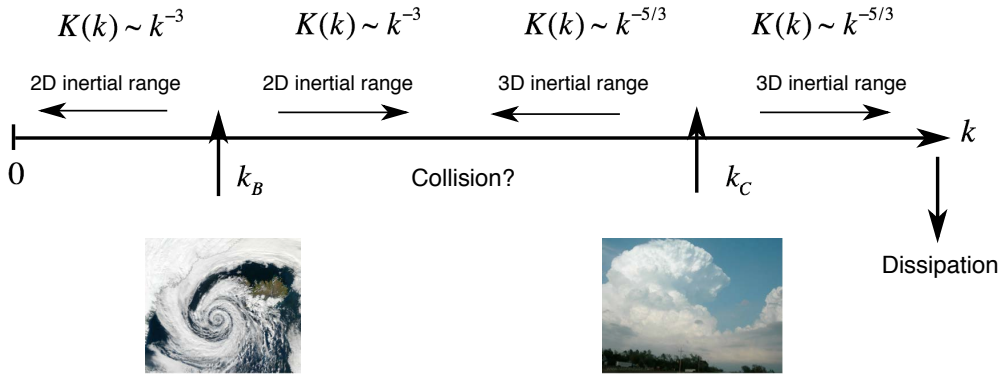


Figure 13.6: Cartoon illustrating the kinetic energy spectrum in the Earth's atmosphere, as implied by dimensional analysis. Baroclinic instability adds kinetic energy at wave number k_B, and convection adds kinetic energy at wave number k_C. See text for details.

Now look at Fig. 13.6, which is a generalized and refined version of Fig. 13.3. In Fig. 13.6, four distinct inertial subranges are indicated. The "low" wave number k_B denotes the scale at which baroclinic instability acts as a kinetic energy source. The "high" wave number k_C denotes the scale at which convection acts as a kinetic energy source. For $k > k_C$, we have a three-dimensional inertial range with $K(k) \sim k^{-5/3}$, and an energy cascade to smaller scales, ultimately feeding kinetic energy dissipation. For k slightly less than , we have a three-dimensional inertial range with and energy flowing upscale. For k slightly greater than k_B, we have a two-dimensional inertial subrange with $K(k) \sim k^{-3}$ and energy

flowing down-scale. For k less than k_B, energy flows upscale, away from the baroclinic energy source. Clearly, the shape of the spectrum has to change somewhere between k_B and k_C, where the word "Collision" appears in the figure.

Lilly et al. (1998) discussed observations that appear to show such a change in the spectral slope, as seen in Fig. 13.7. The spectrum follows for scales less than about 100 km, and for scales greater than about 100 km. There is a "kink" in the spectrum close to 100 km. A second kink is visible at a scale of several thousand kilometers. The correspondence between Fig. 13.6 and Fig. 13.7 should be clear.

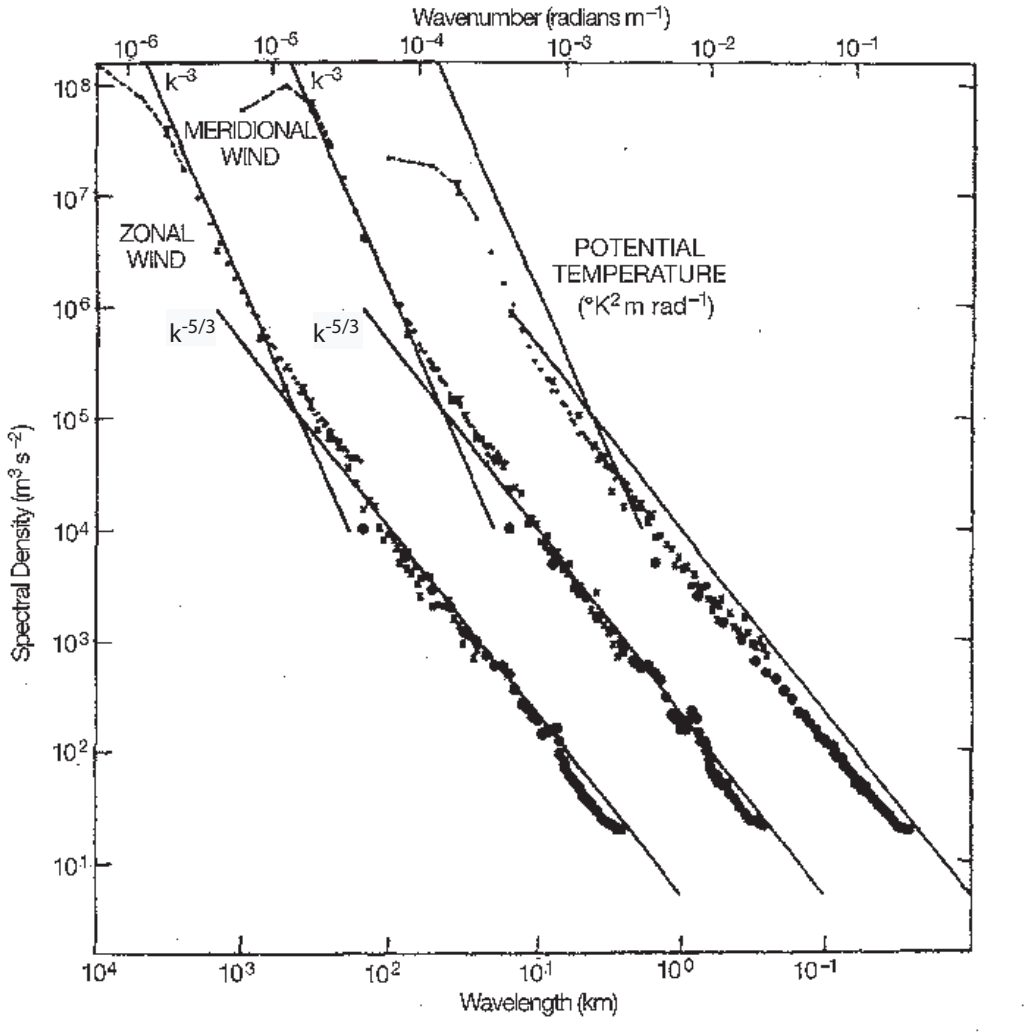


Figure 13.7: Spectra of zonal and meridional winds and potential temperature near the tropopause, obtained from aircraft data over many parts of the world. The meridional wind spectrum is displaced one decade to the right, and the potential temperature spectrum two decades. From Lilly et al. (1998), after Nastrom and Gage (1985).

There is another complication. Rhines (1975) pointed out that sufficiently large vortices will feel the β effect, which exerts a "restoring force" opposing large meridional excursions by the particles making up the vortex. This means that the β effect will tend to resist *meridional* widening of vortices beyond some limit, although they may continue to widen in the zonal direction. Rhines suggested that the eddies will in fact begin to behave as Rossby-wave packets for scales large enough so that the characteristic eddy velocity, U, is comparable to the phase speed of a Rossby wave. Dimensional analysis suggests that this occurs for

$$k \sim k_\beta \equiv \sqrt{\frac{\beta}{U}} \, . \tag{13.46}$$

The length scale corresponding to this wave number is sometimes called the "Rhines length" or the "Rhines barrier." For $k < k_\beta$, further meridional broadening is inhibited, but longitudinal broadening can continue. The eddies therefore become elongated in the zonal direction and ultimately give rise to alternating zonal jets of width k_β^{-1}, like those seen on Jupiter. There have been a number of numerical modeling studies that tend to support this idea (e.g., Huang and Robinson, 1998).

In summary, vorticity and enstrophy are conserved in two-dimensional flow but not in three-dimensional flow. Kinetic energy is conserved under inertial processes in both two-dimensional and three-dimensional flows. Because both energy and enstrophy are conserved in two-dimensional flows, a two-dimensional motion field "has fewer options" than do three-dimensional flows. Because kinetic energy does not cascade in two-dimensional flow, the motion remains smooth and is dominated by "large" eddies.

13.8 Observations of the kinetic energy spectrum

Boer and Shepherd (1983) analyzed observations to examine the spectra of kinetic energy, enstrophy, and available potential energy, and also the exchanges of kinetic energy among various scales. Following a suggestion of Baer (1972), they used the two-dimensional index associated with the spherical harmonics as a measure of scale, much as Blackmon did in the work described in Chapter 7. The vertically integrated spectra obtained by Boer and Shepherd are shown in Fig. 13.8, for kinetic energy and enstrophy only. The slope of the kinetic energy spectrum is plotted as a function of height in Fig. 13.9. A k^{-3} behavior is evident, particularly at the upper levels. Boer and Shepherd evaluated the exchanges of kinetic energy among the various scales, as shown in Fig. 13.10; similar computations were reported by Chen and Wiin-Nielsen (1978). The smaller scales generally experience a kinetic energy cascade towards even smaller scales, as would be expected for three-dimensional turbulence, but the larger scales experience an inverse cascade, as would be expected for two-dimensional turbulence.

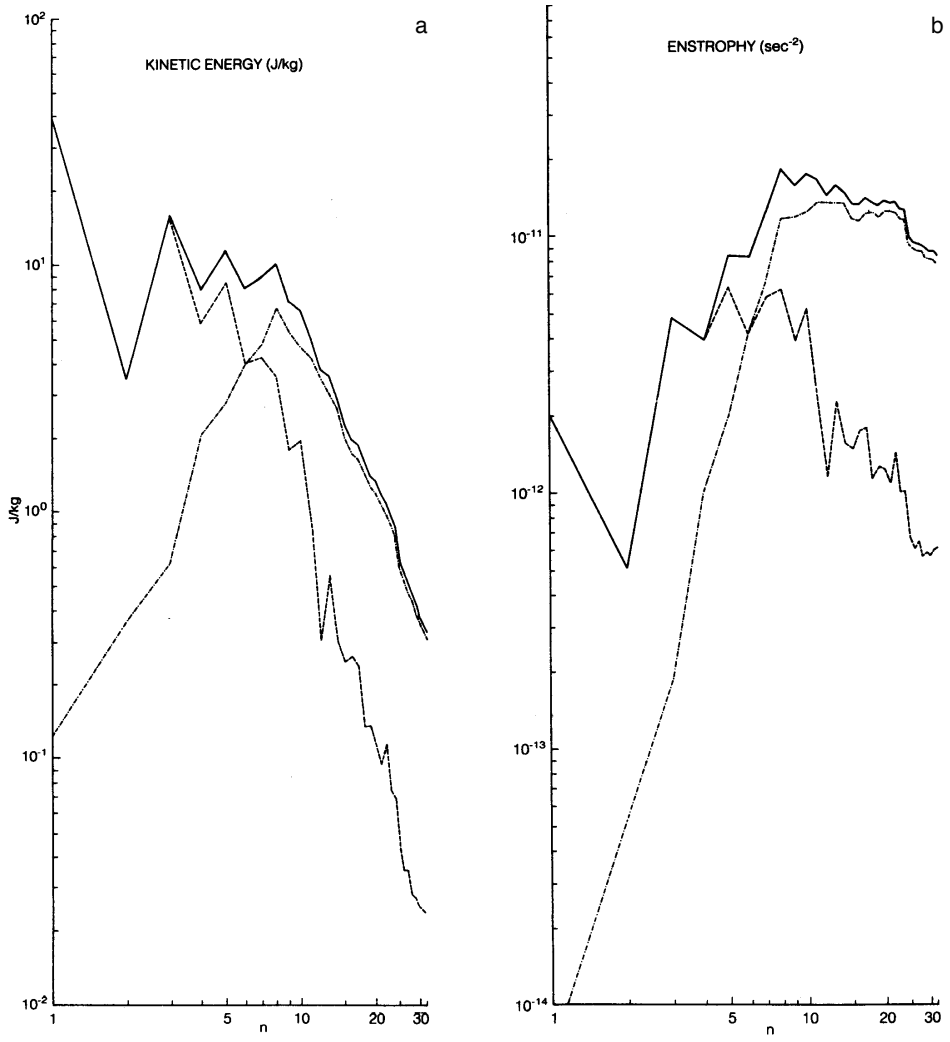


Figure 13.8: The observed vertically integrated spectra of kinetic energy (panel a) and enstrophy (panel b). The solid lines denote total, the dashed stationary, and the dashed-dot transient. From Boer and Shepherd (1983).

It has also been suggested that the formation and maintenance of a blocking high represents an example of up-scale energy transfer. As mentioned in Chapter 11, the formation of a block is associated with the advection of subtropical potential vorticity into middle latitudes. The "agent" that carries out this advection is a rapidly developing cyclone, e.g., over the Gulf Stream (Hoskins et al. 1985). Moreover, it is the interaction of the block with small-scale eddies that allows the block to maintain itself for an extended period (Shutts, 1986).

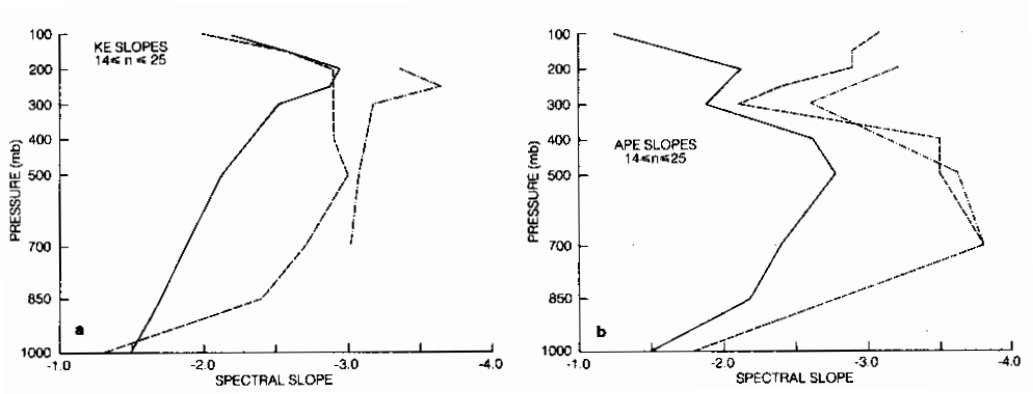


Figure 13.9: Slope of straight line fitted to the kinetic energy spectrum, for two-dimensional indices in the range 14 to 25. The dash-dotted line represents the results of Baer (1972), and the dashed line shows the results of Chen and Wiin-Nielsen (1978). From Boer and Shepherd (1983).

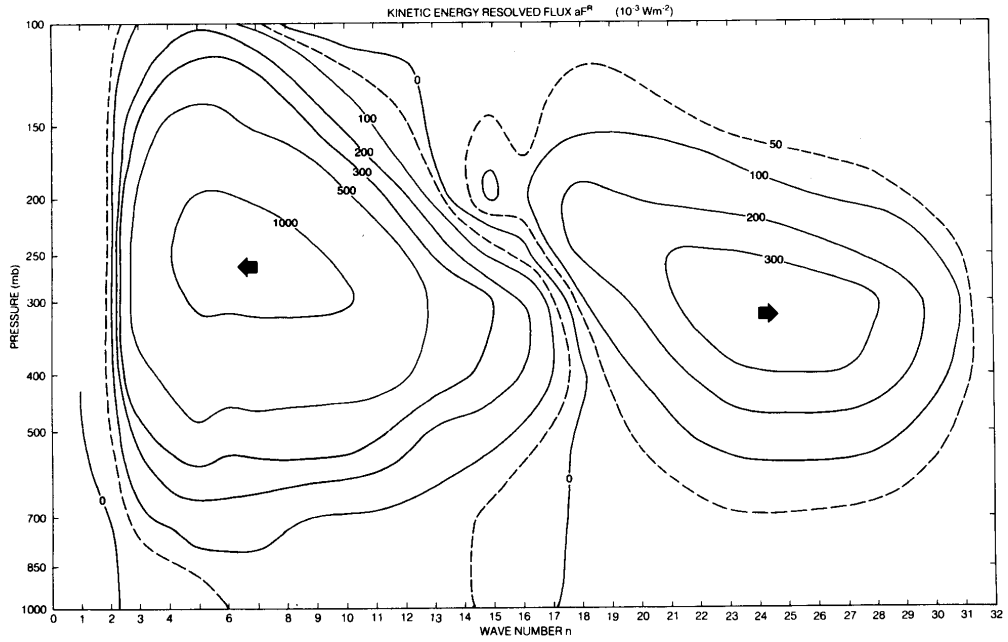


Figure 13.10: Observed nonlinear kinetic energy exchanges as a function of height. From Boer and Shepherd (1983).

13.9 Dissipating enstrophy but not kinetic energy

Sadourny and Basdevant (1985) suggested an interesting approach to representing the effects of quasi-two-dimensional, geostrophic turbulence in the momentum equation. The issue is that in such a flow enstrophy is dissipated but kinetic energy is not. How can we formulate a "frictional" term in the momentum equation that has this property? To examine the idea of Sadourny and Basdevant, we start from the equation of motion in "invariant"

form, using isentropic coordinates:

$$\frac{\partial \mathbf{v}}{\partial t} + q\mathbf{k} \times (\rho_\theta \mathbf{v}) + \nabla \left(\frac{1}{2} \mathbf{v} \cdot \mathbf{v} + s \right) = 0 \,. \tag{13.47}$$

Here q is the potential vorticity, as before. We want to include friction in such a way that it does not affect the kinetic energy. This can be done by introducing a parameter, D, as follows:

$$\frac{\partial \mathbf{V}}{\partial t} + (q - D)\mathbf{k} \times (\rho_\theta \mathbf{v}) + \nabla \left(\frac{1}{2} \mathbf{v} \cdot \mathbf{v} + s \right) = 0 \,. \tag{13.48}$$

We can interpret $q - D$ as a modified potential vorticity. When we dot (13.48) with $\mathbf{v}$ to form the kinetic energy equation, the term involving D drops out, regardless of the actual form of D. This means that D does not contribute to the tendency of the kinetic energy.

We want to choose the form of D in such a way that potential enstrophy is dissipated. The first step is to construct the potential vorticity equation, by taking the curl of (13.48) and using continuity:

$$\frac{Dq}{Dt} = \frac{1}{\rho_\theta} \nabla \cdot (D \rho_\theta \mathbf{v}) \,. \tag{13.49}$$

The right-hand side of (13.49) vanishes, as it should, for $D \equiv 0$. Let the potential enstrophy averaged over an entire isentropic surface (globally) be given by

$$\mathrm{IPV}\,(\theta) = \frac{1}{S} \int_S \int \frac{q^2}{2} \rho_\theta \, dS \,. \tag{13.50}$$

Then (13.49) implies that

$$\frac{d\,\mathrm{IPV}\,(\theta)}{dt} = - \int_S \int (D \rho_\theta \mathbf{v} \cdot \nabla q) dS \,. \tag{13.51}$$

To guarantee dissipation of $\mathrm{IPV}\,(\theta)$, we choose

$$D = \tau \mathbf{v} \cdot \nabla q \,, \tag{13.52}$$

with $\tau > 0$. By substituting (13.52) into (13.51), we find that

$$\frac{d \,\mathrm{IPV}\,(\theta)}{dt} = -\tau \int_S \int \rho_\theta (\mathbf{v} \cdot \nabla q)^2 dS \,, \tag{13.53}$$

This ensures that $\mathrm{IPV}\,(\theta)$ decreases with time. We can recover potential vorticity conservation and potential enstrophy conservation by setting $\tau = 0$. From (13.52), we see that

$$q - D = q - \tau \,(\mathbf{v} \cdot \nabla q) \equiv q_{\text{anticipated}} \,. \tag{13.54}$$

Here $q_{\text{anticipated}}$ can be interpreted as the value of q that we expect or "anticipate" by looking upstream to see what value of q is being advected towards us. Eq. (13.54) is equivalent to

$$\frac{q_{\text{anticipated}} - q}{\tau} = -\mathbf{v} \cdot \nabla q \,, \tag{13.55}$$

For this reason, the idea is referred to as the "anticipated potential vorticity method" to parameterize the effects of geostrophic turbulence in the momentum equation. Sadourny and Basdevant (1985) showed that their approach gives realistic kinetic energy and enstrophy spectra in a numerical model.

The point of this discussion is that it is possible to conceive of specific processes that dissipate potential enstrophy without dissipating kinetic energy. The particular approach suggested by Sadourny and Basdevant (1985) is used here as an example. It was applied by Ringler et al. (2011).

13.10 What does the blender blend?

To the extent that the eddies of the global circulation act like turbulence, it is natural to ask whether they "mix" things, and if so, which things. A variable that is mixed by turbulence is a conserved variable, i.e., it remains unchanged over time following fluid particles. Linear momentum is not very well conserved; it is subject to a variety of non-conservative effects, including the Coriolis acceleration and pressure gradients. The angular momentum $M \equiv$

$a\cos\varphi\,(u+\Omega a\cos\varphi)$, is somewhat more conservative, because it is not affected by the Coriolis acceleration; a little thought shows, however, that uniform angular momentum is impossible unless the uniform value is zero, because a finite angular momentum at the pole would imply infinite zonal wind and vorticity there.

We are thus led to look for evidence that the eddies mix the Ertel potential vorticity, which is a well conserved following fluid particles. In the absence of heating, particles stay on their isentropic surfaces, so we would expect the large-scale turbulence to homogenize potential vorticity (and other conservative variables) along isentropic surfaces. When we showed the zonally averaged PV in Chapter 4, we remarked that the PV is quite homogenous in the troposphere, compared to the stratosphere. Fig. 13.11, from Sun and Lindzen (1994), shows that in the midlatitude troposphere there is some tendency for zonally averaged PV contours to be parallel to zonally averaged potential temperature profiles.

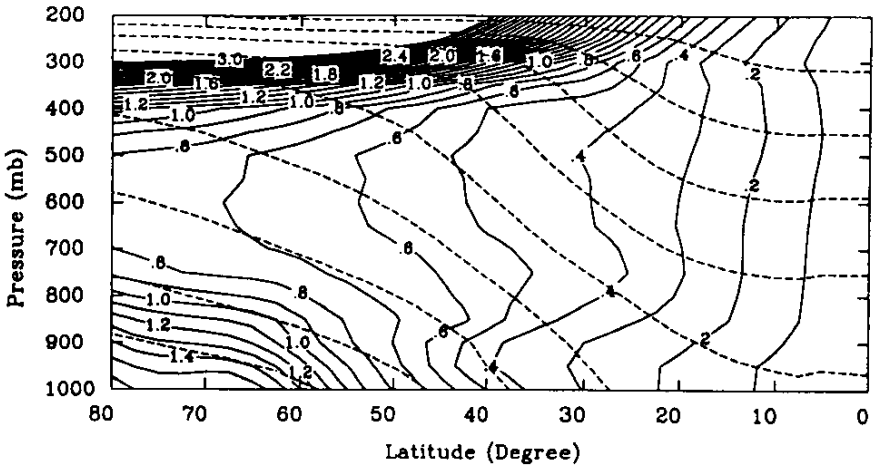


Figure 13.11: Potential vorticity (solid lines) and potential temperature (dashed lines) plotted as functions of pressure and latitude. From Sun and Lindzen (1994).

13.11 Sensitive dependence on initial conditions

Palmer (2024); Berry (2002); Palmer (2022)

When we make a weather forecast, we are solving an "initial value problem." The current state of the atmosphere is the "initial condition." The governing equations are integrated forward in time to predict the future state of the atmosphere.

Is it possible in principle to make a perfect (or arbitrarily accurate) forecast? Broadly speaking, there are three sources of forecast error:

- Properties of the atmosphere itself
  - The Uncertainty Principle of quantum mechanics (not important for this problem)
  - Nonlinearity and instability (very important). As discussed below, these lead to *sensitive dependence on initial conditions*.[1]
- Properties of the observing system
  - Imperfect measurements of the initial conditions (e.g., imperfect thermometers)
  - Imperfect spatial coverage of the initial conditions
  - Mistakes
- Properties of the model
  - Wrong equations
  - Imperfect resolution

We focus on the first and most fundamental source of error, i.e., properties of the atmosphere itself that limit predictability.

Lorenz (1963) discussed the predictability of a "deterministic nonperiodic flow." A system is said to be deterministic if its future evolution is completely determined by a set of rules. Models of the atmosphere (e.g., the primitive equations) are examples of sets of rules. The atmosphere is, therefore, a deterministic system. The behavior of the atmosphere is obviously nonperiodic; its previous history is not repeated. The predictability of periodic flows is a rather boring subject. If the behavior of the atmosphere was periodic, the weather would certainly be predictable!

How does nonperiodic behavior arise? The forcing of the atmosphere by the seasonal and diurnal cycles is at least approximately periodic. For linear systems, periodic forcing always leads to a periodic response. For non-linear systems, however, periodic forcing (including no forcing at all as a special case) can lead to a nonperiodic response. Nonperiodic behavior arises from nonlinearity.

Lorenz (1963) studied an idealized set of nonlinear convection equations, and found that for some values of the parameters all of the steady and periodic solutions are unstable. The model exhibits non-periodic solutions; again, a periodic solution is, by definition, predictable. The equations of the Lorenz's toy model are remarkably simple:

[1] "Had it been another day I might have looked the other way..." – Lennon and McCartney

$$\begin{aligned} \dot{X} &= -\sigma X + \sigma Y \ , \\ \dot{Y} &= -XZ + rX - Y \ , \\ \dot{Z} &= XY - bZ \ . \end{aligned} \tag{13.56}$$

Here

$$\sigma = 10 \ , b = 8/3 \ , \text{ and } r = 24.74 \tag{13.57}$$

are parameters, which are specified before the model is run. The numerical values given in (13.57) are particular choices (not unique ones) that lead to non-periodic behavior. A solution of (13.56) is shown in Fig. 13.12. The state of the model is plotted in a phase space. Most of the time the solution is near one of two "attractors," i.e., at a randomly chosen point in the integration the probability of finding the solution near one of the attractors is high. Occasionally the solution wanders from one attractor to the other. Partly because of the appearance of this plot, the solution is sometimes called the "Butterfly Attractor."

This example illustrates the important point that *even a simple nonlinear system can be unpredictable*. Lack of predictability and complex behavior are not necessarily due to complexity in the definition of the system itself.

Here is why there is a finite limit to deterministic predictability. Two slightly different states of the atmosphere diverge from each other with time because the atmosphere is unstable. See the sketch in Fig. 13.13. This leads to sensitive dependence on the initial conditions, on a scale-by-scale basis. Systems that exhibit sensitive dependence on initial conditions are called "chaotic." The sensitive dependence of the state of the atmosphere on its past history suggests that the flap of a butterfly's wings in China could noticeably change the weather in North America a few days later – a second reason for calling Fig. 13.13 the Butterfly Attractor.

There are many kinds of instability, acting on virtually all spatial scales. Small-scale inflectional instabilities act on scales of meters or less. Buoyant instabilities, including cumulus instability, occur primarily on scales of a few hundred meters to a few kilometers. Baroclinic instability occurs on scales of thousands of kilometers.

Although Lorenz discovered the importance of sensitive dependence on initial conditions, he was not the first to discover it; Poincaré (1912) recognized the phenomenon, and even discussed the fact that it makes long-term weather forecasting impossible. James Clerk Maxwell was also aware, during the 19th century, that because of instabilities deterministic physical laws do not necessarily permit deterministic predictions (Harman, 1998,

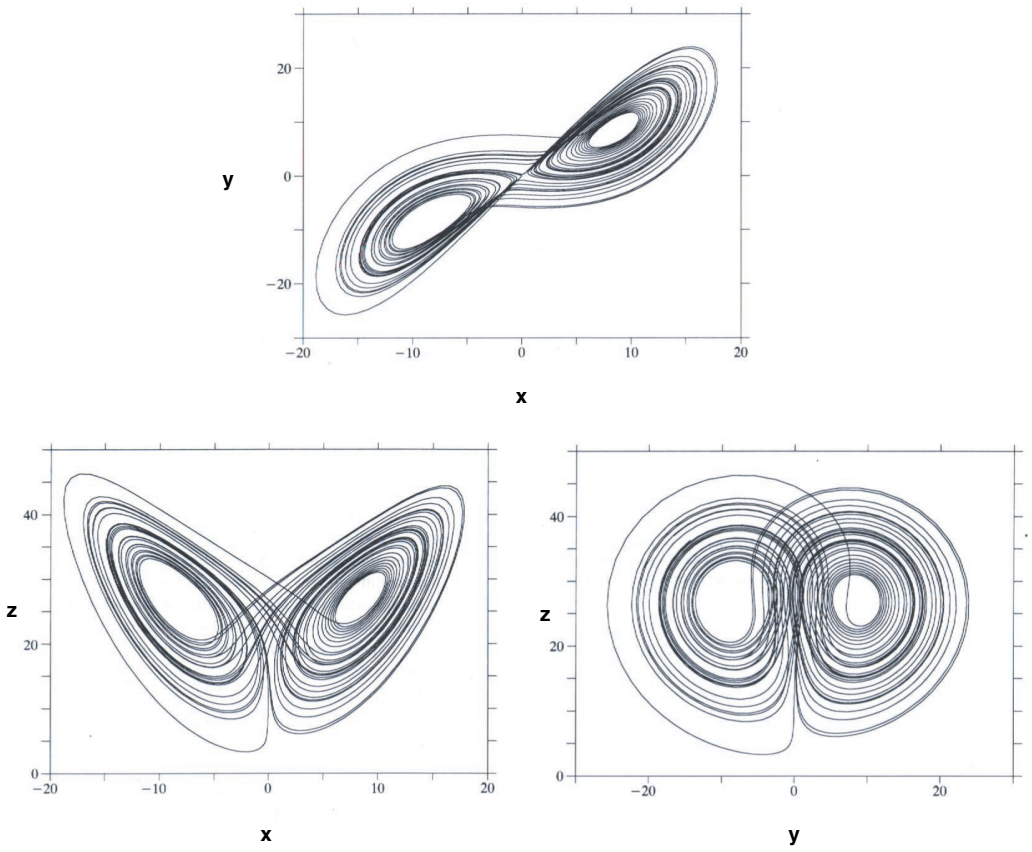


Figure 13.12: The Butterfly Attractor of Lorenz (1963), obtained as the solution of (13.56). From Drazin (1992).

pp. 206-208). Lorenz was the first to realize that the phenomenon of sensitive dependence on initial conditions permits complex unpredictable behavior to occur even in very simple systems. He emphasized the importance of nonlinearity, in addition to instability.

Recall that small-scale eddies produce fluxes that modify larger scales. In this way, errors on small scales can produce errors on larger scales through nonlinear processes, as discussed earlier in this chapter. Recall that scales cannot interact in linear systems.

We conclude that it is the combination of instability and nonlinearity that limits our ability to make skillful forecasts of the largest scales of motion. This idea is illustrated in Fig. 13.14. Both instability and nonlinearity are properties of the atmosphere itself; we cannot make them "go away" by improving our models or our observing systems. It is the properties of the atmosphere itself that lead to an intrinsic limit of deterministic predictability. We can say that sensitive dependence on initial conditions imposes a "predictability time" or "predictability limit" beyond which the weather is unpredictable in principle. This limit is a property of the atmosphere itself, and so it cannot be circumvented by improving our

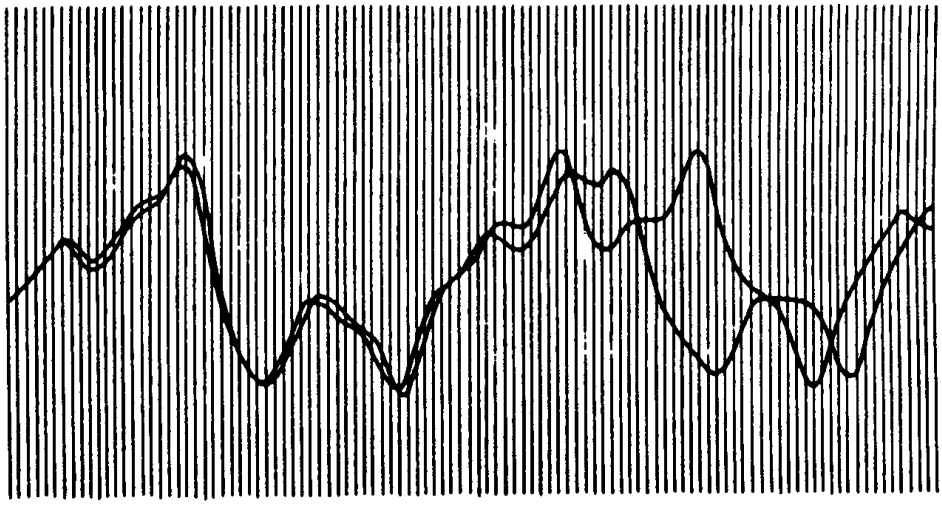


Figure 13.13: How two weather patterns diverge. From nearly the same starting point, Lorenz noticed that two numerical solutions grew farther and farther apart until all resemblance disappeared. From Gleick (2008).

forecast methods; it is not a limitation that is intrinsic to numerical weather prediction or any other specific forecast technique. It applies to all methods.

Errors on smaller scales double (grow proportionately) faster than errors on larger scales, simply because the intrinsic time scales of smaller-scale circulations are shorter. For example, the intrinsic time scale of a buoyant thermal in the boundary layer might be on the order of 20 minutes, that of a thunderstorm circulation might be on the order of one hour, that of a baroclinic eddy might be two or three days, and that of planetary wave number one might be one to two weeks. Because the predictability time is related to the eddy turn-over time, the predictability limit is a function of scale; larger scales are generally more predictable than smaller scales.

When we eliminate errors on smaller spatial scales by adding more observations, the range of our skillful forecasts is increased by a time increment approximately equal to the predictability time of the newly resolved smaller scales. Pushing the initial error down to smaller and smaller spatial scales is, therefore, a strategy for forecast improvement that yields diminishing returns. See Fig. 13.15.

Here is an example: Suppose that you forecast the temperature at your house, and that the forecast error grows exponentially with the length of the forecast. We can write

$$T_{\text{forecast}}(t) = T_{\text{true}}(t) + E_0 e^{\lambda t} \ , \tag{13.58}$$

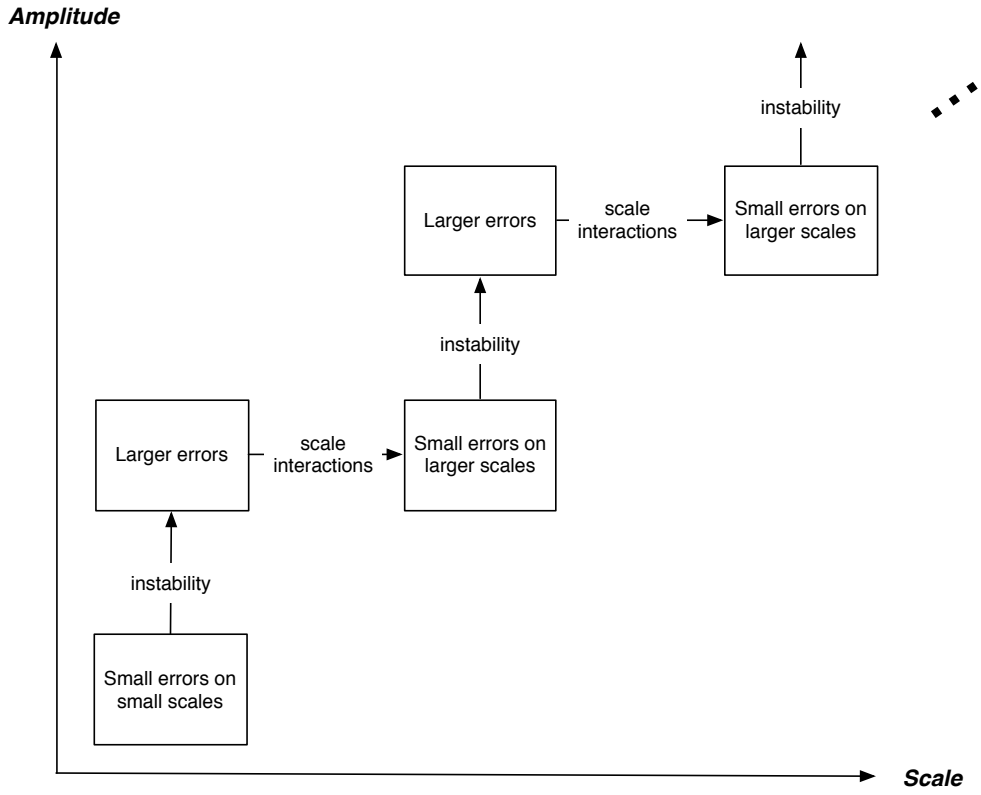


Figure 13.14: The road to chaos. Sketch illustrating the roles of instability in leading to error growth, and of nonlinearity in leading to the spreading of error from small scales to larger scales.

where E_0 is the initial error, t is the length of the forecast, and λ is the exponential growth rate. If $E_0 = 0$ (no initial error), then the error remains zero for all time. We can reduce the error at a given forecast range, say 5 days, by reducing the initial error. Another way of saying this is that, by reducing the initial error, we can postpone the time at which the error attains a certain "reference" value, $E_{\rm ref}$. For example, suppose that

$$E_{\rm ref} = E_0 e^{\lambda t} = \frac{E_0}{2} e^{\lambda (t+\Delta t)} \ . \tag{13.59}$$

According to this formulation, cutting the initial error in half postpones the time when the error grows to by an amount Δt, which is a measure of the gain in forecast skill, and satisfies

$$\Delta t = \frac{\ln(2)}{\lambda} \tag{13.60}$$

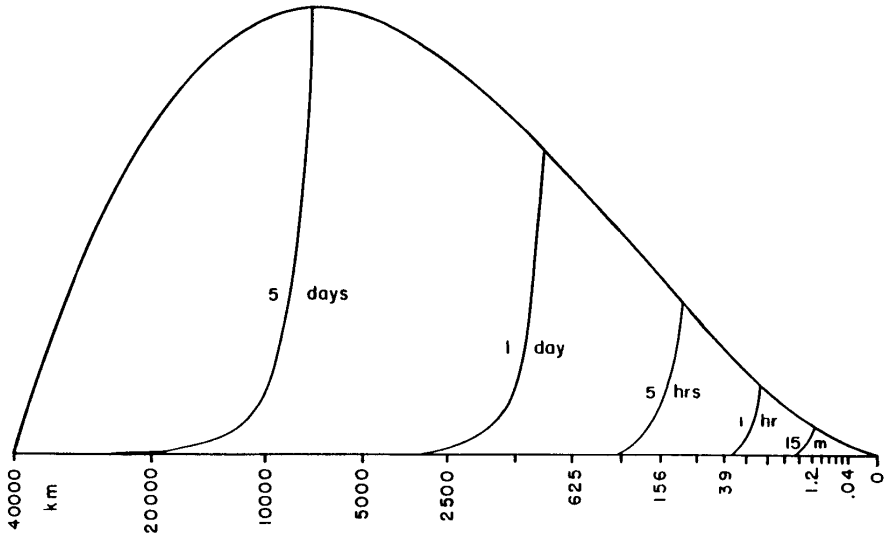


Figure 13.15: A simplified depiction of the energy spectrum $E(k)$ (upper curve), and the error-energy spectra (lower curves) at 15 minutes, 1 hour, 5 hours, 1 day, and 5 days, as interpolated from the result of a numerical study. The lower curves coincide with the upper curve, to the right of their intersections with the upper curve. Areas are proportional to energy. From Lorenz (1969).

A large value of Δt is "good," because it means that the error grows slowly. In order to have a large Δt, we need a small error growth rate, λ.

The growth rate is expected to be larger (faster) for smaller spatial scales, and smaller (slower) for larger spatial scales. Qualitatively, we can write

$$\lambda = V/d\,, \tag{13.61}$$

where V is an "eddy velocity scale" and d is the linear spatial scale (e.g., the radius or "size") of the eddy. We can interpret $1/\lambda$ as the time required to travel distance d at speed V, so it is roughly the time required for an air parcel to make one trip around the eddy. We will call this the "eddy turnover time." Substituting into our formula for Δt gives

$$\Delta t = \frac{d\ln(2)}{V}\,. \tag{13.62}$$

This shows that the gain in forecast skill is less for small values of d, for a given value of V.

Most of the error resides on small spatial scales because those scales are inadequately sampled. The sampling error can be reduced by increasing the density of the observing network. In two dimensions, quadrupling the number of observations cuts the average distance between stations in half. Such an improvement in the observing network would reduce d, i.e., the linear spatial scale of the smallest eddies that are "resolved" by the observing network, by a factor of two.

Suppose that, starting from initial observing network, we improve the network by quadrupling the number of points. This increases the cost of the network by a factor of four, and reduces d by a factor of two. We gain forecast skill by an amount $(\Delta t)_1$. Then we quadruple the number of points again, increasing the cost of the network by another factor of four. The forecast improvement is measured by $(\Delta t)_2 = \frac{1}{2}(\Delta t)_1$, i.e., it is only half as large as $(\Delta t)_1$. Further costly improvements to the network will result in progressively smaller improvements in forecast skill. At some point, further improvements to the network are not practical.

As discussed below, estimates show that small errors on the smallest spatial scales can grow in both amplitude and scale to significantly contaminate the largest scales (comparable to the radius of the Earth) in about two to three weeks. Some aspects of atmospheric behavior may nevertheless be predictable on longer time scales. This is particularly true if they are forced by slowly changing external influences. An obvious example is the seasonal cycle. Another example is the statistical character of the weather anomalies associated with long-lasting sea surface temperature anomalies, such as those due to El Niño. This point will be discussed further, later in this chapter.

At this point, we can offer a definition of turbulence: A turbulent circulation is one that is not predictable on the time scale of interest. For example, midlatitude winter storms can be considered as turbulent eddies if we are interested in seasonal time scales, but they behave as highly predictable, orderly circulations if we are doing a one-day forecast. From this point of view, turbulence is in the eye of the beholder.

As an interesting analogy, the orbits of the planets and moons of our Solar System are well known to be highly predictable; for example, eclipses can be predicted very accurately many years in advance. Nevertheless, the Solar System is known to be chaotic on longer time scales (e.g., Laskar, 1989; Sussman and Wisdom, 1992). On relatively short time scales (up to millions of years!), the motions of the planets are predictable, like laminar flow. On longer time scales, the motions of the planets are unpredictable, like turbulent flow.

13.12 Quantifying the limits of predictability

Three approaches to determine the limits of predictability were discussed by Lorenz (1969):

13.12.1 The dynamical approach

In the dynamical approach, two or more model solutions are produced, starting from similar but not quite identical initial conditions. This procedure is similar to what Lorenz did by accident, when he discovered sensitive dependence on initial conditions. The model is put in place of the atmosphere; no real data is used. Problems with this approach are: 1) truncation error; 2) imperfect equations; 3) lack of information about very small scales. Studies of this type suggest that the doubling time for small errors with spatial scales of a few hundred km is about 5 days. This implies that the limit of predictability is about two weeks.

One of the earliest examples of the dynamical approach is the study of Charney et al. (1965). They used several atmospheric global circulation models (GCMs) to study the growth of small perturbations. Fig. 13.16 shows some of their results, obtained with an early version of the UCLA GCM. The root-mean-square temperature error grows in both hemispheres, but more rapidly in the winter hemisphere, where the circulation is more unstable. The error does not continue to grow indefinitely; it stops growing at the point where the "forecast" is no better than a guess. We say that the error has "saturated." A suitable guess might consist of a state of the system chosen at random from a very long record of such states; think of pulling a weather map at random out of a huge file full of many such maps.

A later example of the dynamical approach is discussed by Shukla (1981, 1985). The results shown in Fig. 13.17 are based on computations with a GCM, as reported by Shukla (1985). The model is used to perform multiple simulations, differing only through very small perturbations of the initial conditions. The two panels on the left side of the figure show the growth and saturation of errors in the sea level pressure for winter and summer. The winter errors grow more rapidly than the summer ones. The errors are large in middle latitudes, where baroclinic instability is active, especially in the winter. The errors are much smaller in the tropics.

Recall from Chapter 3, however, that the tropical sea level pressure normally does not vary much. This means that a small error in the tropics can be important. To take this into account, the two panels on the right show the errors normalized by the temporal standard deviation of the sea level pressure. From this perspective, we see that the tropical errors actually grow more rapidly than those of middle latitudes, and saturate at about the same (normalized) values. At all latitudes the saturation values are close to one. This means that the errors stop growing when they become as large as the temporal standard deviation based on day-to-day variability.

