An Introduction To
The General Circulation of the Atmosphere

David A. Randall
Department of Atmospheric Science
Colorado State University

Spring 2005
About the Cover

This image (actually pieced together from multiple images) was acquired on December 8, 1992, by NASA's Galileo spacecraft, as it swung by Earth on its way out to Jupiter. It shows a wonderful but physically impossible view of the Southern Hemisphere. A substantial fraction of the image should be in darkness, even though the image depicts a time near the summer solstice of the Southern Hemisphere. This view was created by patching together a mosaic of several images taken by Galileo over a 24-hour period, and remapping them as they would be seen from above the pole. South America, Africa, and Australia are respectively seen at the middle left, upper right, and lower right.

Of particular interest are the beautiful cloud patterns associated with extratropical cyclones in the storm track ringing Antarctica. This picture is reminiscent of photos of “dishpan” experiments, in which aspects of the general circulation are simulated in a rotating, differentially heated laboratory tank.
### Announcements

<table>
<thead>
<tr>
<th>Subject:</th>
<th>The general circulation of the atmosphere</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text:</td>
<td>Class notes</td>
</tr>
<tr>
<td>Course grade:</td>
<td>Based entirely on homework, including some projects</td>
</tr>
<tr>
<td>Schedule:</td>
<td>Classes will be missed occasionally. Make-ups will be scheduled. A calendar showing class meetings will be distributed early in the semester, and past experience suggests that not many changes will be made.</td>
</tr>
</tbody>
</table>
### Handy Numbers

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radius of the Earth</td>
<td>$6.37 \times 10^6$ m</td>
</tr>
<tr>
<td>Angular velocity of the Earth’s rotation</td>
<td>$7.29 \times 10^{-5}$ s$^{-1}$</td>
</tr>
<tr>
<td>Latent heat of condensation at 0 °C</td>
<td>$2.52 \times 10^6$ J kg$^{-1}$</td>
</tr>
<tr>
<td>Globally averaged surface air temperature</td>
<td>288 K</td>
</tr>
<tr>
<td>Globally averaged precipitable water</td>
<td>25 mm (= 25 kg m$^{-2}$)</td>
</tr>
<tr>
<td>Annual mean incident solar radiation</td>
<td>340 W m$^{-2}$</td>
</tr>
<tr>
<td>Global albedo</td>
<td>0.30</td>
</tr>
<tr>
<td>Outgoing longwave radiation</td>
<td>240 W m$^{-2}$</td>
</tr>
<tr>
<td>Stefan-Boltzman constant</td>
<td>$5.67 \times 10^8$ W m$^{-2}$ K$^{-4}$</td>
</tr>
<tr>
<td>Globally averaged precipitation rate</td>
<td>3 mm day$^{-1}$</td>
</tr>
<tr>
<td>$c_p$ for dry air</td>
<td>1000 J kg$^{-1}$ K$^{-1}$</td>
</tr>
<tr>
<td>$R$ for dry air</td>
<td>287 J kg$^{-1}$ K$^{-1}$</td>
</tr>
<tr>
<td>Globally averaged surface pressure</td>
<td>984 mb</td>
</tr>
<tr>
<td>Acceleration of Earth's gravity</td>
<td>9.81 m s$^{-2}$</td>
</tr>
<tr>
<td>Density of air near sea level</td>
<td>1.2 kg m$^{-3}$</td>
</tr>
<tr>
<td>Molecular viscosity of air</td>
<td>$1.5 \times 10^{-5}$ m$^2$ s$^{-1}$</td>
</tr>
</tbody>
</table>
General References
(on reserve in ATS Library)


Preface

This is an introductory course on the general circulation of the atmosphere, a subject that is closely tied to atmospheric dynamics. A course on dynamics tends to focus on basic physical concepts and methods for their analysis, however, while a course on the general circulation must focus on what the atmosphere actually does, and why. Graduate-level studies in atmospheric dynamics are essential as preparation for this course, and most students will learn some additional dynamics in the process of taking this course.

It is difficult to draw a line between the general circulation and climate. The two subjects appear to be growing closer together, as the roles of heating and dissipation in the general circulation emerge as key issues. Clearly, such topics as monsoons, the hydrologic cycle, and the planetary energy budget can be included under either “climate” or “general circulation,” although perhaps with different slants. Climate is the bigger subject. This course skirts the edges of physical climatology.

There is way too much material to cover in one semester, so I have had to make choices. This particular course emphasizes the role of moisture in the workings of the general circulation, and makes frequent use of models as tools to aid understanding. I have also discussed energetics in some detail, and have included a chapter on the general circulation as turbulence, including an extended section on the chaotic nature of the circulation.

This is a graduate-level course, and it is an elective. The level of difficulty is set so as to maximize the potential benefit to the strongest students in the class; if they work very hard, they should just barely be able to master the material. It will be worth the effort. This is a beautiful subject.

A lot of work is involved in putting together these course notes. For the past several years, Michelle Beckman has very professionally made many additions and corrections to the notes. She also developed and applied the formatting that you see, and combined the many separate documents into a unified “book.” The notes would be much less useful without her contributions.

Mark Branson, Don Dazlich, Kelley Wittmeyer, and Mike Kelly have ably assisted in the production of some of the figures. Mike Kelly, Cara-Lyn Lappen Katherine Harris, Stefan Tulich, Anning Cheng, Mike Toy, Kyle Wiens, and
Cristiana Stan, and Jason Furtado performed beautifully as Teaching Assistants for the course, and both the students and I learned as a result of their efforts.

Finally, I am grateful to the students who have taken this course over the years; their questions and suggestions have led to major improvements.

David Randall
January 19, 2005 10:49 am
An Introduction to the General Circulation of the Atmosphere

CHAPTER 1 Introduction 1

The nature of the subject ................................................................. 1
A brief overview ............................................................................. 3
Fasten your seatbelts .................................................................... 8

CHAPTER 2 What makes it go? 9

The Earth’s radiation budget: An “upper boundary condition” on the general circulation ................................................. 9
Surface boundary conditions ................................................................. 6
Energy and moisture budgets of the surface and atmosphere .................. 21
Summary ......................................................................................... 27
Problems ......................................................................................... 28

CHAPTER 3 An overview of the observations 29

Introduction ....................................................................................... 29
The global distribution of atmospheric mass ...................................... 32
Zonal wind ......................................................................................... 42
Meridional wind ............................................................................... 46
Geopotential height .......................................................................... 50
Vertical velocity and the mean meridional circulation. ......................... 54
Angular momentum .......................................................................... 63
Temperature ...................................................................................... 69
A view in potential temperature coordinates ..................................... 74
The global distribution of water vapor ................................................. 80
Precipitation ...................................................................................... 85
Surface fluxes due to turbulence ......................................................... 89
A quick and superficial introduction to the effects of large-scale eddies on the zonally averaged flow ............................................. 96
A preliminary interpretation of the observations ................................. 102
Lots of questions .............................................................................. 108
Problems ......................................................................................... 109
# CHAPTER 4  Conservation of momentum and energy  

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
</tr>
</tbody>
</table>

- Introduction ................................................................. 111
- Conservation of momentum on a rotating sphere .................. 111
- Conservation of kinetic energy and potential energy .......... 116
- Conservation of thermodynamic energy .............................. 120
- Conservation of total energy ............................................. 122
- Static energies ................................................................. 126
- Entropy ................................................................................. 127
- Approximations ..................................................................... 131
- The mechanical energy equation in other vertical coordinate systems ........ 132
- The effects of turbulence .................................................. 133
- Summary ................................................................................. 135
- Problems ................................................................................ 136

# CHAPTER 5  The mean meridional circulation  

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>145</td>
</tr>
</tbody>
</table>

- The observed meridional transports of energy and moisture .......... 145
- A simple theory of the Hadley circulation ................................ 151
- Extension to other planetary atmospheres .................................. 160
- Particle trajectories on the sphere: A partial explanation of “bandedness” .......... 163
- Summary ................................................................................. 169
- Problems ................................................................................ 169

# CHAPTER 6  An overview of the effects of radiation and convection  

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>171</td>
</tr>
</tbody>
</table>

- Convective energy transports ............................................. 171
- Radiative-convective equilibrium .......................................... 173
- The observed vertical structure of the atmosphere, and the mechanisms  
  of vertical energy transport ....................................................... 180
- More on moist convection .................................................... 188
- Summary ................................................................................. 203
- Problems ................................................................................ 203

# CHAPTER 7  The Energy Cycle  

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>207</td>
</tr>
</tbody>
</table>

- Available potential energy .................................................. 207
- The gross static stability ....................................................... 213
Examples: The available potential energies of three simple systems ..........218
The APE associated with static instability ..................................................218
The APE associated with meridional temperature gradients ..................218
The APE associated with surface pressure variations ............................223
Variance budgets ......................................................................................224
Generation of available potential energy, and its conversion into kinetic energy ..230
The governing equations for the eddy kinetic energy, zonal kinetic energy,
and total kinetic energy .............................................................................235
Observations of the energy cycle .................................................................238
The role of heating .....................................................................................241
Summary ..................................................................................................242
Problems ..................................................................................................242

CHAPTER 8

Planetary-scale waves and other eddies

Introduction .................................................................................................247
Free and forced small-amplitude oscillations of a thin spherical atmosphere ....249
Perturbation equations ..................................................................................249
Free oscillations of the first and second kinds ............................................253
Observations of stationary and transient eddies in middle latitudes ..........259
Theory of orographically forced stationary waves ........................................266
Tropical waves ..........................................................................................272
The response of the tropical atmosphere to stationary heat sources and sinks .....283
Monsoons .................................................................................................288
The Walker Circulation ............................................................................297
The Madden-Julian Oscillation ..................................................................306
Summary ..................................................................................................311
Problems ..................................................................................................312

CHAPTER 9

Wave–Mean Flow Interactions

Interactions and non-interactions of gravity waves with the mean flow ........317
Vertical propagation of planetary waves .....................................................320
Vertical and meridional fluxes due to planetary waves ............................328
Sudden warmings .....................................................................................335
Eliassen-Palm Theorem-Reprise ...............................................................336
The Eliassen-Palm theorem in isentropic coordinates ................................346
Potential vorticity fluxes ...........................................................................353
The quasi-biennial oscillation ..................................................................355
Blocking ..................................................................................................358
Summary ..................................................................................................364
Problems ..................................................................................................364
CHAPTER 10  The general circulation as turbulence  365

Energy and enstrophy cascades ................................................................. 365
Nonlinearity and scale interactions ........................................................... 369
Two-dimensional turbulence ..................................................................... 370
Quasi-two-dimensional turbulence ............................................................. 372
Dimensional analysis of the kinetic energy spectrum ................................. 375
Observations of the kinetic energy spectrum ............................................... 380
The general circulation as a blender .......................................................... 381
  What does the blender blend? ................................................................. 381
  Dissipating enstrophy but not kinetic energy ......................................... 383
  The Gent-McWilliams theory of tracer transports along isentropic surfaces 384
The limits of deterministic weather prediction ............................................. 387
Quantifying the limits of predictability ....................................................... 393
  The dynamical approach ..................................................................... 393
  The empirical approach .................................................................... 394
  The dynamical-empirical approach ....................................................... 395
Climate prediction ................................................................................... 402
The World’s Simplest GCM ...................................................................... 405
Pushing the attractors around .................................................................. 410
Summary ................................................................................................. 412
Problems ................................................................................................. 413

CHAPTER 11  Tropical atmosphere–ocean interactions  417

Introduction ............................................................................................ 417
The Walker Circulation ............................................................................ 418
The relationship between the Walker Circulation
  and the sea surface temperature ............................................................ 424
Theories of the Walker Circulation ........................................................... 426
El Niño and the Southern Oscillation ....................................................... 432
  Sea surface temperature and thermocline slope .................................... 435
Summary ................................................................................................. 441
References and Bibliography .................................................................... 463
CHAPTER 1 | Introduction

Copyright 2005 David A. Randall

1.1 The nature of the subject

The atmosphere circulates. This circulation is global in extent. It is constantly fighting against friction, but is sustained by thermal forcing, which ultimately comes from the Sun.

The Sun provides on the average about 340 W m⁻² of incident energy, of which a tiny fraction is converted into the kinetic energy of the general circulation. Additional, “primordial” energy leaks out of the Earth’s interior, but this rate is only about 0.08 W m⁻² (Sclater et al., 1980; Bukowinski, 1999).

The externally imposed thermal forcing is strongly influenced by the atmospheric circulation itself, e.g., as clouds form and disappear. This coupling, between the circulation and the heating that drives it, is a major complication that makes the general circulation much more interesting than it would be otherwise.

The conservation equations that govern the behavior of the atmosphere can be used to formulate balance requirements that the general circulation must satisfy. In a time average, the net energy flux at the top of the atmosphere must vanish, the rates of evaporation and precipitation must balance, and the total angular momentum of the atmosphere-ocean-solid Earth system must be invariant, apart from gravitational interactions with the Moon and other extraterrestrial bodies. This check-book approach to the general circulation emphasizes the sources, sinks, and transports of energy, moisture, and angular momentum. We will discuss the general circulation from this classical perspective.

It is important to 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. One purpose of this course is to introduce and analyze these and other phenomena of the atmospheric general circulation.

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

The general circulation of the Earth’s atmosphere and the atmospheric component
An Introduction to the General Circulation of the Atmosphere

Introduction

An Introduction to the General Circulation of the Atmosphere

Much general circulation research today overlaps strongly with the study of global-scale air-sea interactions, including the now universally recognized phenomenon of El Niño, as well as a number of other processes that generally manifest themselves on time scales of years to decades. We therefore briefly discuss selected aspects of the ocean circulation, including the physics of both the upper ocean and the deep thermohaline circulation.

In addition, we occasionally compare the general circulation of the Earth’s...
atmosphere with those of other planets in our solar system. Such comparisons are becoming increasingly useful as our knowledge and understanding of the solar system rapidly expands; they serve to emphasize not only certain similarities among the planetary circulations, but also the numerous ways in which the circulation of the Earth’s atmosphere is, in our experience to date, unique.

1.2 A brief overview

Here is a qualitative, highly simplified overview of the general circulation, just to give you a feeling for the lay of the land.

It is conventional and useful, although somewhat arbitrary, to divide the atmosphere into parts. For purposes of this quick sketch of the general circulation, we will divide the atmosphere up vertically and meridionally, only briefly mentioning the longitudinally varying structures.

Starting at the bottom, the layer of air in direct contact with the Earth’s surface is called the “planetary boundary layer,” or PBL. The air in the PBL is turbulent, and in ways that we will discuss later the turbulence produces rapid exchanges of “sensible heat” (essentially temperature), moisture, and momentum between the atmosphere and the surface. The most important exchanges are of moisture, upward into the atmosphere via

Figure 1.2: This figure shows lidar backscatter from aerosols and clouds. The figure has been created using data from LITE, a lidar that flew on the space shuttle in 1994. The lidar cannot penetrate through thick clouds, which explains the vertical black stripes in the figure. The reddish layer just above the Earth’s surface is the PBL, which is visible because the aerosol concentration decreases sharply upward at the PBL top.
evaporation from the surface, and of momentum, via friction. The surface moisture flux is a key energy input to the general circulation, and surface friction is the primary mechanism that dissipates the kinetic energy of the general circulation. It is only a slight exaggeration to say that the sources and sinks of energy for the general circulation are at the Earth’s surface, at the base of the PBL. The depth of the PBL varies dramatically in space and time, but a ball-park value to remember is 1 km.

Above the PBL is the “free troposphere.” The troposphere actually includes the PBL, so it is conventional to say “free” troposphere to distinguish the part of the troposphere that is not in the PBL. The free troposphere is characterized by positive but modest static stability, i.e., the potential temperature increases upward at a moderate rate. The depth of the troposphere varies strongly with latitude and season.

The troposphere is surmounted by the stratosphere; the boundary between the two is called the tropopause. The height of the tropopause varies from 15 km or so in the tropics, to 8 km or so near the poles. The stratosphere has a much higher static stability than the troposphere, due to radiative heating associated with the absorption of solar ultraviolet radiation, primarily in the upper and middle stratosphere. The lower stratosphere has a nearly isothermal stratification; higher in the stratosphere the temperature actually increases upward. The summer-hemisphere stratosphere is very quiet dynamically, and is filled with easterlies, with warm air over the pole. The winter-hemisphere stratosphere is much more active, with very cold air over the pole.

Even though stratosphere is very dry, its moisture budget is quite interesting.

The top of the stratosphere, i.e., the stratopause, is about 65 km above the Earth’s surface. Regions above the stratopause will not be discussed in this course.

For meteorological purposes, the tropics is the region from about 15° S to 15° N. The tropical troposphere undergoes deep moist convection. What this means is that 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 nearer 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 falling rain inject free-tropospheric air into the PBL. In the convectively active parts of the tropics, the air is slowly rising in an average sense, and this average rising motion is closely but subtly related to the strong but very localized updrafts of the convective clouds.