Fig. 13.18 is taken from Shukla (1981). For numerical experiments on the growth of errors, similar to those discussed above, the figure shows how error growth varies with zonal wave number. In each panel, the solid curve shows the growth of error in the numerical model, and the dashed curve shows the corresponding growth of error when the "forecast" used is simply persistence. We can say that the model forecast is no longer skillful when it

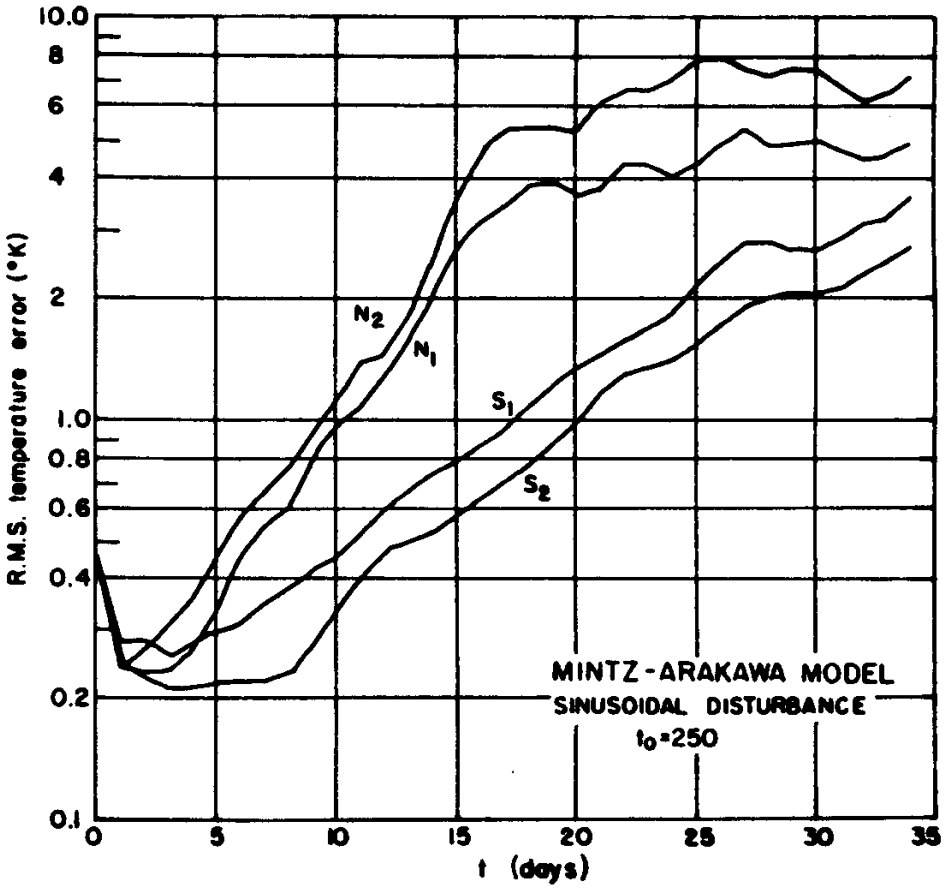


Figure 13.16: Root mean-square temperature error in January simulations performed with the two-level Mintz-Arakawa model. “N” and “S” denote the Northern and Southern Hemispheres, respectively. The subscripts “1” and “2” denote the two model levels. From Charney et al. (1966).

is no better than a forecast based on persistence. After 30 days, the numerical model still has some skill at the lower wave numbers, which correspond to the largest spatial scales. For higher wave numbers, the model’s skill disappears more quickly.

13.12.2 The empirical approach

In the empirical approach, the atmosphere is put in place of a model; the atmosphere itself is used to predict the atmosphere. Lorenz (1969) examined the observational records of the 200, 500, and 850 hPa heights, for the Northern Hemisphere only. He searched for pairs of similar states, or “analogs,” occurring within one month of the same day of the year. He chose 30 December 63 and 13 January 65 as the best available analogs within a five-year record.

The results of the empirical approach show that the doubling time for errors on the

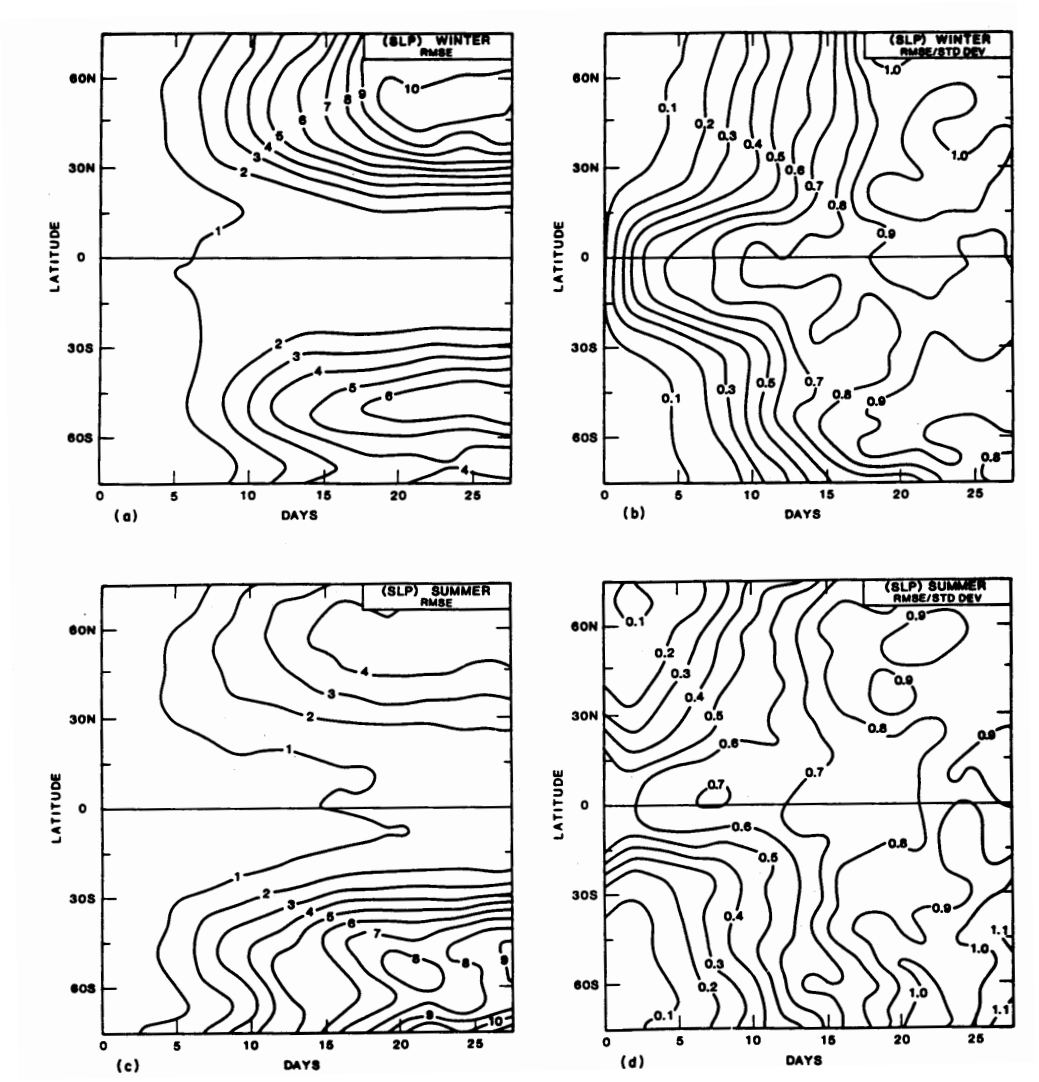


Figure 13.17: IPVonal average of root-mean-square error (RMSE; left panels) and ratio (RMSE/STD; right panels) of root mean square and standard deviation of daily values for sea-level pressure. (a) RMSE and (b) RMSE/STD for six pairs of control and perturbation runs during winter; (c) RMSE and (d) RMSE/STD for three pairs of runs during summer. From Shukla (1985).

smallest resolved scales is less than eight days. They are thus reasonably consistent with the results of the dynamical approach.

Among the problems with the empirical approach are: 1) it is hard to find "good" analogs, in that the smallest error is about 1/2 of the average error; 2) we cannot experiment with the initial error, because we have to take what nature gives us; and 3) the data cannot be used to study the growth of errors on very small scales, simply because such scales are not adequately observed.

Recently, machine learning has been used to produce forecast systems that have been

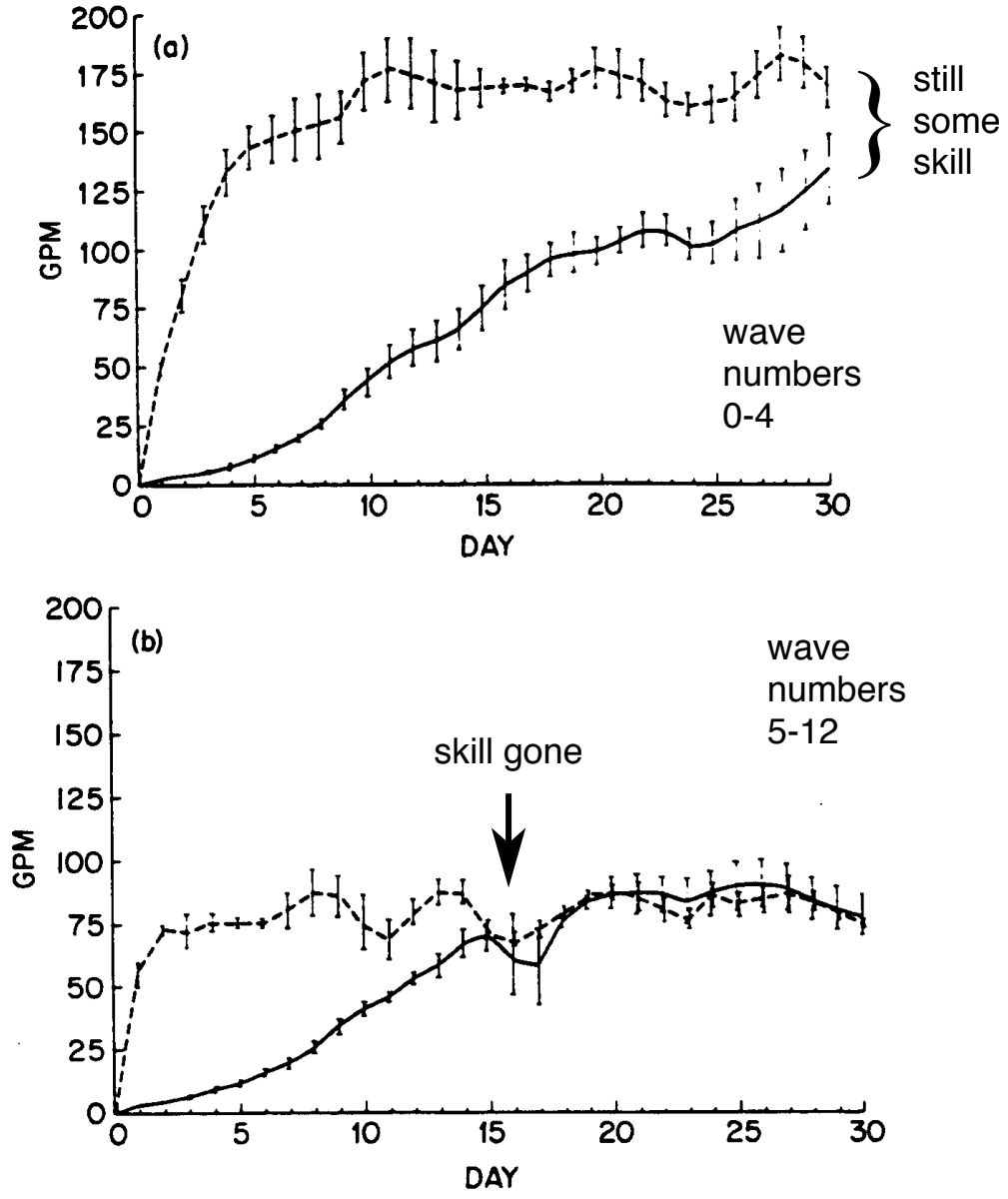


Figure 13.18: Root-mean-square error (solid line) averaged for six pairs of control and perturbation runs and averaged for latitude belt 40-60°N for 500-hPa height for (a) wave numbers 0-4 and (b) wave numbers 5-12. Dashed line is the persistence error averaged for the three control run. Vertical bars denote the standard deviation of the error values. From Shukla (1981).

"trained" on many years of weather data. These systems are modern examples of the empirical approach, and they are quite successful. It's a shame that Lorenz did not live long enough to see this development.

13.12.3 The dynamical-empirical approach

The "dynamical-empirical" approach is the most difficult of the three to understand. The first version, as described by Lorenz (1969), uses a model to develop an equation for the time change of "error energy." We then Fourier transform this error energy equation, so that the spectrum of the kinetic energy appears. The key step is to prescribe this kinetic energy spectrum from observations, down to very small scales. That's where the empiricism comes in. We then use the resulting partially empirical equation to draw conclusions about error growth. Lorenz (1969) used this approach to infer that errors in the smallest scales amplify most quickly and soon dominate. Like the dynamical and empirical approaches, the dynamical-empirical approach suggests that the limit of deterministic predictability is about two to three weeks.

A more modern example of the dynamical-empirical approach was described by Lorenz (1982). Suppose that we make a large number of one-day and two-day forecasts, as shown in Fig. 13.19. Let $z_{i,1}$ be the 1-day forecast for z on day i, and let $z_{i,2}$ be the 2-day forecast for z on day i (the same day). Let $E_{1,2}$ be the root-mean-square (rms) difference between the 1-day and 2-day forecasts for the same day, averaged over the globe and over all verification days:

$$E_{1,2}^2 = N^{-1} S^{-1} \sum_{i=1}^{N} \int_S (z_{i,1} - z_{i,2})^2 dS \,. \tag{13.63}$$

Here the integral is over the area, S, and the sum is over the verification days, which are distinguished by subscript i. We can interpret $E_{1,2}$ as the average or typical growth of the forecast error between the first and second days of the forecasts. If all one- and two-day forecasts were perfect, $E_{1,2}$ would be zero. If the one-day forecasts were perfect but the two-day forecasts were not, $E_{1,2}$ would be positive.

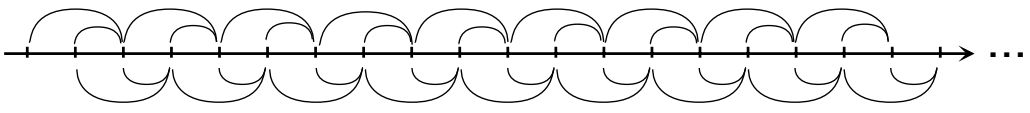


Figure 13.19: A long sequence of one-day and two-day forecasts.

More generally, we can compute the rms error growth between the j-day forecast and the k-day forecast, still verifying on the same day (i.e., day i), from

$$E_{j,k}^2 = N^{-1} S^{-1} \sum_{i=1}^{N} \int_S \left(z_{i,j} - z_{i,k} \right)^2 dS \,. \tag{13.64}$$

We assume without loss of generality that $k \geq j$. It should be clear that $E_{j,k}$ compares j-day forecasts with k-day forecasts for the same day. If $E_{j,k} > 0$, the implication is that k-day forecasts are less skillful, on the average, than j-day forecasts. This is, of course, expected. Note that $E_{0,k}$ compares "0-day forecasts," which are actually analyses rather than forecasts, with k-day forecasts. If the k-day forecasts were perfect, $E_{0,k}$ would be zero. Because of the intrinsic limit of predictability and the inevitable small errors of the initial conditions, however, even a perfect model will give $E_{0,k} > 0$, for $k > 0$. Also, $E_{0,k}$ increases initially as k increases, but then it "saturates" when the k-day forecast becomes no better than a guess, as shown in Fig. 13.20. Even if the model's simulated climate is perfect, $E_{0,k}$ will be positive because of the limit of deterministic predictability.

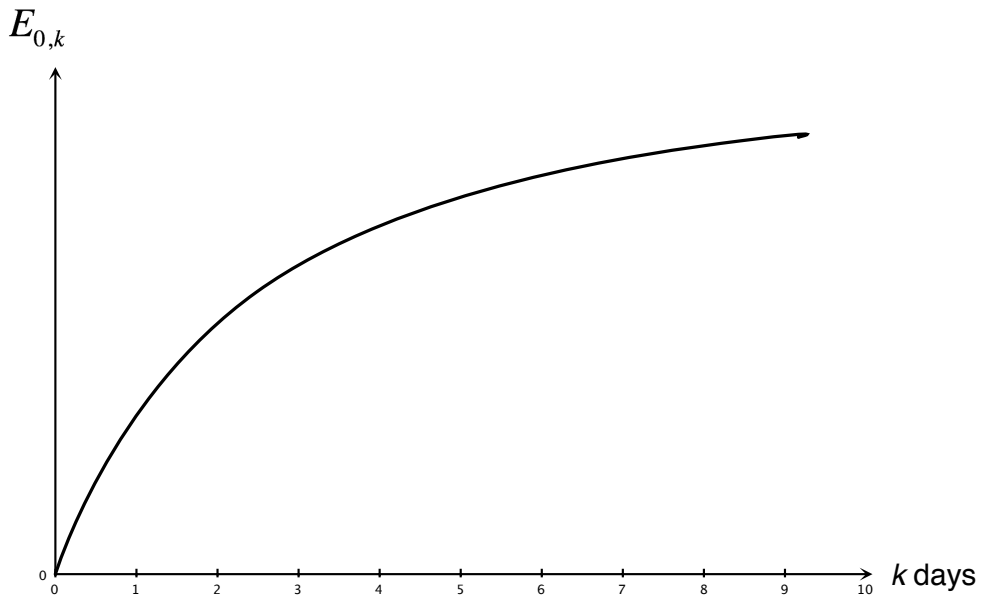


Figure 13.20: The growth of with time. Rapid initial growth is followed by saturation.

As k becomes large, we expect $E_{j,k}$ to saturate, no matter what the value of j is. For $k \to \infty$, you might expect that "should be" independent of j. It is not, at least not for "small" j (< 20). The reason is that the model's climate is different from the real climate. Since the model is started from real data, a j-day forecast looks like the real world when j is small. As j increases, the model goes to its own climate, and so the forecast increasingly departs from the ensemble of real-world states. Note, however, that $E_{j,k}$ becomes independent of j when both j and k are large, because then we are comparing the model's climate with itself.

A special case, illustrated in Fig. 13.21, compares $(k-1)$-day forecasts with k-day forecasts for the same day. For example, for $k=10$, $E_{9,10}$ compares nine-day forecasts with ten-day forecasts for the same day. Of course, nine-day forecasts and ten-day forecasts are both pretty bad (usually), and so generally speaking they will be quite different from each other, even though they are supposed to represent the weather on the same day. For this reason, as $k \to \infty$, $E_{k-1,k}$ is not small. Instead, as $k \to \infty$, we expect $E_{k-1,k}$ to approach a constant, which is a measure of the variability of z in the model's climate. The same is true for any other $E_{j,k}$, when $k-j$ is held fixed as k increases.

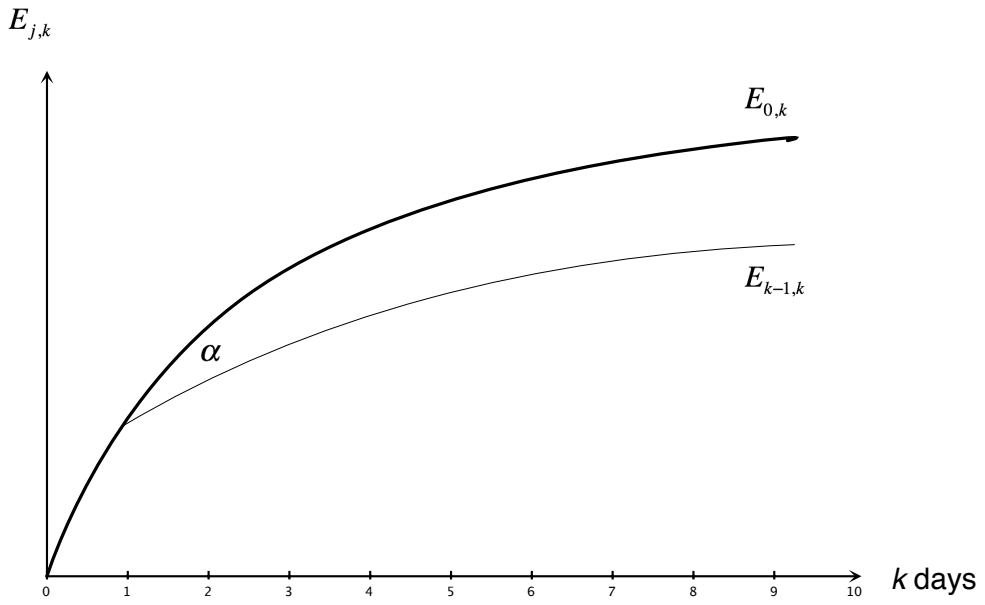


Figure 13.21: The growth of $E_{0,k}$ and $E_{k-1,k}$ with time.The thin line starts at $k=1$ because $E_{k-1,k}$ is not defined for $k<1$. The angle α measures how much the model can be improved by reducing the systematic error.

For a perfect model, we would have

$$\lim_{k\to\infty} E_{0,k} = \lim_{k\to\infty} E_{k-1,k} \ . \tag{13.65}$$

because the model's climate would be identical to the true climate. For an imperfect model, on the other hand, we expect

$$\lim_{k\to\infty} E_{0,k} > \lim_{k\to\infty} E_{k-1,k} \ . \tag{13.66}$$

because the model's climate differs from the true climate. This is a key point.

To explore these ideas with real data, Lorenz chose 1 Dec 80 - 10 March 1981 (00 days) as the verification dates. Because he worked with 100 verification days, i.e., $N = 100$ in (13.64). He used an archive of forecasts and analyses performed with the ECMWF model. Since ECMWF routinely makes ten-day forecasts, he had 100 one-day forecasts, 100 two-day forecasts, etc., out to 100 ten-day forecasts. He plotted $E_{j,k}$ as shown in Fig. 13.22. Each point in the figure represents an average over 100 pairs of forecasts.

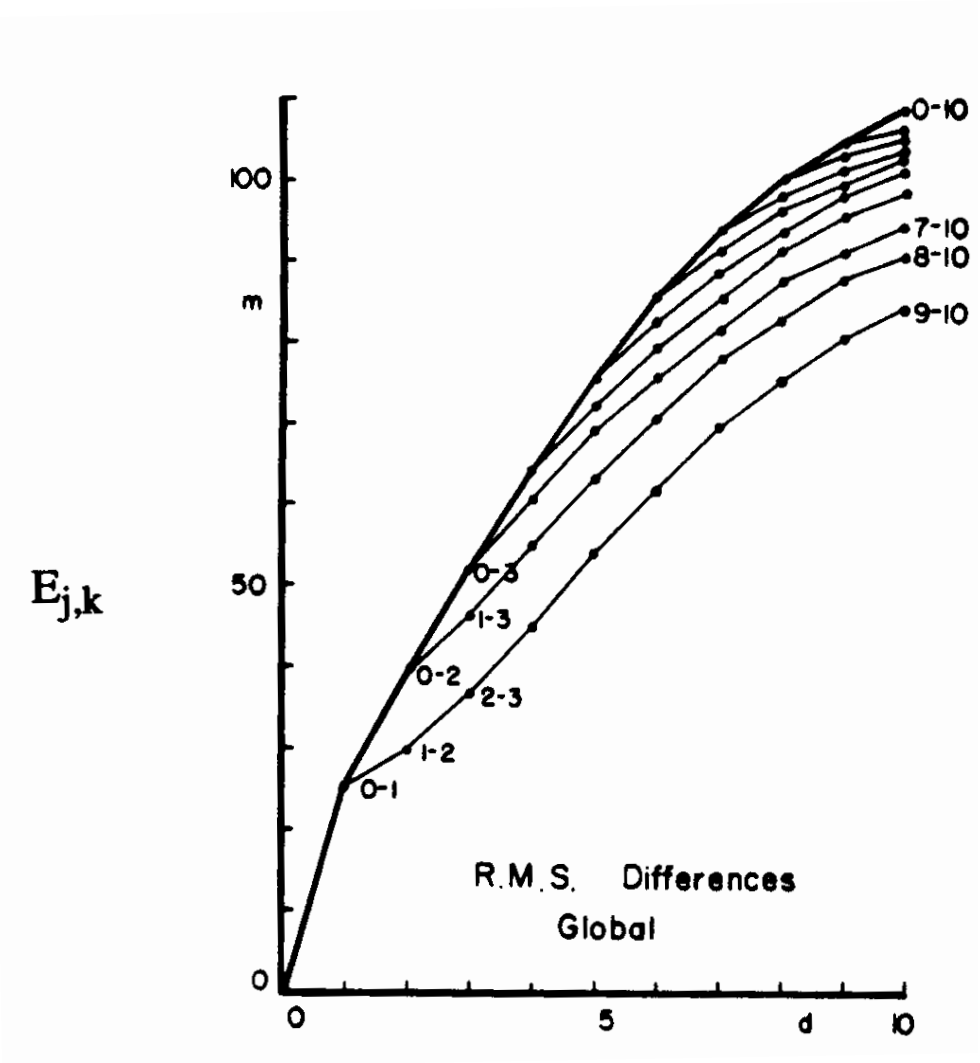


Figure 13.22: Global root-mean-square 500-mb height differences, in meters, between j-day and k-day forecasts made by the ECMWF operational model for the same day, for $j < k$, plotted as functions of k, out to $k = 10$. Values of (j,k) are shown beside some of the points. The heavy curve connects values of $E_{0,k}$. The thin curves connect values of $E_{j,k}$ for constant $k - j$. From Lorenz (1982).

Lorenz noticed that the thin curves in Fig. 13.22 appear to run nearly parallel to each other, in the sense that $dE/dk \cong f(E)$. This means that the rate of error growth is mostly determined by the size of the error, as expected given that the error saturates as $k \to \infty$. Of course we know that the size of the error cannot be the only factor involved, but it can be the dominant one.

Lorenz wanted to estimate the growth rate of very small errors in forecasts performed with a perfect model, even though he did not have a perfect model. To do this, he essentially fit a curve to the error growth rates in the ECMWF forecasts. He hypothesized that, for a perfect model

$$\frac{dE}{dk} = aE - bE^2 \,. \tag{13.67}$$

Here $E \equiv E_{j,k}$, where $k - j =$ constant, and so $dE/dk \cong \left(E_{j+1,k+1} - E_{j,k}\right)$ per day. The second index minus the first index is the same for both E's, so we are following the "thin curves" in Figure 13.22. Eq. (13.67) is consistent with the hypothesis that the rate of error growth is determined by the size of the error. For small E, the exponential growth rate is a. For larger E, saturation occurs so that

$$E_{sat} = a/b \,. \tag{13.68}$$

This means that a/b is the maximum error of the perfect model. Note from (13.67) that if the initial error is zero, there will be no error growth. That is what we expect from a "perfect model."

We want to deduce a, the growth rate of small errors, from the data, even though the data do not contain truly small errors. We can do this by fitting the function $f(E) = aE - bE^2$ [i.e., the right-hand side of (13.67)] against dE/dk as evaluated from the data. See Fig. 13.23. In performing this curve fit, we find the "best" values of a and b in the sense that the rms error is minimized. This calculation is an example of the dynamical-empirical approach to determine the growth rate of small errors, because both a model and data have been used.

Fig. 13.24 shows the same data as Fig. 13.22, plotted in a different way. The dots correspond to the thin curves in Fig. 13.22, and so represent the "perfect model." The X's correspond to the thick curve in Fig. 13.22, and so represent the "imperfect model." The conclusion of this exercise is that the doubling time for small errors on the smallest scales resolved by the model is 2.4 days.

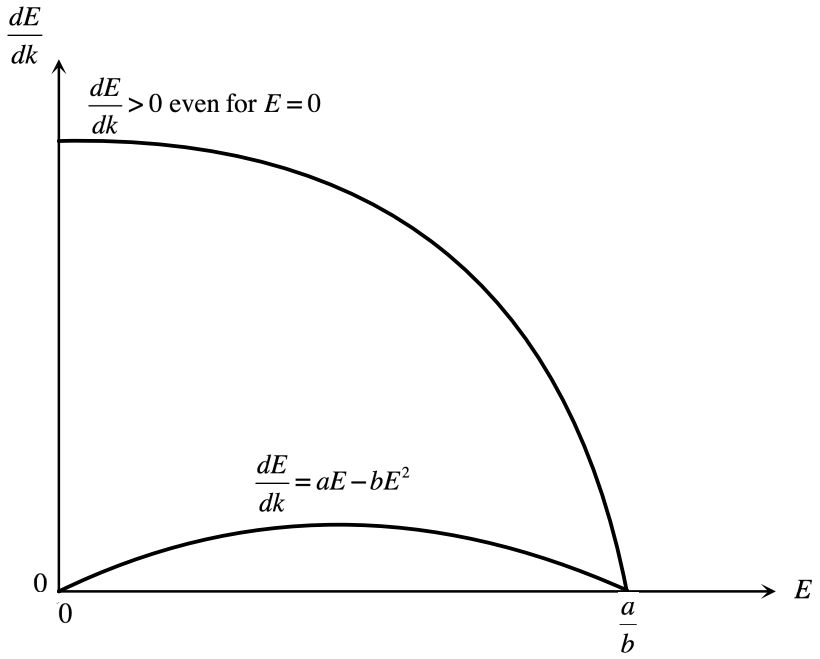


Figure 13.23: Diagram illustrating the expected variation of the error growth rate with the magnitude of the error. The lower curve shows the expected behavior for a perfect model, and the upper curve shows the expected behavior for a real, imperfect model.

In a later publication, Lorenz discussed a set of more skillful forecasts made with an improved model and initialized using better observations. He found that the heavy curve Fig. 13.22 shifted to lower values, but the thin curves also came down. The gap between them was cut in half. The downward shift of the heavy curve, by itself, suggests that the newer model's simulated climate was more realistic than that of the older model. Alternatively, it could mean that the analyses are more realistic. The downward shifts of the thin lines suggest that the newer model's weather is less active than that of the old model.

13.13 Ensemble forecasting

Since 1992, major weather prediction centers, including ECMWF and the U.S. National Centers for Environmental Prediction, have run forecast ensembles as a key element of their operational systems. A forecast ensemble consists of multiple forecasts that are started from slightly different initial conditions. The approach is useful because the different ensemble members diverge as the forecasts proceed, because of sensitive dependence on initial conditions. For any given location, the magnitude of the "spread" is an indication of the reliability of the forecast; a larger spread inspires less confidence. Fig. 13.25 shows an example of an ensemble of forecasts performed at the National Centers for Environmental Prediction.

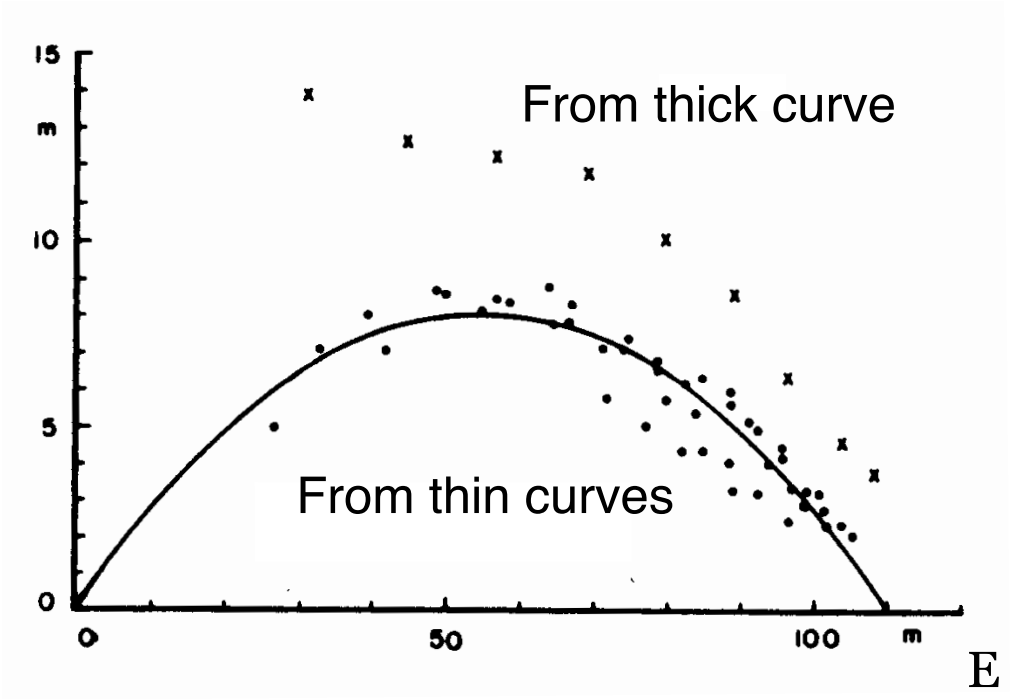


Figure 13.24: Increases in global root-mean-square 500-hPa height differences, , plotted against average height differences, in meters, for each one-day segment of each thin curve in Fig. 13.22 (large dots), and increases plotted against average differences, for each one-day segment of the heavy curve in Fig. 13.22 (crosses). The parabola of "best fit" to the large dots is shown. From Lorenz (1982).

13.14 The response of the atmosphere to changes in sea surface temperature

Over the past few decades, it has become clear that the statistics of the weather respond in predictable ways to changes in sea surface temperature and other surface boundary conditions. Fig. 13.26 shows an early example of research on this topic, taken from Lau (1985). The figure shows temporal variations of monthly indices of the observed sea surface temperature (top) and the simulated zonal wind at 200 and 950 hPa, precipitation, east-west sea level pressure gradient across the South pacific, 200 hPa height and an index of what is called the "Pacific-North American pattern," as obtained in two 15-year model simulations performed with an atmospheric GCM. The two runs are shown in the left- and right-hand panels, respectively. The smooth curves superimposed on the various time series were obtained using a kind of running mean.

The imposed sea surface temperature changes were exactly the same in the two runs, which differed only through perturbations of their initial conditions. The main point here is that the two runs give very similar results, showing that the statistics indicated were con-

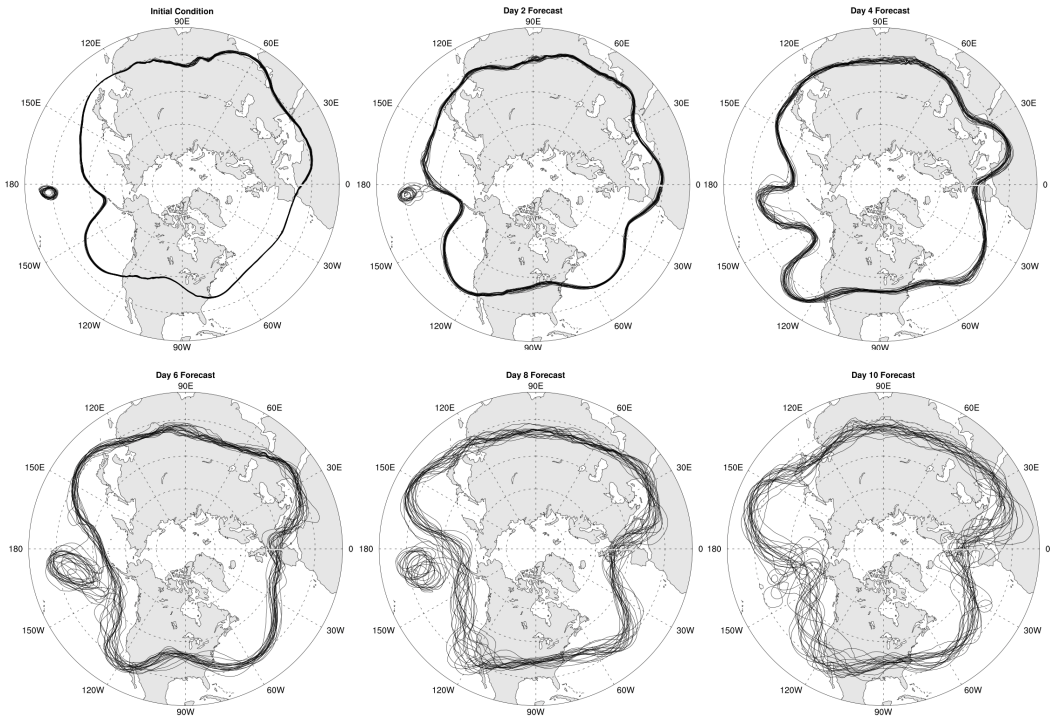


Figure 13.25: Actual forecasts of the shapes of two selected height contours for a pressure surface in the middle troposphere. The plots look down on the North Pole, and North America is at the bottom center of each panel. The forecast times shown are the initial conditions (top left), and then, proceeding to the right across the upper row, and then left to right across the bottom row, every two days after that, out to ten days. The two heavy green curves show the climatological positions of the height contours, which change very little over ten days. The two heavy yellow and grey curves show the predicted positions of the contours in "control runs" with a high-resolution version of the forecast model. the thin blue and red curves show the positions of the contours as predicted in an ensemble of forecasts that are started from initial conditions that differ very slightly from those of the control run. The ensemble is run with a lower-resolution version of the same model. For obvious reasons, these are called "spaghetti diagrams."

trolled by the year-to-year variations of the sea surface temperature, and were predictable far beyond the limit of deterministic predictability for individual weather events.

Lau's (1985) study was ground-breaking and computationally expensive at the time, but similar calculations have been routine since the 1990s, in the context of the Atmospheric Model Intercomparison Project, or AMIP (Gates, 1992). Today, AMIP simulations are a routine way of evaluating the performance of GCMs.

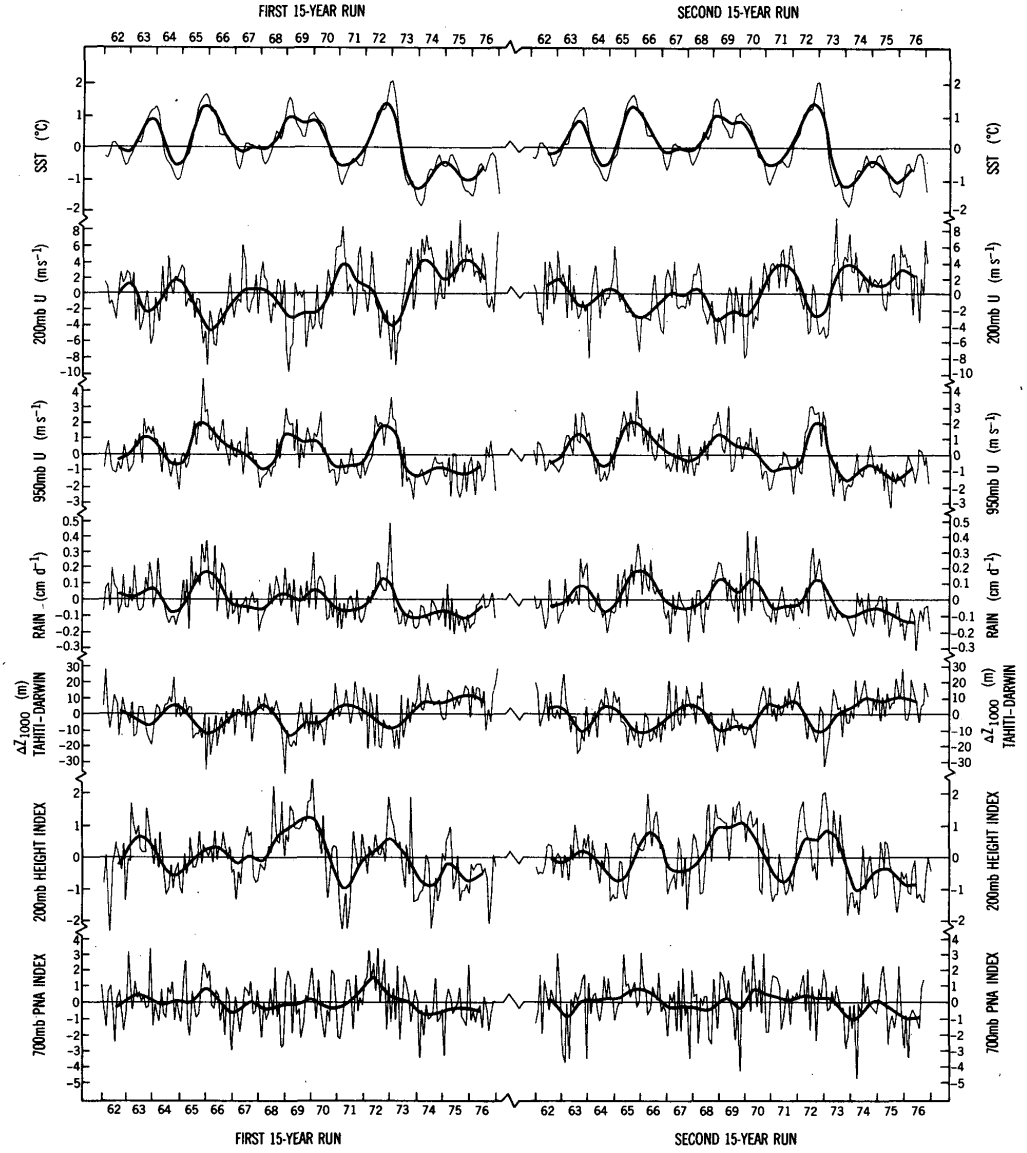


Figure 13.26: Temporal variations of monthly indices of sea surface temperature, zonal wind at 200 and 950 hPa, precipitation, east-west sea level pressure gradient across the South Pacific, 200 hPa height and the Pacific-North American pattern, for the first (left half of figure) and second (right half) 15-year model runs. The smooth curves superimposed on these time series are obtained using a running mean. From Lau (1985).

13.15 Segue

A circulation is turbulent to the extent that its predictability time is shorter than the time scale of interest to us. The fundamental cause of turbulence is inflectional instability. Vortex stretching leads to a kinetic-energy cascade in three-dimensional turbulence. In contrast, enstrophy cascades in two-dimensional turbulence, while kinetic energy anti-

cascades. The Rhines scale limits the meridional growth of the turbulent vortices, which become elongated into pairs of zonal jets.

Deterministic weather prediction is impossible beyond about two weeks. It may be possible to predict some statistics of the weather for a month or a season ahead, because of the external forcing due to persistent boundary anomalies. This appears to be most feasible in the tropics, because of the favorable signal-to-noise ratio. Longer term climate anomalies are also predictable provided that they are driven by an external forcing which is itself predictable, and provided that the response to this forcing is large enough to detect against the climatic noise due to natural variability and competing external forcings.

The ocean circulation is also chaotic but has a much longer predictability time than the atmosphere; therefore, to the extent that the sea surface temperature is predictable, and to the extent that the statistics of the weather are influenced by the sea surface temperature, the statistics of the weather are predictable beyond the predictability limit of the atmosphere itself.

Even a chaotic system responds in a statistically predictable way to sufficiently strong external forcing; this is why summers are predictably warm and winters are predictably cold. The climate responds in a predictable way to a sufficiently strong external forcing, provided of course that the forcing itself is predictable. Chapter 11 deals with the ongoing and future response of the global circulation to an external forcing that is gradually becoming stronger.

13.16 Problems

1. (a) Show that in two-dimensional flow the variance of a passive scalar, per unit wave number, varies as k^{-1}.

   (b) Show that in two-dimensional flow the geopotential variance, per unit wave number, varies as k^{-5}.

2. Consider an idealized state in which a two-dimensional atmosphere is at rest with respect to the rotating Earth. Imagine that, starting from this state of rest, the absolute vorticity is mixed throughout the atmosphere so that it becomes uniform. Plot the resulting meridional distributions of the zonal wind and the absolute angular momentum per unit mass. Compare the angular momentum of the uniform-vorticity state with the angular momentum of the resting state.