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 that follow from the smallness of the Coriolis parameter in the tropics. The moisture and wind fields are more active, however. The tropics is home to a variety of distinctive traveling waves and vortices, which organize the

1. Although the general circulation is solar-powered, the solar radiation is primarily absorbed at the Earth’s surface, and so the Sun’s energy is provided to the atmosphere indirectly, largely in the form of latent heat, i.e., water vapor. This is discussed in detail later.
convective clouds on scales of hundreds to thousands of kilometers. Finally, the tropics
has 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 tropics is home to El Niño, La Niña, and the Southern Oscillation, or “ENSO,”
which is a dramatic, quasi-regular oscillation of the ocean-atmosphere system, with a
period of a few years. In an El Niño, the sea-surface temperatures warm in the eastern
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 physical processes that give rise to ENSO involve both the ocean
and the atmosphere, but its dramatic effects on the atmosphere must be included in a
discussion of the atmospheric general circulation.

The tropical stratosphere is home to an amazing periodic reversal of the zonal
wind, with a period slightly longer than two years, called the Quasi-Biennial Oscillation,
or QBO.

The subtropics in each hemisphere is roughly the region between 15° 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 subtropics is home to the major
deserts of the world. 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.
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.

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 that have scales of a few thousand km, and that grow through baroclinic instability. These baroclinic eddies transport energy and moisture poleward and upward, primarily in the winter hemisphere, but also to some extent in the summer hemisphere. They transport westerly momentum meridionally, and also downward, where it is consumed by surface friction in the PBL. The baroclinicity that supports the baroclinic eddies manifests itself in strong midlatitude temperature and surface pressure variations; this is dynamically possible only where the Coriolis parameter is sufficiently large, i.e., outside the tropics. The baroclinic

Figure 1.5: A beautiful baroclinic eddy over the North Atlantic Ocean in winter.
waves produce massive cloud systems and strong precipitation. The midlatitude atmosphere also contains strong “stationary waves” associated with both mountains and land-sea contrast.

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 and often blanketed by extensive cloudiness, while the South Pole is on a mountainous continent that is largely a desert. The polar regions are home to prominent “annular modes,” which fluctuate on a variety of time scales, almost uniformly in longitude. The annular modes are seen in both the stratosphere and troposphere, and are now known to be comparable in importance to ENSO, in terms of their contributions to the overall variability of the general circulation. The polar stratosphere in winter is occasionally disturbed by “Sudden Warmings,” which are dramatic changes in temperature (and wind) that are particularly frequent in the Northern Hemisphere.

As discussed later, 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 overall flow of energy in the atmosphere is upward into the atmosphere from the surface in the tropics and subtropics, then meridionally and further upward by the Hadley Circulation in the tropics and the baroclinic eddies in middle latitudes, and finally outward to space via infrared radiation at all latitudes, but especially in the subtropics. In the tropics, the atmosphere acquires angular momentum from the continents and oceans, through surface friction. The atmospheric general circulation carries the angular momentum to higher latitudes, where surface friction “puts it back”
into the continents and oceans.

1.3 Fasten your seatbelts

The study of the thermally driven general circulation of the atmosphere naturally brings together concepts from all areas of atmospheric science. We will be discussing general-circulation phenomena that involve large-scale dynamics, convection, turbulence, cloud processes, and radiative transfer. Most of all, we will be discussing the interactions among these processes. This is good, because often these various topics are presented as if they were somehow neatly separated from each other.

Because the general circulation is global and spans all seasons, the concepts of atmospheric science must be applied across a 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 over the tropical jungles. In courses on boundary-layer meteorology, however, the discussion is typically limited to relatively simple horizontally uniform conditions, such as might be encountered on a summer’s day in Kansas. In the study of the general circulation, you will confront a much broader range of conditions.

For these reasons, when we study the general circulation, we quickly run up against the limits of our understanding in all aspects of atmospheric science, and so we are led to push those limits outward. This makes the study of the general circulation a particularly challenging and exciting field -- as you are about to see for yourselves.
CHAPTER 2  What makes it go?

Copyright 2005 David A. Randall

2.1 The Earth’s radiation budget: An “upper boundary condition” on the general circulation

Radiation is (almost) the only mechanism by which the Earth can exchange energy with the rest of the Universe \(^1\). The solar energy flux at the mean radius of the Earth's orbit is about 1370 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 1370 W m\(^{-2}\) is the same as 1.37 GW km\(^{-2}\), i.e., equivalent to 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 Earth’s albedo is near 0.30, independent of season; this number has been known to better than 10% accuracy only since the 1970s. The energy absorbed by the Earth is

Table 2.1: Summary of the annually averaged top-of-the-atmosphere radiation budget.

<table>
<thead>
<tr>
<th>Incident solar radiation</th>
<th>340 W m(^{-2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absorbed solar radiation</td>
<td>240 W m(^{-2})</td>
</tr>
<tr>
<td>Planetary albedo</td>
<td>0.30</td>
</tr>
<tr>
<td>Outgoing longwave radiation</td>
<td>240 W m(^{-2})</td>
</tr>
<tr>
<td>Brightness temperature of the Earth</td>
<td>255 K</td>
</tr>
</tbody>
</table>

\(^1\) Actually, there are two other mechanisms, namely gravitational interactions between the Earth and other objects (i.e., tidal effects), and the capture of extraterrestrial material by the Earth.
Here $S_{abs}$ is the average absorbed solar energy per unit area. $S$ is the “solar constant,” $a$ is the radius of the Earth, and $\alpha$ is the planetary albedo. Note that this rate of energy absorption is referred to the total surface area of the Earth, i.e. \(4\pi a^2\), rather than to the cross-sectional or “projected” area presented to the solar beam, which is four times smaller.

The most important upper boundary condition on the general circulation of the atmosphere is the incident solar radiation. It is basically determined by the Earth’s orbital parameters, (see Fig. 2.1) including the obliquity, eccentricity, and the dates of the equinoxes, as well as the Earth’s rotation rate. These all vary over geologic time (e.g. Crowley and North, 1991.

As a matter of common experience, the insolation varies both diurnally and seasonally. At a given moment, the insolation 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 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. In addition, however, the length of day increases at higher latitudes, and this obviously tends to make the daily-mean insolation increase.
Near the poles, the length-of-day effect dominates, so that the insolation begins to increase again. That is why there is a minimum of the insolation in the mid-latitudes of the summer hemisphere.

As discussed later, 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 (“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 (“polar day”). As is well known, these effects arise from the sun - Earth geometry. In addition, the distance from the sun to the Earth varies with time of year, resulting in a few percent more globally averaged insolation in January than in July. The month of maximum insolation varies over geologic time. According to the astronomical theory of the ice ages, extensive glaciation is favored when the minimum insolation occurs during the Northern Hemisphere winter, because the Northern Hemisphere contains about

---

*Figure 2.2: The seasonal variation of the zonally (or diurnally) averaged insolation at the top the atmosphere. The units are W m\(^{-2}\).*
twice as much land as the Southern Hemisphere (e.g., Crowley and North, 1991).

Of course, the infrared radiation emitted by the Earth is very nearly equal to the solar radiation absorbed by the Earth, i.e. it is about 240 W m$^{-2}$. This near balance between absorbed solar radiation and emitted terrestrial radiation has been directly confirmed by analysis of satellite data. The balance is observed to hold within a few Watts per square meter, which is the within the uncertainty of the measurements.

Fig. 2.3 shows aspects of the Earth's radiation budget as observed in the Earth Radiation Budget Experiment (ERBE; Barkstrom et al., 1989). The zonally averaged incident (i.e. incoming) solar radiation at the top of the atmosphere varies seasonally in response to the Earth's motion around 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 snow and ice as well as cloudiness, but it tends to have a weak secondary maximum in the tropics, again 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 this energy is transported by the atmosphere, and the rest is transported by the oceans.

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 we can regard the general 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 motion field. It is thus a drastic oversimplification to regard these fields as simple forcing functions; they are bound up with the circulation itself.