3. Consider Lorenz's butterfly model:

$$\begin{aligned} \dot{X} &= -\sigma X + \sigma Y \ , \\ \dot{Y} &= -XZ + rX - Y \ , \\ \dot{Z} &= XY - bZ \ . \end{aligned} \tag{13.69}$$

Here σ, b and r are constants. Assume $r \geq 1$, $b \geq 1$, and $\sigma \geq 1$.

(a) Find the three equilibrium solutions for X, Y, and Z.

(b) Discuss how the stability of the simplest equilibrium depends on the values of σ, b, and r.

Chapter 14

The effects of changes in external forcing

14.1 Forced chaotic systems

We can define "weather" as the instantaneous distribution of the atmospheric state variables, and "climate" as the long-term statistical properties of the same variables. The ocean, land-surface, cryosphere, and biosphere influence the climate. Together with the atmosphere, they form the climate system. All parts of the climate system interact, on sufficiently long time scales.

If long-range weather prediction is impossible, how can climate prediction be contemplated at all? If all memory of the initial conditions is "forgotten," what is the point of solving an initial value problem? Two factors make seasonal forecasting and/or climate change prediction possible. First, the Earth system has components with very long memories, notably the ocean.

Second, the system responds in systematic and statistically predictable ways to changes in the external forcing. This means that it is possible to make a prediction without solving an initial-value problem! A good example is the seasonal cycle. If we predict systematic differences in weather between summer and winter at a given location, is that prediction based on solution of an initial value problem? As a second example, suppose that the Sun's energy output decreased significantly. If we predict that a general cooling of the Earth's climate would ensue, is that prediction based on solution of an initial value problem? Predictions can be based on expected changes in external forcing.

14.1.1 Pushing the attractors around

As discussed in Chapter 13, Lorenz showed that the weather is unpredictable beyond a few weeks because the atmosphere behaves chaotically. It is important to understand, however, that chaotic systems respond in predictable ways to changes in external forcing. A simple example of this was constructed by Tim Palmer.

Palmer (1993, 1999) views the climate system as a collection of attractors in a phase space. For example, one attractor might represent an El Niño state, while another represents an La Niña state. The climate itself is a set of statistics that tells what the attractors look like and how frequently each is visited. Palmer argues that slowly varying external forcings can alter the frequency with which each attractor is visited, and that this is how climate change manifests itself. For example, some climate states may have frequent and/or persistent El Niños, while others have very few El Niños.

To illustrate these ideas, Palmer used a modified version of Lorenz's butterfly model:

$$\begin{aligned} \dot{X} &= -\sigma X + \sigma Y + f_0 \cos\theta \ , \\ \dot{Y} &= -XZ + rX - Y + f_0 \sin\theta \ , \\ \dot{Z} &= XY - bZ \ . \end{aligned} \tag{14.1}$$

Here f_0 is a forcing that tries to push X and Y in the direction of the angle θ in the (X,Y) plane. If we put $f_0 = 0$, then (14.1) reduces to Lorenz's butterfly model. Fig. 14.1 shows how different values of θ affect the probability density function (PDF) of the solution in the (X,Y) plane. The maxima of the PDF are the attractors of the model. As θ changes, the locations of the maxima of the PDFs do not change much. This means that the attractors are insensitive to θ. Note, however, that the maxima become stronger or weaker as θ changes. This means that by varying the forcing we can cause some attractors to be visited more often, and others to be visited less often. There is some tendency for the solution to be pushed in the direction in which the forcing acts.

In summary, Palmer's modified butterfly model illustrates that chaotic systems respond in predictable ways to changes in external forcing. The real climate system is currently responding to changes in external forcing that arise from anthropogenic modifications of the atmosphere's composition.

14.2 Transitive and intransitive systems

Lorenz (1976, 1990) discussed the possibility that the climate system can have multiple attractors for a single set of external parameters. Such a system is said to be "intransitive." In contrast, if the system has only one attractor for a given set of external parameters, it is said to be "transitive."

For an intransitive system, the initial condition determines which attractor is manifested. It is possible for an intransitive system to jump between attractors.

Lorenz explored instranstivity using what he called "The World's Simplest GCM" (Lorenz, 1984, 1990). It is described by the following equations:

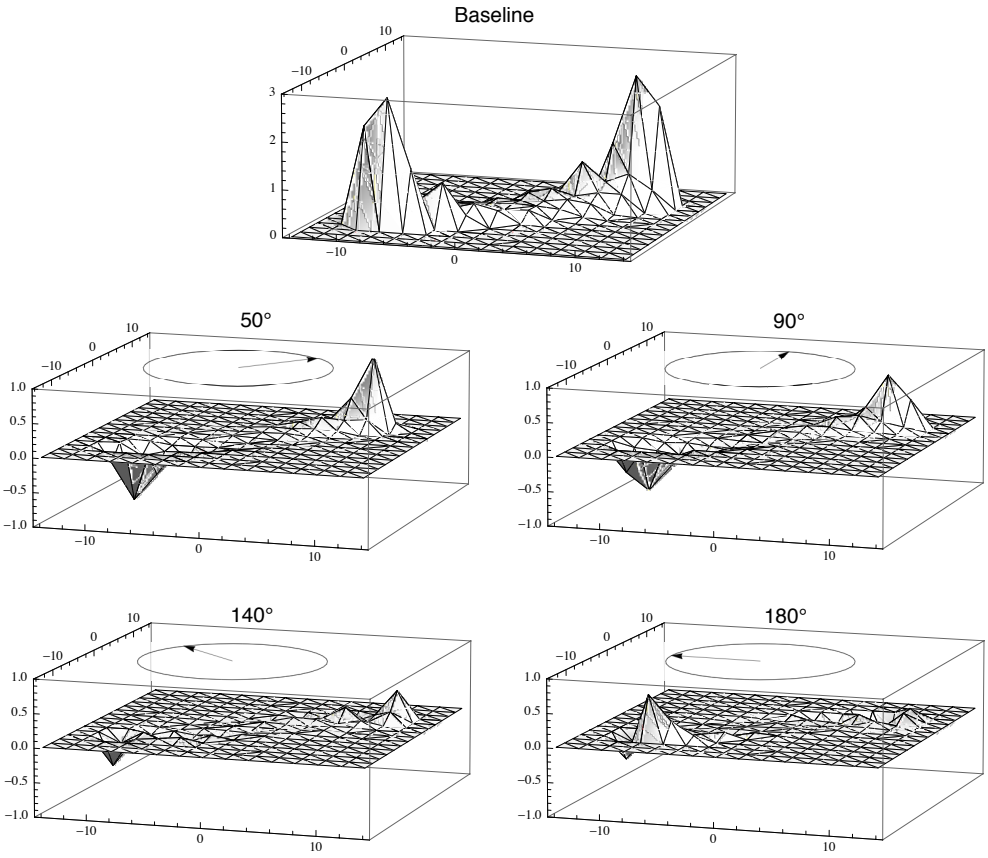


Figure 14.1: The impacts of various imposed forcings on the PDF of the butterfly model state vector with running time mean, in the (X,Y) plane. From Palmer (1999).

$$\begin{aligned}\dot{X} &= -Y^2 - Z^2 - aX + aF \ , \\ \dot{Y} &= XY - bXZ - Y + G \ , \\ \dot{Z} &= bXY + XZ - Z \ .\end{aligned} \tag{14.2}$$

Here the various symbols are interpreted as follows:

- X represents the strength of the westerlies,
- Y and Z are the sine and cosine components of a planetary wave train,
- F represents the meridional heating contrast (a "forcing"),
- G represents land-sea contrast (another forcing), and

- a and b are parameters whose values are set in advance.

For $F = G = 0$ (no forcing), the model has the trivial steady solution $X = Y = Z = 0$. For $G = 0$ we can have the steady solution $X = F$ with $Y = Z = 0$. This eddy-free solution can be unstable, however, and the instability can lead to non-zero values of Y and Z. We interpret "large" $F\,(\sim 8)$ as "winter," and "small" $F\,(\sim 6)$ as "summer." Compare (14.2) with the butterfly model given by (13.56).

For fixed F, there can be one, two, or three steady solutions, depending on the value of G, as shown in Fig. 14.2. Here $F = 2$, so this is "super summer." The steady solutions shown in Fig. 14.2 are found most easily by fixing X and solving for G, Y, and Z.

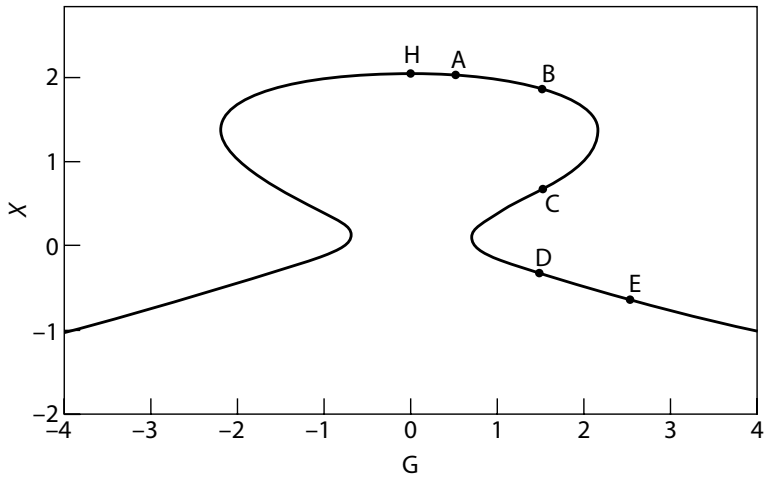


Figure 14.2: Steady solutions of (14.2). The vertical axis is X, and the horizontal axis is G. The figure is constructed for $a = 0.25$, $b = 4.0$, and $F = 2.0$. From Lorenz (1984).

Fig. 14.3 demonstrates that the solutions of the model are sensitively dependent on the initial conditions. The three solutions shown are for the same values of the external parameters, but have slightly different initial conditions. The solutions clearly diverge after some time.

Fig. 14.4 shows two "summer" solutions started from slightly different initial conditions. Note that these two summers are both fairly regular in appearance, but nevertheless appear quite different from each other, indicating that the model is capable of producing two "kinds" of summers: active summers (like the one in the lower panel) and inactive summers (like the one in the upper panel).

Similarly, Fig. 14.5 shows two "winter" solutions started from different initial conditions. The two winters are highly irregular and look much the same, suggesting that all model winters are essentially equivalent; the model makes only one "kind" of winter.

Fig. 14.6 shows results from a six-year simulation, in which two different kinds of

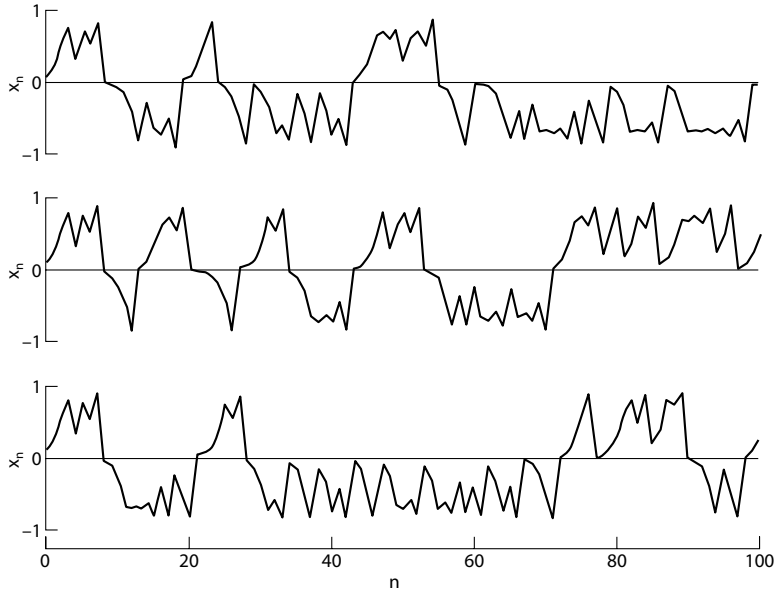


Figure 14.3: Solutions of (14.2) for X over 100 time steps. The initial values of X are 0.0999 (upper), 0.1000 (middle), and 0.1001 (lower). The straight-line segments joining consecutive points are included to make the chronological order easier to see. This figure illustrates sensitive dependence on initial conditions. From Lorenz (1976).

summers occur. The interpretation is very simple and interesting. Winters are chaotic. The model locks into either an active summer or an inactive summer, based on the "initial conditions" at the end of the winter, which are essentially random. When winter returns, all information about the previous summer is obliterated by nonlinear scrambling. The dice are rolled again at the beginning of the next summer.

Define σ as the (nondimensional) standard deviation of X within the period July through September. Active summers have large σ, and inactive summers have small σ. We can calculate one value of σ for each summer, or in other words one for each year. Fig. 14.7 shows the variations of σ (dimensionless) with time in years, in a 100-year numerical solution of (14.2), for the conditions of Fig. 14.6. Active and inactive summers alternate, at irregular intervals.

14.3 Anthropogenic climate change

Our understanding of the global circulation of the atmosphere is very incomplete, but improving at a brisk pace. The last several decades have brought huge advances. During the past 20 years or so, global circulation research has become a sub-discipline within the much broader field of climate research. At the same time, the scientific foundations of global circulation research have become deeper and more intellectually challenging.

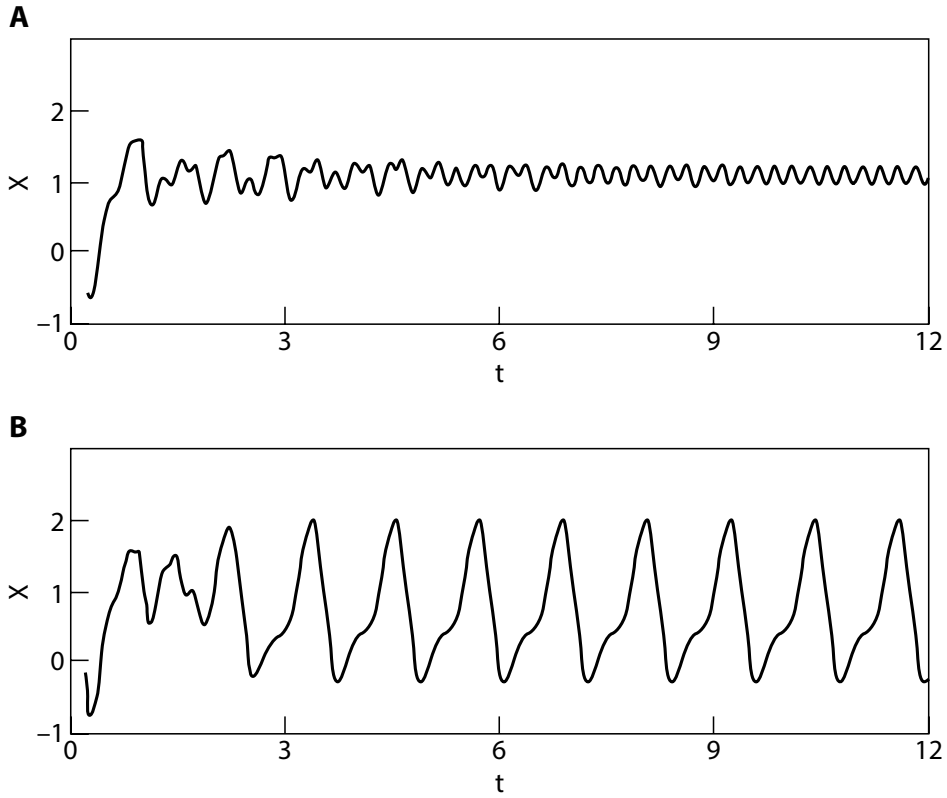


Figure 14.4: (a) The variations of X (dimensionless) with t (months) in a numerical solution of (14.2), with $a = 0.25$, $b = 4.0$, $F = 6.0$, and G=1.0 (summer conditions). The initial state is (2.4, 1.0, 0). (b) The same as panel (a), except that the initial state is (2.5, 1.0, 0). From Lorenz (1990).

There is overwhelming evidence that the Earth's climate is rapidly changing due to the accumulation of anthropogenic greenhouse gases in the atmosphere (Stocker, 2014, hereafter called the IPCC assessment report). In particular, the carbon dioxide concentration of the atmosphere has increased by more than 50% since the middle of the 19th century. In the latter part of the 20th century, there was a sharp decrease in stratospheric ozone due to anthropogenic increases in long-lived chlorine and bromine compounds citepdouglass2014antarctic. Humans have also drastically increased the aerosol load of the atmosphere through various processes including combustion and agriculture, and we have altered the albedo, roughness, and other properties of the land surface through agriculture, city-building, and so on.

As discussed below, these changes have pushed the Earth's radiation budget out of balance. As a result, the global circulation of the atmosphere will be significantly different in the coming decades, and some changes have already become apparent in the observations. Looking on the bright side, it should be possible to learn more about how the circulation works by studying its response to the ongoing anthropogenic perturbations. This closing chapter briefly discusses current trends and expectations for the future.

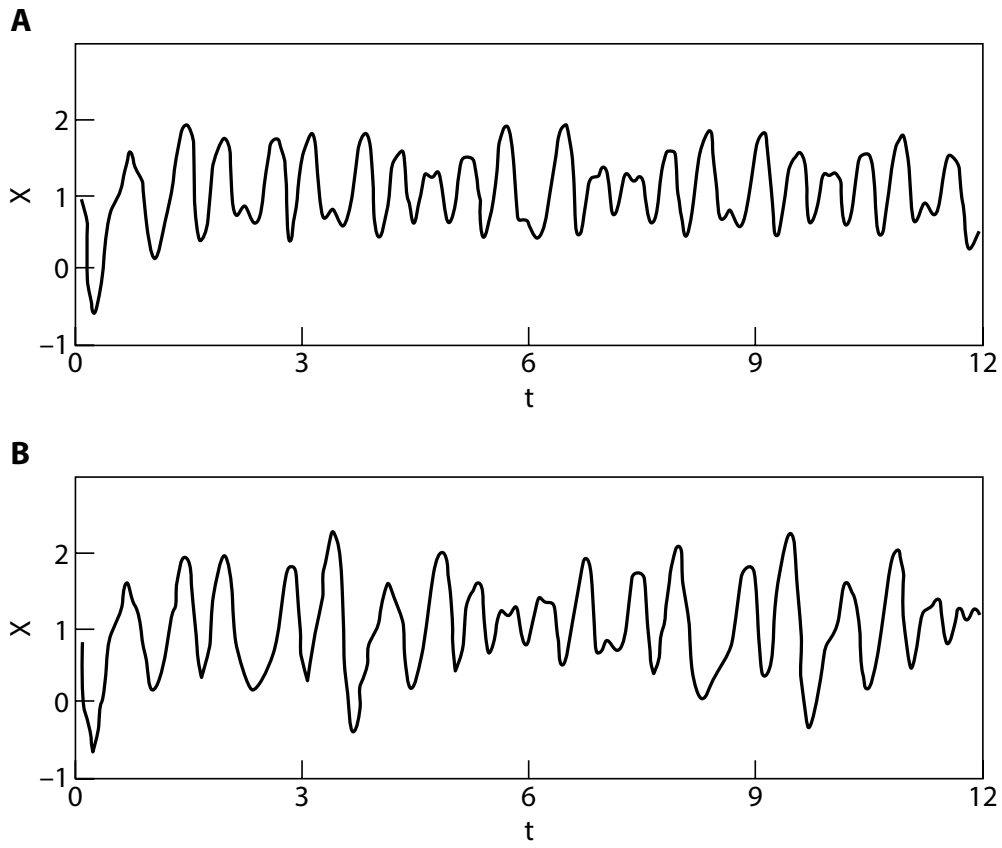


Figure 14.5: (a) The same as Fig. 14.5, except that $F = 8.0$ (winter conditions). (b) The same as panel a, except that the initial state is (2.5, 1.0, 0). From Lorenz (1990).

The 1500-page bulk of the IPCC assessment report contains an assessment of only the most recent research on this very hot topic. It is impossible to cover all of the interesting aspects of the changing circulation in this chapter. I have chosen a few topics. Interested readers should consult the IPCC report for a comprehensive discussion.

As discussed by citelevitus2000warming, levitus2009global, levitus2012world and Loeb et al. (2012), in recent decades the oceans have been accumulating thermal energy at a rate consistent with the satellite-derived imbalance of about 0.5 W m^{-2} in the Earth's radiation budget. This has led to a general warming of the Earth's surface and troposphere, on the order of 0.8 K since the mid-nineteenth century. The near-surface warming is strongest in the winter, and greater at night than during the day (Vinnikov et al., 2002; Vose et al., 2005). The Arctic surface air temperature has warmed about twice as much as the global mean (Screen et al., 2012), with the largest increase in winter. The Arctic warming has been accompanied by a substantial reduction in late-summer Arctic sea ice extent and a year-round decrease in sea ice thickness (Stroeve et al., 2014).

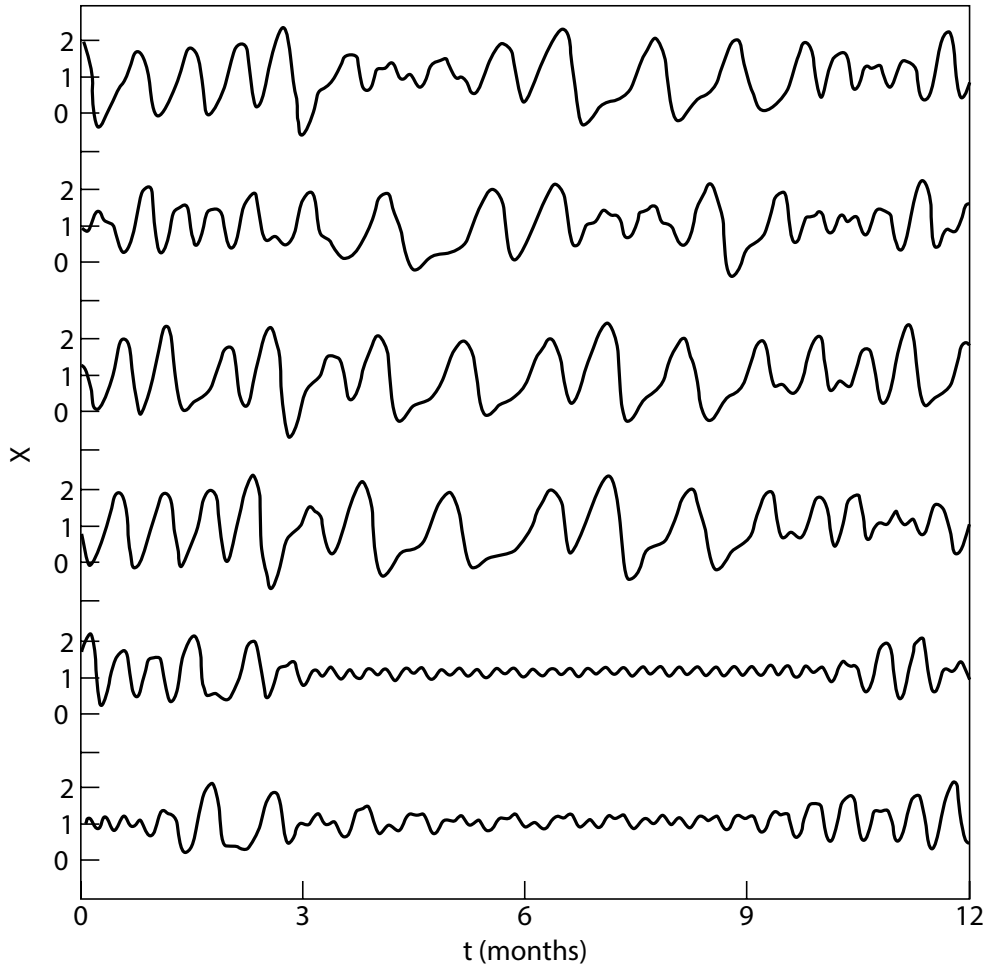


Figure 14.6: The variations of X (dimensionless) with t (months) in a six-year numerical solution of (14.2), with $a = 0.25$, $b = 4.0$, $F = 7 + 2\cos 2\pi t/tau$, and $G = 1.0$, where $\tau = 12$ months. Each row begins on 1 January, and (except for the first) each row is a continuation of the previous one. From Lorenz (1990).

The troposphere above the surface is also warming (Karl et al. 2006). In fact the tropical upper troposphere is expected to warm faster than the surface (e.g., Lorenz and DeWeaver, 2007) because, as discussed in Chapter 6, the tropical lapse rate is constrained by convection to be close to the saturated moist adiabatic lapse rate, which decreases at warmer temperatures. A decrease of the lapse rate means an increase in the (dry) static stability, which inhibits baroclinic eddies. In addition, a decrease in the lapse rate is a negative feedback that tends to limit changes in the Earth's temperature, because infrared emission to space is relatively efficient from the upper troposphere (Bony et al., 2006).

Meanwhile, the stratosphere is observed to be cooling, about twice as fast as the surface is warming (e.g., Thompson et al., 2012). An increase in stratospheric carbon dioxide

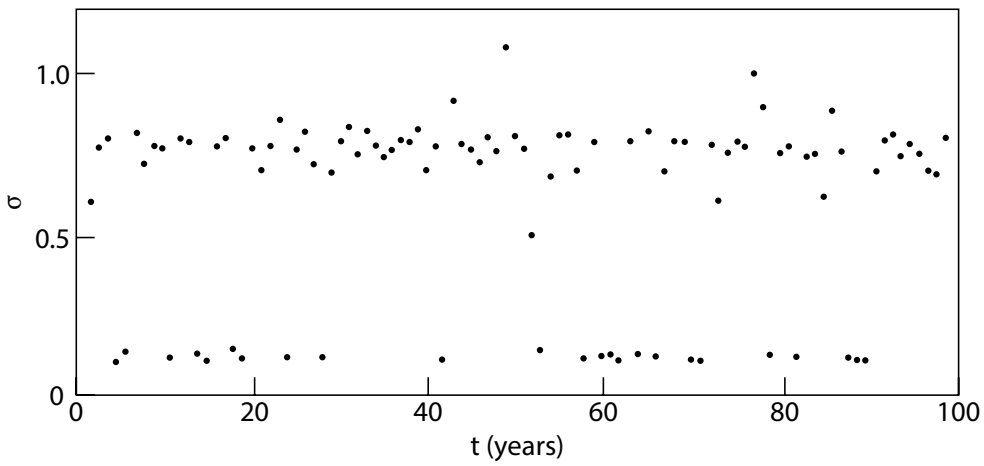


Figure 14.7: The variations of σ (dimensionless) with t (years) in a 100-year numerical solution of (14.2), for the conditions of Fig. 14.6. Here σ is the standard deviation of X within the summer season, i.e., July through September. From Lorenz (1990).

promotes cooling because infrared photons emitted from the stratosphere can easily escape to space. A decrease in stratospheric ozone concentrations also promotes cooling, because ozone warms by very efficiently absorbing solar ultraviolet radiation.

Both cooling of the stratosphere and warming of the troposphere favor an increase in tropopause height (Hoskins, 2003; Lu et al., 2009), and a variety of observations show that the height of the tropopause has been increasing (Highwood and Hoskins, 1998; Highwood et al., 2000; Randel et al., 2000; Seidel et al., 2001; Gettelman et al., 2001; Seidel and Randel, 2006; Schmidt et al., 2008; Austin and Reichler, 2008; Son et al., 2009). Tropical tropopause temperatures are decreasing (Wang et al., 2012). Lu et al. (2009) present evidence that the increasing tropopause height has been due more to the cooling of the stratosphere than to the warming of the troposphere.

So now we have a complicated picture: The tropospheric meridional temperature gradient has decreased (especially in the Northern Hemisphere), the tropospheric static stability has increased, and the tropopause height has increased. Each of these factors can be expected to influence both baroclinic eddy activity and the width of the Hadley circulation, which in turn are linked to each other because the poleward limit of the Hadley regime more or less coincides with the equatorward limit of the baroclinic eddy regime. Observations show that there has been a widening of the Hadley cells (Seidel and Randel, 2007; Hu and Fu, 2007; Lu et al., 2007, 2009; Birner, 2010), which includes a poleward shift of the subtropical subsidence that inhibits precipitation. Observations also show a poleward shift of the westerly jet streams (e.g., Strong and Davis, 2007; Seidel et al., 2008; Archer and Caldeira, 2008; Barton and Ellis, 2009; Fu and Lin, 2011) and the baroclinic eddies (Cornes and Jones, 2011), at least in the Northern Hemisphere.

As mentioned in Chapter 2, for typical current surface temperatures the saturation vapor pressure increases at the rate of 7% per Kelvin (Held and Soden, 2006). The slope will become even larger in a future, warmer climate. Because the actual vapor pressure near the surface is closely coupled to the saturation vapor pressure at the sea surface, a warmer atmosphere will contain much more water vapor. Observations show an upward trend in the atmosphere's water vapor content over the past few decades (Trenberth et al., 2005; Wentz et al., 2007; Jin et al., 2007).

On the other hand, the globally averaged rate of precipitation is strongly linked to the rate at which the atmosphere cools radiatively; again, this was mentioned in Chapter 2. Increases in temperature, water vapor, and CO_2 concentrations favor stronger emission of infrared radiation, i.e., more intense radiative cooling. The precipitation rate is also constrained by the supply of energy available to evaporate water from the surface. Increased downward emission of radiation to the Earth's surface can encourage faster evaporation, and is therefore consistent with stronger precipitation.

For these reasons, the hydrologic cycle is expected to run faster in a warmer climate, and observations suggest that this is already occurring. A particularly interesting line of evidence is based on observations of sea surface salinity. Salinity is increasing in the subtropics where precipitation is weak, in decreasing in the tropics where precipitation is strong (e.g., Durack and Wijffels, 2010; Durack et al., 2012). These changes are consistent with more evaporation in the subtropics and more precipitation in the tropics. According to Durack et al. (2012), the speed of the hydrologic cycle is increasing at the very rapid rate of 8% per Kelvin of surface warming, which is about the same as the rate of increase of the surface saturation vapor pressure, as mentioned above.

On the other hand, most models predict that the speed of the hydrologic cycle will increase by a much smaller 2% per Kelvin (e.g., Vecchi and Soden, 2007), determined by the rate at which atmospheric radiative cooling increases; see Fig. 14.8. It has been suggested that if the precipitation rate increases much more slowly than the atmospheric water vapor content, then weaker Hadley and Walker circulations will meet the needs of the hydrologic cycle (Held and Soden, 2006). This hypothesis is based on the idea that the precipitation rate is given by the atmospheric water content times a measure of the strength of the circulation. The conclusion is that the atmospheric circulation will run more slowly in a warmer, more humid climate (Vecchi and Soden, 2007; Tokinaga et al., 2012; Hsu and Li, 2012; Hsu et al., 2013; Kitoh et al., 2013). There is evidence that the global monsoon circulations have been weakening (Zhou et al., 2008; Annamalai et al., 2013). On the other hand, l'Heureux et al. (2013) report a strengthening of the Walker circulation (see also England et al., 2014). There is also some evidence that the annular modes will be different in a warmer world (Thompson et al., 2000; Previdi and Liepert, 2007).

As discussed in Chapter 10, the physical processes that give rise to the MJO are still not entirely clear, and some of today's global atmospheric models have difficulty in simulating it. Climate-change simulations with models that are able to simulate the present-day MJO

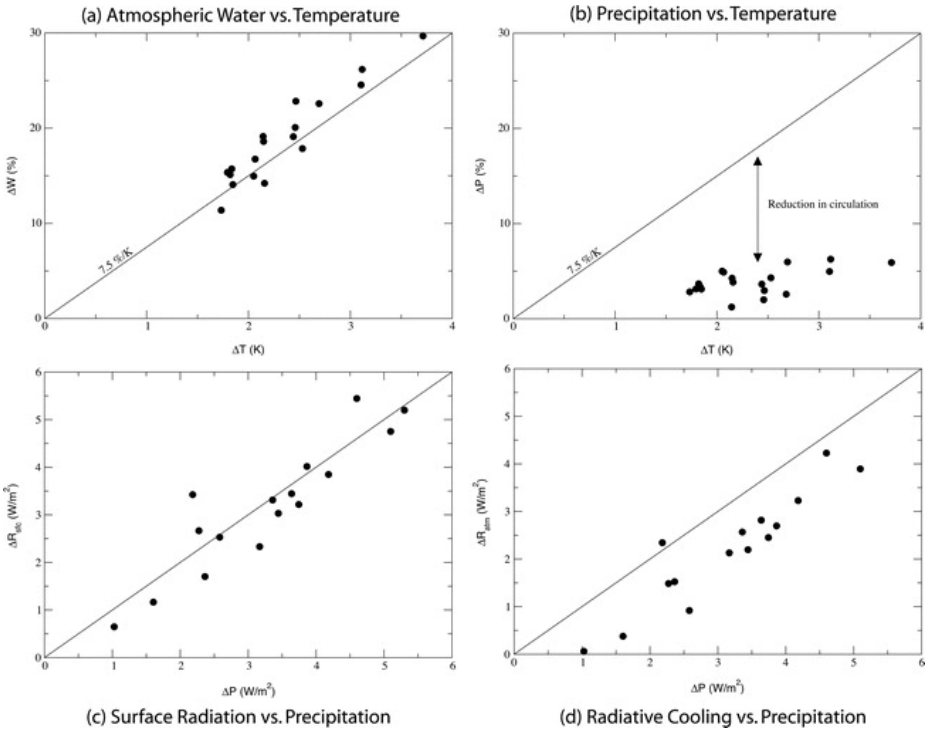


Figure 14.8: Scatterplots of the change in global-mean quantities for each of the models participating in the Coupled Model Intercomparison Project 5 for which the necessary data were available. Results are shown for (a) temperature vs column-integrated water vapor (%); (b) temperature vs precipitation (%); (c) precipitation (W $m^{-}2$) vs net downward radiation at the surface; and (d) precipitation (W $m^{-}2$) vs net radiative cooling of the atmosphere. The differences are computed by subtracting the decadal means from the first 10 yr and last 10 yr of the twenty-first century, as simulated by climate models. Panels (a) and (b) reproduce Figs. 2.a and 2.b of Held and Soden (2006). From Vecchi and Soden (2007).

with some fidelity suggest that the MJO will be much more active in the warmer, more humid world of the future (e.g., Schubert et al., 2013; Arnold et al., 2014). If true, this will have major consequences for the tropical climate system and the people who live in that part of the world.

In short, the global circulation of the atmosphere is changing rapidly and in a wide variety of ways. These changes are expected to continue and accelerate in the coming decades.

14.4 Problems

1. The equations of "The World's Simplest GCM" Lorenz (1984) are given by

$$\begin{aligned}\dot{X} &= -Y^2 - Z^2 - aX + aF \ , \\ \dot{Y} &= XY - bXZ - Y + G \ , \\ \dot{Z} &= bXY + XZ - Z \ .\end{aligned} \tag{14.3}$$

Program the model using the fourth-order Runge-Kutta time-differencing scheme, which is discussed in many numerical analysis books. Use a time step of $\Delta t = 1/30$. Following Lorenz, we interpret a time unit as corresponding to 4 hours, so that six time units correspond to one day. The time step is thus 4/30 hours.

We define two slightly different versions of the model. The first, which we think of as the "real world," uses $a = 0.25$, $b = 4$, $F = 8$, and $G = 1$. The second, which we think of as the "forecast model," is identical to the first, except that $b = 4.01$. The code is set up in "real world" mode.

The "standard" initial conditions are $X = 2.5$, $Y = 1.0$, and $Z = 0$.

(a) Perform two identical 100-day simulations to demonstrate that the results from the "real world" are perfectly reproducible.

(b) Do a test to show that the "real world" exhibits sensitive dependence on initial conditions. How long does it take before two simulations that start from slightly different initial conditions are as different as two states chosen at random from a long simulation with the model?

(c) Make a 100-day simulation of the "real world" and save the results for X, Y, and Z once per simulated day. Starting from each of the first 90 days of this "real world" weather record, make a ten-day forecast by running the "forecast model." Save the results from each of these forecasts, once per simulated day. Use these results to produce a plot similar to Fig. 13.22, with at least the curves $E_{0,k}$ and $E_{k-1,k}$. (You may want to plot additional curves.) Use X as the analysis variable, just as Lorenz (1982) used the 500 hPa height.

(d) Still using X as the analysis variable, plot dE/dk versus E, as in Fig. 13.24, for both the "perfect model" and the "imperfect model." Estimate the value of a, the error e-folding rate.

Appendix A

Vectors, Coordinates, and Coordinate Transformations

A.1 Physical laws and coordinate systems

For the present discussion, we define a "coordinate system" as a tool for describing positions in space. Coordinate systems are human inventions, and therefore are not part of physics, although they can be used to describe physics. For obvious reasons, spherical coordinates are particularly useful in geophysics.

Any physical law should be expressible in a form that makes no reference to any particular coordinate system; we certainly do not expect that the laws of physics change when we switch from spherical coordinates to cartesian coordinates! Nevertheless, it is useful to understand how physical laws can be expressed in different coordinate systems, and in particular how various quantities "transform" as we change from one coordinate system to another.

A.2 Scalars, vectors, and tensors

Tensors can be defined without reference to any particular coordinate system. A tensor is simply "out there," and has a meaning that is the same whether we happen to be working in spherical coordinates, or Cartesian coordinates, or whatever. Tensors are, therefore, just what we need to formulate physical laws.

The simplest kind of tensor, called a "tensor of rank 0," is a scalar, which is represented by a single number – essentially a magnitude with no direction. An example of a scalar is temperature. Not all quantities that are represented by a single number are scalars, because not all of them are defined without reference to any particular coordinate system. An example of a (single) number that is not a scalar is the longitudinal component of the wind, which is defined with respect to a particular coordinate system, i.e., spherical coordinates.

A scalar is expressed in exactly the same way regardless of what coordinate system may be in use to describe non-scalars in a problem. For example, if someone tells you

the temperature in Fort Collins, you don't have to ask whether they are using spherical coordinates or some other coordinate system, because it makes no difference at all.

Vectors are "tensors of rank 1;" a vector can be represented by a magnitude and one direction. An example is the wind vector. In atmospheric science, vectors are normally either three-dimensional or two-dimensional, but in principle they have any number of dimensions. A scalar can be considered to be a vector in a one-dimensional space.

A vector can be expressed in a particular coordinate system by an ordered list of numbers, which are called the "components" of the vector. The components have meaning only with respect to the particular coordinate system. More or less by definition, the number of components needed to describe a vector is equal to the number of dimensions in which the vector is "embedded."

We can define "unit vectors" that point in each of the coordinate directions. A vector can then be written as the vector sum of each of the unit vectors times the "component" associated with the unit vector. In general, the directions in which the unit vectors point depend on position.

Unit vectors are always non-dimensional; here we are using the word "dimension" to refer to physical quantities, such as length, time, and mass. Because the unit vectors are non-dimensional, all components of a vector must have the same dimensions as the vector itself.

Spatial coordinates may or may not have the dimensions of length. In the familiar Cartesian coordinate system, the three coordinates, (x, y, z), each have dimensions of length. In spherical coordinates, (λ, φ, r), where λ is longitude, φ is latitude, and r is distance from the origin, the first two coordinates are non-dimensional angles, while the third has the dimension of length.

When we change from one coordinate system to another, an arbitrary vector $\mathbf{V}$ transforms according to

$$\mathbf{V}' = \mathbf{M}\mathbf{V}. \tag{A.1}$$

Here $\mathbf{V}$ is the representation of the vector in the first coordinate system (i.e., $\mathbf{V}$ is the list of the components of the vector in the first coordinate system), $\mathbf{V}'$ is the representation the vector in the second coordinate system, and $\mathbf{M}$ is a "rotation matrix" that maps $\mathbf{V}$ onto $\mathbf{V}'$. The rotation matrix used to transform a vector from one coordinate system to another is a property of the two coordinate systems in question; it is the same for all vectors, but it *does* depend on the particular coordinate systems involved, so it is not a tensor.

The transformation rule (A.1) is actually part of the definition of a vector, i.e., a vector must, by definition, transform from one coordinate system to another via a rule of the form

(A.1) . It follows that not all ordered lists of numbers are vectors. For example, the list

(mass of the moon, distance from Fort Collins to Denver)

is not a vector.

Now let $\mathbf{V}$ be the a vector representing the three-dimensional velocity of a particle in the atmosphere. The Cartesian and spherical representations of are

$$\mathbf{V} = \dot{x}\mathbf{i} + \dot{y}\mathbf{j} + \dot{z}\mathbf{k} \tag{A.2}$$

$$\mathbf{V} = \dot{\lambda} r \cos\varphi \mathbf{e}_\lambda + r\dot{\varphi}\mathbf{e}_\varphi + \dot{r}\mathbf{e}_r \tag{A.3}$$

Here a "dot" denotes a Lagrangian time derivative, i.e., a time derivative following a moving particle, $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are unit vectors in the cartesian coordinate system, and $\mathbf{e}_\lambda$, $\mathbf{e}_\varphi$, and $\mathbf{e}_r$ are unit vectors in the spherical coordinate system. Eqs. (A.2) and (A.3) both describe the same vector, $\mathbf{V}$, i.e., the meaning of $\mathbf{V}$ is independent of the coordinate system that is chosen to represent it.

Vectors are considered to be tensors of rank one, and scalars are tensors of rank zero. The number of directions associated with a tensor is called the "rank" of the tensor. In principle, the rank can be arbitrarily large, but in atmospheric science we rarely meet tensors with ranks higher than two.

A tensor of rank 2 that is important in atmospheric science is the advective flux of momentum, which has a magnitude and "two directions." One of the directions is associated with the advected momentum itewslf, and the other is associated with the direction in which the momentum is being transported. The advective flux of momentum can be written as $\rho\mathbf{V}\otimes\mathbf{V}$, where ρ is the density of the air, $\mathbf{V}$ is the wind vector, and $\otimes$ is the dyadic or outer product.

Like a vector, a tensor of rank 2 can be expressed in a particular coordinate system, i.e., we can define the "components" of the tensor with respect to a particular coordinate system. The components of a tensor of rank 2 can be arranged in the form of a two-dimensional matrix, in contrast to the components of a (column or row) vector, which form an ordered one-dimensional list. When we change from one coordinate system to another, a tensor of rank 2 transforms according to

$$\mathbf{T}' = \mathbf{MTM}^{-1} \tag{A.4}$$

where $\mathbf{T}$ is the representation of a rank-2 tensor in the first coordinate system, $\mathbf{T}'$ is the representation of the same tensor in the second coordinate system, $\mathbf{M}$ is the matrix introduced in Eq. (1) above, and $\mathbf{M}^{-1}$ is its inverse.

A.3 Differential operators

Several familiar differential operators can be defined without reference to any coordinate system. This makes them more fundamental than, for example, $\partial/\partial x$, where x is a particular spatial coordinate. The coordinate-independent operators that we need most often for atmospheric science (and for most other branches of physics) are:

- the gradient, denoted by $\nabla\alpha$, where α is an arbitrary scalar;
- the divergence, denoted by $\nabla\cdot\mathbf{V}$, where $\mathbf{V}$ is an arbitrary vector;
- the curl, denoted by $\nabla\times\mathbf{V}$, and
- the Laplacian, given by $\nabla^2 A = \nabla\cdot(\nabla A)$.

Note that the gradient and curl are vectors, while the divergence is a scalar. The gradient operator accepts scalars as "input," while the divergence and curl operators consume vectors.

In discussions of two-dimensional motion, it is often convenient to introduce an additional operator called the Jacobian, denoted by

$$\begin{aligned} J(\alpha,\beta) \equiv &\quad \mathbf{k}\cdot(\nabla\alpha\times\nabla\beta) \\ = &\quad \mathbf{k}\cdot\nabla\times(\alpha\nabla\beta) \\ = &-\mathbf{k}\cdot\nabla\times(\beta\nabla\alpha)\,. \end{aligned} \tag{A.5}$$

Here the gradient operators are understood to produce vectors in the two-dimensional space, α and β are arbitrary scalars, and $\mathbf{k}$ is a unit vector perpendicular to the two-dimensional surface. The second and third lines of (A.5) can be derived with the use of vector identities found in a table later in this appendix.