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

$$\frac{\partial E}{\partial t} = \mathbf{N} \cdot \mathbf{G}.$$  (2.2)

Here $E$ is the energy per unit area stored in the column; $t$ is time; $N$ is the net radiation input at the top of the atmosphere, with dimensions of energy per unit time per unit area (e.g. W m$^{-2}$); and the energy transport, $\mathbf{G}$, is a vector with both latitudinal and longitudinal components, and dimensions of energy per unit length per unit time. Later in the course we will discuss how $\mathbf{G}$ can be computed from measurements, and what the various contributions to $\mathbf{G}$ actually are, but for now we just recognize that it represents the movement of energy due to both the winds and the ocean currents.
2.1 The Earth’s radiation budget: An “upper boundary condition” on the general circulation

Figure 2.3: The zonally averaged incident solar radiation, albedo, outgoing longwave radiation, and net radiation at the top of the atmosphere, as observed in the Earth Radiation Budget Experiment (Barkstrom 1989).
Suppose that we average (2.2) over a time interval $\Delta t$:

$$\frac{E(t + \Delta t) - E(t)}{\Delta t} = \overline{N} - \nabla \cdot \overline{G}.$$  \hspace{1cm} (2.3)

Here the overbar represents the time average. Because the Earth is close to energy balance, $E(t + \Delta t)$ and $E(t)$ cannot be wildly different from each other; this means that the numerator on the left-hand side of (2.3) is bounded within a finite range, regardless of how large $\Delta t$ is. Therefore, as $\Delta t$ 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. This means that energy storage inside the atmosphere-ocean/land surface at particular locations can be neglected if the time-averaging interval is long enough; the minimum time required for such an average 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

$$\nabla \cdot \overline{G} = \overline{N}.$$  \hspace{1cm} (2.4)

The global mean of $\nabla \cdot G$ must be exactly zero, at each instant, because the global mean of the divergence of any vector is zero (see Problem 1 at the end of this chapter). This means that the only way that (2.4) can be satisfied everywhere is if the global mean of $N = 0$, i.e., if the Earth is in energy balance.

Now break $G$ into its zonal and meridional components, i.e.

$$G = G_\lambda e_\lambda + G_\phi e_\phi.$$  \hspace{1cm} (2.5)

Here $e_\lambda$ and $e_\phi$ are unit vectors pointing towards the east and north, respectively. We expand the divergence operator in spherical coordinates (see Appendix) as follows:

$$\nabla \cdot G = \frac{1}{a \cos \phi} \frac{\partial G_\lambda}{\partial \lambda} + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} (G_\phi \cos \phi).$$  \hspace{1cm} (2.6)

Here $a$ is the radius of the Earth. Multiply (2.6) by $a \cos \phi$ and integrate with respect to longitude, around a latitude circle, to obtain

$$\int_0^{2\pi} (\nabla \cdot \overline{G})a \cos \phi \, d\lambda = \frac{\partial}{\partial \phi} \int_0^{2\pi} \overline{G_\phi \cos \phi} \, d\lambda = \int_0^{2\pi} \overline{Na \cos \phi} \, d\lambda.$$  \hspace{1cm} (2.7)

Note that the zonal derivative has dropped out as a result of the integration. Further integration of (2.7) with respect to latitude, from the South Pole ($\phi = -\pi/2$) to an arbitrary latitude $\phi$, gives
The Earth’s radiation budget: An “upper boundary condition” on the general circulation of the atmosphere

\[ \bar{T}(\varphi) - \bar{T}(-\pi/2) = \int_0^{2\pi} \int_0^\varphi \bar{N} a^2 \cos \varphi' d\lambda d\varphi', \quad (2.8) \]

where \( T(\varphi) = \int G a \cos \varphi \, d\lambda \). It should be clear that \( T(-\pi/2) = T(\pi/2) = 0 \). If this were not true, a finite amount of energy per unit time would be flowing into or out of a “point” of zero mass. The right-hand side of (2.8) is just the area integral of \( \bar{N} \), over the portion of the Earth extending from the South Pole to latitude \( \varphi \). The dimensions of \( \bar{T}(\varphi) \) are energy per unit time, e.g., Watts. When \( \varphi \), the upper limit of meridional integration in (2.8), is set to \( \pi/2 \), the right-hand side of (2.8) simply reduces to the global mean of \( \bar{N} \), and the left-hand side reduces to zero.

Fig. 2.4 shows a plot of \( \bar{T}(\varphi) \), based on measurements collected in the Earth Radiation Budget Experiment (ERBE; Barkstrom et al., 1989). The poleward energy transport in both hemispheres is clearly apparent. The shape of the curve is roughly like \( \sin(2\varphi) \).

Figure 2.4: The poleward energy transport by the atmosphere and ocean combined, as inferred from the observed annually averaged net radiation at the top of the atmosphere. A Petawatt is \( 10^{15} \) W.

We can say that the “job” of the general circulations of the atmosphere and oceans is to carry out this meridional energy transport. 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 balanced the absorbed solar

radiation locally. Such a hypothetical state is referred to as “radiative-convective

equilibrium;” modeling studies of radiative-convective equilibrium will be discussed later

in this course. Radiative-convective equilibrium would presumably imply much warmer

temperatures in the tropics, and much colder temperatures at the poles. The general

circulation of the atmosphere and oceans thus has a moderating effect on the global

distribution of temperature, tending to warm the higher latitudes and cool the tropics. As

we will see, however, these same thermal contrasts represent a source of energy (called

“available potential energy”) that makes the global circulations of the atmosphere and

oceans possible.

Much further discussion of the observations and theory of energy transports by

the atmosphere and oceans is given later in this course.

2.2 Surface boundary conditions

The lower boundary conditions on the general circulation of the atmosphere

strongly affect the regional characteristics of the circulation. The most obvious examples

are such basic aspects of physical geography as the distribution of land and sea, and the

locations of the Earth’s mountain ranges shown in Fig. 2.5. These are in part

“mechanical” boundary conditions that are independent of season. Note, however, that

the land-sea distribution and the locations of “permanent” (or, more accurately, non-

seasonal) land ice (e.g., Antarctica and Greenland) strongly affect the surface albedo.

Orography can also provide a thermal forcing, in that the surface of a mountain or

an elevated plateau can have a temperature quite different from that of the surrounding air

at the same height. For instance, during the northern summer the Tibetan plateau

produces a “warm spot” in the middle troposphere, and this represents an important

aspect of the thermal forcing that produces the Indian summer monsoon.

“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.

One of the most important properties of the Earth’s surface is that roughly 70% of

it is permanently wet, and so represents a huge source of moisture. The heat capacity of

the ocean is enormous, so that for some purposes (and on sufficiently short time scales,

e.g. a few days or weeks) the sea surface temperature (SST) can be considered to be a

“fixed” lower boundary condition on the atmosphere. The SST is in this sense the

simplest example of a thermal lower boundary condition. The geographical distribution

of SST fluctuates seasonally, and shows significant variations with both longitude and

latitude, as shown in Fig. 2.6. 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 later, the pattern of upwelling is very closely related to the

An Introduction to the General Circulation of the Atmosphere
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 atmospheric general 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 variability of the SST is a major complication hindering our understanding of the atmosphere and ocean as a coupled system.
The distribution of sea ice (Fig. 2.7) 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 such 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. The Arctic ocean is ice-covered all year, while the North Atlantic and the Southern Oceans experience seasonal melting. Of course, the thickness of the ice also varies both geographically and seasonally. In addition, several percent of open water typically occurs, especially when the ice is thin. This open water is often in the form of cracks called “leads.” The open water in leads is often much warmer.
2.2 Surface boundary conditions

Figure 2.6: a) Sea surface temperature distribution for January. The contour interval is 2 K, and values greater than 26 °C are shaded. b) Same for July. c) The sea surface temperature for March, minus the sea surface temperature for September. The contour interval is 2 K. The zero line is heavy, and negative values are indicated by dashed contours.
than the ice nearby, especially in winter. Under such conditions, the large-scale average sensible and latent heat fluxes are typically dominated by the contributions from the leads, even though leads 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. For a perspective on atmosphere-ocean-land ice interactions, see Randall et al. (1998).

The pattern of vegetation on the land surface affects the atmosphere in very complicated ways. A simplified summary of the observed distribution of vegetation on the land surface is given in Fig. 2.8. 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.

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, also shown in Fig. 2.8, is mainly confined to Antarctica and Greenland, in the present climate. The distribution of land ice can vary dramatically on time scales of thousands of years and longer (e.g. Imbrie and Imbrie, 1979).

Some aspects of the atmospheric general circulation can be regarded as more or less direct responses to the various boundary “forcings” mentioned above. Examples include the equator-to-pole energy flux, 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. Examples include baroclinic

---

**Figure 2.8:** A simplified depiction of the distribution of vegetation on the land surface. The resolution is one degree. “Permanent” land ice is shown in white.
waves and organized moist convection.

2.3 **Energy and moisture budgets of the surface and atmosphere**

The planetary radiation budget has already been briefly discussed. We now consider the energy and moisture budgets of the Earth’s surface and the atmosphere. It is a shocking fact that we do not know enough about the globally averaged surface energy budget to do more than sketch rough annual mean values, as shown in Table 2.2. None of the numbers in the table is known to better than 20% accuracy. Of the 240 W m\(^{-2}\) that is absorbed by the Earth-atmosphere system, 176 W m\(^{-2}\) is absorbed by the Earth’s surface. Thus only about 240 - 176 = 64 W m\(^{-2}\) of solar radiation is absorbed by the atmosphere. That is only about 1/4 of the total solar radiation absorbed by the Earth-atmosphere system. It should be noted, however, that the partitioning of the absorbed solar radiation between the atmosphere and the Earth’s surface is currently a matter of some controversy.

<table>
<thead>
<tr>
<th>Absorbed solar (SW)</th>
<th>176 W m(^{-2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downward infrared (LW(\downarrow))</td>
<td>312 W m(^{-2})</td>
</tr>
<tr>
<td>Upward infrared (LW(\uparrow))</td>
<td>-385 W m(^{-2})</td>
</tr>
<tr>
<td>Net longwave (LW)</td>
<td>-73 W m(^{-2})</td>
</tr>
<tr>
<td>Net radiation</td>
<td>103 W m(^{-2})</td>
</tr>
<tr>
<td>Latent heat (LH)</td>
<td>-79 W m(^{-2})</td>
</tr>
<tr>
<td>Sensible heat (SH)</td>
<td>-24 W m(^{-2})</td>
</tr>
</tbody>
</table>

The surface receives a total (LW\(\downarrow\) + SW; see notation defined in Table 2.2) of 488 W m\(^{-2}\), which is given back in the form of LW\(\uparrow\), LH and SH. By far the largest of these is LW\(\uparrow\). Keep in mind that the oceans can transport energy from one place to another, so that the energy absorbed by the oceans is not necessarily given back in the same place where it is absorbed. Also, the large heat capacity of the upper ocean allows energy storage on seasonal time scales. In contrast, the continents cannot transport energy internally at a significant rate, and their limited heat capacity forces near energy balance, everywhere, on time scales longer than a few days (at most).

The net radiative heating of the surface, which amounts to 103 W m\(^{-2}\), is balanced primarily by evaporative cooling of the surface at the rate of 79 W m\(^{-2}\). As discussed below, moisture is of comparable importance in the energy budget of the atmosphere.
The globally averaged energy budget of the atmosphere is shown in Table 2.3.

**Table 2.3:** The globally and annually averaged energy budget of the atmosphere. A positive sign means that the atmosphere is warmed.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Absorbed solar radiation (240 - 176)</strong></td>
<td>64 W m(^{-2})</td>
</tr>
<tr>
<td><strong>Net emitted terrestrial radiation (-240 + 73)</strong></td>
<td>-167 W m(^{-2})</td>
</tr>
<tr>
<td><strong>Net radiative heating</strong></td>
<td>-103 W m(^{-2})</td>
</tr>
<tr>
<td><strong>Latent heat input</strong></td>
<td>79 W m(^{-2})</td>
</tr>
<tr>
<td><strong>Sensible heat input</strong></td>
<td>24 W m(^{-2})</td>
</tr>
</tbody>
</table>

Again, most of the numbers in Table 2.2 and Table 2.3 are only rough estimates. One interpretation of Table 2.3 is that 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 is that required to allow the necessary infrared emission.

We can also say that the net radiative cooling of the atmosphere, at the rate of -103 W m\(^{-2}\), is primarily balanced by the latent energy source due to surface evaporation. Of course, this 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 is 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.

The preceding discussion suggests a second interpretation of the atmospheric energy budget: to a first approximation, the globally averaged precipitation rate 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, and is controlled mainly by dynamical processes.

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.

Taken together, these points suggest a negative feedback loop which tends to regulate the strength of the hydrologic cycle. To see how this works, suppose that we have 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 increase the rate of cirrus cloud production, which will reduce the rate at which the atmosphere is radiatively cooled. In order to restore atmospheric energy balance, it will be necessary for the hydrologic cycle to slow down, i.e., the initial perturbation will be damped. Fowler and Randall (1994) give further discussion.

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 circulating 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. In short, the atmospheric circulation carries energy upward as well as poleward. As discussed later, convective clouds play a particularly important role in the upward energy transport.

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.

The seasonal variations of the surface shortwave and longwave radiation at a particular station are illustrated in Fig. 2.9, which shows the variations of the upward and downward shortwave (SW) and longwave (LW) near-surface radiation at the Boulder Atmospheric Observatory (BAO) tower, near Boulder Colorado. (The tower is located about 1 km west of Interstate 25, and slightly south of Colorado Route 52.) The data cover the three years 1986 - 89. The seasonal cycle is clearly evident. High frequency fluctuations are primarily due to cloudiness. Note that the upward solar radiation has maxima during the winter. These are associated with the increased albedo of the ground following snow storms.

Data like those shown in Fig. 2.9 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.

Fig. 2.10 shows the meridional distribution of the solar radiation absorbed by the
Earth’s surface, as a function of latitude, and for January, July, and the annual mean. The seasonal changes associated with the movement of the Earth in its orbit are clearly visible. These have already been discussed, earlier in this chapter. There are additional features that require some explanation. For example, near 50°N in July there is a minimum of the surface absorbed solar radiation. This is associated with cloudiness in those latitudes, and indicates that the clouds are having a major impact on the energy budget of the ocean in those latitudes. A weaker tropical minimum occurs for the same reason. The annual mean curve is fairly symmetrical about the Equator, but shows a minimum near 10°N associated with the ITCZ. 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 Fig. 2.11. Here there are some real surprises. Although the ocean temperatures are warmer in summer than in winter, the net longwave cooling of the surface is actually stronger in winter than in summer! This occurs despite the fact that the surface emission is essentially proportional to \( T^4 \), and so must be considerably larger in summer than in winter. 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 overwhelms the increase in the upward component, giving a net decrease in surface infrared cooling from winter to summer. 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.

Fig. 2.12 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. Clearly the surface must cool by non-radiative means.
Fig. 2.13 shows the zonally averaged net latent heat flux, as reported by Esbensen and Kushnir (1981). Positive values represent a moistening of the atmosphere and a cooling of the surface. Clearly 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.

Figure 2.13: The zonally averaged surface latent heat flux, positive upward, based on ECMWF analyses for 1987. Note: These data are not true observations, although they are based on observations.
Fig. 2.14 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.

![Sensible Heat Flux](image)

**Figure 2.14:** The zonally averaged surface sensible heat flux, positive upward, based on ECMWF analyses for 1987. Note: These data are not true observations, although they are based on observations.

### 2.4 Summary

This chapter provides a summary 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 transports in more detail; for now we simply note that they must occur.

We have also discussed the lower boundary conditions on the atmosphere, which are provided by the Earth’s surface. These involve many complex factors including the distributions of continents and oceans, the arrangements of the mountains, and the distribution of vegetation.

We also discussed the moisture budget of the Earth’s surface and the atmosphere. Among the most important points to emerge from this analysis 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 have also noted the important effects of water vapor and clouds on the Earth’s radiation budget, and the effects of ice, snow, and the land-sea distribution on the radiative properties of the Earth’s surface. These facts make it clear that the hydrologic cycle plays a very central role.
in the general circulation of the atmosphere, and in the Earth’s energy balance. This point will be made again later in a variety of ways.

We must also discuss exchanges of momentum between the atmosphere and the Earth’s surface; this is postponed until later.

Problems

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

\[
\int_{S} \nabla \cdot \mathbf{Q} \, dS = 0,
\]

where the integral is taken over a closed surface. We assume that \( \mathbf{Q} \) is everywhere tangent to the surface, i.e. it “lies in” the surface. Note: This shows that the globally averaged divergence of any “horizontal” vector field is zero.

b) Also prove that

\[
\int_{S} \mathbf{k} \cdot (\nabla \times \mathbf{Q}) \, dS = 0
\]

where the integral is taken over a closed surface. Here \( \mathbf{k} \) is a unit vector everywhere perpendicular to the surface. Note: This shows that the globally averaged vorticity is zero.

2. Suppose that at 40° N the northward energy transport shown in Fig. 2.4 is entirely due to the atmosphere, and is produced by the combination of a northward wind around half of the latitude circle, and a compensating southward wind in the other half of the latitude circle, with uniform speeds of 5 m s\(^{-1}\) each. For simplicity, suppose that these currents fill the entire depth of the atmosphere, and that the surface pressure is a uniform 1000 mb. Assume that the temperature is zonally uniform within each current. Compute the implied temperature difference between the northward and southward flows.