A definition of the gradient operator that does not make reference to any coordinate system is:

$$\nabla\alpha \equiv \lim_{V\to 0}\left[\frac{1}{V}\oint_S \mathbf{n}\alpha\, dS\right]\,, \tag{A.6}$$

where S is the surface bounding a volume V, and $\mathbf{n}$ is the outward normal on S. Here the terms "volume" and "bounding surface" are used in the following generalized sense: In a

three-dimensional space, "volume" is literally a volume, and "bounding surface" is literally a surface. In a two-dimensional space, "volume" means an area, and "bounding surface" means the curve bounding the area. In a one-dimensional space, "volume" means a curve, and "bounding surface" means the end points of the curve. The limit in (A.6) is one in which the volume and the area of its bounding surface shrink to zero.

As an example, consider a Cartesian coordinate system on a plane, with unit vectors $\mathbf{i}$ and $\mathbf{j}$ in the x and y directions, respectively. Consider a "box" of width Δx and height Δy, as shown in Figure A.1. We can write

$$\begin{aligned}\nabla A \equiv \lim_{(\Delta x,\Delta y)\to 0} \Bigg\{ \frac{1}{\Delta x \Delta y} \Bigg[& A\left(x_0+\frac{\Delta x}{2}, y_0\right)\Delta y\mathbf{i} + A\left(x_0, y_0+\frac{\Delta y}{2}\right)\Delta x\mathbf{j} \\ & -A\left(x_0-\frac{\Delta x}{2}, y_0\right)\Delta y\mathbf{i} - A\left(x_0, y_0-\frac{\Delta y}{2}\right)\Delta x\mathbf{j}\Bigg]\Bigg\} \\ & = \frac{\partial A}{\partial x}\mathbf{i} + \frac{\partial A}{\partial y}\mathbf{j}\,.\end{aligned} \tag{A.7}$$

This is the answer that we expect.

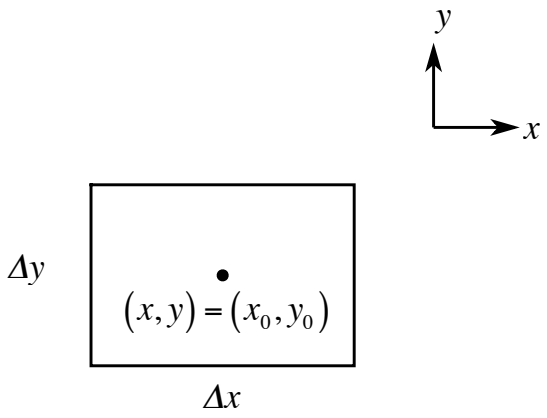


Figure A.1: A rectangular box in a planar two-dimensional space, with center at (x_0, y_0), width Δx, and height Δy.

Definitions of the divergence and curl operators that do not make reference to any coordinate system are:

$$\nabla \cdot \mathbf{Q} \equiv \lim_{V\to 0}\left[\frac{1}{V}\oint_S \mathbf{n}\cdot\mathbf{Q}\,dS\right] \tag{A.8}$$

$$\nabla \times \mathbf{Q} \equiv \lim_{V \to 0} \left[\frac{1}{V} \oint_S \mathbf{n} \times \mathbf{Q} dS \right] \tag{A.9}$$

It is possible to work through exercises similar to (A.7) for these operators too. You might want to try it yourself, to see if you understand.

Finally, the Jacobian on a two-dimensional surface can be defined by

$$J(\alpha, \beta) = \lim_{C \to 0} \left[\oint_C \alpha \nabla \beta \cdot \mathbf{t} dl \right], \tag{A.10}$$

where $\mathbf{t}$ is a unit vector that is tangent to the bounding curve C.

A.4 Vector identities

Many useful identities relate the divergence, curl, and gradient operators. Most of the following identities can be found in any mathematics reference manual, e.g., Beyer (1984). As before, let α and β be arbitrary scalars, let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be arbitrary vectors, and let $\mathbf{T}$ be an arbitrary tensor of rank 2. Then:

$$\nabla \times (\nabla \alpha) = 0 \tag{A.11}$$

$$\nabla \cdot (\nabla \times \mathbf{A}) = 0 \tag{A.12}$$

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \tag{A.13}$$

$$\nabla \cdot (\alpha \mathbf{A}) = \alpha (\nabla \cdot \mathbf{A}) + \mathbf{A} \cdot \nabla \alpha \tag{A.14}$$

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = (\nabla \times \mathbf{A}) \cdot \mathbf{B} - (\nabla \times \mathbf{B}) \cdot \mathbf{A} \tag{A.15}$$

$$\nabla \times (\alpha \mathbf{A}) = \nabla \alpha \times \mathbf{A} + \alpha (\nabla \times \mathbf{A}) \tag{A.16}$$

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C} = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A}) \tag{A.17}$$

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B} (\mathbf{C} \cdot \mathbf{A}) - \mathbf{C} (\mathbf{A} \cdot \mathbf{B}) \tag{A.18}$$

$$\nabla \times (\mathbf{A} \times \mathbf{B}) = \mathbf{A} (\nabla \cdot \mathbf{B}) - \mathbf{B} (\nabla \cdot \mathbf{A}) - (\mathbf{A} \cdot \nabla) \mathbf{B} + (\mathbf{B} \cdot \nabla) \mathbf{A} \tag{A.19}$$

$$\nabla (\mathbf{A} \cdot \mathbf{B}) = (\mathbf{A} \cdot \nabla) \mathbf{B} + (\mathbf{B} \cdot \nabla) \mathbf{A} + \mathbf{A} \times (\nabla \times \mathbf{B}) + \mathbf{B} \times (\nabla \times \mathbf{A}) \tag{A.20}$$

$$\begin{aligned} J(\alpha, \beta) \equiv \mathbf{k} \cdot (\nabla \alpha \times \nabla \beta) = &\quad \mathbf{k} \cdot \nabla \times (\alpha \nabla \beta) \\ &= -\mathbf{k} \cdot \nabla \times (\beta \nabla \alpha) \\ &= -\mathbf{k} \cdot (\nabla \beta \times \nabla \alpha) \end{aligned} \tag{A.21}$$

$$\nabla^2 \mathbf{A} \equiv (\nabla \cdot \nabla) \mathbf{A} = \nabla (\nabla \cdot \mathbf{A}) - \nabla \times (\nabla \times \mathbf{A}) \tag{A.22}$$

$$\nabla \cdot (\mathbf{A} \otimes \mathbf{B}) = (\mathbf{A} \cdot \nabla) \mathbf{B} + (\mathbf{B} \cdot \nabla) \mathbf{A} \tag{A.23}$$

$$\nabla \cdot (\alpha \mathbf{T}) = (\nabla \alpha) \cdot \mathbf{T} + \alpha (\nabla \cdot \mathbf{T}) \tag{A.24}$$

In (A.23), $\otimes$ denotes the outer or dyadic product of two vectors, which yields a tensor of rank 2.

A special case of (A.20) is

$$\frac{1}{2} \nabla (\mathbf{A} \cdot \mathbf{A}) = (\mathbf{A} \cdot \nabla) \mathbf{A} + \mathbf{A} \times (\nabla \times \mathbf{A}) \tag{A.25}$$

This identity is used to write the advection terms of the momentum equation in alternative forms.

Identity (A.22) says that the Laplacian of a vector is the gradient of the divergence of the vector, minus the curl of the curl of the vector. The first term involves only the divergent part of the wind field, and the second term involves only the rotational part. Eq. (A.22) can be used, for example, in a parameterization of momentum diffusion.

A.5 Spherical coordinates

A.5.1 Vector operators in spherical coordinates

The gradient, divergence, curl, Laplacian, and Jacobian operators can be expressed in spherical coordinates as follows:

$$\nabla \alpha = \left(\frac{1}{r \cos \varphi} \frac{\partial \alpha}{\partial \lambda}, \frac{1}{r} \frac{\partial \alpha}{\partial \varphi}, \frac{\partial \alpha}{\partial r} \right) \tag{A.26}$$

$$\nabla \cdot \mathbf{A} = \frac{1}{r \cos \varphi} \frac{\partial A_\lambda}{\partial \lambda} + \frac{1}{r \cos \varphi} \frac{\partial}{\partial \varphi} \left(A_\varphi \cos \varphi \right) + \frac{1}{r^2} \frac{\partial}{\partial r} \left(A_r r^2 \right) \tag{A.27}$$

$$\nabla \times \mathbf{A} = \left\{ \frac{1}{r} \left[\frac{\partial A_r}{\partial \varphi} - \frac{\partial}{\partial r} \left(r A_\varphi \right) \right], \frac{1}{r} \frac{\partial}{\partial r} \left(r A_\lambda \right) - \frac{1}{r \cos \varphi} \frac{\partial A_r}{\partial \lambda}, \frac{1}{r \cos \varphi} \left[\frac{\partial A_\varphi}{\partial \lambda} - \frac{\partial}{\partial \varphi} \left(A_\lambda \cos \varphi \right) \right] \right\} \tag{A.28}$$

$$\nabla^2\alpha = \frac{1}{r^2\cos^2\varphi}\frac{\partial^2\alpha}{\partial\lambda^2} + \frac{1}{r^2\cos\varphi}\frac{\partial}{\partial\varphi}\left(\frac{\partial\alpha}{\partial\varphi}\cos\varphi\right) + \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\alpha}{\partial r}\right) \tag{A.29}$$

$$J(\alpha,\beta) = \frac{1}{r^2\cos\varphi}\left(\frac{\partial\alpha}{\partial\lambda}\frac{\partial\beta}{\partial\varphi} - \frac{\partial\beta}{\partial\lambda}\frac{\partial\alpha}{\partial\varphi}\right) \tag{A.30}$$

The divergence operator can be expanded as

$$\nabla\cdot\mathbf{A} = \frac{1}{r\cos\varphi}\frac{\partial A_\lambda}{\partial\lambda} + \frac{1}{r\cos\varphi}\frac{\partial}{\partial\varphi}\left(A_\varphi\cos\varphi\right) + \frac{\partial A_r}{\partial r} + \frac{2A_r}{r}. \tag{A.31}$$

Because the Earth's atmosphere is very thin compared to the radius of the Earth, the last term of (A.31) is negligible, and the divergence operator can be approximated by

$$\nabla\cdot\mathbf{A} \cong \frac{1}{a\cos\varphi}\frac{\partial A_\lambda}{\partial\lambda} + \frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(A_\varphi\cos\varphi\right) + \frac{\partial A_r}{\partial r}. \tag{A.32}$$

Note that r has been replaced by a in the first two terms. In this book, we normally use (A.32) rather than (A.31), largely because it is conventional to do so, although it is not clear that the approximation (A.32) actually makes our work simpler. The approximation would not be applicable to a deep atmosphere, such as that of a star, or of Jupiter.

A.5.2 Horizontal and vertical vectors in spherical coordinates

The unit vectors in spherical coordinates are denoted by $\mathbf{e}_\lambda$ pointing towards the east, $\mathbf{e}_\varphi$ pointing towards the north, and $\mathbf{e}_r$ pointing outward from the origin (in geophysics, outward from the center of the Earth).

A useful result that is a special case of (A.19) is

$$\mathbf{e}_r\cdot[\nabla\times(\alpha\mathbf{e}_r\times\mathbf{H})] = \alpha\nabla\cdot\mathbf{H}, \tag{A.33}$$

where $\mathbf{H}$ is an arbitrary horizontal vector. Similarly, a useful special case of (A.15) is

$$\nabla \cdot (\alpha \mathbf{e}_r \times \mathbf{H}) = -\alpha \mathbf{e}_r \cdot (\nabla \times \mathbf{H}) \tag{A.34}$$

If a three-dimensional vector $\mathbf{V}$ is separated into a horizontal vector and a vertical vector, as in

$$\mathbf{V} = \mathbf{v} + w\mathbf{e}_r, \tag{A.35}$$

then (A.28)) can be written as

$$\boxed{\nabla \times (\mathbf{v} + w\mathbf{e}_r) = \nabla_r \times \mathbf{v} + \mathbf{e}_r \times \left[\frac{1}{r}\frac{\partial}{\partial r}(r\mathbf{v}) - \nabla_r w\right].} \tag{A.36}$$

In case $\mathbf{V}$ is the velocity, the first term on the right-hand side of (A.36) is the vertical component of the vorticity, and the second term is the horizontal vorticity vector. Eq. (A.36) shows that the curl of a purely vertical vector is minus $\mathbf{e}_r$ crossed with the horizontal gradient of the magnitude of that vector. The three-dimensional curl of a purely horizontal vector has both a vertical part, given by $\nabla_r \times \mathbf{v}$, and a horizontal part, given by $\mathbf{e}_r \times \left[\frac{1}{r}\frac{\partial}{\partial r}(r\mathbf{v}) - \nabla_r w\right]$. The *two-dimensional* curl of a horizontal vector has only a vertical component, namely $\nabla_r \times \mathbf{v}$.

Finally, the 3D curl of the 3D curl of a 3D vector is given by

$$\nabla \times [\nabla \times (\mathbf{v} + w\mathbf{e}_r)] = \nabla_r \times \boldsymbol{\eta} + \mathbf{e}_r \times \left[\frac{1}{r}\frac{\partial}{\partial r}(r\boldsymbol{\eta}) - \nabla_r \zeta\right], \tag{A.37}$$

where

$$\boldsymbol{\eta} \equiv \mathbf{e}_r \times \left[\frac{1}{r}\frac{\partial}{\partial r}(r\mathbf{v}) - \nabla_r w\right], \tag{A.38}$$

and

$$\zeta \equiv \mathbf{e}_r \cdot (\nabla_r \times \mathbf{v}). \tag{A.39}$$

A.5.3 Derivation of the gradient operator in spherical coordinates

Consider how the two-dimensional version of (A.26) can be derived from (A.6). Figure A.2 illustrates the problem. Here we have replaced r by a, the radius of the Earth. The angle θ depicted in the figure arises from the gradual rotation of $\mathbf{e}_\lambda$ and $\mathbf{e}_\varphi$, the unit vectors associated with the spherical coordinates, as the longitude changes; the directions of $\mathbf{e}_\lambda$ and $\mathbf{e}_\varphi$ in the center of the area element, where ∇A is defined, are different from their respective directions on either east-west wall of the area element. Inspection of Figure A.2 shows that θ satisfies

$$\begin{aligned}\sin\theta &= \frac{-\frac{1}{2}\left[a\cos\left(\varphi+d\varphi\right)-a\cos\varphi\right]d\lambda}{ad\varphi}\\ &\to -\frac{1}{2}\left(\frac{\partial}{\partial\varphi}\cos\varphi\right)d\lambda\\ &= \frac{1}{2}\sin\varphi d\lambda\,.\end{aligned} \tag{A.40}$$

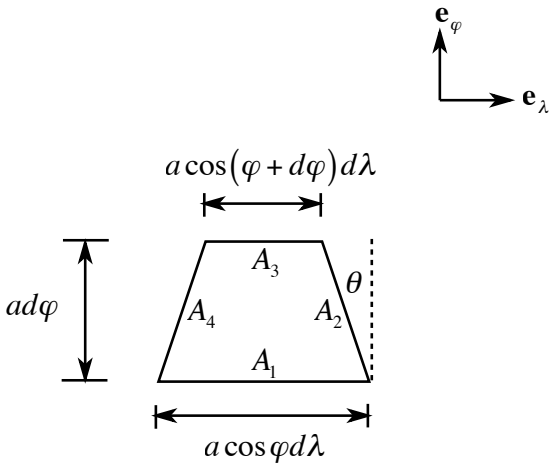


Figure A.2: A patch of the sphere, with longitudinal width $a\cos\varphi d\lambda$, and latitudinal height $ad\varphi$.

The angle θ is of "differential" or infinitesimal size. Nevertheless, it is needed in the derivation of (A.26). The line integral in (A.6) can be expressed as

$$\begin{aligned}\frac{1}{\text{Area}}\oint A\mathbf{n}dl =&\frac{1}{a^2\cos\varphi d\lambda d\varphi}\left[-\mathbf{e}_\varphi A_1 a\cos\varphi d\lambda+\mathbf{e}_\lambda A_2\cos\theta a d\varphi+\mathbf{e}_\varphi A_2\sin\theta a d\varphi\right.\\&\left.+\mathbf{e}_\varphi A_3 a\cos(\varphi+d\varphi)-\mathbf{e}_\lambda A_4\cos\theta a d\varphi+\mathbf{e}_\varphi A_4\sin\theta a d\varphi\right]\\=&\mathbf{e}_\lambda\frac{(A_2-A_4)\cos\theta}{a\cos\varphi d\lambda}\\&+\mathbf{e}_\varphi\left\{\frac{[A_3\cos(\varphi+d\varphi)-A_1\cos\varphi]d\lambda+(A_2+A_4)\sin\theta d\varphi}{a\cos\varphi d\lambda d\varphi}\right\}\end{aligned} \tag{A.41}$$

Note how the angle θ has entered here. Put $\cos\theta\to 1$ and $\sin\theta\to\frac{1}{2}\sin\varphi d\lambda$ to obtain

$$\begin{aligned}\frac{1}{\text{Area}}\oint\alpha\mathbf{n}dl=&\mathbf{e}_\lambda\frac{(A_2-A_4)}{a\cos\varphi d\lambda}+\mathbf{e}_\varphi\left\{\left[\frac{A_3\cos(\varphi+d\varphi)-A_1\cos\varphi}{a\cos\varphi d\varphi}\right]+\left(\frac{A_2+A_4}{2}\right)\frac{\sin\varphi}{a\cos\varphi}\right\}\\\to&\mathbf{e}_\lambda\frac{1}{a\cos\varphi}\frac{\partial A}{\partial\lambda}+\mathbf{e}_\varphi\left[\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}(A\cos\varphi)+\frac{A\sin\varphi}{a\cos\varphi}\right]\\=&\mathbf{e}_\lambda\frac{1}{a\cos\varphi}\frac{\partial A}{\partial\lambda}+\mathbf{e}_\varphi\frac{1}{a}\frac{\partial A}{\partial\varphi},\end{aligned} \tag{A.42}$$

which agrees with the two-dimensional version of (A.26).

Similar (but more straightforward) derivations can be given for (A.27) - (A.30).

A.5.4 Applying vector operators to the unit vectors in spherical coordinates

Using (A.11) - (A.13), we can prove the following about the unit vectors in spherical coordinates:

$$\nabla\cdot\mathbf{e}_\lambda=0, \tag{A.43}$$

$$\nabla\cdot\mathbf{e}_\varphi=-\frac{\tan\varphi}{r}, \tag{A.44}$$

$$\nabla \cdot \mathbf{e}_r = \frac{2}{r}, \tag{A.45}$$

$$\nabla \times \mathbf{e}_\lambda = \frac{\mathbf{e}_\varphi}{r} + \frac{\tan\varphi}{r}\mathbf{e}_r, \tag{A.46}$$

$$\nabla \times \mathbf{e}_\varphi = -\frac{\mathbf{e}_\lambda}{r}, \tag{A.47}$$

$$\nabla \times \mathbf{e}_r = 0. \tag{A.48}$$

The following relations are useful when working with the momentum equation in spherical coordinates:

$$(\mathbf{v}\cdot\nabla)\,\mathbf{e}_\lambda = \frac{u\sin\varphi}{r}\mathbf{e}_\varphi - \frac{u\cos\varphi}{r}\mathbf{e}_r\,, \tag{A.49}$$

$$(\mathbf{v}\cdot\nabla)\,\mathbf{e}_\varphi = -\frac{u\sin\varphi}{r}\mathbf{e}_\lambda - \frac{v\sin\varphi}{r}\mathbf{e}_r, \tag{A.50}$$

$$(\mathbf{v}\cdot\nabla)\,\mathbf{e}_r = \frac{\mathbf{v}}{r}\,. \tag{A.51}$$

Here $\mathbf{v}$ is the horizontal wind vector.

A.6 Solid body rotation

As an example of the application of (A.28), the vertical component of the vorticity is

$$\zeta = \frac{1}{r\cos\varphi}\left[\frac{\partial v}{\partial \lambda} - \frac{\partial}{\partial \varphi}(u\cos\varphi)\right] \tag{A.52}$$

For the case of pure solid body rotation of the atmosphere about the Earth's axis of rotation, we have

$$u = \dot{\lambda} r\cos\varphi \text{ and } v = 0, \tag{A.53}$$

where $\dot{\lambda}$ is independent of φ (because that is what "solid body rotation" means). Substitution of (A.53) into (A.52) gives

$$\begin{aligned}\zeta =& \frac{-1}{r\cos\varphi}\frac{\partial}{\partial\varphi}\left(\dot{\lambda} r\cos^2\varphi\right)\\ =& 2\dot{\lambda}\sin\varphi\ ,\end{aligned} \tag{A.54}$$

which looks just like the familiar Coriolis parameter.

A.7 Formulas that are useful for two-dimensional flow

Consider the special case of two-dimensional flow. Two useful identities are

$$\nabla_r \times (\mathbf{e}_r \times \nabla_r \alpha) = \mathbf{e}_r \nabla_r^2 \alpha, \tag{A.55}$$

and

$$\nabla_r \cdot (\mathbf{e}_r \times \nabla_r \alpha) = 0. \tag{A.56}$$

Also for two-dimensional flow, the Laplacian of a vector can be written in a very simple way. Let $\zeta \mathbf{e}_r \equiv \nabla_r \times \mathbf{v}$ and $\delta \equiv \nabla_r \cdot \mathbf{v}$. Then (A.22) reduces to

$$\nabla_r^2 \mathbf{v} = \nabla_r \delta - \nabla_r \times (\zeta \mathbf{e}_r) \tag{A.57}$$

Using (A.28), we can write

$$\begin{aligned}\nabla_r \times (\zeta \mathbf{e}_r) &= \left\{ \frac{1}{r}\frac{\partial \zeta}{\partial \varphi},\ -\frac{1}{r\cos\varphi}\frac{\partial \zeta}{\partial \lambda},\ 0 \right\} \\ &= -\mathbf{e}_r \times \nabla_r \zeta\ .\end{aligned} \tag{A.58}$$

Then (A.57) becomes

$$\nabla_r^2 \mathbf{v} = \nabla_r \delta + \mathbf{e}_r \times \nabla_r \zeta. \tag{A.59}$$

A.8 Vertical coordinate transformations

A.8.1 Basics

Consider two vertical coordinates, denoted by "z" and "$\hat{z}$," respectively. Although the symbol z suggests height, no such implication is intended here; z and $\hat{z}$ can be any variables at all, so long as they vary monotonically with height (and with each other). For example, z could be pressure and $\hat{z}$ could be potential temperature. We assume that we have a rule telling how to compute $\hat{z}$ for a given value of z, and vice versa. For example, we could define $\hat{z} \equiv z - z_S(x,y)$, where $z_S(x,y)$ is the distribution of z along the Earth's surface.

Now consider the variation of an arbitrary dependent variable, f, with the independent variables x and z, as sketched in Figure A.3. Our goal is to relate $(\partial f/\partial x)_{\hat{z}}$ to $(\partial f/\partial x)_z$. Here x is not necessarily a spatial coordinate; it could be time. With reference to Fig. A.3, we can write

$$\begin{aligned}\frac{f_B - f_A}{x_2 - x_1} &= \left(\frac{f_C - f_A}{x_2 - x_1}\right) - \left(\frac{f_C - f_B}{x_2 - x_1}\right) \\ &= \left(\frac{f_C - f_A}{x_2 - x_1}\right) - \left(\frac{f_C - f_B}{z_2 - z_1}\right)\left(\frac{z_2 - z_1}{x_2 - x_1}\right)\ .\end{aligned} \tag{A.60}$$

Taking the limit as the increments become small, we find that

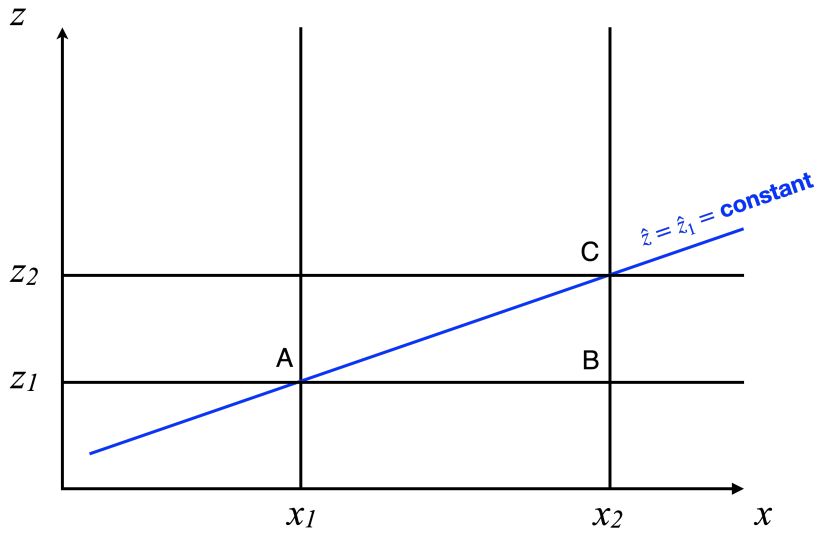


Figure A.3: A sketch used in the derivation of (A.60).

$$\boxed{\left(\frac{\partial f}{\partial x}\right)_z = \left(\frac{\partial f}{\partial x}\right)_{\hat{z}} - \left(\frac{\partial f}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_{\hat{z}} .} \tag{A.61}$$

A.8.2 Some useful operators

Starting from (A.61), we can show that the horizontal gradient satisfies

$$\begin{aligned} \nabla_z f =& \nabla_{\hat{z}} f - \frac{\partial f}{\partial z} \nabla_{\hat{z}} z \\ =& \nabla_{\hat{z}} f - \frac{\partial}{\partial z} \left(f \nabla_{\hat{z}} z \right) \\ =& \nabla_{\hat{z}} f - \frac{\partial}{\partial \hat{z}} \left(\frac{\partial \hat{z}}{\partial z} f \nabla_{\hat{z}} z \right) . \end{aligned} \tag{A.62}$$

Here we have used

$$\frac{\partial}{\partial z} \left(\nabla_{\hat{z}} z \right) = \nabla_{\hat{z}} \left(\frac{\partial z}{\partial z} \right) = \nabla_{\hat{z}} (1) = 0 . \tag{A.63}$$

The three-dimensional gradient of f can therefore be written as

$$\begin{aligned}
\nabla f =& \nabla_z f + \mathbf{k}\frac{\partial f}{\partial z} \\
=& \nabla_{\hat{z}} f - \frac{\partial}{\partial \hat{z}}\left(\frac{\partial \hat{z}}{\partial z} f \nabla_{\hat{z}} z\right) + \mathbf{k}\frac{\partial f}{\partial z} \\
=& \nabla_{\hat{z}} f - \frac{\partial}{\partial \hat{z}}\left(\frac{\partial \hat{z}}{\partial z} f \nabla_{\hat{z}} z\right) + \frac{\partial \hat{z}}{\partial z}\mathbf{k}\frac{\partial f}{\partial \hat{z}} \\
=& \nabla_{\hat{z}} f - \frac{\partial}{\partial \hat{z}}\left(\frac{\partial \hat{z}}{\partial z} f \nabla_{\hat{z}} z\right) + \frac{\partial}{\partial \hat{z}}\left(\frac{\partial \hat{z}}{\partial z}\mathbf{k} f\right) \\
=& \nabla_{\hat{z}} f + \frac{\partial}{\partial \hat{z}}\left[\frac{\partial \hat{z}}{\partial z}(-f\nabla_{\hat{z}} z + \mathbf{k} f)\right] .
\end{aligned} \tag{A.64}$$

Analogous identities apply with other operators. For example, the curl of an arbitrary horizontal vector $\mathbf{v}$ is given by

$$\begin{aligned}
\nabla_z \times \mathbf{v} =& \nabla_{\hat{z}} \times \mathbf{v} + \frac{\partial \mathbf{v}}{\partial z} \times \nabla_{\hat{z}} z \\
=& \nabla_{\hat{z}} \times \mathbf{v} + \frac{\partial \mathbf{v}}{\partial \hat{z}}\frac{\partial \hat{z}}{\partial z} \times \nabla_{\hat{z}} z .
\end{aligned} \tag{A.65}$$

The curl of an arbitrary three-dimensional vector $\mathbf{V} = \mathbf{v} + w\mathbf{k}$ can therefore be written (in Cartesian coordinates) as

$$\begin{aligned}
\nabla_z \times \mathbf{V} =& \nabla_{\hat{z}} \times \mathbf{v} + \frac{\partial \mathbf{v}}{\partial z} \times \nabla_{\hat{z}} z - \mathbf{k} \times \nabla_z w \\
=& \nabla_{\hat{z}} \times \mathbf{v} + \frac{\partial \mathbf{v}}{\partial z} \times \nabla_{\hat{z}} z - \mathbf{k} \times \left[\nabla_{\hat{z}} w - \frac{\partial}{\partial \hat{z}}\left(\frac{\partial \hat{z}}{\partial z} w \nabla_{\hat{z}} z\right)\right] .
\end{aligned} \tag{A.66}$$

Similarly, the divergence of an arbitrary horizontal vector is

$$\nabla_z \cdot \mathbf{v} = \nabla_{\hat{z}} \cdot \mathbf{v} - \frac{\partial \mathbf{v}}{\partial z} \cdot \nabla_{\hat{z}} z . \tag{A.67}$$

The divergence of an arbitrary three-dimensional vector can therefore be written (in Cartesian coordinates) as

$$\begin{aligned}\nabla\cdot\mathbf{V} =& \nabla_z\cdot\mathbf{v}+\frac{\partial w}{\partial z}\\ =& \nabla_{\hat{z}}\cdot\mathbf{v}-\frac{\partial\mathbf{v}}{\partial z}\cdot\nabla_{\hat{z}}z+\frac{\partial w}{\partial z}\\ =& \nabla_{\hat{z}}\cdot\mathbf{v}+\frac{\partial}{\partial z}\left(-\mathbf{v}\cdot\nabla_{\hat{z}}z+w\right)\\ =& \nabla_{\hat{z}}\cdot\mathbf{v}+\frac{\partial}{\partial\hat{z}}\left[-\frac{\partial\hat{z}}{\partial z}\left(\mathbf{v}\cdot\nabla_{\hat{z}}z-w\right)\right]\ .\end{aligned} \tag{A.68}$$

A.9 Concluding summary

Scalars, vectors, and tensors have meanings without reference to any particular coordinate system, although they can be expressed in coordinate systems.

This brief overview is intended mainly as a refresher for students who learned these concepts once upon a time, but may have not thought about them for awhile. We have also included many useful equations that are not readily available elsewhere, even on the Web.

Appendix B

Dimensional Analysis, Scale Analysis, and Similarity Theories

B.1 Dimensions and units

So far as we know, nature can be described using the four "dimensions" or "primary quantities" of length, time, mass, and electric charge. Here the word "dimension" is used to refer to aspects of nature that are independent in the sense that they are not inter-convertible. For example, length cannot be re-scaled somehow as a mass, and time cannot be re-scaled as an electric charge.

All physical quantities can be expressed as combinations of these four dimensions. For example, velocity can be expressed as length divided by time, and energy can be expressed as mass times length squared divided by time squared.

Temperature is an anomaly; there has been a debate about whether or not it should be included as a primary quantity independent of mass, length, and time. Temperature is a statistic of the molecular motions, and can be defined in terms of the average kinetic energy of the molecular motions per unit mass. In atmospheric science, however, temperature is usually treated as a fifth primary quantity.

Units are different from dimensions. The various primary quantities are measured using units, which can be defined in very arbitrary ways. For example, length can be measured using meters, furlongs, or the size of Henry VIII's foot. Today, scientists almost always use the metric or "SI" system of units, abbreviated from the French "Système International."

It is possible to define natural or fundamental units (e.g., Wilczek, 2005, 2006a,b), but these are not convenient for use in atmospheric science.

The starting point for our discussion is that physical principles must be independent of the choice of units. For example, Newton's law , which says that force equals mass times acceleration, must predict the same physical phenomena whether we use English units or

SI units.

We need some definitions:

1. *Physical quantity*: A conceptual property of a physical system, which can be expressed numerically in terms of one or more standards. Example: The radius of the Earth.

2. *Primary quantities*: A set of quantities (hereafter called *q*s) chosen arbitrarily for the description of a problem, subject to the constraint that the units of measurement chosen for the quantities can be assigned independently. Primary quantities are sometimes called "fundamental quantities." Examples: Length, time, mass.

3. *Dimension*: The relationship of a derived physical quantity to whatever primary quantities have been selected. Example: velocity = length / time.

4. *Standard*: An arbitrary reference measure adopted for purposes of communication. Example: meter.

5. *Unit*: An arbitrary fraction or multiple of a standard, used to avoid inconveniently large (or small) numbers. Example: kilometer.

6. *Extraneous standard*: A standard that is irrelevant for a particular problem. Example: The length of Henry VIII's foot, which is irrelevant except when Henry VIII visits a shoe store.

7. *Extraneous unit*: A unit based on an extraneous standard. Example: mile = 5280 ft.

8. *Dimensionless quantity*: A quantity that is expressed in units derivable from the problem (rather than in extraneous units). There are no intrinsically dimensionless quantities. A quantity is dimensionless or not only with respect to a particular problem. Example: Rossby number.

9. *Dimensional analysis*: The process of removing extraneous information from a problem by forming dimensionless groups.

10. *Nondimensionalization*: Conversion of a system of dimensional equations to a system that contains only nondimensional quantities.

11. *Scale analysis, sometimes called "scaling"*: Using chosen numerical values for the dimensional parameters to compare the orders of magnitude of various terms of a system of non-dimensional equations. The "chosen numerical values" are relevant to the particular problem to which we are trying to apply the equations. The purpose of a scale analysis is to identify terms of the equations that can be neglected for the particular application. A scale analysis can only be performed when the governing equations are known *and* a particular application of the equations has been chosen. For example, the equation of motion can be applied to large-scale dynamics,

for which the Coriolis terms are important, or small-scale dynamics, for which the Coriolis terms may be negligible.

12. *Similar systems*: Those for which the relevant dimensionless quantities have identical values, even though the dimensional quantities may be very different. For example, a small model airplane in a wind tunnel can be dynamically similar to a full-size airplane flying near the tropopause.

13. *Similarity theory*: A theory based on the hypothesis that functional relationships exist among the nondimensional parameters describing a physical system. The functions themselves are not given by the similarity theory, and must be determined empirically. Similarity theories can be useful when the desired functions cannot be derived from the governing equations.

B.2 Consistent use of dimensional and nondimensional quantities

It makes no sense to add or subtract quantities that have different dimensions. You cannot add a length to a mass.

All terms of an equation must have the same dimensions. This can be called the requirement of dimensional homogeneity.

An exponent must be nondimensional, because it is the number (pure number) of times that something is multiplied by itself.

Similarly, the arguments of (inputs to) mathematical functions must be nondimensional. Examples include exponentials, logarithms and trigonometric functions. You can take the sine of 2 but you cannot take the sine of 2 meters.

It is not unusual to see dimensional quantities as the arguments of logarithms, even though it makes no sense. For example, you might see something like this:

$$\frac{1}{\theta}\frac{D\theta}{Dt} = \frac{D}{Dt}\left(\ln\theta\right) ,$$

where θ is the potential temperature. A more correct (but longer) statement is this:

$$\frac{1}{\theta}\frac{D\theta}{Dt} = \frac{D}{Dt}\left[\ln\left(\frac{\theta}{\theta_{ref}}\right)\right] ,$$

where θ_{ref} is a constant reference value. It seems that people just don't want to be bothered with including θ_{ref} and then saying "where θ_{ref} is a constant reference value." Too much trouble.

B.3 The Buckingham Pi Theorem

The fundamental theorem of dimensional analysis is due to Buckingham, and is stated here without proof:

If the equation

$$\phi\left(q_1, q_2 \ldots, q_n\right) = 0 \tag{B.1}$$

is the only relationship among the q_i, and if it holds for any arbitrary choice of units in which the $q_1, q_2 \ldots, q_n$ are measured, then (B.1) can be written in the form

$$\phi\left(\pi_1, \pi_2 \ldots, \pi_m\right) = 0\,, \tag{B.2}$$

where the $\pi_1, \pi_2 \ldots, \pi_m$ are independent nondimensionless products or ratios of the qs. Furthermore, if k is the minimum number of primary quantities necessary to express the dimensions of the qs, then

$$m = n - k\,. \tag{B.3}$$

Since $k > 0$, (B.3) implies that $m < n$. According to (B.3), the number of dimensionless products is the number of dimensional parameters minus the number of primary quantities.

Another way of writing (B.2) is

$$\phi'\left(\pi_1, \pi_2, \ldots, \pi_m; 1, \ldots, 1\right) = 0\,, \tag{B.4}$$

where the number of "1s" appearing in the argument list is k. Clearly the 1's carry no information about the functional relationship among the qs, so we can just omit them, as was done in (B.2). In (B.4), the 1s clearly represent "extraneous" information, which entered the problem through the extraneous units of the qs. By dropping the extraneous information, we simplify the problem. We don't actually change the problem. We just boil it down to its essentials.

The choice of the qs can be made by inspection of the governing equations (if known), or by inspection of the physical system.

The dimensions of the qs can be determined in terms of chosen primary quantities. The primary quantities can be chosen arbitrarily, provided that their units can be assigned

independently. It is necessary to choose enough primary quantities to ensure that nondimensional combinations can be formed in all cases.

B.4 An example of dimensional analysis

Consider thermal convection of a shallow fluid in a laboratory tank, with no mean flow or rotation. The linearized governing equations are:

$$\left(\frac{\partial}{\partial t} - \nu\nabla^2\right) u' = -\frac{\partial}{\partial x}\left(\frac{p'}{\rho_0}\right) ,$$

$$\left(\frac{\partial}{\partial t} - \nu\nabla^2\right) v' = -\frac{\partial}{\partial y}\left(\frac{p'}{\rho_0}\right) ,$$

$$\left(\frac{\partial}{\partial t} - \nu\nabla^2\right) w' = -\frac{\partial}{\partial z}\left(\frac{p'}{\rho_0}\right) + g\alpha T' ,$$

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z} = 0 ,$$

$$\left(\frac{\partial}{\partial t} - \kappa\nabla^2\right) T' = -w'\Gamma .$$

Here x, y, and z are the spatial coordinates, u, v, and w are the corresponding components of the velocity vector, T is temperature, p is pressure, ρ_0 is a constant reference-state density, g is the acceleration of gravity, α is a thermal expansion coefficient, ν is the molecular viscosity, κ is the molecular conductivity, and $\Gamma \equiv -\partial\overline{T}/\partial z$ is the rate at which temperature increases upward in the mean state. An overbar denotes the mean state, and a prime denotes the departure from the mean state.

By inspection of the equations, we see that the six dimensional parameters of the problem are

$$g, \Gamma, h, \nu, \kappa, \text{ and } \alpha .$$

These parameters are determined by the design of the laboratory experiment, e.g., the depth of the tank and the choice of convecting fluid (water, air, oil, etc.). We don't include the dependent variables u, v, w, p, or T in the list, because they are part of the solution, and so depend on the parameters listed above. We don't include ρ_0 in the list because it appears only in the combination p'/ρ_0, and so we just think of p'/ρ_0 as one of the dependent variables.

Table B.1: The dimensions of the dimensional quantities used in the nondimensionalization of the equations that govern thermal advection.

Quantity	L	T	Θ
g	1	-2	0
α	0	0	-1
Γ	-1	0	1
h	1	0	0
ν	2	-1	0
κ	2	-1	0

As primary quantities, we choose length (L), time (T), and temperature (Θ). Since there are three primary quantities and six dimensional parameters, the Π-theorem tells us that we should be able to eliminate $6-3=3$ pieces of extraneous information. The dimensions of the qs are shown in Table B.1. As our pertinent (non-extraneous) length scale, we choose h. Forming products, we systematically eliminate the "lengths" from our set of quantities, as shown in Table B.2.

Table B.2: As in Table B.1, but with the dimension of length eliminated.

Quantity	L	T	Θ
gh^{-1}	0	-2	0
α	0	0	-1
Γh	0	0	1
1	0	0	0
νh^{-2}	0	-1	0
κh^{-2}	0	-1	0

As our time scale, we use $h^2\nu^{-1}$. (We could just as well use $h^2\kappa^{-1}$, and the final result would be the same.) Again forming products, we obtain the results shown in Table B.3.

Obviously, to complete the procedure, we simply form the product $\Gamma\alpha h$. The final result is shown in Table B.4.

Table B.3: As in Table B.2, but with the dimension of time eliminated.

Quantity	L	T	Θ
$gh^3\nu^{-2}$	0	0	0
α	0	0	-1
Γh	0	0	1
1	0	0	0
1	0	0	0
$\kappa\nu^{-1}$	0	0	0

Table B.4: As in Table B.3, but with all dimensional quantities eliminated.

Quantity	L	T	Θ
$gh^3\nu^{-2}$	0	0	0
$\Gamma\alpha h$	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
$\kappa\nu^{-1}$	0	0	0

There are three 1s in the header column of Table B.4. This means that three pieces of extraneous information have been eliminated, as promised by the Π-theorem. We have the three nondimensional combinations

$$x_1 \equiv gh^3\nu^{-2} \,, \tag{B.5}$$

the Prandtl number, which is defined by

$$\mathrm{Pr} \equiv \kappa\nu^{-1} \,, \tag{B.6}$$

and

$$x_2 \equiv \Gamma \alpha h \,, \tag{B.7}$$

It is conventional to define the Rayleigh number, Ra, *by*

$$\begin{aligned} \mathrm{Ra} =& \mathrm{Pr}^{-1} x_1 x_2 \\ =& \kappa^{-1} \nu g h^3 \nu^{-2} \Gamma \alpha h \\ =& \frac{g \alpha \Gamma h^4}{\kappa \nu} \,. \end{aligned} \tag{B.8}$$

We can, alternatively, regard Pr, Ra, and x_1 (say) as our three nondimensional combinations.

Chandrasekhar (2013) shows that only two dimensionless combinations matter for the convection problem, namely Pr and Ra. Why, then, have we found three? In the governing equations, g and α appear only in the combination $g\alpha$, so they don't have to be separately included in our list of qs. Because this reduces the number of qs by one, without reducing the number primary quantities, the Π-theorem tells us that the number of dimensionless combinations will also be reduced by one. The lesson here is that we should not enter dimensional quantities separately if they enter the equations only in one particular combination.

B.5 Scale analysis

In atmospheric science and oceanography, scale analysis is very widely used to justify various approximations that can aid in the solution of the governing equations. Scale analysis always starts from a set of governing equations.

Here is an important example: Quasi-geostrophic theory was derived and justified by Charney (1990) on the basis of a scale analysis. We now perform a simple scale analysis of the equation of motion, in the spirit of Charney's work. The equation of motion can be written in simplified form as

$$\frac{D\mathbf{V}}{Dt} + f\mathbf{k} \times \mathbf{V} = -\nabla_p \phi \,. \tag{B.9}$$

Here $\mathbf{V}$ is the horizontal wind vector. We have omitted friction, for simplicity. The three terms included in (B.9) suffice to describe the evolution of the large-scale horizontal wind throughout most of the atmosphere.

We can nondimensionalize (B.9) through a straightforward, almost mechanical procedure, as follows. Let U be a velocity scale, and write

$$\mathbf{V} = U\widehat{\mathbf{V}} \,,$$

where the "hat" notation denotes a nondimensional variable. Similarly, let L be a (dimensional) length scale. We can then construct a time scale as $T = L/U$, and write

$$t = T\widehat{t} = (L/U)\widehat{t} \,,$$

and

$$\frac{D}{Dt} = \frac{1}{T}\frac{D}{D\widehat{t}} = \frac{U}{L}\frac{D}{D\widehat{t}} ,$$

Finally, let $f = f_0\widehat{f}$, where f_0 is a suitably chosen representative (dimensional) value of the Coriolis parameter. Note that to nondimensionalize (B.9) we do not need scales for mass or temperature.