3. a) Suppose that 1 W m\(^{-2}\) is supplied to a column of water 100 m deep. 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 on the top of the atmosphere were used to warm the ocean uniformly, how long would it take to increase the temperature of the entire ocean by 1 K?
CHAPTER 3
An overview of the observations

Copyright 2005 David A. Randall

3.1 Introduction

This chapter is intended to provide a “quick look” at the observed seasonally varying general circulation of the atmosphere, with a few additional comments about atmosphere-ocean interactions. Selected fields will be shown and described. Lots of questions will be raised, but for the most part the answers will be deferred until later chapters. Most of the fields shown in this chapter are based on analyses performed at the European Centre for Medium Range Weather Forecasts (ECMWF; see the 1997 ECMWF report in the reference list at the end of this chapter).

The observations discussed in this chapter show many things, but among the most important phenomena that can be glimpsed are the tropical Hadley and Walker circulations, the monsoons, stationary planetary waves, and some aspects of the hydrologic cycle.

We need some equations for use in interpreting the observations. The table below gives some that we will refer to in this chapter; you should already be familiar with them. In writing these equations, pressure has been used as the vertical coordinate.

<table>
<thead>
<tr>
<th>Physical Principle</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conservation of horizontal momentum</td>
<td>( \frac{DV}{Dt} + \left( f + \frac{u \tan \phi}{a} \right) k \times \mathbf{V} = -\nabla_p \phi + g \frac{\partial F}{\partial p} ) (3.1)</td>
</tr>
<tr>
<td>Hydrostatic equation</td>
<td>( \frac{\partial \phi}{\partial p} = -\alpha ) (3.2)</td>
</tr>
<tr>
<td>Conservation of mass</td>
<td>( \nabla_p \cdot \mathbf{V} + \frac{\partial \omega}{\partial p} = 0 ) (3.3)</td>
</tr>
</tbody>
</table>
### Physical Principle

| Conservation of thermodynamic energy | \[ c_p \frac{DT}{Dt} = \omega \alpha + g \frac{\partial F}{\partial p} + Q_{rad} + Q_{lat} \] (3.4) |

Eq. (3.4) can also be written in terms of potential temperature:

\[ \frac{D\theta}{Dt} = g \frac{\partial F}{\partial p} + \left( \frac{\theta}{c_p T} \right) (Q_{rad} + Q_{lat}) \] (3.5)

### Conservation of water vapor

\[ \frac{Dq_v}{Dt} = g \frac{\partial F}{\partial p} - \frac{Q_{lat}}{L} \] (3.6)

By using the continuity equation, (3.6) can be transformed to “flux form:"

\[ \left( \frac{\partial q_v}{\partial t} \right)_p + \nabla_p \cdot (\nabla q_v) + \frac{\partial}{\partial p} (\omega q_v) = g \frac{\partial F}{\partial p} - \frac{Q_{lat}}{L} \] (3.7)

### Equation of state

\[ p\alpha = RT \] (3.8)

The symbols used in these equations are defined as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V )</td>
<td>The horizontal wind vector</td>
</tr>
<tr>
<td>( t )</td>
<td>time</td>
</tr>
<tr>
<td>( \frac{D}{Dt} )</td>
<td>The three-dimensional Lagrangian time derivative, i.e. the time derivative following a particle, which satisfies</td>
</tr>
<tr>
<td></td>
<td>[ \frac{D(\cdot)}{Dt} \equiv \left[ \left( \frac{\partial}{\partial t} \right)_p + V \cdot \nabla_p + \omega \frac{\partial}{\partial p} \right] (\cdot) ]</td>
</tr>
<tr>
<td>( f )</td>
<td>The Coriolis parameter, ( 2\Omega \sin \varphi )</td>
</tr>
<tr>
<td>( k )</td>
<td>A unit vector pointing upwards. Note that the direction denoted by “upwards” varies from place to place on the Earth.</td>
</tr>
</tbody>
</table>
### 3.1 Introduction

With the approximation of geostrophic balance, the momentum equations reduce to

\[ 0 = f v_y - \frac{1}{a \cos \phi} \left( \frac{\partial \phi}{\partial \lambda} \right)_p, \]  

(3.9)

<table>
<thead>
<tr>
<th><strong>Variable</strong></th>
<th><strong>Definition</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>( u, v )</td>
<td>The zonal and meridional wind components. (The meridional component is not actually used in the equations given above, but is defined here for completeness.)</td>
</tr>
<tr>
<td>( \lambda, \varphi )</td>
<td>Longitude and latitude (longitude is not actually used above)</td>
</tr>
<tr>
<td>( \Omega )</td>
<td>The angular velocity of the Earth’s rotation</td>
</tr>
<tr>
<td>( \phi )</td>
<td>Geopotential</td>
</tr>
<tr>
<td>( a )</td>
<td>The radius of the Earth</td>
</tr>
<tr>
<td>( F_x )</td>
<td>The upward turbulent flux of quantity ( x )</td>
</tr>
<tr>
<td>( p )</td>
<td>Pressure, used here as a vertical coordinate</td>
</tr>
<tr>
<td>( g )</td>
<td>Acceleration of gravity</td>
</tr>
<tr>
<td>( L )</td>
<td>Latent heat of condensation</td>
</tr>
<tr>
<td>( R )</td>
<td>The specific “gas constant” of dry air</td>
</tr>
<tr>
<td>( T )</td>
<td>Temperature</td>
</tr>
<tr>
<td>( \omega )</td>
<td>The vertical “pressure velocity”</td>
</tr>
<tr>
<td>( c_p )</td>
<td>The specific heat of air at constant pressure</td>
</tr>
<tr>
<td>( \theta )</td>
<td>Potential temperature</td>
</tr>
<tr>
<td>( q_v )</td>
<td>The mixing ratio of water vapor</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>The specific volume of the air, which is the inverse of the density, and is equal to ( RT/p )</td>
</tr>
<tr>
<td>( Q_{rad} )</td>
<td>The radiative heating rate</td>
</tr>
<tr>
<td>( Q_{lat} )</td>
<td>The latent heating rate</td>
</tr>
</tbody>
</table>
32 An overview of the observations

\[ 0 = -fu_g - \frac{1}{a} \left( \frac{\partial \phi}{\partial \phi} \right)_p. \]  \hspace{1cm} (3.10)

Here \( u_g \) and \( v_g \) are the zonal and meridional components of the geostrophic wind, respectively. By combining (3.9) and (3.10) with the hydrostatic equation, (3.2), and the equation of state, (3.8), we can derive the thermal wind equations:

\[ \frac{\partial u_g}{\partial p} = \frac{1}{fpa} \left( \frac{\partial T}{\partial \phi} \right)_p, \]  \hspace{1cm} (3.11)

\[ \frac{\partial v_g}{\partial p} = \frac{-R}{fpa \cos \phi} \left( \frac{\partial T}{\partial \lambda} \right)_p. \]  \hspace{1cm} (3.12)

These relationships between the vertical derivatives of the winds and the horizontal derivatives of the temperature will be used in the following discussion of the observations.

3.2 The global distribution of atmospheric mass

Mass is arguably the most fundamental quantity in any physical description of the atmosphere. The density, \( \rho \), is defined as the mass of air per unit volume:

\[ \rho \equiv \frac{M}{V}, \]  \hspace{1cm} (3.13)

and the specific volume is its inverse:

\[ \alpha \equiv \frac{1}{\rho} = \frac{V}{M}. \]  \hspace{1cm} (3.14)

Let \( \mathbf{V} \equiv (u, v, w) \) denote the three-dimensional wind vector. The vector \( \rho \mathbf{V} \) is the flux of mass across a unit area. Consider a volume \( V \) fixed in space, through which fluid can flow (Fig. 3.1). Conservation of mass is expressed by

\[ \frac{d}{dt} \int_V \rho dV = \int_V \frac{\partial \rho}{\partial t} dV = -\int_{\partial V} \rho \mathbf{V} \cdot d\mathbf{\sigma}, \]  \hspace{1cm} (3.15)

where \( d\mathbf{\sigma} \) is an outward normal vector whose magnitude is that of a (differential) element of the bounding surface. Using Gauss' Theorem, we can rewrite (3.15) as

\[ \int_V \left[ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) \right] dV = 0. \]  \hspace{1cm} (3.16)
3.2 The global distribution of atmospheric mass

Since the control volume is arbitrary, we conclude that

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0.$$  (3.17)

We can also express mass conservation in terms of a Lagrangian framework, in which we consider time rates of change following particles. For this purpose, we use the Lagrangian (particle-following) time derivative, \( \frac{D}{Dt} \), which can be expanded as

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla .$$  (3.18)

To demonstrate this, let the change of \( \psi(x, y, z, t) \) following a particle be expanded as

$$\frac{D\psi}{Dt} \cdot \Delta t = \psi(x + \Delta x, y + \Delta y, z + \Delta z, t + \Delta t) - \psi(x, y, z, t).$$  (3.19)

Because \( \frac{\Delta x}{\Delta t} \equiv u \), \( \frac{\Delta y}{\Delta t} \equiv v \), and \( \frac{\Delta z}{\Delta t} \equiv w \), we obtain

$$\frac{D\psi}{Dt} = \frac{\partial \psi}{\partial t} + \mathbf{V} \cdot \nabla \psi .$$  (3.20)
Eq. (3.20) describes what happens to a particular particle, and so it is valid in any frame of reference. Note, however, that the individual terms on the right-hand side of (3.20) will take different values in coordinate systems that are in relative motion.

Using (3.18), we can rewrite (3.17) as

$$\frac{DP}{Dt} + \rho \nabla \cdot \mathbf{V} = 0,$$

or as

$$\nabla \cdot \mathbf{V} = -\frac{1}{\rho} \frac{DP}{Dt} = -\frac{D\alpha}{\alpha}.$$

For a fluid whose density cannot be changed, so that \( \frac{DP}{Dt} = 0 \), Eq. (3.22) implies that \( \nabla \cdot \mathbf{V} = 0 \), i.e., the three-dimensional flow is nondivergent.

The mass budget of a column of air can be derived from (3.17) as follows. Integrate with respect to height, through the entire atmospheric column, to obtain

$$\int_{z_S}^{\infty} \frac{\partial \rho}{\partial t} \, dz + \int_{z_S}^{\infty} \nabla \cdot (\rho \mathbf{V}) \, dz = 0.\quad (3.23)$$

Here \( z_S \) is the surface height, which varies in space and, less obviously, with time. Taking the integral inside the derivatives, and taking into account the spatial and possible temporal variability of \( z_S \), we obtain

$$\frac{\partial}{\partial t} \left( \int_{z_S}^{\infty} \rho \, dz \right) + \nabla \cdot \left( \int_{z_S}^{\infty} \rho \mathbf{V}_H \, dz \right) + \rho_s \left( \frac{\partial z_S}{\partial t} + \mathbf{V}_S \cdot \nabla z_S - w_S \right) = 0.\quad (3.24)$$

Here \( \mathbf{V}_H \) is the horizontal velocity vector, subscript \( S \) denotes the Earth’s surface, \( \mathbf{V}_S \) is the horizontal velocity vector evaluated at the Earth’s surface, \( w \) is the vertical velocity.

The expression \( \rho_s \left( \frac{\partial z_S}{\partial t} + \mathbf{V}_S \cdot \nabla z_S - w_S \right) \) represents the flux of mass across the Earth’s surface. We assume that in fact no mass crosses the Earth’s surface:

$$\rho_s \left( \frac{\partial z_S}{\partial t} + \mathbf{V}_S \cdot \nabla z_S - w_S \right) = 0.\quad (3.25)$$

In writing (3.24), we have also assumed that there is no mass flux across the top of the
3.2 The global distribution of atmospheric mass

atmosphere.

Normally, of course, the height of the Earth’s surface can be assumed to be independent of time, in which case (3.25) reduces to

\[ w_S = V_S \cdot \nabla z_S. \] (3.26)

We could keep the \( \frac{\partial z_S}{\partial t} \) term of (3.25), e.g., if we wanted to study the effects of earthquakes. In that case, we would specify \( \frac{\partial z_S}{\partial t} \) to describe how the lower boundary moved, and we would be able to compute the atmospheric motions induced by the earthquake.

With the use of (3.25), (3.24) reduces to

\[ \frac{\partial}{\partial t} \left( \int_{z_s}^{\infty} \rho \, dz \right) + \nabla \cdot \left( \int_{z_s}^{\infty} \rho \mathbf{V}_H \, dz \right) = 0, \] (3.27)

which expresses conservation of mass for the air column. According to (3.27), 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.

Finally we note that to an excellent approximation a vertical increment of pressure is proportional to the corresponding vertical increment of mass of air per unit area above the surface, i.e.

\[ dp = -\rho g dz. \] (3.28)

Use of (3.28) in (3.27) gives

\[ \frac{\partial p_S}{\partial t} + \nabla \cdot \left( \int_{0}^{p_S} \mathbf{V}_H \, dp \right) = 0. \] (3.29)

This is called the surface-pressure-tendency equation.

It is interesting to note that there is a non-zero flux of mass across the Martian surface, near the poles. The Martian atmosphere consists almost entirely of carbon dioxide. 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 carbon dioxide returns to gaseous form (e.g., Lewis and Prinn, 1984; Zent, 1996).

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 given in Fig. 3.2. The figure shows surface pressure plotted against surface elevation, for 00Z (i.e., midnight in Greenwich England) on January 1, 2001. The main point of the figure is that a
An overview of the observations

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. A secondary point is that the data fall roughly onto two curves. The lower curve consists mainly of Northern Hemisphere (winter) points, and the upper curve consists mainly of Southern Hemisphere (summer) points. This illustrates the fact that, over land, the surface pressure tends to be higher in winter and lower in summer. The scatter of the data about the curves indicates the dynamical variability of the surface pressure. The range of variability of the surface pressure is particularly large when the surface height of zero, which means that it is large for the oceans.

Fig. 3.3 shows the January and July monthly mean maps of sea level pressure, respectively. These maps are much smoother than weather maps plotted for particular observation times, because the moving highs and lows that represent individual weather systems have been smoothed out by the time-averaging. As already discussed, time averages are denoted by overbars, so the time-averaged sea-level pressure can be represented by the symbol \( \bar{p}_{SL} \). Fig. 3.4 shows the corresponding zonally averaged distributions of the sea level pressure for January and July. Following a tradition that goes back to the 1940s, we will use square brackets to denote zonal means:

\[
[(\,\,\,)] = \frac{1}{2\pi} \int_0^{2\pi} (\,\,\,) d\lambda .
\]  

(3.30)
Figure 3.3: Sea level pressure maps. The contour interval is 3 mb.

Using this notation, the quantity plotted in Fig. 3.4 is $\bar{p}_{SL}$. Note that we have applied both an overbar and square brackets.

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, as shown in Fig. 3.5. The maps show the departure of the surface temperature from its zonal mean at each latitude. The departure of a quantity from its zonal mean at each latitude is called the “eddy” component of that quantity, and is denoted by a star:
Fig. 3.5 thus shows the time average of the eddy component of the surface temperature, for January and July. This quantity can be denoted by the symbol $T_S^*$. We see in Fig. 3.5 that in middle latitudes, especially in the Northern Hemisphere, the surface air temperature is colder over the continents than the oceans in winter, and vice versa in summer. Particularly cold temperatures occur over the eastern sides of the Northern Hemisphere continents in January. This is 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, and it has cooled all along the way.

Comparison of Fig. 3.3 and Fig. 3.5 shows that there is a tendency for high surface pressure to be associated with cold surface temperatures, and vice versa. To understand this, recall that the surface pressure measures the amount of mass in the column. Cold air has a higher density than warm air, so that a “pile” of cold air of a given geometrical thickness will contain more mass than a “pile” of warm air of the same geometrical thickness. This can be seen mathematically by combining the ideal gas law with the hydrostatic equation to obtain

$$p_S = \int_{z_S}^{\infty} \left( \frac{pg}{RT} \right) dz.$$  \hspace{1cm} (3.32)
Fig. 3.4 shows that for both January and July there is a tendency for high sea-level pressure to occur in the subtropics. The 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.

Throughout the year, there is a belt of low pressure in the tropics. In view of
geostrophic balance, the decrease of sea-level pressure from the sub-tropics towards the Equator implies tropical easterlies near the surface.

Because the sea-level pressure generally decreases from the sub-tropics to middle latitudes, the geostrophic relation leads us to expect a tendency for surface westerlies on the poleward side of the sub-tropical highs. Similarly, the relatively high pressure over the poles suggests surface easterlies in high latitudes.