Making the various substitutions into (B.9), we can rewrite it as

$$\frac{U^2}{L}\frac{D\widehat{\mathbf{V}}}{D\widehat{t}} + f_0\widehat{f}\mathbf{k}\times U\widehat{\mathbf{V}} = -\nabla_p\phi \,. \tag{B.10}$$

Dividing through by f_0U, we find that

$$\mathrm{Ro}\frac{D\widehat{\mathbf{V}}}{D\widehat{t}} + \widehat{f}\mathbf{k}\times\widehat{\mathbf{V}} = -\frac{\nabla_p\phi}{f_0U} \,, \tag{B.11}$$

where Ro is the (nondimensional) Rossby number. Equation (B.11) is the nondimensional form of (B.9).

In converting the dimensional equation (B.9) to the nondimensional form (B.11) we have not really changed anything. The problem to be solved remains the same. Nondimensionalization does remove extraneous information, however, and that simplifies analysis. Nondimensionalization also reveals nondimensional parameters of physical importance, such as the Rossby number. When we nondimensionalized the equations governing

Rayleigh convection, the nondimensional Rayleigh and Prandtl numbers emerged in much the same way that the Rossby number emerged above.

As mentioned earlier, two systems are said to be "similar" if their nondimensional parameters have the same values, even though their dimensional parameters are quite different. The most familiar example is a wind-tunnel, in which small models are used to investigate the aerodynamic properties of full-scale aircraft. We can also construct laboratory analogues of the large-scale atmospheric motions. The scale analysis presented above suggests that, in order to be a useful analogue of the atmosphere, the laboratory system must be designed to have the same Rossby number as the atmosphere.

Nondimensionalization is the first step of a scale analysis. The second step is to choose the numerical values of the scales. They are chosen so that the nondimensional dependent variables are of order one, for the problem of interest. For example, suppose that we want to investigate large-scale midlatitude motions in the Earth's atmosphere. We choose U to be 10 m s^{-1}, to be 10^6 m, and f_0 to be 10^{-4} s^{-1}, because these are about the right size for midlatitude large-scale motions.

Using these scales, we find that Ro $\cong 0.1$. This means that the (leading) acceleration term of (B.11) is an order of magnitude smaller than the Coriolis term. In order for the equation to be satisfied, the Coriolis term has to be balanced by the only remaining term, which represents the pressure-gradient force. Then (B.11) reduces to

$$\widehat{f}\mathbf{k}\times\widehat{\mathbf{V}} = -\frac{\nabla_p\phi}{f_0 U}\,, \tag{B.12}$$

which is a statement of geostrophic balance. Since $\nabla_p \sim 1/L$, we conclude that the horizontal variations of ϕ on pressure surfaces, denoted by $\delta\phi$, satisfy

$$\delta\phi \sim f_0 U L\,. \tag{B.13}$$

The scale analysis has thus allowed us to deduce the order of magnitude of $\delta\phi$.

The conclusions drawn above depend on our choices for the numerical values of the dimensional scales U, L, and f_0. If we were interested in a different meteorological phenomenon, such as individual turbulent eddies in the boundary layer, we would make different choices for the scales, and reach different conclusions.

B.6 Similarity theories

Scale analysis makes use of the equations that describe a physical system. Unfortunately, there are many cases in which we don't know those equations.

For example: The wind produces and experiences a drag force as the air moves near the Earth's surface. We hypothesize that there exists (i.e., it is possible to find) a "formula" that tells us how to compute the drag given the wind speed, the lapse rate of temperature, the roughness of the ground (or ocean), and perhaps several other dimensional quantities. We don't know how to derive the formula from the equation of motion and the other basic physical principles of our science, but we believe that the formula exists.

Similarity theories aim to find such formulas by means of hypotheses, which could be described as (hopefully) inspired guesses. We start by writing down the list of dimensional parameters that appear in the basic governing equations, including the boundary conditions. In general, this would include such things as the spatial coordinates and time, as well as parameters like the Earth's rotation rate and the acceleration of gravity. The list of dimensional parameters can be very long, especially when you consider that it could in principle include detailed information about such things as the Earth's topography.

We want to make the list shorter. As a first step, we can hypothesize, often with good reason, that some of the dimensional quantities are irrelevant to the problem at hand. For example, I'm pretty sure that the height of Mt. Everest is irrelevant to the frictional force that the air experiences as it flows over Kansas City. As a less ludicrous example, we might omit the density of water vapor from the list of dimensional parameters to be used in our search for the drag formula.

Having settled on a reasonably short list of dimensional parameters, we then nondimensionalize, and the Buckingham Pi Theorem tells us that this will yield an even shorter list of nondimensional parameters. We then assert that there are functional relationships among the nondimensional combinations of interest. If the number of parameters is small enough we have a reasonable hope of finding the functional relationships empirically. This is exactly what has been done in developing the famous Monin-Obukhov similarity theory, which provides very useful empirical formulas for determining the surface fluxes of momentum and sensible heat in terms of the mean wind and temperature profiles near the Earth's surface.

As a second example, which connects with the discussion of the preceding section, we can imagine two laboratory convection experiments in which h, ν, κ, $g\alpha$ and Γ are all different. We would like to find a formula that relates the upward heat flux in the convection tank to the nondimensional parameters of the problem, namely Pr and Ra. We don't know how to derive such a formula from first principles. A possible similarity hypothesis is that if the nondimensional parameters Pr and Ra are the same in the two experiments, then the (nondimensional) heat flux will also be the same.

This turns out to be true. If the suitable nondimensionalized heat fluxes from many "similar" experiments are plotted in nondimensional form, all of the data fall neatly onto families of curves. For example, for a given value of Pr, we can plot the experimentally determined nondimensional heat flux against Ra. For each value of Pr, the data fall onto orderly curves. The similarity theory does not tell us the shapes of the curves; it only asserts that the curves exist. The shapes of the curves have to be determined empirically, but at least the similarity theory tells us what to look for. If the same data were plotted in dimensional form, a "scattered" set of points would result; no order would be apparent.

A famous example of similarity analysis was provided by G. I. Taylor (Taylor, 1950a,b), who analyzed a time sequence of photographs of a nuclear explosion. Using his (formidable) intuition, he chose the qs for the problem that are shown in Table B.5. The five dimensional parameters in the table can be described using the three primary quantities length, time, and mass. Taylor used dimensional analysis to identify the following two nondimensional parameters:

$$\frac{\rho_0 R^5}{E t^2} \text{ and } \frac{p_0 R^3}{E} \, . \tag{B.14}$$

He could then assert that

$$\frac{\rho_0 r^5}{E t^2} = f\left(\frac{p_0 R^3}{E}\right) \, . \tag{B.15}$$

where f is a function to be determined empirically.

Table B.5: The dimensional quantities used by G. I. Taylor in his analysis of nuclear explosions.

Symbol	Definition	Representative value or first guess
R	Radius of wave front	10^2 m
t	Time	10^{-2} s
p_0	Ambient pressure	10^5 Pa
ρ_0	Ambient density	1 kg m^{-3}
E	Energy released	10^{14} J

Actually, Taylor did better than that. With the numerical values given in the third column of Table B.5, including a "first guess" at the value of E, Taylor estimated that

$$\frac{\rho_0 R^5}{E t^2} = 1 \ , \tag{B.16}$$

and

$$\frac{p_0 R^3}{E} \ll 1 \ . \tag{B.17}$$

The nuclear fireball, with its huge internal pressure, does not give a damn about the relatively puny ambient pressure. Taylor concluded that the second parameter is physically irrelevant, implying that (B.15) can be replaced by

$$\frac{\rho_0 R^5}{E t^2} = 1 \ . \tag{B.18}$$

where A is expected to be close to one. This implies that

$$R^5 = \left(\frac{E}{\rho_0} \right) t^2 \ . \tag{B.19}$$

The factor E/ρ_0 is, of course, independent of time. Using published (unclassified) magazine photos of the explosion, Taylor was able to confirm that R^5 increased in proportion to t^2. He then used (B.19) to estimate E. His accurate estimate embarrassed the government, which had not declassified the amount of energy released in the explosion.

B.7 Summary

Dimensional reasoning is very common in atmospheric science and engineering. It is used in several different ways.

Mass, length, time, and temperature suffice to describe the physical quantities used in most atmospheric science work. The Buckingham Pi Theorem tells us that a physical problem can be described most concisely by using only nondimensional combinations of the dimensional variables. Dimensional analysis can be used to identify such combinations.

Scale analysis goes further, by using chosen dimensional scales, i.e., numerical values expressed in units, to determine the dominant terms of the nondimensionalized equations that describe a physical system. The scales are chosen so that the nondimensional unknowns are of order one for the problem of interest. Scale analysis always starts from the known governing equations, so it can only be carried out if the governing equations are known.

Finally, similarity theories are used to find empirical relationships among the nondimensional parameters that characterize a physical system. Similarity theories are used when the equations that are sought cannot be derived from first principles.

Similar systems are those for which the relevant nondimensional parameters take the same values. The concept of similar systems is relevant to both scale analysis and similarity theories.

Fig. B.1 is an attempt to summarize the logical relationships between dimensional analysis, scale analysis, and similarity theories.

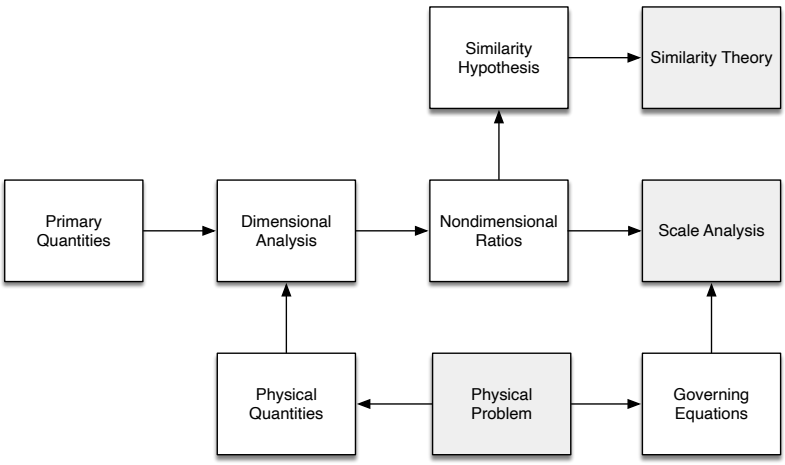


Figure B.1: An attempt to summarize the logical connections among the various topics discussed in this essay. Everything flows from the physical problem, and the end-points are similarity theory and scale analysis.

Appendix C

Derivation of the Isentropic System

C.1 The horizontal pressure-gradient force

The horizontal pressure-gradient force per unit volume can be written as

$$\nabla_z p = \nabla_\theta p - \frac{\partial p}{\partial z}\nabla_\theta z\,, \tag{C.1}$$

and the horizontal gradient of the geopotential as

$$\begin{aligned}\nabla_z \phi =& \nabla_\theta \phi - \frac{\partial \phi}{\partial z}\nabla_\theta z \\ =& \nabla_\theta \phi - g\nabla_\theta z\,.\end{aligned} \tag{C.2}$$

The first term on the right-hand side of (C.1) can be rewritten, using (4.67), (4.68), and the ideal gas law:

$$\begin{aligned}\nabla_\theta p =& \frac{p}{\kappa}\frac{\nabla_\theta \Pi}{\Pi} \\ =& \rho\theta\nabla\left(c_p T\right) \\ =& \rho\left(\nabla_\theta s - g\nabla_\theta z\right)\,.\end{aligned} \tag{C.3}$$

Here s is the dry static energy, which was introduced earlier in this chapter in the context of energy conservation. It may be confusing to see the same variable in the expression

for the horizontal pressure gradient force as expressed in isentropic coordinates. Yes, it really is the same variable, playing two very different roles in our equations. In the context of the pressure gradient force, the dry static energy is sometimes called the "Montgomery potential."

C.2 The vertical pressure-gradient force

Similarly, we can write minus the vertical pressure-gradient force per unit volume as

$$\begin{aligned}\frac{\partial p}{\partial z} =& \frac{p}{\kappa\Pi}\frac{\partial \Pi}{\partial z} \\ =& \rho\theta\frac{\partial \Pi}{\partial z} \\ =& \rho\left[\frac{\partial}{\partial z}(\Pi\theta) - \Pi\frac{\partial \theta}{\partial z}\right] \\ =& \rho\left[\frac{\partial}{\partial z}(c_p T) - \Pi\frac{\partial \theta}{\partial z}\right] \\ =& \rho\left(\frac{\partial s}{\partial z} - g - \Pi\frac{\partial \theta}{\partial z}\right) \\ =& \rho\left[\frac{\partial \theta}{\partial z}\left(\frac{\partial s}{\partial \theta} - \Pi\right) - g\right] .\end{aligned} \tag{C.4}$$

Using (C.3) and (C.4) in (C.1), we find that

$$\begin{aligned}\nabla_z p =& \rho\left(\nabla_\theta s - g\nabla_\theta z\right) - \rho\left[\frac{\partial \theta}{\partial z}\left(\frac{\partial s}{\partial \theta} - \Pi\right) - g\right]\nabla_\theta z \\ =& \rho\left[\nabla_\theta s - \frac{\partial \theta}{\partial z}\left(\frac{\partial s}{\partial \theta} - \Pi\right)\nabla_\theta z\right] .\end{aligned} \tag{C.5}$$

With the use of (C.2), (C.4) and (C.5), we can rewrite (4.70) and (4.71) as

$$\frac{D\mathbf{v}}{Dt} + 2\left[\boldsymbol{\Omega}_\varphi \times (w\mathbf{e}_r) + \boldsymbol{\Omega}_z \times \mathbf{v}\right] = -\left(\nabla_\theta \phi - g\nabla_\theta z\right) - \left[\nabla_\theta s - \frac{\partial \theta}{\partial z}\left(\frac{\partial s}{\partial \theta} - \Pi\right)\nabla_\theta z\right] - \mathscr{F}_{\mathbf{v}} , \tag{C.6}$$

and

$$\frac{Dw}{Dt} + 2\left(\boldsymbol{\Omega}_h \times \mathbf{v}\right) \cdot \mathbf{e}_r = -\frac{\partial \theta}{\partial z}\left(\frac{\partial s}{\partial \theta} - \Pi\right) - \mathscr{F}_w \,. \tag{C.7}$$

In both of these equations, all factors of ρ have cancelled out. Inspection of (C.7) shows that in isentropic coordinates the hydrostatic equation can be written as

$$\frac{\partial s}{\partial \theta} - \Pi = 0 \,. \tag{C.8}$$

C.3 The continuity equation

Next, we transform the continuity equation, (4.72). The isentropic pseudodensity is defined by

$$\rho_\theta \equiv \rho \frac{\partial z}{\partial \theta} \,. \tag{C.9}$$

Differentiation of (4.120) gives

$$\left(\frac{\partial \rho_\theta}{\partial t}\right)_\theta = \frac{\partial z}{\partial \theta}\left(\frac{\partial \rho}{\partial t}\right)_\theta + \rho \frac{\partial}{\partial \theta}\left(\frac{\partial z}{\partial t}\right)_\theta \,, \tag{C.10}$$

and

$$\nabla_\theta \cdot (\rho_\theta \mathbf{v}) = \frac{\partial z}{\partial \theta} \nabla_\theta \cdot (\rho \mathbf{v}) + \rho_\theta \mathbf{v} \cdot \frac{\partial}{\partial \theta} \nabla_\theta z \tag{C.11}$$

Here we have used the coordinate transformation methods described in Appendix A.

Using (A.61), we can write

$$\left(\frac{\partial \rho}{\partial t}\right)_z = \left(\frac{\partial \rho}{\partial t}\right)_\theta - \frac{\partial \theta}{\partial z}\frac{\partial \rho}{\partial \theta}\left(\frac{\partial z}{\partial t}\right)_\theta \tag{C.12}$$

and

$$\nabla_z \cdot (\rho \mathbf{v}) = \nabla_\theta \cdot (\rho \mathbf{v}) - \frac{\partial \theta}{\partial z}\left[\frac{\partial}{\partial \theta}(\rho \mathbf{v})\right] \cdot \nabla_\theta z \,. \tag{C.13}$$

Using these results, the continuity equation Eq. (4.72) can be written as

$$\left(\frac{\partial \rho}{\partial t}\right)_\theta - \frac{\partial \theta}{\partial z}\frac{\partial \rho}{\partial \theta}\left(\frac{\partial z}{\partial t}\right)_\theta + \nabla_\theta \cdot (\rho \mathbf{v}) - \frac{\partial \theta}{\partial z}\left[\frac{\partial}{\partial \theta}(\rho \mathbf{v})\right] \cdot \nabla_\theta z + \frac{\partial \theta}{\partial z}\frac{\partial}{\partial \theta}(\rho w) = 0 \,, \tag{C.14}$$

or

$$\frac{\partial z}{\partial \theta}\left(\frac{\partial \rho}{\partial t}\right)_\theta - \frac{\partial \rho}{\partial \theta}\left(\frac{\partial z}{\partial t}\right)_\theta + \frac{\partial z}{\partial \theta}\nabla_\theta \cdot (\rho \mathbf{v}) - \left[\frac{\partial}{\partial \theta}(\rho \mathbf{v})\right] \cdot \nabla_\theta z + \frac{\partial}{\partial \theta}(\rho w) = 0 \,. \tag{C.15}$$

Substituting (C.10) and (C.11) into (C.15), we obtain

$$\begin{aligned}\left(\frac{\partial \rho_\theta}{\partial t}\right)_\theta - \rho \frac{\partial}{\partial \theta}\left(\frac{\partial z}{\partial t}\right)_\theta - \frac{\partial \rho}{\partial \theta}\left(\frac{\partial z}{\partial t}\right)_\theta + \nabla_\theta \cdot (\rho_\theta \mathbf{v}) - \rho \mathbf{v} \cdot \frac{\partial}{\partial \theta}(\nabla_\theta z) \\ - \left[\frac{\partial}{\partial \theta}(\rho \mathbf{v})\right] \cdot \nabla_\theta z + \frac{\partial}{\partial \theta}(\rho w) = 0 \,,\end{aligned} \tag{C.16}$$

or, after collecting terms,

$$\left(\frac{\partial \rho_\theta}{\partial t}\right)_\theta + \nabla_\theta \cdot (\rho_\theta \mathbf{v}) - \frac{\partial}{\partial \theta}\left\{\rho\left[\left(\frac{\partial z}{\partial t}\right)_\theta + \mathbf{v} \cdot \nabla_\theta z - w\right]\right\} = 0 \,. \tag{C.17}$$

C.4 The vertical mass flux

Referring back to the discussion of Eqs. (4.13) and (4.15), we recognize

$$\mu \equiv -\rho \left[\left(\frac{\partial z}{\partial t} \right)_\theta + \mathbf{v} \cdot \nabla_\theta z - w \right] \tag{C.18}$$

as the upward mass flux across an isentropic surface. The minus sign appears in (C.18) because an increase in the height of an isentropic surface corresponds to mass moving "downward" relative to that surface. The definition (C.18) allows us to simplify the continuity equation (C.17) to

$$\left(\frac{\partial \rho_\theta}{\partial t} \right)_\theta + \nabla_\theta \cdot (\rho_\theta \mathbf{v}) + \frac{\partial \mu}{\partial \theta} = 0. \tag{C.19}$$

C.5 The equation for an intensive scalar

In a similar way, the conservation equation for an intensive scalar, (4.75), becomes

$$\left[\frac{\partial}{\partial t} (\rho_\theta A) \right]_\theta + \nabla_\theta \cdot (\rho_\theta \mathbf{v} A) + \frac{\partial}{\partial \theta} (\mu A) = \rho_\theta S_A \ . \tag{C.20}$$

The advective form corresponding to (C.20) can be obtained by combining (C.19) with (C.20):

$$\rho_\theta \left(\frac{\partial A}{\partial t} \right)_\theta + \rho_\theta \mathbf{v} \cdot \nabla_\theta A + \mu \frac{\partial A}{\partial \theta} = \rho_\theta S_A \ . \tag{C.21}$$

As a special case of (C.21), the advective form of the potential temperature equation is

$$\mu = \rho_\theta \dot{\theta} \ , \tag{C.22}$$

where we have used $\dot{\theta} = S_\theta$.

C.6 Final forms of the prognostic equations

Using (C.22), we can now rewrite (C.18), (C.19) and (C.20) as

$$\rho\left[\left(\frac{\partial z}{\partial t}\right)_\theta + \mathbf{v}\cdot\nabla_\theta z - w\right] = -\rho_\theta\dot{\theta}\ , \tag{C.23}$$

$$\left(\frac{\partial \rho_\theta}{\partial t}\right)_\theta + \nabla_\theta\cdot(\rho_\theta\mathbf{v}) + \frac{\partial}{\partial\theta}\left(\rho_\theta\dot{\theta}\right) = 0\ , \tag{C.24}$$

and

$$\left[\frac{\partial}{\partial t}(\rho_\theta A)\right]_\theta + \nabla_\theta\cdot(\rho_\theta\mathbf{v}A) + \frac{\partial}{\partial\theta}\left(\rho_\theta\dot{\theta}A\right) = \rho_\theta S_A\ , \tag{C.25}$$

respectively. The Lagrangian time derivative is given by

$$\frac{D}{Dt} = \left(\frac{\partial}{\partial t}\right)_\theta + \mathbf{v}\cdot\nabla_\theta + \dot{\theta}\frac{\partial}{\partial\theta}\ . \tag{C.26}$$

Appendix D

Means, eddies, and transients

The time average of the product of two variables v and T can be expanded as follows:

$$\begin{aligned}\overline{vT}^t =& \overline{(\overline{v}^t + v')\left(\overline{T}^t + T'\right)}^t \\ =& \overline{\overline{v}^t\overline{T}^t + v'T' + \overline{v}^t T' + v'\overline{T}^t}^t \\ \cong& \overline{v}^t\overline{T}^t + \overline{v'T'}^t \ .\end{aligned} \tag{D.1}$$

As indicated, the last line of (D.1) is only approximate, unless the averaging interval is infinite, or unless the time averages are taken over discrete, non-overlapping blocks of time. The product $\overline{v'T'}^t$ is the temporal covariance of v and T. We can decompose the first term of (D.1) into its zonal mean and eddy components:

$$\begin{aligned}\overline{v}^t\overline{T}^t =& \overline{\left(\overline{v}^\lambda + v^*\right)}\,\overline{\left(\overline{T}^\lambda + T^*\right)}^t \\ =& \left(\overline{v}^{t,\lambda} + \overline{v^*}^t\right)\left(\overline{T}^{t,\lambda} + \overline{T^*}^t\right) \\ =& \overline{v}^{t,\lambda}\overline{T}^{t,\lambda} + \overline{v^*}^t\overline{T^*}^t + \overline{v}^{t,\lambda}\overline{T}^* + \overline{v^*}^t\overline{T}^{t,\lambda} \ .\end{aligned} \tag{D.2}$$

It follows that

$$\overline{\overline{v}^t\overline{T}^t}^\lambda = \overline{v}^{t,\lambda}\overline{T}^{t,\lambda} + \overline{\overline{v^*}^t\overline{T^*}^t}^\lambda \ . \tag{D.3}$$

Similarly,

$$\overline{v'T'}^{\lambda,t} = \overline{\overline{v}^{\lambda}\overline{T'}^{\lambda}}^{t} + \overline{v'^{*}T'^{*}}^{\lambda,t} \tag{D.4}$$

Both (D.3) and (D.4) are exact. Finally, substitution into the zonal average of (D.1) gives

$$\overline{vT}^{t,\lambda} = \overline{v}^{t,\lambda}\overline{T}^{t,\lambda} + \overline{\overline{v^{*}}^{t}\overline{T^{*}}^{t}}^{\lambda} + \overline{\overline{v'}^{\lambda}\overline{T'}^{\lambda}}^{t} + \overline{v'^{*}T'^{*}}^{\lambda,t}. \tag{D.5}$$

"Recipes" for some statistics of interest are given in Table D.1.

Table D.1: Notation used to denote time averages, departures from time averages, zonal averages, and departures from zonal averages.

1	$\overline{v}^{t}$, $\overline{T}^{t}$	Time mean fields
2	$\overline{vT}^{t}$	Time average of a product
3	$\overline{v}^{\lambda}\overline{T}^{\lambda}$	Product of zonal means
4	$\overline{v}^{t}\overline{T}^{t}$	Product of time averages
5	$\overline{v'T'}^{t} = \overline{vT}^{t} - \overline{v}^{t}\overline{T}^{t}$	Total transient
6	$\overline{v}^{t,\lambda}\overline{T}^{t,\lambda}$	Stationary symmetric
7	$\overline{\overline{v^{*}}^{t}\overline{T^{*}}^{t}}^{\lambda} = \overline{\overline{v}^{t}\overline{T}^{t}}^{\lambda} - \overline{v}^{t,\lambda}\overline{T}^{t,\lambda}$	Stationary eddy
8	$\overline{\overline{v'}^{\lambda}\overline{T'}^{\lambda}}^{t}$	Transient symmetric
9	$\overline{v'^{*}T'^{*}}^{t,\lambda}$	Transient eddy
10	$\overline{v^{*}T^{*}}^{t,\lambda}$	Total eddy

Here are some examples of quantities in each category. Suppose that v is the meridional wind, and T is temperature. Then vT can be called the meridional flux of temperature, which is defined at each point in space. Applying a time average, we get $\overline{vT}^{t}$, which is the time-averaged meridional flux of temperature. A product of time averages, such as $\overline{v}^{t}\overline{T}^{t}$, can represent the flux of temperature due to the time-averaged meridional wind and the time-averaged temperature. This quantity does not include additional fluxes that can arise from fluctuations on time scales shorter than the averaging interval.

The product $\overline{v}^{\lambda}\overline{T}^{\lambda}$ is the instantaneous contribution to the meridional flux of temperature that comes from the zonally averaged meridional wind and temperature, e.g., the Hadley circulation. The total *eddy* temperature flux can be written as the total flux minus $\overline{v}^{\lambda}\overline{T}^{\lambda}$. The transient symmetric flux, shown in row 8 of the table, arises from the temporally varying part of the zonally averaged circulation. The transient eddy flux, shown in row 9 of the table, comes from the parts of v and T that vary both zonally and in time.

Fig. D.1 provides an example. The top map shows the eddy part of the 850 hPa meridional wind averaged over the month of January, i.e., $\overline{v^{*}}^{t}$; the corresponding map of the full field was given earlier. The middle panel shows a similar plot for the eddy temperature field at 850 hPa, i.e., $\overline{T^{*}}^{t}$; again, a map of the full field was presented earlier. The eddy meridional wind field looks similar to the full meridional wind field simply because the zonal mean of the meridional wind is fairly small at all latitudes, especially away from the Equator. In contrast, the eddy temperature field looks very different from the full temperature field; the eddy field is lumpy, while the full field has a strong tendency towards east-west stripes. The stripes are "removed" when the zonal mean is subtracted from the full field to construct the eddy field. The bottom panel of Fig. D.1 shows a map of the product, $\overline{v^{*}}^{t}\overline{T^{*}}^{t}$. To its right is the corresponding zonal mean plot showing $\overline{\overline{v^{*}}^{t}\overline{T^{*}}^{t}}^{\lambda}$, i.e., the zonally averaged meridional temperature flux due to stationary eddies. The flux is largest in the middle latitudes of the Northern Hemisphere; as discussed later, strong stationary eddies are produced, in part, by flow over topography in the Northern Hemisphere in winter.

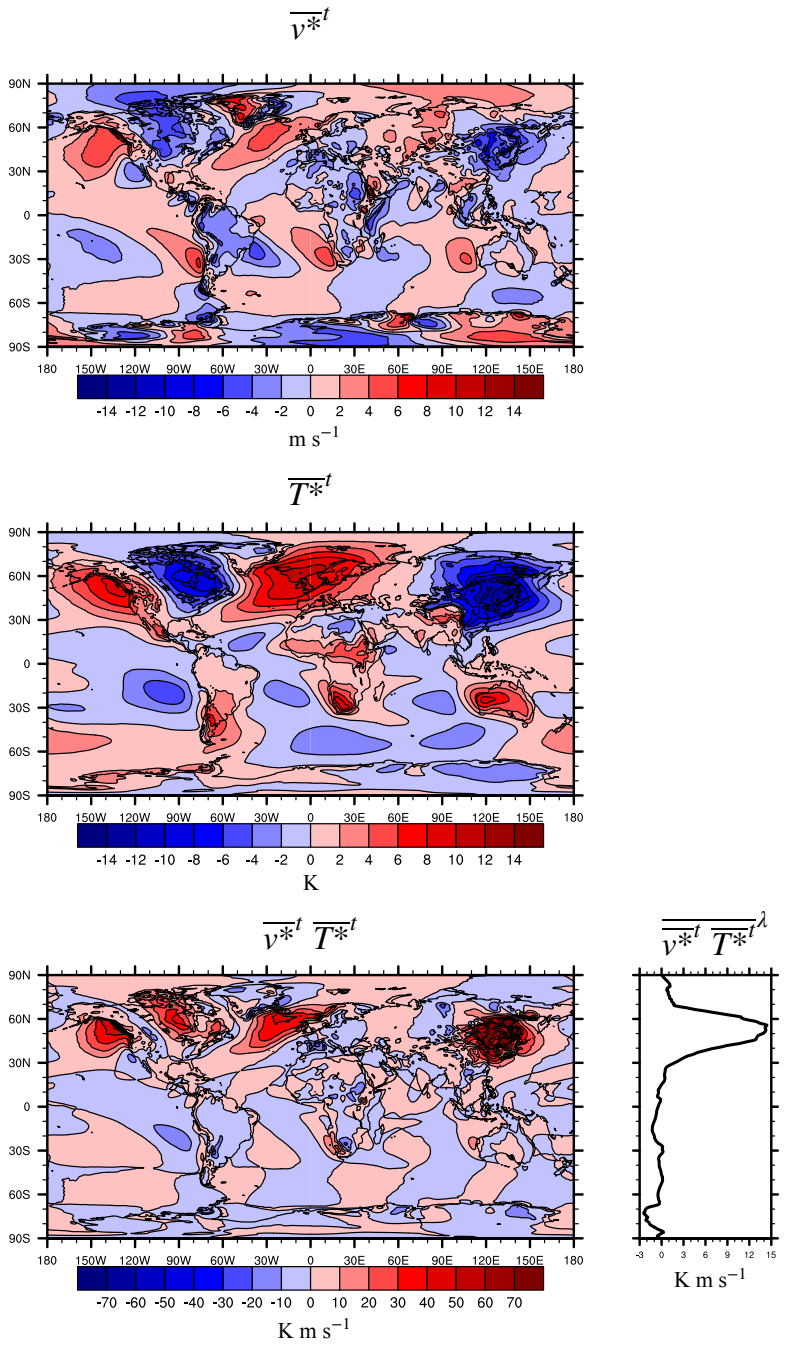


Figure D.1: An example of the decompositions of fields into their zonal mean and eddy parts, and the zonal mean of an eddy product. The top and center panels show maps of the stationary eddy parts of the January-mean meridional wind and temperature at 850 hPa, respectively. The bottom panel shows the product. The zonal mean of the product is shown on the bottom right. See text for details.

Appendix E

Why is the Dissipation Rate Positive?

As shown in shown in Chapter 4, the rate of kinetic energy dissipation, per unit mass, is

$$\delta \equiv -\alpha\,(\mathbf{S}\cdot\nabla)\cdot\mathbf{V}\,, \tag{E.1}$$

where $\mathbf{S}$ is the stress tensor. We will prove that

$$\delta \geq 0\,. \tag{E.2}$$

For simplicity, consider a Cartesian coordinate system applied to some small volume of the atmosphere. The coordinates are (x,y,z), and the corresponding velocity components are (u,v,w). The stress tensor can be written as a matrix:

$$\mathbf{S} = \begin{bmatrix} 0 & F_{v,x} & F_{w,x} \\ F_{u,y} & 0 & F_{w,y} \\ F_{u,z} & F_{v,z} & 0 \end{bmatrix}\,. \tag{E.3}$$

As an example to explain the notation, $F_{u,y}$ is the flux of u-momentum in the y-direction. The diagonal elements of the matrix are set to zero because they would represent "normal stresses" (e.g., the pressure), and dissipation comes only from shearing stresses, which "deform" the fluid elements in the same way that dough is deformed when it is kneaded or paint is deformed when it is stirred. It can be shown that the stress tensor has to be symmetric about its diagonal, i.e., $F_{v,x} = F_{u,y}$, $F_{w,x} = F_{u,z}$, and $F_{w,y} = F_{v,z}$. If this were not

true, the stresses would exert a finite torque on an infinitesimal air particle. Invoking this symmetry, we can rewrite (E.3) as

$$\mathbf{S} = \begin{bmatrix} 0 & F_{u,y} & F_{u,z} \\ F_{u,y} & 0 & F_{v,z} \\ F_{u,z} & F_{v,z} & 0 \end{bmatrix} . \tag{E.4}$$

The divergence operator can be written as a "row" vector, i.e.,

$$\nabla\cdot = \begin{bmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \end{bmatrix} , \tag{E.5}$$

Combining (E.4) and (E.5) we find that

$$\nabla\cdot\mathbf{S} = \left(\frac{\partial F_{v,x}}{\partial y} + \frac{\partial F_{w,x}}{\partial z}\right)\mathbf{i} + \left(\frac{\partial F_{u,y}}{\partial x} + \frac{\partial F_{w,y}}{\partial z}\right)\mathbf{j} + \left(\frac{\partial F_{u,z}}{\partial x} + \frac{\partial F_{v,z}}{\partial y}\right)\mathbf{k} , \tag{E.6}$$

where $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are the unit vectors in the x, y, and z directions, respectively. The application of (E.4) to (E.5) to obtain (E.6) illustrates how and why the divergence of a tensor is a vector. Similarly,

$$\mathbf{S}\cdot\mathbf{V} = (F_{v,x}v + F_{w,x}w)\,\mathbf{i} + (F_{u,y}u + F_{w,y}w)\,\mathbf{j} + (F_{u,z}u + F_{v,z}v)\,\mathbf{k} , \tag{E.7}$$

is the (vector) flux of kinetic energy due to work done by friction on the air "next door." For example, $F_{u,z}u$ is the component of the energy exchange vector that acts in the z-direction (hence, multiplied by $\mathbf{k}$) due to the work done as friction transports u-momentum in the z-direction.

The dissipation rate, $-\alpha\,(\mathbf{S}\cdot\nabla)\cdot\mathbf{V}$, can be constructed as follows: We start by writing

$$\begin{aligned}\mathbf{S}\cdot\nabla &= \begin{bmatrix} 0 & F_{v,x} & F_{w,x} \\ F_{u,y} & 0 & F_{w,y} \\ F_{u,z} & F_{v,z} & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix} \\ &= \begin{bmatrix} \left(F_{v,x}\frac{\partial}{\partial y} + F_{w,x}\frac{\partial}{\partial z}\right)\mathbf{i} & \left(F_{u,y}\frac{\partial}{\partial x} + F_{w,y}\frac{\partial}{\partial z}\right)\mathbf{j} & \left(F_{u,z}\frac{\partial}{\partial x} + F_{v,z}\frac{\partial}{\partial y}\right)\mathbf{k} \end{bmatrix} .\end{aligned} \tag{E.8}$$

We call $\mathbf{S}\cdot\nabla$ an "operator," because it can be dotted with vectors, i.e., it "eats" vectors. Dotting it with $\mathbf{V}$, we obtain

$$\begin{aligned}(\mathbf{S}\cdot\nabla)\cdot\mathbf{V} &= \left(F_{v,x}\frac{\partial}{\partial y} + F_{w,x}\frac{\partial}{\partial z}\right)u + \left(F_{u,y}\frac{\partial}{\partial x} + F_{w,y}\frac{\partial}{\partial z}\right)v + \left(F_{u,z}\frac{\partial}{\partial x} + F_{v,z}\frac{\partial}{\partial y}\right)w \\ &= F_{u,y}\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right) + F_{u,z}\left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right) + F_{v,z}\left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}\right) ,\end{aligned} \tag{E.9}$$

The second line of (E.9) is obtained by using the symmetry of the stress tensor, which was discussed above.

To finish our proof that the dissipation rate is non-negative, we have to formulate the stresses in terms of the wind components. Air is an example of a "Newtonian fluid," for which the stress is proportional to the spatial derivatives of the motion (called the "rate of strain") as follows:

$$F_{v,x} = F_{u,y} = -\mu\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right) , \tag{E.10}$$

$$F_{w,x} = F_{u,z} = -\mu\left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right) , \tag{E.11}$$

$$F_{w,y} = F_{v,z} = -\mu\left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}\right) . \tag{E.12}$$

Here μ is the positive, nearly constant molecular viscosity, which is a "material property," rather than a property of the motion. Eqs. (E.10) - (E.12) are examples of "stress-strain relationships." As an example, (E.11) says that an upward increase of u (a "strain") will tend to favor a negative (downward) value of the stress $F_{u,z}$. The flow of momentum is from "fast" to "slow," thus tending to homogenize the momentum over time. The jargon is that such a flux is "down-gradient." Substituting (E.10) - (E.12) into the second line of (E.9), we find that

$$(\mathbf{S}\cdot\nabla)\cdot\mathbf{V} = -\mu\left[\left(\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\right)^2+\left(\frac{\partial w}{\partial x}+\frac{\partial u}{\partial z}\right)^2+\left(\frac{\partial w}{\partial y}+\frac{\partial v}{\partial z}\right)^2\right], \tag{E.13}$$

which establishes (E.2). It also shows that dissipation occurs only when shear is present.

Appendix F

The Moist Adiabatic Lapse Rate

As discussed in Chapter 4, the moist static energy,

$$h \equiv c_p \mathrm{T} + gz + Lq_v \tag{F.1}$$

is approximately conserved under moist adiabatic processes. The saturation moist static energy is given by

$$h_{\mathrm{sat}} \equiv c_p \mathrm{T} + gz + Lq_{\mathrm{sat}}(\mathrm{T}, p) \ , \tag{F.2}$$

where q_{sat} is the saturation mixing ratio, which depends, as indicated, on temperature and pressure. Differentiating (F.2) with respect to height gives

$$\frac{\partial h_{\mathrm{sat}}}{\partial z} = c_p \frac{\partial T}{\partial z} + g + L\left[\left(\frac{\partial q_{\mathrm{sat}}}{\partial \mathrm{T}}\right)_p \frac{\partial T}{\partial z} + \left(\frac{\partial q_{\mathrm{sat}}}{\partial p}\right)_{\mathrm{T}} \frac{\partial \bar{p}}{\partial z}\right] \ , \tag{F.3}$$

where

$$\gamma \equiv \frac{L}{c_p}\left(\frac{\partial q_{\mathrm{sat}}}{\partial T}\right)_p \tag{F.4}$$

is evaluated using the mean-state temperature and pressure. The nondimensional parameter γ is positive and of order one. With the use of the hydrostatic equation, Eq. (F.3) can be rearranged to

$$\frac{\partial h_{\rm sat}}{\partial z} = c_p \left\{ (1+\gamma) \frac{\partial T}{\partial z} + \frac{g}{c_p} \left[1 - L \left(\frac{\partial q_{\rm sat}}{\partial p} \right)_{\rm T} \bar{\rho} \right] \right\} \ . \tag{F.5}$$

The saturation moist static energy is conserved by a saturated parcel in the absence of mixing and radiative heating. Setting $\partial h_{\rm sat}/\partial z = 0$ in (F.5) therefore leads to an expression for the moist adiabatic lapse rate:

$$\begin{aligned} \Gamma_m &\equiv - \left(\frac{\partial T}{\partial z} \right)_{\text{moist adiabatic}} \\ &= \frac{(g/c_p)}{(1+\gamma)} \left[1 - L \left(\frac{\partial q_{\rm sat}}{\partial p} \right)_T \rho \right] \ . \end{aligned} \tag{F.6}$$

To go further, we need formulas for $(\partial q_{\rm sat}/\partial p)_T$ and γ.

The saturation mixing ratio is defined by

$$q_{\rm sat}(T,p) \equiv \frac{(\rho_v)_{\rm sat}}{\rho_d} \ , \tag{F.7}$$

where $(\rho_v)_{\rm sat}$ is the saturation vapor density, and ρ_d is the density of dry air. Both water vapor and dry air obey the equation of state, with the same temperature but different gas constants:

$$p_d = \rho_d R_d T \text{ and } e = \rho_v R_v T \ . \tag{F.8}$$

Here e is the vapor pressure, and $R_v \cong 4.62$ J kg^{-1} K^{-1}. Taking the ratio of the two equations in (F.8), the temperatures cancel and we find that

$$\frac{e}{p_d} = \frac{\rho_v R_v}{\rho_d R_d} \ , \tag{F.9}$$

which is equivalent to

$$\begin{aligned} q \equiv & \rho_v / \rho_d \\ = & \varepsilon e / p_d \ , \end{aligned} \tag{F.10}$$

where

$$\varepsilon \equiv R_d / R_v \cong 0.622 \ . \tag{F.11}$$

Throughout the troposphere, $p_d \gg e$, so we can approximate (F.10) by

$$q \cong \frac{\varepsilon e}{p} \ , \tag{F.12}$$

and similarly

$$q_{\text{sat}}(T,p) \cong \frac{\varepsilon e_{\text{sat}}(T)}{p} \ . \tag{F.13}$$

From (F.13), we see that

$$\left(\frac{\partial q_{\text{sat}}}{\partial p} \right)_T \cong - \frac{q_{\text{sat}}(T,p)}{p} < 0 \ . \tag{F.14}$$

and

$$\begin{aligned} \gamma \equiv & \frac{L}{c_p} \left(\frac{\partial q_{\text{sat}}}{\partial T} \right)_p \\ = & \frac{L}{c_p} \frac{\varepsilon}{p} \frac{d}{dT} e_{\text{sat}}(T) \ . \end{aligned} \tag{F.15}$$

To finish, we need to use the Clausius-Clapeyron equation. A commonly used approximate form of that equation is

$$\frac{d}{dT}e_{\text{sat}}(T) \cong \frac{Le_{\text{sat}}(T)}{R_v T^2} \ . \tag{F.16}$$

Substitution of (F.16) into (F.15) gives

$$\begin{aligned} \gamma \equiv & \frac{L}{c_p}\left(\frac{\partial q_{\text{sat}}}{\partial T}\right)_p \\ = & \frac{L^2 q_{\text{sat}}}{c_p R_v T^2} \ . \end{aligned} \tag{F.17}$$

Then (F.6) can be written as

$$\boxed{\Gamma_m = \Gamma_d \left[\frac{1+\frac{Lq_{\text{sat}}}{R_d T}}{1+\frac{L^2 q_{\text{sat}}}{c_p R_v T^2}}\right] ,} \tag{F.18}$$

where

$$\begin{aligned} \Gamma_d = & -\left(\frac{\partial T}{\partial z}\right)_{\text{dry adiabatic}} \\ = & \frac{g}{c_p} \end{aligned} \tag{F.19}$$

is the dry adiabatic lapse rate. The denominator of (F.18) is larger than the numerator, so $\Gamma_m < \Gamma_d$, although at cold temperatures $\Gamma_m \to \Gamma_d$. For example, with a pressure of 1000 hPa and a temperature of 288 K, we find that $\Gamma_m = 4.67$ K km^{-1}. As the temperature increases for a given pressure, the moist adiabatic lapse rate decreases.

Appendix G

Eddy Kinetic Energy and Zonal Kinetic Energy

We define the eddy kinetic energy per unit mass by

$$KE \equiv \frac{1}{2}\left[\overline{(u*)^2}^{\lambda} + \overline{(v*)^2}^{\lambda}\right] \tag{G.1}$$

and the zonal kinetic energy by

$$KZ \equiv \frac{1}{2}\left[\left(\overline{u}^{\lambda}\right)^2 + \left(\overline{v}^{\lambda}\right)^2\right] \tag{G.2}$$

It follows that

$$\overline{K}^{\lambda} = KZ + KE \tag{G.3}$$

All three quantities in Eq. (G.3) are independent of longitude.

To derive equations that govern KE and KZ, we start from the zonal and meridional equations of motion in flux form:

$$\begin{aligned}\frac{\partial u}{\partial t} + \frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda}(uu) + \frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}(vu\cos\varphi) + \frac{\partial}{\partial p}(\omega u)\\ = \frac{uv\tan\varphi}{a} + fv - \frac{1}{a\cos\varphi}\frac{\partial\phi}{\partial\lambda} + g\frac{\partial F_u}{\partial p}\end{aligned} \tag{G.4}$$

$$\begin{aligned}&\frac{\partial v}{\partial t}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\lambda}(uv)+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}(vv\cos\varphi)+\frac{\partial}{\partial p}(\omega v)\\&\quad=-\frac{u^2\tan\varphi}{a}-fu-\frac{1}{a}\frac{\partial\phi}{\partial\varphi}+g\frac{\partial F_v}{\partial p}\end{aligned}\tag{G.5}$$

We also need mass continuity:

$$\frac{1}{a\cos\varphi}\frac{\partial u}{\partial\lambda}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}(v\cos\varphi)+\frac{\partial\omega}{\partial p}=0\tag{G.6}$$

Zonal averaging (G.4) - (G.6) gives

$$\begin{aligned}&\frac{\partial\overline{u}^\lambda}{\partial t}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left\{\left(\overline{v}^\lambda\overline{u}^\lambda+\overline{v^*u^*}^\lambda\right)\cos\varphi\right\}+\frac{\partial}{\partial p}\left(\overline{\omega}^\lambda\overline{u}^\lambda+\overline{\omega^*u^*}^\lambda\right)\\&\quad=\left(\overline{v}^\lambda\overline{u}^\lambda+\overline{v^*u^*}^\lambda\right)\frac{\tan\varphi}{a}+f\overline{v}^\lambda+g\frac{\partial\overline{F_u}^\lambda}{\partial p},\end{aligned}\tag{G.7}$$

$$\begin{aligned}&\frac{\partial\overline{v}^\lambda}{\partial t}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left\{\left(\overline{v}^\lambda\overline{v}^\lambda+\overline{v^*v^*}^\lambda\right)\cos\varphi\right\}+\frac{\partial}{\partial p}\left(\overline{\omega}^\lambda\overline{v}^\lambda+\overline{\omega^*v^*}^\lambda\right)\\&\quad=-\left(\overline{u}^\lambda\overline{u}^\lambda+\overline{u^*u^*}^\lambda\right)\frac{\tan\varphi}{a}-f\overline{u}^\lambda-\frac{1}{a}\frac{\partial\overline{\phi}^\lambda}{\partial\varphi}+g\frac{\partial\overline{F_v}^\lambda}{\partial p}.\end{aligned}\tag{G.8}$$

and

$$\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v}^\lambda\cos\varphi\right)+\frac{\partial\overline{\omega}^\lambda}{\partial p}=0.\tag{G.9}$$

We use the zonally averaged continuity equation, (G.9), to convert (G.7) and (G.8) to advective form:

$$\begin{aligned}
&\frac{\partial \overline{u}^{\lambda}}{\partial t}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial \overline{u}^{\lambda}}{\partial \varphi}+\overline{\omega}^{\lambda}\frac{\partial \overline{u}^{\lambda}}{\partial p}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{v^{*}u^{*}}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial p}\overline{\omega^{*}u^{*}}^{\lambda}\\
&\qquad=\left(\overline{v}^{\lambda}\overline{u}^{\lambda}+\overline{v^{*}u^{*}}^{\lambda}\right)\frac{\tan\varphi}{a}+f\overline{v}^{\lambda}+g\frac{\partial \overline{F_{u}}^{\lambda}}{\partial p},
\end{aligned} \tag{G.10}$$

$$\begin{aligned}
&\frac{\partial \overline{v}^{\lambda}}{\partial t}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial \overline{v}^{\lambda}}{\partial \varphi}+\overline{\omega}^{\lambda}\frac{\partial \overline{v}^{\lambda}}{\partial p}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{v^{*}v^{*}}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial p}\overline{\omega^{*}v^{*}}^{\lambda}\\
&\qquad=-\left(\overline{u}^{\lambda}\overline{u}^{\lambda}+\overline{u^{*}u^{*}}^{\lambda}\right)\frac{\tan\varphi}{a}-f\overline{u}^{\lambda}-\frac{1}{a}\frac{\partial \overline{\phi}^{\lambda}}{\partial \varphi}+g\frac{\partial \overline{F_{v}}^{\lambda}}{\partial p}.
\end{aligned} \tag{G.11}$$

Next, multiply (G.10) by $\overline{u}$, and (G.11) by $\overline{v}$, and add the results, to obtain

$$\begin{aligned}
&\frac{\partial}{\partial t}KZ+\frac{\overline{v}^{\lambda}}{a}\frac{\partial}{\partial \varphi}KZ+\overline{\omega}^{\lambda}\frac{\partial}{\partial p}KZ\\
&+\frac{\overline{u}^{\lambda}}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{v^{*}u^{*}}^{\lambda}\cos\varphi\right)+\frac{\overline{v}^{\lambda}}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{v^{*}v^{*}}^{\lambda}\cos\varphi\right)\\
&+\overline{u}^{\lambda}\frac{\partial}{\partial p}\overline{\omega^{*}u^{*}}^{\lambda}+\overline{v}^{\lambda}\frac{\partial}{\partial p}\overline{\omega^{*}v^{*}}^{\lambda}\\
&=\overline{u}^{\lambda}\left(\overline{v}^{\lambda}\overline{u}^{\lambda}+\overline{v^{*}u^{*}}^{\lambda}\right)\frac{\tan\varphi}{a}-\overline{v}^{\lambda}\left(\overline{v}^{\lambda}\overline{u}^{\lambda}+\overline{u^{*}u^{*}}^{\lambda}\right)\frac{\tan\varphi}{a}\\
&-\frac{\overline{v}^{\lambda}}{a}\frac{\partial \overline{\phi}^{\lambda}}{\partial \varphi}+\overline{u}^{\lambda}g\frac{\partial \overline{F_{u}}^{\lambda}}{\partial p}+\overline{v}^{\lambda}g\frac{\partial \overline{F_{v}}^{\lambda}}{\partial p}.
\end{aligned} \tag{G.12}$$

Use (G.9) to go back to flux form, and simplify, to obtain

$$\begin{aligned}
&\frac{\partial}{\partial t}KZ+\frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{v}^{\lambda}KZ\cos\varphi\right)+\frac{\partial}{\partial p}\left(\overline{\omega}^{\lambda}KZ\right)\\
&+\frac{\overline{u}^{\lambda}}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{v^{*}u^{*}}^{\lambda}\cos\varphi\right)+\frac{\overline{v}^{\lambda}}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{v^{*}v^{*}}^{\lambda}\cos\varphi\right)\\
&+\overline{u}^{\lambda}\frac{\partial}{\partial p}\overline{\omega^{*}u^{*}}^{\lambda}+\overline{v}^{\lambda}\frac{\partial}{\partial p}\overline{\omega^{*}v^{*}}^{\lambda}\\
&=\left(\overline{u}^{\lambda}\overline{v^{*}u^{*}}^{\lambda}-\overline{v}^{\lambda}\overline{u^{*}u^{*}}^{\lambda}\right)\frac{\tan\varphi}{a}-\frac{\overline{v}^{\lambda}}{a}\frac{\partial \overline{\phi}^{\lambda}}{\partial \varphi}+\overline{u}^{\lambda}g\frac{\partial \overline{F_{u}}^{\lambda}}{\partial p}+\overline{v}^{\lambda}g\frac{\partial \overline{F_{v}}^{\lambda}}{\partial p}.
\end{aligned} \tag{G.13}$$

The terms on the second line of (G.13) can be manipulated as follows:

$$\begin{aligned}
&\frac{\overline{u}^\lambda}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^*u^*}^\lambda\cos\varphi\right)+\frac{\overline{v}^\lambda}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^*v^*}^\lambda\cos\varphi\right)+\overline{u}^\lambda\frac{\partial}{\partial p}\overline{\omega^*u^*}^\lambda+\overline{v}^\lambda\frac{\partial}{\partial p}\overline{\omega^*v^*}^\lambda\\
&=\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\left(\overline{u}^\lambda\overline{v^*u^*}^\lambda+\overline{v}^\lambda\overline{v^*v^*}^\lambda\right)\cos\varphi\right]-\frac{\overline{v^*u^*}^\lambda}{a}\frac{\partial\overline{u}^\lambda}{\partial\varphi}-\frac{\overline{v^*v^*}^\lambda}{a}\frac{\partial\overline{v}^\lambda}{\partial\varphi}\\
&+\frac{\partial}{\partial p}\left(\overline{u}^\lambda\overline{\omega^*u^*}^\lambda+\overline{v}^\lambda\overline{\omega^*v^*}^\lambda\right)-\overline{\omega^*u^*}^\lambda\frac{\partial\overline{u}^\lambda}{\partial p}-\overline{\omega^*v^*}^\lambda\frac{\partial\overline{v}^\lambda}{\partial p}\,.
\end{aligned}\tag{G.14}$$

Finally, we write

$$\begin{aligned}
\frac{\overline{v}^\lambda}{a}\frac{\partial\overline{\phi}^\lambda}{\partial\varphi}&=\frac{\overline{v}^\lambda\cos\varphi}{a\cos\varphi}\frac{\partial\overline{\phi}^\lambda}{\partial\varphi}\\
&=\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v}^\lambda\overline{\phi}^\lambda\cos\varphi\right)-\frac{\overline{\phi}^\lambda}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v}^\lambda\cos\varphi\right)\\
&=\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v}^\lambda\overline{\phi}^\lambda\cos\varphi\right)+\overline{\phi}^\lambda\frac{\partial\overline{\omega}^\lambda}{\partial p}\\
&=\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v}^\lambda\overline{\phi}^\lambda\cos\varphi\right)+\frac{\partial}{\partial p}\left(\overline{\omega}^\lambda\overline{\phi}^\lambda\right)-\overline{\omega}^\lambda\frac{\partial\overline{\phi}^\lambda}{\partial p}\\
&=\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v}^\lambda\overline{\phi}^\lambda\cos\varphi\right)+\frac{\partial}{\partial p}\left(\overline{\omega}^\lambda\overline{\phi}^\lambda\right)+\overline{\omega}^\lambda\overline{\alpha}^\lambda\,.
\end{aligned}\tag{G.15}$$

Substituting back into (G.13), and combining terms, we find that

$$
\begin{aligned}
&\frac{\partial}{\partial t}KZ \\
&+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\left(\overline{v}^{\lambda}KZ+\overline{u}^{\lambda}\overline{v^{*}u^{*}}^{\lambda}+\overline{v}^{\lambda}\overline{v^{*}v^{*}}^{\lambda}+\overline{v}^{\lambda}\overline{\phi}^{\lambda}\right)\cos\varphi\right] \\
&+\frac{\partial}{\partial p}\left(\overline{\omega}^{\lambda}KZ+\overline{u}^{\lambda}\overline{\omega^{*}u^{*}}^{\lambda}+\overline{v}^{\lambda}\overline{\omega^{*}v^{*}}^{\lambda}+\overline{\omega}^{\lambda}\overline{\phi}^{\lambda}\right) \\
&=\frac{\overline{v^{*}u^{*}}^{\lambda}}{a}\frac{\partial\overline{u}^{\lambda}}{\partial\varphi}+\frac{\overline{v^{*}v^{*}}^{\lambda}}{a}\frac{\partial\overline{v}^{\lambda}}{\partial\varphi}+\overline{\omega^{*}u^{*}}^{\lambda}\frac{\partial\overline{u}^{\lambda}}{\partial p}+\overline{\omega^{*}v^{*}}^{\lambda}\frac{\partial\overline{v}^{\lambda}}{\partial p} \\
&+\left(\overline{u}^{\lambda}\overline{v^{*}u^{*}}^{\lambda}-\overline{v}^{\lambda}\overline{u^{*}u^{*}}^{\lambda}\right)\frac{\tan\varphi}{a} \\
&-\overline{\omega}^{\lambda}\overline{\alpha}^{\lambda} \\
&+\overline{u}^{\lambda}g\frac{\partial\overline{F_{u}}^{\lambda}}{\partial p}+\overline{v}^{\lambda}\frac{\partial\overline{F_{v}}^{\lambda}}{\partial p}.
\end{aligned}
\tag{G.16}
$$

This is the zonal kinetic energy equation.

To derive the eddy kinetic energy equation, we return to the equations of motion. Use the continuity equation, (G.6), to convert (G.4) and (G.5) to advective form:

$$
\frac{\partial u}{\partial t}+\frac{u}{a\cos\varphi}\frac{\partial u}{\partial\lambda}+\frac{v}{a}\frac{\partial u}{\partial\varphi}+\omega\frac{\partial u}{\partial p}=\frac{uv\tan\varphi}{a}+fv-\frac{1}{a\cos\varphi}\frac{\partial\phi}{\partial\lambda}+g\frac{\partial F_{u}}{\partial p} \tag{G.17}
$$

$$
\frac{\partial v}{\partial t}+\frac{u}{a\cos\varphi}\frac{\partial v}{\partial\lambda}+\frac{v}{a}\frac{\partial v}{\partial\varphi}+\omega\frac{\partial v}{\partial p}=-\frac{u^{2}\tan\varphi}{a}-fu-\frac{1}{a}\frac{\partial\phi}{\partial\varphi}+g\frac{\partial F_{v}}{\partial p} \tag{G.18}
$$

Subtract (G.10) and (G.11) from (G.17) and (G.18), respectively, to obtain:

$$
\begin{aligned}
&\frac{\partial u^{*}}{\partial t}+\frac{u}{a\cos\varphi}\frac{\partial u^{*}}{\partial\lambda}+\left(\frac{v}{a}\frac{\partial u}{\partial\varphi}-\frac{\overline{v}^{\lambda}}{a}\frac{\partial\overline{u}^{\lambda}}{\partial\varphi}\right)+\left(\omega\frac{\partial u}{\partial p}-\overline{\omega}^{\lambda}\frac{\partial\overline{u}^{\lambda}}{\partial p}\right) \\
&-\left[\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^{*}u^{*}}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial p}\overline{\omega^{*}u^{*}}^{\lambda}\right] \\
&=\frac{uv\tan\varphi}{a}-\frac{\left(\overline{v}^{\lambda}\overline{u}^{\lambda}+\overline{v^{*}u^{*}}^{\lambda}\right)\tan\varphi}{a}+fv^{*}-\frac{1}{a\cos\varphi}\frac{\partial\phi^{*}}{\partial\lambda}+g\frac{\partial F_{u}{}^{*}}{\partial p},
\end{aligned}
\tag{G.19}
$$

$$\begin{aligned}
&\frac{\partial v^*}{\partial t} + \frac{u}{a\cos\varphi}\frac{\partial v^*}{\partial \lambda} + \left(\frac{v}{a}\frac{\partial v}{\partial \varphi} - \frac{\overline{v}^\lambda}{a}\frac{\partial \overline{v}^\lambda}{\partial \varphi}\right) + \left(\omega\frac{\partial v}{\partial p} - \overline{\omega}^\lambda\frac{\partial \overline{v}^\lambda}{\partial p}\right) \\
&- \left[\frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{v^*v^*}^\lambda\cos\varphi\right) + \frac{\partial}{\partial p}\overline{\omega^*v^*}^\lambda\right] \\
&= -\frac{u^2\tan\varphi}{a} + \frac{\left(\overline{u}^\lambda\overline{u}^\lambda + \overline{u^*u^*}^\lambda\right)\tan\varphi}{a} - fu^* - \frac{1}{a}\frac{\partial \phi^*}{\partial \varphi} + g\frac{\partial {F_v}^*}{\partial p}\,.
\end{aligned} \tag{G.20}$$

Expanding the nonlinear terms, we get:

$$\begin{aligned}
&\frac{\partial u^*}{\partial t} + \left(\frac{\overline{u}^\lambda + u^*}{a\cos\varphi}\right)\frac{\partial u^*}{\partial \lambda} + \frac{\left(\overline{v}^\lambda + v^*\right)}{a}\frac{\partial}{\partial \varphi}\left(\overline{u}^\lambda + u^*\right) + \left(\overline{\omega}^\lambda + \omega^*\right)\frac{\partial}{\partial p}\left(\overline{u}^\lambda + u^*\right) \\
&- \left(\frac{\overline{v}^\lambda}{a}\frac{\partial \overline{u}^\lambda}{\partial \varphi} + \overline{\omega}^\lambda\frac{\partial \overline{u}^\lambda}{\partial p}\right) - \left[\frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{v^*u^*}^\lambda\cos\varphi\right) + \frac{\partial}{\partial p}\overline{\omega^*u^*}^\lambda\right] \\
&= \left[\left(\overline{u}^\lambda + u^*\right)\left(\overline{v}^\lambda + v^*\right) - \left(\overline{v}^\lambda\overline{u}^\lambda + \overline{v^*u^*}^\lambda\right)\right]\frac{\tan\varphi}{a} + fv^* - \frac{1}{a\cos\varphi}\frac{\partial \phi^*}{\partial \lambda} + g\frac{\partial {F_u}^*}{\partial p}\,,
\end{aligned} \tag{G.21}$$

$$\begin{aligned}
&\frac{\partial v^*}{\partial t} + \frac{\left(\overline{u}^\lambda + u^*\right)}{a\cos\varphi}\frac{\partial v^*}{\partial \lambda} + \frac{\left(\overline{v}^\lambda + v^*\right)}{a}\frac{\partial}{\partial \varphi}\left(\overline{v}^\lambda + v^*\right) + \left(\overline{\omega}^\lambda + \omega^*\right)\frac{\partial}{\partial p}\left(\overline{v}^\lambda + v^*\right) \\
&- \left(\frac{\overline{v}^\lambda}{a}\frac{\partial \overline{v}^\lambda}{\partial \varphi} + \overline{\omega}^\lambda\frac{\partial \overline{v}^\lambda}{\partial p}\right) - \left[\frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(\overline{v^*v^*}^\lambda\cos\varphi\right) + \frac{\partial}{\partial p}\overline{\omega^*v^*}^\lambda\right] \\
&= \left[-\left(\overline{u}^\lambda + u^*\right)^2 + \left(\overline{u}^\lambda\overline{u}^\lambda + \overline{u^*u^*}^\lambda\right)\right]\frac{\tan\varphi}{a} - fu^* - \frac{1}{a}\frac{\partial \phi^*}{\partial \varphi} + g\frac{\partial {F_v}^*}{\partial p}\,.
\end{aligned} \tag{G.22}$$

Rearrange and simplify:

$$
\begin{aligned}
&\left(\frac{\partial}{\partial t}+\frac{\overline{u}^{\lambda}}{a\cos\varphi}\frac{\partial}{\partial\lambda}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial}{\partial\varphi}+\overline{\omega}^{\lambda}\frac{\partial}{\partial p}\right)u^{*}\\
&+\left(\frac{u^{*}}{a\cos\varphi}\frac{\partial}{\partial\lambda}+\frac{v^{*}}{a}\frac{\partial}{\partial\varphi}+\omega^{*}\frac{\partial}{\partial p}\right)u^{*}+\frac{v^{*}}{a}\frac{\partial\overline{u}^{\lambda}}{\partial\varphi}+\omega^{*}\frac{\partial\overline{u}^{\lambda}}{\partial p}\\
&-\left[\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^{*}u^{*}}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial p}\overline{\omega^{*}u^{*}}^{\lambda}\right]\\
&=\left[\left(\overline{u}^{\lambda}v^{*}+u^{*}\overline{v}^{\lambda}+v^{*}u^{*}\right)-\overline{v^{*}u^{*}}^{\lambda}\right]\frac{\tan\varphi}{a}+fv^{*}-\frac{1}{a\cos\varphi}\frac{\partial\phi^{*}}{\partial\lambda}+g\frac{\partial {F_u}^{*}}{\partial p}\,,
\end{aligned}
\tag{G.23}
$$

$$
\begin{aligned}
&\left(\frac{\partial}{\partial t}+\frac{\overline{u}^{\lambda}}{a\cos\varphi}\frac{\partial}{\partial\lambda}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial}{\partial\varphi}+\overline{\omega}^{\lambda}\frac{\partial}{\partial p}\right)v^{*}\\
&+\left(\frac{u^{*}}{a\cos\varphi}\frac{\partial}{\partial\lambda}+\frac{v^{*}}{a}\frac{\partial}{\partial\varphi}+\omega^{*}\frac{\partial}{\partial p}\right)v^{*}+\frac{v^{*}}{a}\frac{\partial\overline{v}^{\lambda}}{\partial\varphi}+\omega^{*}\frac{\partial\overline{v}^{\lambda}}{\partial p}\\
&-\left[\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^{*}v^{*}}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial p}\overline{\omega^{*}v^{*}}^{\lambda}\right]\\
&=\left[-\left(2\overline{u}^{\lambda}u^{*}+u^{*}u^{*}\right)+\overline{u^{*}u^{*}}^{\lambda}\right]\frac{\tan\varphi}{a}-fu^{*}-\frac{1}{a}\frac{\partial\phi^{*}}{\partial\varphi}+g\frac{\partial {F_v}^{*}}{\partial p}\,.
\end{aligned}
\tag{G.24}
$$

Now multiply (G.23) by u^{*}, and (G.24) by v^{*}:

$$
\begin{aligned}
&\left(\frac{\partial}{\partial t}+\frac{\overline{u}^{\lambda}}{a\cos\varphi}\frac{\partial}{\partial\lambda}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial}{\partial\varphi}+\overline{\omega}^{\lambda}\frac{\partial}{\partial p}\right)\frac{u^{*2}}{2}\\
&+\left(\frac{u^{*}}{a\cos\varphi}\frac{\partial}{\partial\lambda}+\frac{v^{*}}{a}\frac{\partial}{\partial\varphi}+\omega^{*}\frac{\partial}{\partial p}\right)\frac{u^{*2}}{2}+\frac{u^{*}v^{*}}{a}\frac{\partial\overline{u}^{\lambda}}{\partial\varphi}+u^{*}\omega^{*}\frac{\partial\overline{u}^{\lambda}}{\partial p}\\
&-u^{*}\left[\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^{*}u^{*}}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial p}\overline{\omega^{*}u^{*}}^{\lambda}\right]\\
&=u^{*}\left[\left(\overline{u}^{\lambda}v^{*}+u^{*}\overline{v}^{\lambda}+v^{*}u^{*}\right)-\overline{v^{*}u^{*}}^{\lambda}\right]\frac{\tan\varphi}{a}+fu^{*}v^{*}-\frac{u^{*}}{a\cos\varphi}\frac{\partial\phi^{*}}{\partial\lambda}+u^{*}g\frac{\partial {F_u}^{*}}{\partial p}\,,
\end{aligned}
\tag{G.25}
$$

$$
\begin{aligned}
&\left(\frac{\partial}{\partial t}+\frac{\overline{u}^{\lambda}}{a\cos\varphi}\frac{\partial}{\partial\lambda}+\frac{\overline{v}^{\lambda}}{a}\frac{\partial}{\partial\varphi}+\overline{\omega}^{\lambda}\frac{\partial}{\partial p}\right)\frac{v^{*2}}{2}\\
&+\left(\frac{u^{*}}{a\cos\varphi}\frac{\partial}{\partial\lambda}+\frac{v^{*}}{a}\frac{\partial}{\partial\varphi}+\omega^{*}\frac{\partial}{\partial p}\right)\frac{v^{*2}}{2}+\frac{v^{*}v^{*}}{a}\frac{\partial\overline{v}^{\lambda}}{\partial\varphi}+v^{*}\omega^{*}\frac{\partial\overline{v}^{\lambda}}{\partial p}\\
&-v^{*}\left[\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^{*}v^{*}}^{\lambda}\cos\varphi\right)+\frac{\partial}{\partial p}\overline{\omega^{*}v^{*}}^{\lambda}\right]\\
&=v^{*}\left[-\left(2\overline{u}^{\lambda}u^{*}+u^{*}u^{*}\right)+\overline{u^{*}u^{*}}^{\lambda}\right]\frac{\tan\varphi}{a}-fv^{*}u^{*}-\frac{v^{*}}{a}\frac{\partial\phi^{*}}{\partial\varphi}+v^{*}g\frac{\partial{F_{v}}^{*}}{\partial p}\,.
\end{aligned}
\tag{G.26}
$$

Next, use the zonal mean continuity equation, (G.9), to convert to flux form the terms of (G.25) and (G.26) that represent advection by the zonally averaged winds, and use the eddy continuity equation,

$$
\frac{1}{a\cos\varphi}\frac{\partial u^{*}}{\partial\lambda}+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(v^{*}\cos\varphi\right)+\frac{\partial\omega^{*}}{\partial p}=0 \tag{G.27}
$$

to similarly rewrite the terms that represent advection by the eddy winds. Taking the zonal means of the results, we get

$$
\begin{aligned}
&\frac{1}{2}\frac{\partial}{\partial t}\overline{u^{*}u^{*}}^{\lambda}+\frac{1}{2}\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\left(\overline{v}^{\lambda}\overline{u^{*}u^{*}}^{\lambda}+\overline{v^{*}u^{*2}}^{\lambda}\right)\cos\varphi\right]+\frac{1}{2}\frac{\partial}{\partial p}\left(\overline{\omega}^{\lambda}\overline{u^{*}u^{*}}^{\lambda}+\overline{\omega^{*}u^{*2}}^{\lambda}\right)\\
&=-\frac{\overline{v^{*}u^{*}}^{\lambda}}{a}\frac{\partial\overline{u}^{\lambda}}{\partial\varphi}-\overline{\omega^{*}u^{*}}^{\lambda}\frac{\partial\overline{u}^{\lambda}}{\partial p}\\
&+\left(\overline{u}^{\lambda}\overline{v^{*}u^{*}}^{\lambda}+\overline{u^{*}u^{*}}^{\lambda}\overline{v}^{\lambda}+\overline{u^{*}v^{*}u^{*}}^{\lambda}\right)\frac{\tan\varphi}{a}+f\overline{v^{*}u^{*}}^{\lambda}-\overline{\frac{u^{*}}{a\cos\varphi}\frac{\partial\phi^{*}}{\partial\lambda}}^{\lambda}+\overline{u^{*}g\frac{\partial{F_{u}}^{*}}{\partial p}}^{\lambda}\,,
\end{aligned}
\tag{G.28}
$$

$$\frac{1}{2}\frac{\partial\overline{v^*v^*}^\lambda}{\partial t}+\frac{1}{2}\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\left(\overline{v}^\lambda\overline{v^*v^*}^\lambda+\overline{v^*v^*v^*}^\lambda\right)\cos\varphi\right]+\frac{1}{2}\frac{\partial}{\partial p}\left(\overline{\omega}^\lambda\overline{v^*v^*}^\lambda+\overline{\omega^*v^{*2}}^\lambda\right)$$
$$=-\frac{\overline{v^*v^*}^\lambda}{a}\frac{\partial\overline{v}^\lambda}{\partial\varphi}-\overline{\omega^*v^*}^\lambda\frac{\partial\overline{v}^\lambda}{\partial p}$$
$$-\left(2\overline{u}^\lambda\overline{v^*u^*}^\lambda+\overline{u^*u^*v^*}^\lambda\right)\frac{\tan\varphi}{a}-f\overline{v^*u^*}^\lambda-\overline{\frac{v^*}{a}\frac{\partial\phi^*}{\partial\varphi}}^\lambda+g\overline{v^*\frac{\partial F_v{}^*}{\partial p}}^\lambda\quad. \tag{G.29}$$

Now add (G.28) and (G.29) to obtain:

$$\frac{\partial}{\partial t}KE+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left\{\left[\overline{v}^\lambda KE+\frac{1}{2}\overline{v^*\left(u^{*2}+v^{*2}\right)}^\lambda\right]\cos\varphi\right\}+\frac{\partial}{\partial p}\left[\overline{\omega}^\lambda KE+\frac{1}{2}\overline{\omega^*\left(u^{*2}+v^{*2}\right)}^\lambda\right]$$
$$=-\frac{\overline{v^*u^*}^\lambda}{a}\frac{\partial\overline{u}^\lambda}{\partial\varphi}-\frac{\overline{v^*v^*}^\lambda}{a}\frac{\partial\overline{v}^\lambda}{\partial\varphi}-\overline{\omega^*u^*}^\lambda\frac{\partial\overline{u}^\lambda}{\partial p}-\overline{\omega^*v^*}^\lambda\frac{\partial\overline{v}^\lambda}{\partial p}$$
$$+\left(-\overline{u}^\lambda\overline{v^*u^*}^\lambda+\overline{u^*u^*}^\lambda\overline{v}^\lambda\right)\frac{\tan\varphi}{a}-\overline{\frac{u^*}{a\cos\varphi}\frac{\partial\phi^*}{\partial\lambda}}^\lambda-\overline{\frac{v^*}{a}\frac{\partial\phi^*}{\partial\varphi}}^\lambda+g\overline{u^*\frac{\partial F_u{}^*}{\partial p}}^\lambda+g\overline{v^*\frac{\partial F_v{}^*}{\partial p}}^\lambda\quad. \tag{G.30}$$

Finally, by analogy with (G.15), the pressure-gradient terms of (G.30) can be rewritten as

$$\overline{\frac{u^*}{a\cos\varphi}\frac{\partial\phi^*}{\partial\lambda}}^\lambda+\overline{\frac{v^*}{a}\frac{\partial\phi^*}{\partial\varphi}}^\lambda=\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left(\overline{v^*\phi^*}^\lambda\cos\varphi\right)+\frac{\partial}{\partial p}\overline{\omega^*\phi^*}^\lambda+\overline{\omega^*\alpha^*}^\lambda \tag{G.31}$$

Substitution of (G.31) into (G.30) gives

$$
\begin{aligned}
&\frac{\partial}{\partial t}KE \\
&+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left\{\left[\overline{v}^{\lambda}KE+\frac{1}{2}\overline{v^{*}\left(u^{*2}+v^{*2}\right)}^{\lambda}+\overline{v^{*}\phi^{*}}^{\lambda}\right]\cos\varphi\right\} \\
&+\frac{\partial}{\partial p}\left[\overline{\omega}^{\lambda}KE+\frac{1}{2}\overline{\omega^{*}\left(u^{*2}+v^{*2}\right)}^{\lambda}+\overline{\omega^{*}\phi^{*}}^{\lambda}\right] \\
&=-\frac{\overline{v^{*}u^{*}}^{\lambda}}{a}\frac{\partial\overline{u}^{\lambda}}{\partial\varphi}-\frac{\overline{v^{*}v^{*}}^{\lambda}}{a}\frac{\partial\overline{v}^{\lambda}}{\partial\varphi}-\overline{\omega^{*}u^{*}}^{\lambda}\frac{\partial\overline{u}^{\lambda}}{\partial p}-\overline{\omega^{*}v^{*}}^{\lambda}\frac{\partial\overline{v}^{\lambda}}{\partial p} \\
&+\left(-\overline{u}^{\lambda}\overline{v^{*}u^{*}}^{\lambda}+\overline{u^{*}u^{*}}^{\lambda}\overline{v}^{\lambda}\right)\frac{\tan\varphi}{a}-\overline{\omega^{*}\alpha^{*}}^{\lambda} \\
&+\overline{gu^{*}\frac{\partial F_{u}{}^{*}}{\partial p}}^{\lambda}+\overline{gv^{*}\frac{\partial F_{v}{}^{*}}{\partial p}}^{\lambda}\quad.
\end{aligned}
\tag{G.32}
$$

All of the transport terms, including the pressure-work terms, are combined on the second and third lines of (G.32). The terms on the fourth line represent gradient production, i.e., the conversion between the kinetic energy of the mean flow and that of the eddies. This conversion increases the eddy kinetic energy when the eddy momentum flux is "down the gradient," i.e., when it is from higher mean momentum to lower mean momentum. The $\overline{\omega^{*}\alpha^{*}}^{\lambda}$ term represents conversion between eddy kinetic energy and eddy available potential energy.

The (possibly surprising) metric terms in the equations for KZ and KE arise because we have defined "eddies" in terms of departures from the zonal mean, so that the latitude-longitude coordinate system is implicit in the very definition of KE. The metric terms cancel when we add the equations for KZ and KE to obtain the equation for the zonally averaged total kinetic energy, $\overline{K}$:

$$
\begin{aligned}
&\frac{\partial\overline{K}^{\lambda}}{\partial t} \\
&+\frac{1}{a\cos\varphi}\frac{\partial}{\partial\varphi}\left[\left(\overline{v}^{\lambda}\overline{K}^{\lambda}+\frac{1}{2}\overline{v^{*}\left(u^{*2}+v^{*2}\right)}^{\lambda}+\overline{u}^{\lambda}\overline{v^{*}u^{*}}^{\lambda}+\overline{v}^{\lambda}\overline{v^{*}v^{*}}^{\lambda}+\overline{v}^{\lambda}\overline{\phi}^{\lambda}+\overline{v^{*}\phi^{*}}^{\lambda}\right)\cos\varphi\right] \\
&+\frac{\partial}{\partial p}\left[\overline{\omega}^{\lambda}\overline{K}^{\lambda}+\frac{1}{2}\overline{\omega^{*}\left(u^{*2}+v^{*2}\right)}^{\lambda}+\overline{u}^{\lambda}\overline{\omega^{*}u^{*}}^{\lambda}+\overline{v}^{\lambda}\overline{\omega^{*}v^{*}}^{\lambda}+\overline{\phi}^{\lambda}\overline{\omega}^{\lambda}+\overline{\omega^{*}\phi^{*}}^{\lambda}\right] \\
&=-\overline{\omega}^{\lambda}\overline{\alpha}^{\lambda}-\overline{\omega^{*}\alpha^{*}}^{\lambda} \\
&+\overline{gu^{*}\frac{\partial F_{u}{}^{*}}{\partial p}}^{\lambda}+\overline{gv^{*}\frac{\partial F_{v}{}^{*}}{\partial p}}^{\lambda}+\overline{u}^{\lambda}g\frac{\partial}{\partial p}\overline{F_{u}}^{\lambda}+\overline{v}^{\lambda}\frac{\partial}{\partial p}\overline{F_{v}}^{\lambda}\quad.
\end{aligned}
\tag{G.33}
$$

Appendix H

Spherical Harmonics

The spherical surface harmonics are convenient functions for representing the distribution of geophysical quantities over the surface of the spherical Earth.

We consider a three-dimensional space and look for solutions of Laplace's differential equation, which is

$$\nabla^2 S = 0 \,. \tag{H.1}$$

The operator ∇^2 can be expanded as:

$$\nabla^2 S = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial S}{\partial r}\right) + \nabla_h^2 S = 0 \tag{H.2}$$

Here r is the distance from the origin, and ∇_h^2 is the Laplacian on a two-dimensional surface of constant r, i.e., on a spherical surface. The form of ∇_h^2 is discussed below. Inspection of (H.2) suggests that S should be proportional to a power of r, as in

$$S = r^n Y_n \,. \tag{H.3}$$

The Y_n are called spherical surface harmonics of order n. We assume that they are independent of radius. This is a "separability" assumption. The subscript n is attached to Y_n to remind us that it corresponds a particular exponent in the radial dependence of S.

In order for S to remain finite as $r \to 0$, we need $n \geq 0$. Since $n = 0$ would mean that S is independent of radius, we conclude that n must be a positive integer. Using

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial S}{\partial r}\right)=\frac{n\left(n+1\right)}{r^2}S\,, \tag{H.4}$$

which follows immediately from (H.3), we can rewrite (H.1) as

$$\boxed{\nabla_h^2 Y_n+\frac{n\left(n+1\right)}{r^2}Y_n=0\,.} \tag{H.5}$$

Before continuing with the separation of variables, it is important to point out that all of the quantities appearing in (H.5) have meaning without the need for any particular coordinate system on the two-dimensional spherical surface. We are going to use spherical (i.e., longitude and latitude) coordinates below, but we do not need them to write down (H.5).

At this point we make an analogy with a trigonometric functions. Suppose that we have a "doubly periodic" function $W\left(x,y\right)$ defined on a plane, with the usual Cartesian coordinates x and y. As an example, let

$$W\left(x,y\right)=A\sin\left(kx\right)\cos\left(ly\right)\ . \tag{H.6}$$

where A is an arbitrary constant. In Cartesian coordinates the two-dimensional Laplacian of W is

$$\begin{aligned}\nabla_h^2 W&=\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)W\\&=-\left(k^2+l^2\right)W\ ;.\end{aligned} \tag{H.7}$$

Comparing (H.5) and (H.7), we see that they are closely analogous. In particular, $n\left(n+1\right)/r^2$ in (H.5) corresponds to k^2+l^2 in (H.7). This shows that $n\left(n+1\right)$ is proportional to a "total horizontal wave number" on the sphere. For a given value of r, larger n means larger total horizontal wave number, which implies a smaller horizontal scale.

We now expand the "horizontal Laplacian" in spherical coordinates as

$$\nabla_h^2 S=\frac{1}{r^2\cos\varphi}\frac{\partial}{\partial\varphi}\left(\cos\varphi\frac{\partial S}{\partial\varphi}\right)+\frac{1}{r^2\cos^2\varphi}\frac{\partial^2 S}{\partial\lambda^2}\,. \tag{H.8}$$

where λ is longitude and φ is latitude. Then (H.5) can be rewritten as

$$\frac{1}{\cos\varphi}\frac{\partial}{\partial\varphi}\left(\cos\varphi\frac{\partial Y_n}{\partial\varphi}\right)+\frac{1}{\cos^2\varphi}\frac{\partial^2 Y_n}{\partial\lambda^2}+n\,(n+1)\,Y_n=0\,. \tag{H.9}$$

Factors of $1/r^2$ have cancelled out in (H.9), and as a result r no longer appears. Nevertheless, n, the exponent of r, is still visible, like the smile of the Cheshire cat. In the following discussion, remember where n comes from.

Next, we separate the longitude and latitude dependence in the $Y_n(\varphi,\lambda)$, i.e., we assume that

$$Y_n(\varphi,\lambda)=\Phi(\varphi)\,\Lambda(\lambda)\ . \tag{H.10}$$

where $\Lambda(\lambda)$ and $\Phi(\varphi)$ are to be determined. By substitution of (H.7) into (H.9), we find that

$$\frac{\cos^2\varphi}{\Phi}\left[\frac{1}{\cos\varphi}\frac{d}{d\varphi}\left(\cos\varphi\frac{d\Phi}{d\varphi}\right)+n\,(n+1)\,\Phi\right]=-\frac{1}{\Lambda}\frac{d^2\Lambda}{d\lambda^2}\,. \tag{H.11}$$

The left-hand side of (H.11) does not depend on λ, and the right-hand side does not depend on φ, so both sides must be equal to a constant, c. Then the longitudinal structure of the solution is governed by

$$\frac{d^2\Lambda}{d\lambda^2}+c\Lambda=0\,. \tag{H.12}$$

It follows that Λ must be a trigonometric function of longitude, i.e.,

$$\Lambda=A_\Lambda\exp(im\lambda)\ \text{ where } m=\sqrt{c},\ , \tag{H.13}$$

and A_Λ is an arbitrary complex constant. The cyclic condition $\Lambda(\lambda+2\pi)=\Lambda(\lambda)$ implies that m must be an integer. We refer to m as the zonal wave number. It is non-dimensional, and can be either positive or negative.

Notice that m has meaning only with respect to a particular spherical coordinate system. In this way m is less fundamental than n, which comes from the radial dependence of the three dimensional function S, and has a meaning that is independent of any particular spherical coordinate system.

The equation for $\Phi(\varphi)$, which determines the meridional structure of the solution, is

$$\frac{1}{\cos\varphi}\frac{d}{d\varphi}\left(\cos\varphi\frac{d\Phi}{d\varphi}\right)+\left[n(n+1)-\frac{m^2}{\cos^2\varphi}\right]\Phi=0\,, \tag{H.14}$$

Note that the zonal wave number, m, appears in this "meridional structure equation," as does the radial exponent, n. The longitude and radius dependencies have disappeared, but the zonal wave number and the radial exponent are still visible.

For convenience, we define a new independent variable to measure latitude,

$$\mu\equiv\sin\varphi\,, \tag{H.15}$$

so that $d\mu=\cos\varphi d\,\varphi$. Then (H.14) can be written as

$$\frac{d}{d\mu}\left[\left(1-\mu^2\right)\frac{d\Phi}{d\mu}\right]+\left[n(n+1)-\frac{m^2}{1-\mu^2}\right]=0\,. \tag{H.16}$$

Eq. (H.16) is simpler than (H.14), in that (H.16) does not involve trigonometric functions of the independent variable. This added simplicity is the motivation for using (H.15). The solutions of (H.16) are called the associated Legendre functions, are denoted by $P_n^m(\mu)$, and are given by

$$\begin{aligned}P_n^m(\mu)=\frac{(2n)!}{2^n n!\,(n-m)!}\left(1-\mu^2\right)^{m/2}\left[\mu^{n-m}-\frac{(n-m)(n-m-1)}{2(2n-1)}\mu^{n-m-2}\right.\\ \left.+\frac{(n-m)(n-m-1)(n-m-2)(n-m-3)}{2\cdot4(2n-1)(2n-3)}\mu^{n-m-4}-\ldots\right]\,.\end{aligned} \tag{H.17}$$

The superscript m and subscript n on $P_n^m(\mu)$ are just "markers" to remind us that m and n appear as parameters in (H.16), denoting the radial exponent and zonal wave number, respectively, of $S(r,\lambda,\varphi)$. The sum in (H.17) is continued out as far as necessary to include all non-negative powers of μ. The factor in square brackets is, therefore, a polynomial of degree $n-m$, and that is why we must require that

$$n \geq m \,, \tag{H.18}$$

Substitution can be used to demonstrate that, when (H.18) is satisfied, the associated Legendre functions are indeed solutions of (H.16).

In view of the leading factor of $\left(1-\mu^2\right)^{m/2}$ in (H.17), the complete function $P_n^m(\mu)$ is a polynomial in μ for even values of m, but not for odd values of m. The functions $P_n^m(\mu)$ are said to be of "order n" and "rank m." Figure A7.1 gives some examples of associated Legendre functions, which you might want to check for their consistency with (H.17). It can be shown that the associated Legendre functions are mutually orthogonal, i.e.,

$$\begin{aligned} &\int_{-1}^{1} P_n^m(\mu)\cdot P_l^m(\mu)\; d\mu = 0, \text{ for } n \neq l \text{ and} \\ &\int_{-1}^{1} \left[P_n^m(\mu)\right]^2 d\mu = \left(\frac{2}{2n+1}\right)\frac{(n+m)!}{(n-m)!}\,. \end{aligned} \tag{H.19}$$

It follows that the functions

$$\sqrt{\left(\frac{2n+1}{2}\right)\frac{(n-m)!}{(n+m)!}}P_n^m(\mu)\,, n = m, m+1, m+2, \ldots \tag{H.20}$$

are mutually orthonormal for $-1 \leq \mu \leq 1$.

Referring back to (H.7), we see that a particular spherical surface harmonic can be written as

$$\boxed{Y_n^m(\mu,\lambda) = P_n^m(\mu)\exp(im\lambda)\,.} \tag{H.21}$$

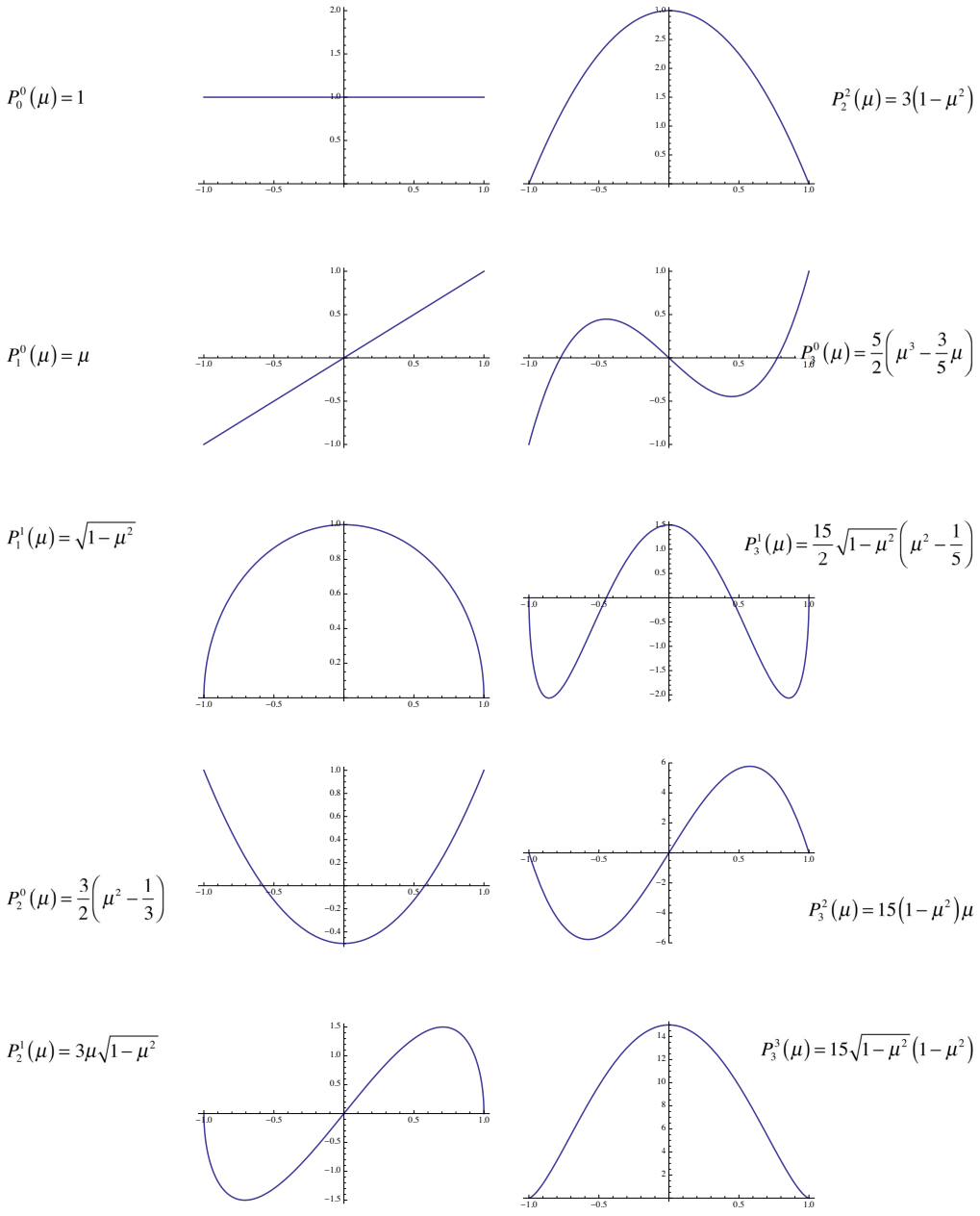


Figure H.1: Algebraic forms and plots of selected associated Legendre functions.

It is the product of an associated Legendre function of μ with a trigonometric function of λ. Note that the arbitrary constant has been set to unity. For a given value of n, there are $2n+1$ spherical harmonics, correspoinding to $m = 0, 1, \ldots n$.

Fig. H.2 shows examples of spherical harmonics of low order, as mapped out onto the longitude-latitude plane. Fig. H.3 gives similar diagrams for $n = 5$ and $m = 0$, 1, 2,...,5, plotted out onto stretched spheres. Fig. H.4 shows some low-order spherical harmonics

mapped onto three-dimensional pseudo-spheres, in which the local radius of the surface of the pseudo-sphere is one plus a constant times the local value of the spherical harmonic.

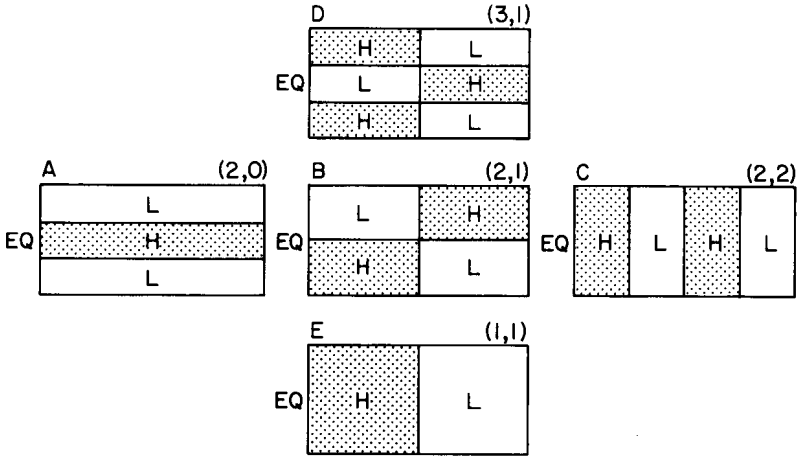


Figure H.2: Examples of low-resolution spherical harmonics, mapped out onto the plane. The horizontal direction in each panel represents longitude, and the vertical direction represents latitude. The numbers in parentheses in each panel are the appropriate values of n and m, in that order. Recall that the number of nodes in the meridional direction is $n-m$. The shading in each panel represents the sign of the field (and all signs can be flipped arbitrarily). You may think of "white" as negative and "stippled" as positive. From Washington and Parkinson (1986).

By using the orthogonality condition (H.19) for the associated Legendre functions, and also the orthogonality properties of the trigonometric functions, we can show that

$$\int_{-1}^{1}\int_{0}^{2\pi} P_n^m(\mu)\exp(im\lambda)\,P_l^{m'}(\mu)\exp\left(im'\lambda\right)d\mu d\lambda = 0 \text{ unless } n=1 \text{ and } m=m'\,, \quad \text{(H.22)}$$

The mean value over the surface of a sphere of the square of a spherical surface harmonic is given by

$$\frac{1}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi}[P_n^m(\mu)\exp(im\lambda)]^2 d\mu d\lambda = \frac{1}{2(2n+1)}\frac{(n+m)!}{(n-m)!} \text{ for } s\neq 0\,. \quad \text{(H.23)}$$

For the special case $m=0$, the corresponding value is $1/(2n+1)$.

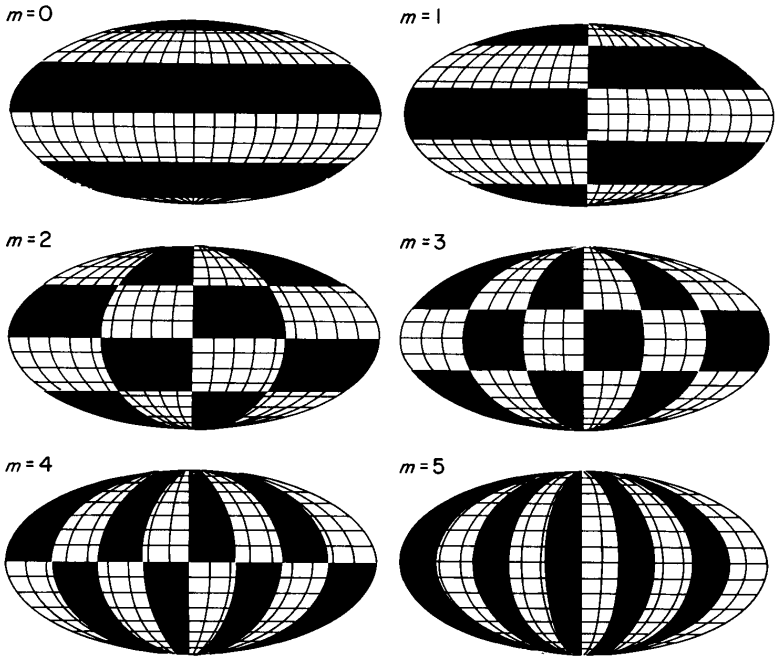


Figure H.3: Alternating patterns of positives and negatives for spherical harmonics with $n = 5$ and $m = 0, 1, 2, \ldots, 5$. From Baer (1972).

For a given n, the mean values given by (H.23) vary greatly with m, which is inconvenient for the interpretation of data. For this reason, it is customary to replace the P_n^m by the "semi-normalized associated Legendre functions," denoted by $\widehat{P}_n^m(\mu)$. These functions are identical with P_n^m when $m = 0$. For $m > 0$, the semi-normalized functions are defined by

$$\widehat{P}_n^m(\mu) \equiv \sqrt{2\frac{(n-m)!}{(n+m)!}} P_n^m(\mu) \ . \tag{H.24}$$

The mean value over the sphere of the square of $\widehat{P}_n^m(\mu)$ is then $1/(2n+1)$, for any values of n and m.

The spherical harmonics can be shown to form a complete orthonormal basis, and so can be used to represent an arbitrary function, $F(\lambda, \varphi)$ of latitude and longitude:

$$F(\lambda, \varphi) = \sum_{m=-\infty}^{\infty} \sum_{n=|m|}^{\infty} T_n^m Y_n^m(\lambda, \varphi) \ . \tag{H.25}$$

Here the T_n^m are the expansion coefficients. Note that the sum over m ranges over both positive and negative values, and that the sum over n is taken so that $n - \|m\| \geq 0$.

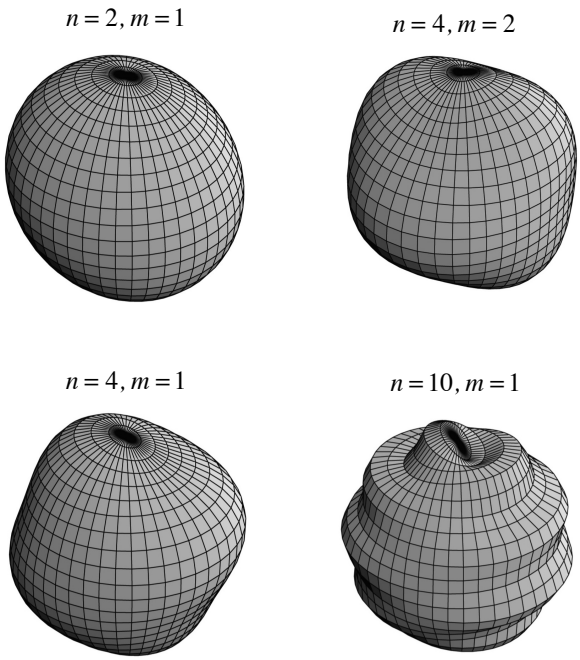


Figure H.4: Selected spherical harmonics mapped onto three-dimensional pseudo-spheres, in which the local radius of the surface of the pseudo-sphere is one plus a constant times the local value of the spherical harmonic.

The sums in (H.25) range over an infinity of terms, but in practice, of course, we have to truncate after a finite number of terms, so that (H.25) is replaced by

$$\overline{F} = \sum_{m=-M}^{M} \sum_{n=|m|}^{N(m)} T_n^m Y_n^m \ . \tag{H.26}$$

Here the overbar is a reminder that the sum is truncated. The sum over n ranges up to $N(m)$, which has to be specified somehow. The sum over m ranges from $-M$ to M. It can be shown that this ensures that the final result is real; this is an important result that you should prove for yourself.

The choice of $N(m)$ fixes what is called the "truncation procedure." There are two commonly used truncation procedures. The first, called "rhomboidal," takes

$$N(m) = M + |m| \ . \tag{H.27}$$

The second, called "triangular," takes

$$N(m) = M\ . \tag{H.28}$$

Triangular truncation has the following beautiful property: In order to actually perform a spherical harmonic transform, it is necessary to adopt a spherical coordinate system (λ, φ). There are of course infinitely many such systems. There is no reason in principle that the coordinates have to be chosen in the conventional way, so that the poles of the coordinate system coincide with the Earth's poles of rotation. The choice of a particular spherical coordinate system is, therefore, somewhat arbitrary. Suppose that we choose two different spherical coordinate systems (tilted with respect to one another in an arbitrary way), perform a triangularly truncated expansion in both, then plot the results. It can be shown that *if triangular truncation is used the two maps will be identical*. This means that the arbitrary orientations of the spherical coordinate systems used had no effect whatsoever on the results obtained. The coordinate system used "disappears" at the end. Triangular truncation is very widely used today, in part because of this nice property.

Fig. H.5 shows an example based on 500 hPa height data, provided originally on a 2.5° longitude-latitude grid. The figure shows how the data look when represented by just a few spherical harmonics (top left), a few more (top right), a moderate number (bottom left) and at full 2.5° resolution. The smoothing effect of severe truncation is clearly visible.

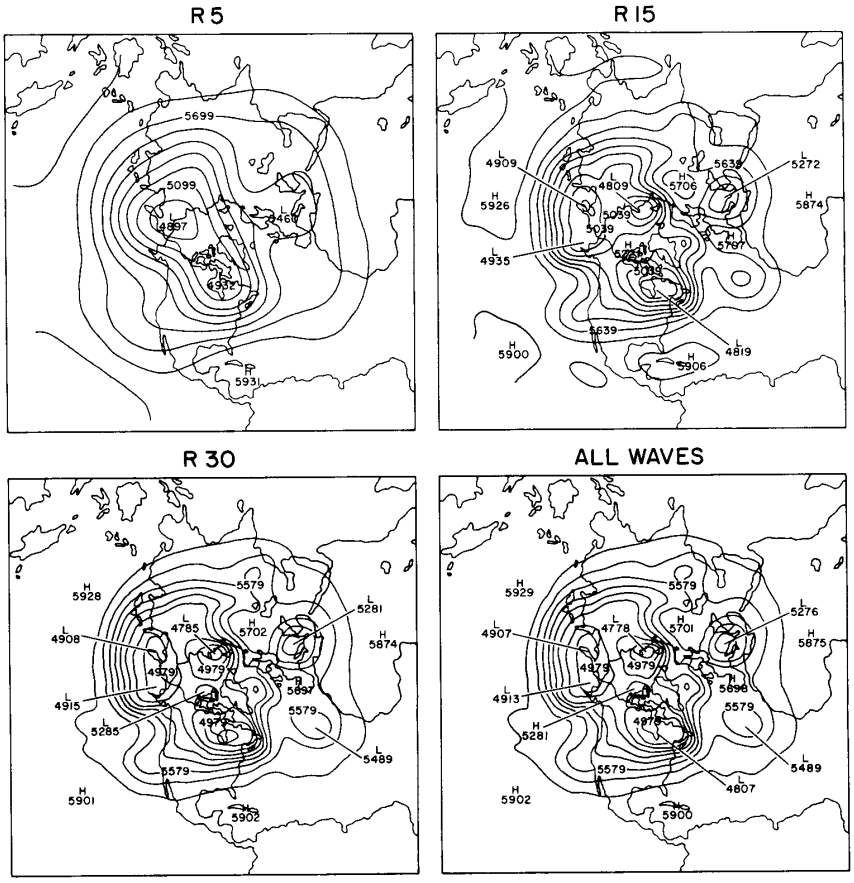


Figure H.5: A demonstration of the effects of various horizontal truncations of 500 hPa geopotential height (m) data. Th e original data are provided on a 2.5° longitude-latitude grid. From Washington and Parkinson (1986).

Appendix I

Hermite Polynomials

The Hermite polynomials are defined by:

$$H_n(y) \equiv (-1)^n e^{y^2} \frac{d^n}{dy^n}\left(e^{-y^2}\right) \text{ for } n \geq 0 \,. \tag{I.1}$$

The algebraic forms and plots of the first six Hermite polynomials are given in Fig. I.1: Note that the even-numbered Hermite polynomials are even functions, and the odd-numbered Hermite polynomials are odd functions.

The Hermite polynomials are orthogonal with respect to e^{-y^2}:

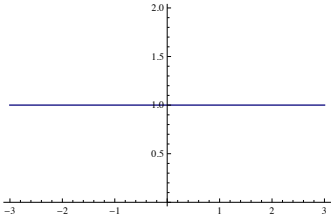
$$\int_{-\infty}^{\infty} H_n(y) H_m(y) e^{-y^2} dy = 0 \text{ for } n \neq m \,, \tag{I.2}$$

$$\int_{-\infty}^{\infty} H_n(y) H_m(y) e^{-y^2} dy = 2^n n! \sqrt{\pi} \text{ for } n = m \,. \tag{I.3}$$

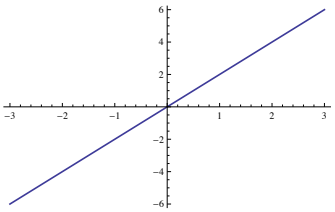
Two useful recursion relations are

$$\frac{d}{dy} H_n(y) = 2n H_{n-1}(y) \text{ for } n \geq 1 \,, \tag{I.4}$$

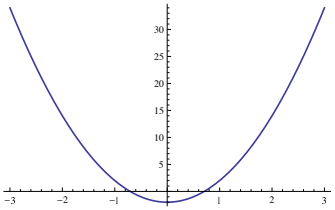
$H_0(y) = 1$



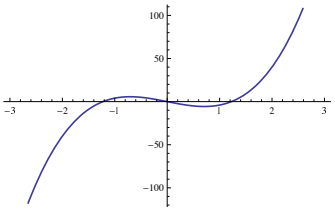
$H_1(y) = 2y$



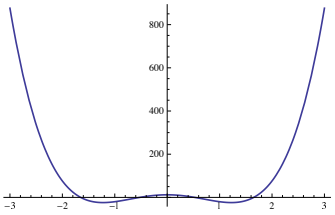
$H_2(y) = 4y^2 - 2$



$H_3(y) = 8y^3 - 12y$



$H_4(y) = 16y^4 - 48y^2 + 12$



$H_5(y) = 32y^5 - 160y^3 + 120y$

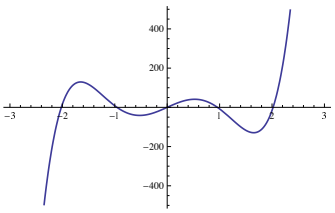


Figure I.1: The algebraic forms and plotted shapes of the first six Hermite polynomials.

and

$$H_{n+1}(y) = 2yH_n(y) - 2nH_{n-1}(y) \ . \tag{I.5}$$

Consider the function

$$\psi_n(y) \equiv \frac{e^{-y^2/2}H_n(y)}{\sqrt{2^n n!\pi^{1/2}}} \ . \tag{I.6}$$

The ψ_n satisfy

$$\left\{ \frac{d^2}{dy^2} + \left[(2n+1) - y^2\right] \right\} \psi_n = 0 \ , \tag{I.7}$$

as can be verified by substitution. Note that for $n \geq 0$, the expression $2n+1$, which appears in (I.7), generates all positive odd integers.

Bibliography

Agnew, D. C., 2025: As the world turns—irregularly. *Physics Today*, **78 (3)**, 54–55.

Aiyyer, A. R., and J. Molinari, 2003: Evolution of mixed rossby–gravity waves in idealized mjo environments. *Journal of the atmospheric sciences*, **60 (23)**, 2837–2855.

Andersen, J. A., and Z. Kuang, 2012: Moist static energy budget of mjo-like disturbances in the atmosphere of a zonally symmetric aquaplanet. *Journal of Climate*, **25 (8)**, 2782–2804.

Andrews, D., and M. McIntyre, 1978: An exact theory of nonlinear waves on a lagrangian-mean flow. *Journal of fluid Mechanics*, **89 (4)**, 609–646.

Andrews, D., and M. E. Mcintyre, 1976: Planetary waves in horizontal and vertical shear: The generalized eliassen-palm relation and the mean zonal acceleration. *Journal of Atmospheric Sciences*, **33 (11)**, 2031–2048.

Andrews, D. G., 1983: A finite-amplitude eliassen-palm theorem in isentropic coordinates. *Journal of Atmospheric Sciences*, **40 (8)**, 1877–1883.

Andrews, D. G., J. R. Holton, and C. B. Leovy, 1987: *Middle atmosphere dynamics*. 40, Academic press.

Arakawa, A., 1969: Parameterization of cumulus convection. *Proceedings of WMO/IUGG Symposium, Numerical Weather Prediction, Japan Meteorological Agency*, Vol. 8, 1–6.

Arakawa, A., 2004: The cumulus parameterization problem: Past, present, and future. *Journal of climate*, **17 (13)**, 2493–2525.

Arakawa, A., and J.-M. Chen, 1986: Closure assumptions in the cumulus parameterization problem. *Journal of the Meteorological Society of Japan. Ser. II*, **64**, 107–131.

Arakawa, A., and C. S. Konor, 2009: Unification of the anelastic and quasi-hydrostatic systems of equations. *Monthly Weather Review*, **137 (2)**, 710–726.

Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment, part i. *Journal of the Atmospheric Sciences*, **31 (3)**, 674–701.

Arnold, N. P., M. Branson, Z. Kuang, D. A. Randall, and E. Tziperman, 2015: Mjo intensification with warming in the superparameterized cesm. *Journal of Climate*, **28 (7)**, 2706–2724.

Arpe, K., Č. Branković, E. Oriol, and P. Speth, 1986: Variability in time and space of energetics from a long series of atmospheric data produced by ecmwf. *Beitraege zur Physik der Atmosphaere*, **59 (2)**, 321–355.

Baldwin, M., and Coauthors, 2001: The quasi-biennial oscillation. *Reviews of Geophysics*, **39 (2)**, 179–229.

Baldwin, M. P., and Coauthors, 2019: 100 years of progress in understanding the stratosphere and mesosphere. *Meteorological Monographs*, **59**, 27–1.

Baldwin, M. P., and Coauthors, 2021: Sudden stratospheric warmings. *Reviews of Geophysics*, **59 (1)**, e2020RG000 708.

Banfield, D., and Coauthors, 2020: The atmosphere of mars as observed by insight. *Nature Geoscience*, **13**, 190–198.

Bangalath, H., and O. Pauluis, 2020: A new mass flux correction procedure for vertically integrated energy transport by constraining mass, energy, and water budgets. *Geophysical Research Letters*, **47 (22)**, e2020GL089 764.

Barnett, T. P., L. Dümenil, U. Schlese, E. Roeckner, and M. Latif, 1989: The effect of eurasian snow cover on regional and global climate variations. *Journal of Atmospheric Sciences*, **46 (5)**, 661–686.

Barriopedro, D., and N. Calvo, 2014: On the relationship between enso, stratospheric sudden warmings, and blocking. *Journal of Climate*, **27 (12)**, 4704–4720.

Bauer, P., A. J. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55.

Bengtsson, L., and J. Shukla, 1988: Integration of space and in situ observations to study global climate change. *Bulletin of the American Meteorological Society*, **69 (10)**, 1130–1143.

Bengtsson, L., and Coauthors, 2007: The need for a dynamical climate reanalysis. *Bulletin of the American Meteorological Society*, **88 (4)**, 495–501.

Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004: Mesoscale weather prediction with the ruc hybrid isentropic–terrain-following coordinate model. *Monthly Weather Review*, **132 (2)**, 473–494.

Bennetts, D., and B. J. Hoskins, 1979: Conditional symmetric instability - a possible explanation for frontal rainbands. *Quarterly Journal of the Royal Meteorological Society*, **105**, 945–962, URL https://api.semanticscholar.org/CorpusID:123005327.

Beres, J. H., R. R. Garcia, B. A. Boville, and F. Sassi, 2005: Implementation of a gravity wave source spectrum parameterization dependent on the properties of convection in the whole atmosphere community climate model (waccm). *Journal of Geophysical Research: Atmospheres*, **110 (D10)**.

Berry, M., 2002: Singular limits. *Physics Today*, **55 (5)**, 10–11.

Betts, A., 1973: Non-precipitating cumulus convection and its parameterization. *Quarterly Journal of the Royal Meteorological Society*, **99 (419)**, 178–196.

Bister, M., and K. A. Emanuel, 1998: Dissipative heating and hurricane intensity. *Meteorology and Atmospheric Physics*, **65 (3)**, 233–240.

Bjerknes, J., 1938: Saturated-adiabatic ascent of air through dry-adiabatically descending environment. *Quart. J. Roy. Meteor. Soc.*, **64**, 325–330.

Bjerknes, J., 1948: Practical application of h. jeffrey's theory of the general circulation. *Jeffreys' theory of the general circulation. Programme et R4sum6 des M6moires, R6union d'Osto, Assoc. M6t6or., UGGI, Geneva*, 13–14.

Bjerknes, J., 1966: A possible response of the atmospheric hadley circulation to equatorial anomalies of ocean temperature. *Tellus A*, **18**, 820–829, URL https://api.semanticscholar.org/CorpusID:140624229.

Bjerknes, J., 1969: Atmospheric teleconnections from the tropical pacific. *Monthly Weather Review*, **97**, 163–172, URL https://api.semanticscholar.org/CorpusID:124582870.

Blackmon, M. L., 1976: A climatological spectral study of the 500 mb geopotential height of the northern hemisphere. *Journal of the Atmospheric Sciences*, **33 (8)**, 1607–1623.

Bleck, R., 1973: Numerical forecasting experiments based on the conservation of potential vorticity on isentropic surfaces. *Journal of Applied Meteorology*, **12 (5)**, 737–752.

Bony, S., and K. A. Emanuel, 2005: On the role of moist processes in tropical intraseasonal variability: Cloud–radiation and moisture–convection feedbacks. *Journal of the atmospheric sciences*, **62 (8)**, 2770–2789.

Bowman, K. P., and P. J. Cohen, 1997: Interhemispheric exchange by seasonal modulation of the Hadley circulation. *Journal of the atmospheric sciences*, **54 (16)**, 2045–2059.

Boyd, J. P., 1976: The noninteraction of waves with the zonally averaged flow on a spherical earth and the interrelationships on eddy fluxes of energy, heat and momentum. *Journal of Atmospheric Sciences*, **33 (12)**, 2285–2291.

Bretherton, C. S., and C. Schär, 1993: Flux of potential vorticity substance: A simple derivation and a uniqueness property. *Journal of the atmospheric sciences*, **50**, 1834–1834.

Bretherton, F. P., 1966: Critical layer instability in baroclinic flows. *Quarterly Journal of the Royal Meteorological Society*, **92 (393)**, 325–334.

Bryan, F. O., 1997: The axial angular momentum balance of a global ocean general circulation model. *Dynamics of atmospheres and oceans*, **25 (3)**, 191–216.

Bukowinski, M. S., 1999: Earth science: Taking the core temperature. *Nature*, **401 (6752)**, 432.

Businger, S., and J. A. Businger, 2001: Viscous dissipation of turbulence kinetic energy in storms. *Journal of the atmospheric sciences*, **58 (24)**, 3793–3796.

Butchart, N., 2014: The brewer-dobson circulation. *Reviews of geophysics*, **52 (2)**, 157–184.

Butchart, N., 2022: The stratosphere: A review of the dynamics and variability. *Weather and Climate Dynamics*, **3 (4)**, 1237–1272.

Carissimo, B., A. Oort, and T. Vonder Haar, 1985: Estimating the meridional energy transports in the atmosphere and ocean. *Journal of Physical Oceanography*, **15 (1)**, 82–91.

Chandrasekhar, S., 2013: *Hydrodynamic and hydromagnetic stability*. Courier Corporation.

Chang, C.-P., Y. Ding, R. H. Johnson, G. N.-c. Lau, B. Wang, and T. Yasunari, 2011: *Global Monsoon System, The: Research And Forecast*, Vol. 5. World Scientific.

Chang, E. K., S. Lee, and K. L. Swanson, 2002: Storm track dynamics. *Journal of climate*, **15 (16)**, 2163–2183.

Chao, W. C., 1987: On the origin of the tropical intraseasonal oscillation. *Journal of Atmospheric Sciences*, **44 (15)**, 1940–1949.

Chapman, S., and R. S. Lindzen, 1969: *Atmospheric tides: thermal and gravitational*, Vol. 15. Springer Science & Business Media.

Charlton, A. J., and L. M. Polvani, 2007: A new look at stratospheric sudden warmings. part i: Climatology and modeling benchmarks. *Journal of climate*, **20 (3)**, 449–469.

Charney, J. G., 1963: A note on large-scale motions in the tropics. *Journal of the Atmospheric Sciences*, **20 (6)**, 607–609.

Charney, J. G., 1971: Geostrophic turbulence. *Journal of the Atmospheric Sciences*, **28 (6)**, 1087–1095.

Charney, J. G., 1973: Planetary fluid dynamics. *Dynamic Meteorology: Lectures Delivered at the Summer School of Space Physics of the Centre National D'Etudes Spatiales, Held at Lannion, France, 7 August–12 September 1970*, Springer, 97–351.

Charney, J. G., 1990: On the scale of atmospheric motions. *The Atmosphere—A Challenge*, Springer, 251–265.

Charney, J. G., and J. G. DeVore, 1979: Multiple flow equilibria in the atmosphere and blocking. *Journal of Atmospheric Sciences*, **36 (7)**, 1205–1216.

Charney, J. G., and P. G. Drazin, 1961: Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *Journal of Geophysical Research*, **66 (1)**, 83–109.

Charney, J. G., and A. Eliassen, 1949: A numerical method for predicting the perturbations of the middle latitude westerlies. *Tellus*, **1 (2)**, 38–54.

Charney, J. G., and J. Shukla, 1981: Predictability of monsoons. *Monsoon dynamics*, 99–110.

Chavas, D. R., and J. Peters, 2023: Static energy deserves greater emphasis in the meteorology community. *Bulletin of the American Meteorological Society*, **104 (10)**, E1918–E1927.

Chavas, D. R., and K. A. Reed, 2019: Dynamical aquaplanet experiments with uniform thermal forcing: System dynamics and implications for tropical cyclone genesis and size. *Journal of the Atmospheric Sciences*, **76 (8)**, 2257–2274.

Chikira, M., 2014: Eastward-propagating intraseasonal oscillation represented by chikira–sugiyama cumulus parameterization. part ii: Understanding moisture variation under weak temperature gradient balance. *Journal of the Atmospheric Sciences*, **71 (2)**, 615–639.

Cripe, D. G., and D. A. Randall, 2001: Joint variations of temperature and water vapor over the midlatitude continents. *Geophysical research letters*, **28 (13)**, 2613–2616.

Cronin, T. W., and D. R. Chavas, 2019: Dry and semidry tropical cyclones. *Journal of the Atmospheric Sciences*, **76 (8)**, 2193–2212.

Crowley, T. J., and G. R. North, 1991: *Paleoclimatology*. New York, NY (United States); Oxford University Press.

Dee, D. P., and Coauthors, 2011: The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, **137 (656)**, 553–597.

Defant, A., 1921: Die zirkulation der atmosphäre in den gemässigten breiten der erde: Grundzüge einer theorie der klimaschwankungen. *Geografiska Annaler*, **3 (3)**, 209–266.

Deland, R. J., 1965: Some observations of the behavior of spherical harmonic waves. *Monthly Weather Review*, **93 (5)**, 307–312.

Denning, A. S., I. Y. Fung, and D. Randall, 1995: Latitudinal gradient of atmospheric co2 due to seasonal exchange with land biota. *Nature*, **376 (6537)**, 240–243.

Denning, A. S., T. Takahashi, and P. Friedlingstein, 1999: Can a strong atmospheric co2 rectifier effect be reconciled with a "reasonable" carbon budget? keynote perspective. *Tellus B*, **51 (2)**, 249–253.

Dickinson, R. E., 1968: Planetary rossby waves propagating vertically through weak westerly wind wave guides. *Journal of Atmospheric Sciences*, **25 (6)**, 984–1002.

Dima, I. M., and J. M. Wallace, 2003: On the seasonality of the Hadley cell. *Journal of the atmospheric sciences*, **60 (12)**, 1522–1527.

Dole, R. M., 1986: The life cycles of persistent anomalies and blocking over the north Pacific. *Advances in Geophysics*, Vol. 29, Elsevier, 31–69.

Dove, H. W., 1837: *Meteorologische Untersuchungen*. Sanderischen Buchhandlung.

Dutton, J. A., 2002: *The ceaseless wind: An introduction to the theory of atmospheric motion*. Courier Corporation.

Edmon Jr., H., B. Hoskins, and M. McIntyre, 1980: Eliassen-palm cross sections for the troposphere. *Journal of Atmospheric Sciences*, **37 (12)**, 2600–2616.

Edouard, S., R. Vautard, and G. Brunet, 1997: On the maintenance of potential vorticity in isentropic coordinates. *Quarterly Journal of the Royal Meteorological Society*, **123 (543)**, 2069–2094.

Egger, J., and K.-P. Hoinka, 2008: The angular momentum budget of the transformed eulerian mean equations. *Journal of the atmospheric sciences*, **65 (10)**, 3305–3314.

Egger, J., W. Metz, and G. Moller, 1986: Forcing of planetary-scale blocking anticyclones by synoptic-scale eddies. *Advances in Geophysics*, Vol. 29, Elsevier, 183–198.

Egger, J., K. Weickmann, and K.-P. Hoinka, 2007: Angular momentum in the global atmospheric circulation. *Reviews of Geophysics*, **45 (4)**.

Eliasen, E., and B. Machenhauer, 1965: A study of the fluctuations of the atmospheric planetary flow patterns represented by spherical harmonics. *Tellus*, **17 (2)**, 220–238.

Eliassen, A., and E. Palm, 1961: On the transfer of energy in stationary mountain waves. geofysike publikasjoner. *Geophysica Norvegica*, **22**.

Eliassen, A., and E. Raustein, 1968: A numerical integration experiment with a model atmosphere based on isentropic surfaces. *Meteorologiske Annaler*, **5**, 45–63.

Emanuel, K., 2001: Contribution of tropical cyclones to meridional heat transport by the oceans. *Journal of Geophysical Research: Atmospheres*, **106 (D14)**, 14 771–14 781.

Emanuel, K., and D. Raymond, 1993: *The presentation of cumulus convection in numerical models, Monogr*. American Meteorological Society.

Emanuel, K., and Coauthors, 2005: *Divine wind: the history and science of hurricanes*. Oxford university press.

Emanuel, K. A., 1979: Inertial instability and mesoscale convective systems. part i: Linear theory of inertial instability in rotating viscous fluids. *Journal of the Atmospheric Sciences*, **36 (12)**, 2425–2449.

Emanuel, K. A., 1982: Inertial instability and mesoscale convective systems. part ii. symmetric cisk in a baroclinic flow. *Journal of Atmospheric Sciences*, **39 (5)**, 1080–1097.

Emanuel, K. A., 1983: On assessing local conditional symmetric instability from atmospheric soundings. *Monthly weather review*, **111 (10)**, 2016–2033.

Emanuel, K. A., 1987a: An air-sea interaction model of intraseasonal oscillations in the tropics. *Journal of Atmospheric Sciences*, **44 (16)**, 2324–2340.

Emanuel, K. A., 1987b: The dependence of hurricane intensity on climate. *Nature*, **326 (6112)**, 483–485.

Fajber, R., A. Donohoe, S. Ragen, K. C. Armour, and P. J. Kushner, 2023: Atmospheric heat transport is governed by meridional gradients in surface evaporation in modern-day earth-like climates. *Proceedings of the National Academy of Sciences*, **120 (25)**, e2217202 120.

Ferrari, R., and R. Plumb, 2003: Residual circulation in the ocean. *Near-Boundary Processes and Their Parameterization: Proc.'Aha Huliko'a Hawaiian Winter Workshop*, Citeseer, 219–228.

Flasar, F. M., and Coauthors, 2004: An intense stratospheric jet on jupiter. *Nature*, **427 (6970)**, 132–135.

Forbes, J. M., and H. B. Garrett, 1979: Theoretical studies of atmospheric tides. *Reviews of Geophysics*, **17 (8)**, 1951–1981.

Fouchet, T., S. Guerlet, D. Strobel, A. A. Simon-Miller, B. Bézard, and F. Flasar, 2008: An equatorial oscillation in saturn's middle atmosphere. *Nature*, **453 (7192)**, 200–202.

Fovell, R., D. Durran, and J. Holton, 1992: Numerical simulations of convectively generated stratospheric gravity waves. *Journal of Atmospheric Sciences*, **49 (16)**, 1427–1442.

Friedson, A. J., 1999: New observations and modelling of a qbo-like oscillation in jupiter's stratosphere. *Icarus*, **137 (1)**, 34–55.

Fuchs, Ž., and D. J. Raymond, 2007: A simple, vertically resolved model of tropical disturbances with a humidity closure. *Tellus A: Dynamic Meteorology and Oceanography*, **59 (3)**, 344–354.

Fulton, S. R., and W. H. Schubert, 1985: Vertical normal mode transforms: Theory and application. *Monthly Weather Review*, **113 (4)**, 647–658.

Garcia, R. R., T. J. Dunkerton, R. S. Lieberman, and R. A. Vincent, 1997: Climatology of the semiannual oscillation of the tropical middle atmosphere. *Journal of Geophysical Research: Atmospheres*, **102 (D22)**, 26 019–26 032.

Geisler, J., 1981: A linear model of the walker circulation. *Journal of Atmospheric Sciences*, **38 (7)**, 1390–1400.

Gelaro, R., and Coauthors, 2017: The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of climate*, **Volume 30 Iss 13**, 5419–5454.

Gent, P. R., and J. C. Mcwilliams, 1990: Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, **20 (1)**, 150–155.

Gill, A. E., 1980: Some simple solutions for heat-induced tropical circulation. *Quarterly Journal of the Royal Meteorological Society*, **106 (449)**, 447–462.

Gleick, J., 2008: *Chaos: Making a new science*. Penguin.

Grabowski, W. W., and M. Moncrieff, 2004: Moisture–convection feedback in the tropics. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **130 (604)**, 3081–3104.

Grell, G. A., Y.-H. Kuo, and R. J. Pasch, 1991: Semiprognostic tests of cumulus parameterization schemes in the middle latitudes. *Monthly weather review*, **119 (1)**, 5–31.

Grose, W. L., and B. J. Hoskins, 1979: On the influence of orography on large-scale atmospheric flow. *Journal of Atmospheric Sciences*, **36 (2)**, 223–234.

Hack, J. J., W. H. Schubert, and P. L. S. Dias, 1984: A spectral cumulus parameterization for use in numerical models of the tropical atmosphere. *Monthly weather review*, **112 (4)**, 704–716.

Hack, J. J., W. H. Schubert, D. E. Stevens, and H.-C. Kuo, 1989: Response of the Hadley circulation to convective forcing in the itcz. *Journal of Atmospheric Sciences*, **46 (19)**, 2957–2973.

Han, J., M. L. Witek, J. Teixeira, R. Sun, H.-L. Pan, J. K. Fletcher, and C. S. Bretherton, 2016: Implementation in the ncep gfs of a hybrid eddy-diffusivity mass-flux (edmf) boundary layer parameterization with dissipative heating and modified stable boundary layer mixing. *Weather and Forecasting*, **31 (1)**, 341–352.

Hansen, A. R., and T.-C. Chen, 1982: A spectral energetics analysis of atmospheric blocking. *Monthly Weather Review*, **110 (9)**, 1146–1165.

Hardiman, S. C., D. G. Andrews, A. A. White, N. Butchart, and I. Edmond, 2010: Using different formulations of the transformed eulerian mean equations and eliassen–palm diagnostics in general circulation models. *Journal of the atmospheric sciences*, **67 (6)**, 1983–1995.

Harper, K. C., 2012: *Weather by the numbers: the genesis of modern meteorology*. mit Press.

Hartmann, D. L., and J. R. Gross, 1988: Seasonal variability of the 40–50 day oscillation in wind and rainfall in the tropics. *Journal of Atmospheric Sciences*, **45 (19)**, 2680–2702.

Hartmann, D. L., and F. Lo, 1998: Wave-driven zonal flow vacillation in the southern hemisphere. *Journal of the Atmospheric Sciences*, **55 (8)**, 1303–1315.

Hartmann, D. L., M. E. Ockert-Bell, and M. L. Michelsen, 1992: The effect of cloud type on earth's energy balance: Global analysis. *Journal of Climate*, 1281–1304.

Haurwitz, B., 1937: The oscillations of the atmosphere. *Gerlands Beitr. Geophys*, **51**, 195–233.

Haynes, P. H., and M. E. McIntyre, 1987: On the evolution of vorticity and potential vorticity in the presence of diabatic heating and frictional or other forces. *Journal of the Atmospheric Sciences*, **44 (5)**, 828–841.

Held, I. M., 1983: Stationary and quasi-stationary eddies in the extratropical troposphere: Theory. *Large-scale dynamical processes in the atmosphere*, **127**, 168.

Held, I. M., and B. J. Hoskins, 1985: Large-scale eddies and the general circulation of the troposphere. *Advances in geophysics*, Vol. 28, Elsevier, 3–31.

Held, I. M., and A. Y. Hou, 1980: Nonlinear axially symmetric circulations in a nearly inviscid atmosphere. *Journal of the Atmospheric Sciences*, **37 (3)**, 515–533.

Held, I. M., and T. Schneider, 1999: The surface branch of the zonally averaged mass transport circulation in the troposphere. *Journal of the atmospheric sciences*, **56 (11)**, 1688–1697.

Held, I. M., and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *Journal of climate*, **19 (21)**, 5686–5699.

Held, I. M., M. Ting, and H. Wang, 2002: Northern winter stationary waves: Theory and modeling. *Journal of climate*, **15 (16)**, 2125–2144.

Hersbach, H., and Coauthors, 2020: The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146 (730)**, 1999–2049.

Hess, S., R. Henry, C. B. Leovy, J. Ryan, and J. E. Tillman, 1977: Meteorological results from the surface of mars: Viking 1 and 2. *Journal of Geophysical Research*, **82 (28)**, 4559–4574.

Hide, R., and J. O. Dickey, 1991: Earth's variable rotation. *Science*, **253 (5020)**, 629–637.

Holloway, C. E., and J. D. Neelin, 2009: Moisture vertical structure, column water vapor, and tropical deep convection. *Journal of the atmospheric sciences*, **66 (6)**, 1665–1683.

Holloway, C. E., S. J. Woolnough, and G. M. Lister, 2013: The effects of explicit versus parameterized convection on the mjo in a large-domain high-resolution tropical case study. part i: Characterization of large-scale organization and propagation. *Journal of the Atmospheric Sciences*, **70 (5)**, 1342–1369.

Holton, J. R., 1973: An introduction to dynamic meteorology. *American Journal of Physics*, **41 (5)**, 752–754.

Holton, J. R., 1980: The dynamics of sudden stratospheric warmings. *Annual Review of Earth and Planetary Sciences*, **8 (1)**, 169–190.

Holton, J. R., and R. S. Lindzen, 1972: An updated theory for the quasi-biennial cycle of the tropical stratosphere. *Journal of Atmospheric Sciences*, **29 (6)**, 1076–1080.

Hoskins, B. J., 1991: Towards a PV-θ view of the general circulation. *Tellus A: Dynamic Meteorology and Oceanography*, **43 (4)**, 27–36.

Hoskins, B. J., I. N. James, and G. H. White, 1983: The shape, propagation and mean-flow interaction of large-scale weather systems. *Journal of Atmospheric Sciences*, **40 (7)**, 1595–1612.

Hoskins, B. J., M. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorological Society*, **111 (470)**, 877–946.

Hough, S. S., 1898: V. on the application of harmonic analysis to the dynamical theory of the tides.—part ii. on the general integration of laplace's dynamical equations. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **62 (191)**, 139–185.

Hsu, Y.-J. G., and A. Arakawa, 1990: Numerical modeling of the atmosphere with an isentropic vertical coordinate. *Monthly Weather Review*, **118 (10)**, 1933–1959.

Hu, Y., D. Yang, and J. Yang, 2008: Blocking systems over an aqua planet. *Geophysical research letters*, **35 (19)**.

Hung, M.-P., J.-L. Lin, W. Wang, D. Kim, T. Shinoda, and S. J. Weaver, 2013: Mjo and convectively coupled equatorial waves simulated by cmip5 climate models. *Journal of Climate*, **26 (17)**, 6185–6214.

Hurrell, J. W., and C. Deser, 2010: North atlantic climate variability: the role of the north atlantic oscillation. *Journal of marine systems*, **79 (3-4)**, 231–244.

Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: An overview of the north atlantic oscillation. *Geophysical Monograph-American Geophysical Union*, **134**, 1–36.

Hurrell, J. W., Y. Kushnir, and M. Visbeck, 2001: The north atlantic oscillation. *Science*, **291 (5504)**, 603–605.

Igel, M. R., and J. A. Biello, 2020: The nontraditional coriolis terms and tropical convective clouds. *Journal of Atmospheric Sciences*, **77 (12)**, 3985–3998.

Imbrie, J., and K. P. Imbrie, 1986: *Ice ages: solving the mystery*. Harvard University Press.

Iribarne, J. V., and W. L. Godson, 1981: *Atmospheric thermodynamics*, Vol. 6. Springer Science & Business Media.

Jayne, S. R., D. Roemmich, N. Zilberman, S. C. Riser, K. S. Johnson, G. C. Johnson, and S. R. Piotrowicz, 2017: The argo program: present and future. *Oceanography*, **30 (2)**, 18–28.

Jeffreys, H., 1926: On the dynamics of geostrophic winds. *Quart. J. Roy. Meteor. Soc*, **52 (217)**, 85–104.

Johnson, D. R., 1989: The forcing and maintenance of global monsoonal circulations: An isentropic analysis. *Advances in Geophysics*, Vol. 31, Elsevier, 43–316.

Johnson, D. R., and L. W. Uccellini, 1983: A comparison of methods for computing the sigma-coordinate pressure gradient force for flow over sloped terrain in a hybrid theta-sigma model. *Monthly Weather Review*, **111 (4)**, 870–886.

Johnson, R. H., 2011: Diurnal cycle of monsoon convection. *The global monsoon system: Research and forecast*, World Scientific, 257–276.

Jones, T. R., and D. A. Randall, 2011: Quantifying the limits of convective parameterizations. *Journal of Geophysical Research: Atmospheres*, **116 (D8)**.

Juckes, M., 2001: A generalization of the transformed eulerian-mean meridional circulation. *Quarterly Journal of the Royal Meteorological Society*, **127 (571)**, 147–160.

Juckes, M., I. James, and M. Blackburn, 1994: The influence of antarctica on the momentum budget of the southern extratropics. *Quarterly Journal of the Royal Meteorological Society*, **120 (518)**, 1017–1044.

Kalnay, E., and Coauthors, 1996: The ncep/ncar 40-year reanalysis project. *Renewable Energy*.

Kao, C.-Y. J., and Y. Ogura, 1987: Response of cumulus clouds to large-scale forcing using the arakawa-schubert cumulus parameterization. *Journal of Atmospheric Sciences*, **44 (17)**, 2437–2458.

Khairoutdinov, M., and K. Emanuel, 2013: Rotating radiative-convective equilibrium simulated by a cloud-resolving model. *Journal of Advances in Modeling Earth Systems*, **5 (4)**, 816–825.

Kiladis, G. N., M. C. Wheeler, P. T. Haertel, K. H. Straub, and P. E. Roundy, 2009: Convectively coupled equatorial waves. *Reviews of Geophysics*, **47 (2)**.

Kim, D., and Coauthors, 2009: Application of mjo simulation diagnostics to climate models. *Journal of Climate*, **22 (23)**, 6413–6436.

Kim, D., and Coauthors, 2014: Process-oriented mjo simulation diagnostic: Moisture sensitivity of simulated convection. *Journal of Climate*, **27 (14)**, 5379–5395.

Klein, S. A., and D. L. Hartmann, 1993: The seasonal cycle of low stratiform clouds. *Journal of climate*, **6 (8)**, 1587–1606.

Klemp, J., and D. Lilly, 1978: Numerical simulation of hydrostatic mountain waves. *Journal of Atmospheric Sciences*, **35 (1)**, 78–107.

Köhler, M., 2006: An integral approach to modeling pbl transports and clouds: Edmf@ ecmwf. *Proceedings of workshop on parametrization of clouds in large-scale models*, 13–15.

Kolmogorov, A. N., 1941: The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. *Cr Acad. Sci. URSS*, **30**, 301–305.

Krishnamurti, T., H. Bedi, and M. Subramaniam, 1989: The summer monsoon of 1987. *Journal of Climate*, 321–340.

Krishnamurti, T. N., and D. Subrahmanyam, 1982: The 30–50 day mode at 850 mb during monex. *Journal of Atmospheric Sciences*, **39 (9)**, 2088–2095.

Krueger, A. J., and R. A. Minzner, 1976: A mid-latitude ozone model for the 1976 US Standard Atmosphere. *Journal of Geophysical Research*, **81 (24)**, 4477–4481.

Kunz, A., P. Konopka, R. Müller, and L. Pan, 2011: Dynamical tropopause based on isentropic potential vorticity gradients. *Journal of Geophysical Research: Atmospheres*, **116 (D1)**.

Landu, K., and E. D. Maloney, 2011: Effect of sst distribution and radiative feedbacks on the simulation of intraseasonal variability in an aquaplanet gcm. *Journal of the Meteorological Society of Japan. Ser. II*, **89 (3)**, 195–210.

Laskar, J., 1989: A numerical experiment on the chaotic behaviour of the solar system. *Nature*, **338 (6212)**, 237–238.

Lau, K.-M., and P. H. Chan, 1986: Aspects of the 40–50 day oscillation during the northern summer as inferred from outgoing longwave radiation. *Monthly Weather Review*, **114 (7)**, 1354–1367.

Lau, N.-C., 1985: Modeling the seasonal dependence of the atmospheric response to observed el ninos in 1962–76. *Monthly Weather Review*, **113 (11)**, 1970–1996.

Lembo, V., F. Fabiano, V. M. Galfi, R. G. Graversen, G. Messori, and Coauthors, 2022: Meridional-energy-transport extremes and the general circulation of northern hemisphere mid-latitudes: dominant weather regimes and preferred zonal wavenumbers. *Weather and Climate Dynamics*, **3 (3)**, 1037–1062.

Leovy, C. B., A. J. Friedson, and G. S. Orton, 1991: The quasiquadrennial oscillation of jupiter's equatorial stratosphere. *Nature*, **354 (6352)**, 380–382.

Lesieur, M., 2008: Introduction to turbulence in fluid mechanics. *Turbulence in Fluids: Fourth Revised and Enlarged Edition*, 1–23.

Lewis, J., and R. Prinn, 1984: Planets and their atmospheres- origin and evolution(book). *Orlando, FL, Academic Press, Inc.(International Geophysics Series.*, **33**.

Li, H., and R. L. Sriver, 2018: Impact of tropical cyclones on the global ocean: Results from multidecadal global ocean simulations isolating tropical cyclone forcing. *Journal of Climate*, **31 (21)**, 8761–8784.

Li, T., L. Wang, M. S. Peng, B. Wang, C. Zhang, W. Lau, and H. chi Kuo, 2018: A paper on the tropical intraseasonal oscillation published in 1963 in a chinese journal. *Bulletin of the American Meteorological Society*.

Lighthill, J., and R. P. Pearce, 1981: *Monsoon dynamics*. Cambridge University Press.

Lilly, D. K., 1968: Models of cloud-topped mixed layers under a strong inversion. *Quarterly Journal of the Royal Meteorological Society*, **94 (401)**, 292–309.

Lilly, D. K., G. Bassett, K. Droegemeier, and P. Bartello, 1998: Stratified turbulence in the atmospheric mesoscales. *Theoretical and computational fluid dynamics*, **11 (3-4)**, 139–153.

Limpasuvan, V., and D. L. Hartmann, 2000: Wave-maintained annular modes of climate variability. *Journal of Climate*, **13 (24)**, 4414–4429.

Lindzen, R. S., 1966: On the theory of the diurnal tide. *Monthly Weather Review*, **94 (5)**, 295–301.

Lindzen, R. S., 1967: Planetary waves on beta-planes. *Monthly Weather Review*, **95 (7)**, 441–451.

Lindzen, R. S., 1990: *Dynamics in atmospheric physics*, Vol. 310. Cambridge University Press Cambridge.

Lindzen, R. S., and J. R. Holton, 1968: A theory of the quasi-biennial oscillation. *Journal of Atmospheric Sciences*, **25 (6)**, 1095–1107.

Lindzen, R. S., and A. V. Hou, 1988: Hadley circulations for zonally averaged heating centered off the equator. *Journal of Atmospheric Sciences*, **45 (17)**, 2416–2427.

Lindzen, R. S., and S. Nigam, 1987: On the role of sea surface temperature gradients in forcing low-level winds and convergence in the tropics. *Journal of Atmospheric Sciences*, **44 (17)**, 2418–2436.

Liu, W. T., 2002: Progress in scatterometer application. *Journal of oceanography*, **58**, 121–136.

Loeb, N. G., J. M. Lyman, G. C. Johnson, R. P. Allan, D. R. Doelling, T. Wong, B. J. Soden, and G. L. Stephens, 2012: Observed changes in top-of-the-atmosphere radiation and upper-ocean heating consistent within uncertainty. *Nature Geoscience*, **5 (2)**, 110–113.

Longuet-Higgins, M. S., 1968: The eigenfunctions of laplace's tidal equation over a sphere. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **262 (1132)**, 511–607.

Lord, S. J., 1982: Interaction of a cumulus cloud ensemble with the large-scale environment. part iii: Semi-prognostic test of the arakawa-schubert cumulus parameterization. *Journal of Atmospheric Sciences*, **39 (1)**, 88–103.

Lord, S. J., and A. Arakawa, 1980: Interaction of a cumulus cloud ensemble with the large-scale environment. part ii. *Journal of Atmospheric Sciences*, **37 (12)**, 2677–2692.

Lorenz, E., 1967: The nature and theory of the general circulation of the atmosphere. *World meteorological organization*, **161**.

Lorenz, E. N., 1951: Seasonal and irregular variations of the northern hemisphere sea-level pressure profile. *Journal of Atmospheric Sciences*, **8 (1)**, 52–59.

Lorenz, E. N., 1955: Available potential energy and the maintenance of the general circulation. *Tellus*, **7 (2)**, 157–167.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *Journal of atmospheric sciences*, **20 (2)**, 130–141.

Lorenz, E. N., 1969: Three approaches to atmospheric predictability. *Bull. Amer. Meteor. Soc*, **50 (3454)**, 349.

Lorenz, E. N., 1976: Nondeterministic theories of climatic change. *Quaternary Research*, **6 (4)**, 495–506.

Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34 (6)**, 505–513.

Lorenz, E. N., 1984: Irregularity: A fundamental property of the atmosphere. *Tellus A*, **36 (2)**, 98–110.

Lorenz, E. N., 1990: Can chaos and intransitivity lead to interannual variability? *Tellus a*, **42 (3)**, 378–389.

Luo, D., and Z. Chen, 2006: The role of land–sea topography in blocking formation in a block–eddy interaction model. *Journal of the atmospheric sciences*, **63 (11)**, 3056–3065.

Ma, C.-C., C. R. Mechoso, A. W. Robertson, and A. Arakawa, 1996: Peruvian stratus clouds and the tropical pacific circulation: A coupled ocean-atmosphere gcm study. *Journal of Climate*, **9 (7)**, 1635–1645.

Madden, R. A., and P. R. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical pacific. *Journal of Atmospheric Sciences*, **28 (5)**, 702–708.

Madden, R. A., and P. R. Julian, 1972: Description of global-scale circulation cells in the tropics with a 40–50 day period. *Journal of Atmospheric Sciences*, **29 (6)**, 1109–1123.

Madden, R. A., and P. R. Julian, 1994: Observations of the 40–50-day tropical oscillation—a review. *Monthly weather review*, **122 (5)**, 814–837.

Maloney, E. D., 2009: The moist static energy budget of a composite tropical intraseasonal oscillation in a climate model. *Journal of Climate*, **22 (3)**, 711–729.

Maloney, E. D., A. H. Sobel, and W. M. Hannah, 2010: Intraseasonal variability in an aquaplanet general circulation model. *Journal of Advances in Modeling Earth Systems*, **2 (2)**.

Manabe, S., and A. J. Broccoli, 2020: *Beyond Global Warming: How Numerical Models Revealed the Secrets of Climate Change*. Princeton University Press.

Manabe, S., J. Smagorinsky, and R. F. Strickler, 1965: Simulated climatology of a general circulation model with a hydrologic cycle. *Monthly Weather Review*, **93 (12)**, 769–798.

Manabe, S., and T. B. Terpstra, 1974: The effects of mountains on the general circulation of the atmosphere as identified by numerical experiments. *Journal of Atmospheric Sciences*, **31 (1)**, 3–42.

Manabe, S., and R. T. Wetherald, 1967: Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *Journal of the Atmospheric Sciences*.

Mapes, B. E., and J. T. Bacmeister, 2012: Diagnosis of tropical biases and the mjo from patterns in the merra analysis tendency fields. *Journal of climate*, **25 (18)**, 6202–6214.

Margot, J.-L., D. B. Campbell, J. D. Giorgini, J. S. Jao, L. G. Snedeker, F. D. Ghigo, and A. Bonsall, 2021: Spin state and moment of inertia of venus. *Nature Astronomy*, 1–8.

Margules, M., 1893: *Luftbewegungen in einer rotirenden Sphäroidschale*. Aus Der Kaiserlich-Königlichen Hof-Und Staatsdruckerei. In Commission Bei F ....

Martius, O., L. Polvani, and H. Davies, 2009: Blocking precursors to stratospheric sudden warming events. *Geophysical Research Letters*, **36 (14)**.

Maruyama, T., and M. Yanai, 1967: Evidence of large-scale wave disturbances in the equatorial lower stratosphere. *Journal of the Meteorological Society of Japan. Ser. II*, **45 (2)**, 196–199.

Masuda, K., 1988: Meridional heat transport by the atmosphere and the ocean: Analysis of FGGE data. *Tellus A*, **40 (4)**, 285–302.

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *Journal of the Meteorological Society of Japan. Ser. II*, **44 (1)**, 25–43.

Matsuno, T., 1970: Vertical propagation of stationary planetary waves in the winter northern hemisphere. *Journal of Atmospheric Sciences*, **27 (6)**, 871–883.

Matsuno, T., 1971: A dynamical model of the stratospheric sudden warming. *Journal of Atmospheric Sciences*, **28 (8)**, 1479–1494.

Matsuno, T., and K. Nakamura, 1979: The eulerian-and lagrangian-mean meridional circulations in the stratosphere at the time of a sudden warming. *Journal of the Atmospheric Sciences*, **36 (4)**, 640–654.

McCarthy, G. D., and Coauthors, 2015: Measuring the Atlantic meridional overturning circulation at 26 n. *Progress in Oceanography*, **130**, 91–111.

McCreary, J., 1981: A linear stratified ocean model of the equatorial undercurrent. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **298 (1444)**, 603–635.

McDowell, J. C., 2018: The edge of space: Revisiting the Karman Line. *Acta Astronautica*, **151**, 668–677.

McFarlane, N., 1987: The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere. *Journal of Atmospheric Sciences*, **44 (14)**, 1775–1800.

McWilliams, J. C., J. B. Weiss, and I. Yavneh, 1994: Anisotropy and coherent vortex structures in planetary turbulence. *Science*, **264 (5157)**, 410–413.

Miller, R. L., 1997: Tropical thermostats and low cloud cover. *Journal of Climate*, **10**, 409–440, URL https://api.semanticscholar.org/CorpusID:14866638.

MOREL, P., 1973: Dynamic meteorology(book). *Dordrecht, D. Reidel Publishing Co., 1973. 628 p.*

Mullen, S. L., 1987: Transient eddy forcing of blocking flows. *Journal of Atmospheric Sciences*, **44 (1)**, 3–22.

Murakami, T., L.-X. Chen, A. Xie, and M. L. Shrestha, 1986: Eastward propagation of 30–60 day perturbations as revealed from outgoing longwave radiation data. *Journal of Atmospheric Sciences*, **43 (10)**, 961–971.

Naegele, A., and D. Randall, 2019: Geographical and seasonal variability of cloud-radiative feedbacks on precipitation. *Journal of Geophysical Research: Atmospheres*, **124 (2)**, 684–699.

Nakamura, H., M. Nakamura, and J. L. Anderson, 1997: The role of high-and low-frequency dynamics in blocking formation. *Monthly Weather Review*, **125 (9)**, 2074–2093.

Nakamura, N., and C. S. Huang, 2018: Atmospheric blocking as a traffic jam in the jet stream. *Science*, **361 (6397)**, 42–47.

Nakazawa, T., 1988: Tropical super clusters within intraseasonal variations over the western pacific. *Journal of the Meteorological Society of Japan. Ser. II*, **66 (6)**, 823–839.

Namias, J., and R. G. Stone, 1940: *An introduction to the study of air mass and isentropic analysis*. American Meteorological Society,.

Nastrom, G., and K. S. Gage, 1985: A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *Journal of the atmospheric sciences*, **42 (9)**, 950–960.

National Oceanic and Atmospheric Administration, 1976: U.S. Standard Atmosphere, 1976. U.S.Government Printing Office, 241 pp.

Neelin, J. D., D. S. Battisti, A. C. Hirst, F.-F. Jin, Y. Wakata, T. Yamagata, and S. E. Zebiak, 1998: Enso theory. *Journal of Geophysical Research: Oceans*, **103 (C7)**, 14 261–14 290.

Neelin, J. D., and I. M. Held, 1987: Modeling tropical convergence based on the moist static energy budget. *Monthly Weather Review*, **115 (1)**, 3–12.

Neelin, J. D., and J.-Y. Yu, 1994: Modes of tropical variability under convective adjustment and the madden–julian oscillation. part i: Analytical theory. *Journal of Atmospheric Sciences*, **51 (13)**, 1876–1894.

Neggers, R. A., 2006: Extending edmf into the statistical modeling of boundary layer clouds. *ECMWF Workshop on Parametrization of clouds in large scale*, 77–92.

Newell, R. E., 1963: Transfer through the tropopause and within the stratosphere. *Quarterly Journal of the Royal Meteorological Society*, **89 (380)**, 167–204.

Newell, R. E., J. W. Kidson, D. Vincent, and G. Boer, 1972: The general circulation of the tropical atmosphere and interactions with extratropical latitudes, volume i. MIT Press, 258pp.

Newell, R. E., Y. Zhu, E. V. Browell, W. G. Read, and J. W. Waters, 1996: Walker circulation and tropical upper tropospheric water vapor. *Journal of Geophysical Research: Atmospheres*, **101 (D1)**, 1961–1974.

Newman, P. A., and E. R. Nash, 2005: The unusual southern hemisphere stratosphere winter of 2002. *Journal of the atmospheric sciences*, **62 (3)**, 614–628.

Noda, A., 2010: A general three-dimensional transformed eulerian mean formulation. *SOLA*, **6**, 85–88.

Ong, H., and P. E. Roundy, 2020: The compressional beta effect: Analytical solution, numerical benchmark, and data analysis. *Journal of the Atmospheric Sciences*, **77 (11)**, 3721–3732.

Oort, A. H., 1989: Angular momentum cycle in the atmosphere–ocean–solid Earth system. *Bulletin of the American Meteorological Society*, **70 (10)**, 1231–1242.

Oort, A. H., and E. M. Rasmusson, 1971: *Atmospheric circulation statistics*, Vol. 5. US Department of Commerce, National Oceanic and Atmospheric Administration ....

Oort, A. H., and T. H. Vonder Haar, 1976: On the observed annual cycle in the ocean-atmosphere heat balance over the northern hemisphere. *Journal of Physical Oceanography*, **6 (6)**, 781–800.

Orton, G. S., and Coauthors, 1991: Thermal maps of jupiter: Spatial organization and time dependence of stratospheric temperatures, 1980 to 1990. *Science*, **252 (5005)**, 537–542.

Orton, G. S., and Coauthors, 1994: Spatial organization and time dependence of jupiter's tropospheric temperatures, 1980-1993. *Science*, **265 (5172)**, 625–631.

Palmer, T., 2024: The real butterfly effect and maggoty apples. *Physics Today*, **77 (5)**, 30–35.

Palmer, T., G. Shutts, and R. Swinbank, 1986: Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parametrization. *Quarterly Journal of the Royal Meteorological Society*, **112 (474)**, 1001–1039.

Palmer, T. N., 1993: Extended-range atmospheric prediction and the lorenz model. *Bulletin of the American Meteorological Society*, **74 (1)**, 49–66.

Palmer, T. N., 1999: A nonlinear dynamical perspective on climate prediction. *Journal of climate*, **12 (2)**, 575–591.

Palmer, T. N., 2022: *The Primacy of Doubt: From climate change to quantum physics, how the science of uncertainty can help predict and understand our chaotic world*. Oxford University Press.

Pan, D.-M., and D. D. Randall, 1998: A cumulus parameterization with a prognostic closure. *Quarterly Journal of the Royal Meteorological Society*, **124 (547)**, 949–981.

Pauluis, O., V. Balaji, and I. M. Held, 2000: Frictional dissipation in a precipitating atmosphere. *Journal of the atmospheric sciences*, **57 (7)**, 989–994.

Pauluis, O., A. Czaja, and R. Korty, 2010: The global atmospheric circulation in moist isentropic coordinates. *Journal of Climate*, **23 (11)**, 3077–3093.

Pauluis, O., T. Shaw, and F. Laliberté, 2011: A statistical generalization of the transformed eulerian-mean circulation for an arbitrary vertical coordinate system. *Journal of the atmospheric sciences*, **68 (8)**, 1766–1783.

Pauluis, O. M., and A. A. Mrowiec, 2013: Isentropic analysis of convective motions. *Journal of the atmospheric sciences*, **70 (11)**, 3673–3688.

Peixoto, J. P., and A. H. Oort, 1992: *Physics of climate*. New York, NY (United States); American Institute of Physics.

Philander, S. G., 1990: *El Niño, La Niña, and the Southern Oscillation*. Academic Press, 293 pp.

Plumb, R., A. McEwan, and Coauthors, 1978: The instability of a forced standing wave in a viscous stratified fluid- a laboratory analogue of the quasi-biennial oscillation. *Journal of the atmospheric sciences*, **35 (10)**, 1827–1839.

Plumb, R. A., 1984: The quasi-biennial oscillation. *Dynamics of the Middle atmosphere*, J. Holton, and T. Matsuno, Eds., URL https://api.semanticscholar.org/CorpusID:126506164.

Plumb, R. A., and R. Ferrari, 2005: Transformed eulerian-mean theory. part i: Nonquasi-geostrophic theory for eddies on a zonal-mean flow. *Journal of Physical Oceanography*, **35 (2)**, 165–174.

Pritchard, M. S., and C. S. Bretherton, 2014: Causal evidence that rotational moisture advection is critical to the superparameterized madden–julian oscillation. *Journal of the Atmospheric Sciences*, **71 (2)**, 800–815.

Provenzale, A., 1999: Transport by coherent barotropic vortices. *Annual review of fluid mechanics*, **31 (1)**, 55–93.

Quiroz, R. S., 1986: The association of stratospheric warmings with tropospheric blocking. *Journal of Geophysical Research: Atmospheres*, **91 (D4)**, 5277–5285.

Ramanathan, V., and W. Collins, 1993: A thermostat in the tropics? *Nature*, **361 (6411)**, 410–411.

Randall, D., M. Khairoutdinov, A. Arakawa, and W. Grabowski, 2003: Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, **84 (11)**, 1547–1564.

Randall, D. A., 1987: Turbulent fluxes of liquid water and buoyancy in partly cloudy layers. *Journal of the Atmospheric Sciences*, **44 (5)**, 850–858.

Randall, D. A., 2013: Beyond deadlock. *Geophysical Research Letters*, **40 (22)**, 5970–5976.

Randall, D. A., and D.-M. Pan, 1993: Implementation of the arakawa-schubert cumulus parameterization with a prognostic closure. *The representation of cumulus convection in numerical models*, Springer, 137–144.

Randall, D. A., Q. Shao, and C.-H. Moeng, 1992: A second-order bulk boundary-layer model. *Journal of the Atmospheric Sciences*, **49 (20)**, 1903–1923.

Raymond, D. J., 2001: A new model of the madden–julian oscillation. *Journal of the atmospheric sciences*, **58 (18)**, 2807–2819.

Reed, R. J., 1962: Some features of the annual temperature regime in the tropical stratosphere. *Monthly Weather Review*, **90 (6)**, 211–215.

Reed, R. J., 1966: Zonal wind behavior in the equatorial stratosphere and lower mesosphere. *Journal of Geophysical Research*, **71 (18)**, 4223–4233.

Reed, R. J., and E. E. Recker, 1971: Structure and properties of synoptic-scale wave disturbances in the equatorial western pacific. *Journal of Atmospheric Sciences*, **28 (7)**, 1117–1133.

Ren, H.-L., F.-F. Jin, J.-S. Kug, J.-X. Zhao, and J. Park, 2009: A kinematic mechanism for positive feedback between synoptic eddies and nao. *Geophysical Research Letters*, **36 (11)**.

Rhines, P. B., 1975: Waves and turbulence on a beta-plane. *Journal of Fluid Mechanics*, **69 (3)**, 417–443.

Richter, J. H., F. Sassi, and R. R. Garcia, 2010: Toward a physically based gravity wave source parameterization in a general circulation model. *Journal of the Atmospheric Sciences*, **67 (1)**, 136–156.

Riehl, H., and J. S. Malkus, 1958: On the heat balance in the equatorial trough zone. *Geophysica*, **6**, 537.

Robinson, T. D., and D. C. Catling, 2014: Common 0.1 bar tropopause in thick atmospheres set by pressure-dependent infrared transparency. *Nature Geoscience*, **7 (1)**, 12–15.

Roemmich, D., and Coauthors, 2009: The argo program: Observing the global ocean with profiling floats. *Oceanography*, **22 (2)**, 34–43.

Rogers, J. C., and H. van Loon, 1982: Spatial variability of sea level pressure and 500 mb height anomalies over the southern hemisphere. *Monthly Weather Review*, **110 (10)**, 1375–1392.

Rosen, R. D., D. A. Salstein, T. M. Eubanks, J. O. Dickey, and J. A. Steppe, 1984: An El Niño signal in atmospheric angular momentum and Earth rotation. *Science*, **225 (4660)**, 411–414.

Rossby, C.-G., 1939: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. *J. Marine Res.*, **2**, 38–55.

Rossby, C.-G., 1941: The scientific basis of modern meteorology. *Climate and Man. Yearbook of Agriculture*, U.S. Government Printing Office, 599–655.

Rossby, C.-G., 1947: On the distribution of angular velocity in gaseous envelopes under the influence of large-scale horizontal mixing processes. *Bulletin of the American Meteorological Society*, **28 (2)**, 53–68.

Rossby, C.-G., and Collaborators, 1937: Isentropic analysis. *Bulletin of the american meteorological society*, **18 (6-7)**, 201–209.

Rueda, V. O. M., 1991: *Tropical-extratropical atmospheric interactions*. University of California, Los Angeles.

Sadourny, R., and C. Basdevant, 1985: Parameterization of subgrid scale barotropic and baroclinic eddies in quasi-geostrophic models: Anticipated potential vorticity method. *Journal of Atmospheric Sciences*, **42 (13)**, 1353–1363.

Salby, M. L., and H. H. Hendon, 1994: Intraseasonal behavior of clouds, temperature, and motion in the tropics. *Journal of Atmospheric Sciences*, **51 (15)**, 2207–2224.

Savijärvi, H., 1988: Global energy and moisture budgets from rawinsonde data. *Monthly Weather Review*, **116 (2)**, 417–430.

Scherhag, R., 1952: Die explosionsartige stratosphaerenerwarmung des spaetwinters 1951/1952. *Ber. Deut. Wetterdienstes*, **6**, 51.

Scherhag, R., 1960: Stratospheric temperature changes and the associated changes in pressure distribution. *Journal of Atmospheric Sciences*, **17 (6)**, 575–583.

Schneider, T., 2006: The general circulation of the atmosphere. *Annual Review of Earth and Planetary Sciences*, **34**, 655–688.

Schoeberl, M. R., 1978: Stratospheric warmings: Observations and theory. *Reviews of Geophysics*, **16 (4)**, 521–538.

Schubert, W., 2004: A generalization of ertel's potential vorticity to a cloudy, precipitating atmosphere. *Meteorologische Zeitschrift (Berlin)*, **13**.

Schubert, W. H., and M. T. Masarik, 2006: Potential vorticity aspects of the mjo. *Dynamics of atmospheres and oceans*, **42 (1-4)**, 127–151.

Schulman, L. L., 1973: On the summer hemisphere Hadley cell. *Quarterly Journal of the Royal Meteorological Society*, **99 (419)**, 197–201.

Sclater, J., C. Jaupart, and D. Galson, 1980: The heat flow through oceanic and continental crust and the heat loss of the Earth. *Reviews of Geophysics*, **18 (1)**, 269–311.

Sellers, P., and Coauthors, 1997: Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science*, **275 (5299)**, 502–509.

Shangguan, M., and W. Wang, 2022: The semi-annual oscillation (SAO) in the upper troposphere and lower stratosphere (UTLS). *Atmospheric Chemistry and Physics Discussions*, 1–25.

Shaw, D., 1978: Meteorology over the tropical oceans. *Nature*, **276 (5687)**, 445–446.

Sherwood, S. C., 1999: Convective precursors and predictability in the tropical western pacific. *Monthly Weather Review*, **127 (12)**, 2977–2991.

Shukla, J., and D. Mooley, 1987: Empirical prediction of the summer monsoon rainfall over india. *Monthly Weather Review*, **115 (3)**, 695–704.

Shutts, G., 1986: A case study of eddy forcing during an Atlantic blocking episode. *Advances in Geophysics*, Vol. 29, Elsevier, 135–162.

Siebesma, A. P., P. M. Soares, and J. Teixeira, 2007: A combined eddy-diffusivity mass-flux approach for the convective boundary layer. *Journal of the atmospheric sciences*, **64 (4)**, 1230–1248.

Sikka, D., and S. Gadgil, 1980: On the maximum cloud zone and the itcz over indian longitudes during the southwest monsoon. *Monthly Weather Review*, **108 (11)**, 1840–1853.

Simmons, A., M. Hortal, G. Kelly, A. McNally, A. Untch, and S. Uppala, 2005: Ecmwf analyses and forecasts of stratospheric winter polar vortex breakup: September 2002 in the southern hemisphere and related events. *Journal of the Atmospheric Sciences*, **62 (3)**, 668–689.

Smith, A. K., R. R. Garcia, A. C. Moss, and N. J. Mitchell, 2017: The semiannual oscillation of the tropical zonal wind in the middle atmosphere derived from satellite geopotential height retrievals. *Journal of the Atmospheric Sciences*, **74 (8)**, 2413–2425.

Smith, R. K., 2013: *The physics and parameterization of moist atmospheric convection*, Vol. 505. Springer Science & Business Media.

Soares, P., P. Miranda, A. Siebesma, and J. Teixeira, 2004: An eddy-diffusivity/mass-flux parametrization for dry and shallow cumulus convection. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **130 (604)**, 3365–3383.

Sobel, A., and E. Maloney, 2012: An idealized semi-empirical framework for modeling the madden–julian oscillation. *Journal of the Atmospheric Sciences*, **69 (5)**, 1691–1705.

Sobel, A., and E. Maloney, 2013: Moisture modes and the eastward propagation of the mjo. *Journal of the Atmospheric Sciences*, **70 (1)**, 187–192.

Sobel, A. H., and C. S. Bretherton, 1999: Development of synoptic-scale disturbances over the summertime tropical northwest pacific. *Journal of the Atmospheric Sciences*, **56 (17)**, 3106–3127.

Solberg, M., 1936: Le mouvement d'inertie de l'atmosphere stable et son role dans le theorie des cyclones. *Union Geodesique et Geophisique internationale VIeme assemblee, Edinburg*, 66–82.

Staniforth, A. N., 2022: *Global Atmospheric and Oceanic Modelling: Fundamental Equations*. Cambridge University Press.

Starr, V. P., 1945: A quasi-lagrangian system of hydrodynamical equations. *Journal of the Atmospheric Sciences*, **2 (4)**, 227–237.

Starr, V. P., 1948: An essay on the general circulation of the Earth's atmosphere. *Journal of Atmospheric Sciences*, **5 (2)**, 39–43.

Stephens, G., and D. O'Brien, 1993: Entropy and climate. I: ERBE observations of the entropy production of the Earth. *Quarterly Journal of the Royal Meteorological Society*, **119 (509)**, 121–152.

Stocker, T., 2014: *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge university press.

Stroeve, J. C., T. Markus, L. Boisvert, J. Miller, and A. Barrett, 2014: Changes in arctic melt season and implications for sea ice loss. *Geophysical research letters*, **41 (4)**, 1216–1225.

Stull, R. B., 2012: *An introduction to boundary layer meteorology*, Vol. 13. Springer Science & Business Media.

Sugiyama, M., 2009: The moisture mode in the quasi-equilibrium tropical circulation model. part i: Analysis based on the weak temperature gradient approximation. *Journal of the Atmospheric Sciences*, **66 (6)**, 1507–1523.

Sussman, G. J., and J. Wisdom, 1992: Chaotic evolution of the solar system. *Science*, **257 (5066)**, 56–62.

Takayabu, Y. N., and T. Nitta, 1993: 3-5 day-period disturbances coupled with convection over the tropical pacific ocean. *Journal of the Meteorological Society of Japan. Ser. II*, **71 (2)**, 221–246.

Tan, Z., C. M. Kaul, K. G. Pressel, Y. Cohen, T. Schneider, and J. Teixeira, 2018: An extended eddy-diffusivity mass-flux scheme for unified representation of subgrid-scale turbulence and convection. *Journal of advances in modeling earth systems*, **10 (3)**, 770–800.

Tapley, B. D., S. Bettadpur, M. Watkins, and C. Reigber, 2004: The gravity recovery and climate experiment: Mission overview and early results. *Geophysical research letters*, **31 (9)**.

Taylor, G. I., 1950a: The formation of a blast wave by a very intense explosion i. theoretical discussion. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **201 (1065)**, 159–174.

Taylor, G. I., 1950b: The formation of a blast wave by a very intense explosion.-ii. the atomic explosion of 1945. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **201 (1065)**, 175–186.

Teixeira, J., K. Suselj, and M. J. Kurowski, 2023: An introduction to the eddy–diffusivity/mass–flux (edmf) approach: A unified turbulence and convection parameterization. *Fast Processes in Large-Scale Atmospheric Models: Progress, Challenges, and Opportunities*, 271–282.

Thayer-Calder, K., and D. A. Randall, 2009: The role of convective moistening in the madden–julian oscillation. *Journal of the Atmospheric Sciences*, **66 (11)**, 3297–3312.

Thompson, D. W., M. P. Baldwin, and S. Solomon, 2005: Stratosphere–troposphere coupling in the southern hemisphere. *Journal of the atmospheric sciences*, **62 (3)**, 708–715.

Thompson, D. W., and E. A. Barnes, 2014: Periodic variability in the large-scale southern hemisphere atmospheric circulation. *Science*, **343 (6171)**, 641–645.

Thompson, D. W., and J. M. Wallace, 1998: The arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical research letters*, **25 (9)**, 1297–1300.

Thompson, D. W., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. part i: Month-to-month variability. *Journal of climate*, **13 (5)**, 1000–1016.

Thompson, D. W., and J. D. Woodworth, 2014: Barotropic and baroclinic annular variability in the southern hemisphere. *Journal of the Atmospheric Sciences*, **71 (4)**, 1480–1493.

Tillman, J. E., N. C. Johnson, P. Guttorp, and D. B. Percival, 1993: The martian annual atmospheric pressure cycle: Years without great dust storms. *Journal of Geophysical Research*, **98**, 10 963–10 971.

Tomas, R. A., and P. J. Webster, 1997: The role of inertial instability in determining the location and strength of near-equatorial convection. *Quarterly Journal of the Royal Meteorological Society*, **123 (542)**, 1445–1482.

Townsend, R. D., and D. R. Johnson, 1985: A diagnostic study of the isentropic zonally averaged mass circulation during the first garp global experiment. *Journal of Atmospheric Sciences*, **42 (15)**, 1565–1579.

Trenberth, K. E., and J. M. Caron, 2001: Estimates of meridional atmosphere and ocean heat transports. *Journal of Climate*, **14 (16)**, 3433–3443.

Trenberth, K. E., J. R. Christy, and J. G. Olson, 1987: Global atmospheric mass, surface pressure, and water vapor variations. *Journal of Geophysical Research: Atmospheres*, **92 (D12)**, 14 815–14 826.

Trenberth, K. E., J. T. Fasullo, and M. A. Balmaseda, 2014: Earth's energy imbalance. *Journal of Climate*, **27 (9)**, 3129–3144.

Trenberth, K. E., J. T. Fasullo, and J. Kiehl, 2009: Earth's global energy budget. *Bulletin of the American Meteorological Society*, **90 (3)**, 311–324.

Tung, K. K., 1986: Nongeostrophic theory of zonally averaged circulation. part i: Formulation. *Journal of Atmospheric Sciences*, **43 (22)**, 2600–2618.

Uppala, S. M., and Coauthors, 2005: The era-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **131 (612)**, 2961–3012.

Vallis, G. K., 2017: *Atmospheric and oceanic fluid dynamics*. Cambridge University Press.

VonderHaar, T. H., and A. H. Oort, 1973: New estimate of annual poleward energy transport by northern hemisphere oceans. *Journal of Physical Oceanography*, **3 (2)**, 169–172.

Wallace, J., 1983: The climatological mean stationary waves: Observational evidence. large-scale dynamical processes in the atmosphere, bj hoskins and rp pearce, eds. Academic Press.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly weather review*, **109 (4)**, 784–812.

Wallace, J. M., and V. Kousky, 1968: Observational evidence of kelvin waves in the tropical stratosphere. *J. atmos. Sci*, **25 (5)**, 900–907.

Wang, B., X. Li, and L. Wu, 1997: Direction of hurricane beta drift in horizontally sheared flows. *Journal of the atmospheric sciences*, **54 (11)**, 1462–1471.

Wang, J., and D. A. Randall, 1994: The moist available energy of a conditionally unstable atmosphere. part ii: Further analysis of gate data. *Journal of Atmospheric Sciences*, **51 (5)**, 703–710.

Webster, P., 1983: Mechanisms of low-frequency monsoon variability. *J. Atmos. Sci*, **40**, 2110–2124.

Webster, P., 1987: The variable and interactive monsoon in monsoons edited by js fein and pl stephens, 269–330, wilwy-interscience publucation. John Wiley and Sons, New York.

Webster, P. J., 1972: Response of the tropical atmosphere to local, steady forcing. *Monthly Weather Review*, **100 (7)**, 518–541.

Webster, P. J., 1994: The role of hydrological processes in ocean-atmosphere interactions. *Reviews of Geophysics*, **32 (4)**, 427–476.

Webster, P. J., V. O. Magana, T. Palmer, J. Shukla, R. Tomas, M. Yanai, and T. Yasunari, 1998: Monsoons: Processes, predictability, and the prospects for prediction. *Journal of Geophysical Research: Oceans*, **103 (C7)**, 14 451–14 510.

Wedi, N. P., and P. K. Smolarkiewicz, 2006: Direct numerical simulation of the plumb–mcewan laboratory analog of the QBO. *Journal of the atmospheric sciences*, **63 (12)**, 3226–3252.

Weickmann, K., and S. Khalsa, 1990: The shift of convection from the indian ocean to the western pacific ocean during a 30–60 day oscillation. *Monthly Weather Review*, **118 (4)**, 964–978.

Wheeler, M., and G. N. Kiladis, 1999: Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber–frequency domain. *Journal of Atmospheric Sciences*, **56 (3)**, 374–399.

White, A. A., and R. Bromley, 1995: Dynamically consistent, quasi-hydrostatic equations for global models with a complete representation of the coriolis force. *Quarterly Journal of the Royal Meteorological Society*, **121 (522)**, 399–418.

Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee III, G. L. Smith, and J. E. Cooper, 1996: Clouds and the Earth's radiant energy system (CERES): An Earth observing system experiment. *Bulletin of the American Meteorological Society*, **77 (5)**, 853–868.

Wielicki, B. A., and Coauthors, 1998: Clouds and the Earth's radiant energy system (CERES): Algorithm overview. *IEEE Transactions on Geoscience and Remote Sensing*, **36 (4)**, 1127–1141.

Wiin-Nielsen, A., J. A. Brown, and M. Drake, 1963: On atmospheric energy conversions between the zonal flow and the eddies. *Tellus*, **15 (3)**, 261–279.

Wilczek, F., 2005: On absolute units, i: Choices. *Physics Today*, **58 (10)**, 12–13.

Wilczek, F., 2006a: On absolute units, ii: Challenges and responses. *Physics Today*, **59 (1)**, 10–11.

Wilczek, F., 2006b: On absolute units, iii: Absolutely not? *Physics Today*, **59 (5)**, 10–11.

Williams, P. D., and Coauthors, 2017: A census of atmospheric variability from seconds to decades. *Geophysical Research Letters*, **44 (21)**, 11–201.

Willoughby, H., 1998: Tropical cyclone eye thermodynamics. *Monthly weather review*, **126 (12)**, 3053–3067.

Wing, A. A., and Coauthors, 2020: Clouds and convective self-aggregation in a multi-model ensemble of radiative-convective equilibrium simulations. *Journal of Advances in Modeling Earth Systems*, **12**.

Winker, D. M., R. H. Couch, and M. McCormick, 1996: An overview of lite: Nasa's lidar in-space technology experiment. *Proceedings of the IEEE*, **84 (2)**, 164–180.

Witek, M. L., J. Teixeira, and G. Matheou, 2011: An integrated tke-based eddy diffusivity/mass flux boundary layer closure for the dry convective boundary layer. *Journal of the Atmospheric Sciences*, **68 (7)**, 1526–1540.

Woollings, T., and Coauthors, 2018: Blocking and its response to climate change. *Current Climate Change Reports*, **4 (3)**, 287–300.

Xie, Y.-B., S.-J. Chen, I.-L. Chang, and Y.-L. Huang, 2018: A preliminarily statistic and synoptic study about the basic currents over southeastern asia and the initiation of typhoons. *Journal of Meteorological Research*, **32**, 661–669.

Xu, K.-M., and A. Arakawa, 1992: Semiprognostic tests of the arakawa-schubert cumulus parameterization using simulated data. *Journal of the atmospheric sciences*, **49 (24)**, 2421–2436.

Xu, K.-M., and K. A. Emanuel, 1989: Is the tropical atmosphere conditionally unstable? *Monthly Weather Review*, **117 (7)**, 1471–1479.

Yanai, M., S. Esbensen, and J.-H. Chu, 1973: Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *J. Atmos. Sci.*, **30**, 611–627.

Yanai, M., and C. Li, 1994: Mechanism of heating and the boundary layer over the tibetan plateau. *Monthly Weather Review*, **122 (2)**, 305–323.

Yanai, M., C. Li, and Z. Song, 1992: Seasonal heating of the tibetan plateau and its effects on the evolution of the asian summer monsoon. *Journal of the Meteorological Society of Japan. Ser. II*, **70 (1B)**, 319–351.

Yasunari, T., 1979: Cloudiness fluctuations associated with the northern hemisphere summer monsoon. *Journal of the Meteorological Society of Japan. Ser. II*, **57 (3)**, 227–242.

Yu, J.-Y., and J. D. Neelin, 1994: Modes of tropical variability under convective adjustment and the madden–julian oscillation. part ii: Numerical results. *Journal of Atmospheric Sciences*, **51 (13)**, 1895–1914.

Zeitlin, V., 2018: Symmetric instability drastically changes upon inclusion of the full coriolis force. *Physics of Fluids*, **30 (6)**, 061 701.

Zent, A. P., 1996: The evolution of the martian climate. *American scientist*, **84 (5)**, 442–451.

Zhang, Z., W. Wang, and B. Qiu, 2014: Oceanic mass transport by mesoscale eddies. *Science*, **345 (6194)**, 322–324.