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 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, which represent “stationary eddies,” are much more prominent 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 Charney (1963), in terms of the differences in dynamical balance between the tropics; specifically, the effects of the Earth’s rotation are much more important than particle accelerations in middle latitudes, and the opposite is true in the tropics. Jule Charney, one of the giants of twentieth-century meteorology, is cited here for the first time in Chapter 2; his name will come up again and again throughout the remainder of this course; a photo of Charney is shown in Fig. 3.6. Charney began his analysis with the equation of motion, (3.1), which can be written in simplified form as

$$\frac{DV}{Dt} + f k \times V = -\nabla \phi . \tag{3.33}$$

Here we have omitted the $\tan \phi$ term and the friction term, for simplicity, so that there are only three terms left. We have also dropped the subscript $p$ on the del operator, to make the notation simpler. The three terms shown in (3.33) represent most of the “action” throughout most of the atmosphere. Their orders of magnitude can be estimated as follows:

$$\frac{DV}{Dt} \sim \frac{V^2}{L} \tag{3.34}$$

$$f k \times V \sim fV \tag{3.35}$$

$$\nabla \phi \sim \frac{\delta \phi}{L} \tag{3.36}$$

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 a typical fluctuation of the geopotential
The global distribution of atmospheric mass

41

3.2 The global distribution of atmospheric mass

height. Note that $\delta V \sim V$, but $\delta \phi$ is generally much less than $\phi$. 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 we use the term “large-scale” in the same way for both regions.

In middle latitudes, the acceleration following a particle, $\frac{DV}{Dt}$, is typically negligible in (3.33) compared to the rotation term. A typical value of $\frac{DV}{Dt}$ can be estimated as

$$\left| \frac{DV}{Dt} \right| \sim \frac{V^2}{L} = \frac{10^2}{10^6} = 10^{-4} \text{ m s}^{-2}.$$  

A representative midlatitude value of the Coriolis parameter is $f_{\text{midlat}} \sim 10^{-4}$ s$^{-1}$, so that $f_{\text{midlat}} V \sim 10^{-3}$ m s$^{-2}$, about one order of magnitude larger than $\frac{DV}{Dt}$. Geostrophic balance is, therefore, approximately satisfied in mid-latitudes, i.e.,

$$f_{\text{midlat}} V \sim \frac{(\delta \phi)_{\text{midlat}}}{L}, \quad (3.37)$$

---

Figure 3.6: Prof. Jule G. Charney, whose accomplishments include scale analyses of both extratropical and tropical motions, development of the quasigeostrophic model, development (in his Ph.D. thesis at UCLA) 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. Because he did so much, Charney’s work is frequently cited in this course.

An Introduction to the General Circulation of the Atmosphere
Here we have added the “midlat” subscript to \( \delta \phi \), just for clarity. According to (3.38), rotation strongly determines the magnitude of midlatitude height fluctuations. In other words, rotation can balance pressure gradients in midlatitudes.

On the Equator the Coriolis parameter vanishes, so we expect that sufficiently close to the Equator geostrophic balance breaks down (a point discussed further in a later chapter), 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):

\[
\frac{V^2}{L} \sim \frac{(\delta \phi)_{\text{tropics}}}{L},
\]

or

\[
(\delta \phi)_{\text{tropics}} \sim V^2.
\]

Comparing (3.38) and (3.40), we see that

\[
\frac{(\delta \phi)_{\text{tropics}}}{(\delta \phi)_{\text{midlat}}} \sim \frac{V}{f_{\text{midlat}} L} \equiv R_{\text{omidlat}},
\]

where \( R_{\text{omidlat}} \) is a midlatitude Rossby number. By substituting the numerical values given above, we find that \( R_{\text{omidlat}} \equiv 0.1 \). Eq. (3.41) therefore tells us that geopotential height fluctuations in the tropics are much smaller than those in middle latitudes.

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. We conclude that if the horizontal pressure gradient is small at all levels, then the horizontal temperature gradients must also be small. Fig. 3.7 illustrates this.

Later in this chapter we will return to Charney’s analysis to interpret the relationships between vertical motion and heating in the tropics.

### 3.3 Zonal wind

Fig. 3.8 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, its dynamical influence extends downward into the troposphere,
and so the stratosphere should be of interest even to students who are mainly focused on the
general 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.

Westerly jets (“jet streams”) are quite prominent, especially in the winter hemisphere.
The main tropospheric winter jet tends to occur at about 30° latitude. The summer jet, which
is weaker, tends to be at about 45° latitude. The midlatitude jet maxima are consistently found
near the 200 mb level.

In the Southern Hemisphere in July, there is a clear minimum in the westerlies near
150 mb, at about 40° S. Above and poleward of this minimum is a stratospheric westerly jet,
called the polar night jet. A similar but weaker jet occurs in the Northern Hemisphere winter.

Weak surface easterlies are found near the poles.

The most striking things about the stratospheric zonal winds are the strong westerly
vortex in the winter hemisphere, often called the “polar night jet,” and the easterlies filling the
summer hemisphere. Note that the polar night jet is separated from the westerly jet at the
tropopause level by a local minimum in the westerlies; nevertheless there is a band of strong
westerlies that extends upwards from the midlatitude troposphere into the high-latitude
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.

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 \( u \) passes through zero, i.e. where surface easterlies meet surface westerlies. If you compare Fig. 3.4 with Fig. 3.8, you will see that in fact the subtropical surface-pressure maxima occur at the latitudes where the zonal component of the surface wind passes through zero.

Fig. 3.9 shows maps of the 850 mb zonal wind for January and July, respectively. Again, keep in mind that many intense small-scale (~ 1000 km) features would appear in

**Figure 3.8:** Latitude–height section of the zonal wind. The contour interval is 5 m s\(^{-1}\). Easterlies are shaded.
daily maps, but have been smoothed out here by time averaging.

![Maps of the 850 mb zonal wind, for January and July. The contour interval used is 3 m s\(^{-1}\). Easterlies are shaded.](image)

The monthly mean maps show very obvious alternating bands of easterlies and westerlies, which qualitatively remind us of Jupiter (see Fig. 3.10), although of course 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, stationary eddies
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. The westerlies in the Northern Hemisphere in winter are particularly strong over the oceans.

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

Fig. 3.11 shows the corresponding maps for 200 mb. The winds are generally stronger aloft than near the surface; this is true of both the zonal mean 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 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 will be discussed later, these regions of mean westerlies allow waves to propagate from middle latitudes into the tropics.

3.4 Meridional wind

Fig. 3.12 shows the latitude-height distribution of the zonally averaged meridional
3.4 Meridional wind

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 an obvious 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 of course associated with the Hadley circulation. Poleward flow is also found near the surface in middle latitudes, with weak
As with most other fields, the seasonal changes of the meridional wind in the midlatitudes of the Southern Hemisphere are quite weak.

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, which is
3.4 Meridional wind

\[
\frac{\partial p_S}{\partial t} = -\nabla \cdot \left[ \int_0^{p_s} \mathbf{v} \, dp \right].
\]  (3.42)

Problem 3 at the end of this chapter asks you to derive (3.42). Averaging (3.42) around a latitude circle gives

\[
\frac{\partial}{\partial t} \left[ p_S \right] = \frac{-1}{a \cos \phi} \frac{\partial}{\partial \phi} \left\{ \int_0^{p_s} \mathbf{v} \, dp \right\} \cos \phi.
\]  (3.43)

Here subscript \( S \) denotes a surface value, and square brackets represent zonal averages:

\[
\left\{ \right\} \equiv \frac{1}{2\pi} \int_0^{2\pi} \left( \right) d\lambda.
\]  (3.44)

In an average over a sufficiently long time, \( \frac{\partial}{\partial t} \left[ p_S \right] \) must become small at each latitude, because \( p_S \) is bounded within a fairly narrow range. It follows that

\[
\frac{\partial}{\partial \phi} \left\{ \int_0^{p_s} \mathbf{v} \, dp \right\} \cos \phi = 0,
\]  (3.45)

where the overbar denotes a time average. This means that \( \left\{ \int_0^{p_s} \mathbf{v} \, dp \right\} \cos \phi \) is independent of latitude. Since \( \cos \phi = 0 \) at both poles, we can conclude that

\[
\left\{ \int_0^{p_s} \mathbf{v} \, dp \right\} = 0 \text{ at all latitudes.}
\]  (3.46)

The physical meaning of (3.46) is very simple. Suppose, for instance, that \( \left\{ \int_0^{p_s} \mathbf{v} \, dp \right\} \) was positive at the Equator. This would mean that, in a zonal and vertical average, air would be flowing out of the Southern Hemisphere and into the Northern Hemisphere. Of course this can really happen at a given instant, but if it continued over time the surface pressure in the Southern Hemisphere would decrease to zero and the surface pressure in the Northern Hemisphere would increase to roughly double its normally observed values. Obviously the pressure-gradient force would resist such a scenario.
Note also that it follows immediately from (3.9) that the zonally averaged meridional component of the geostrophic wind vanishes on each pressure surface:

\[ 0 = [v_g]. \] (3.47)

This shows that all of the zonally averaged meridional circulations, \{[v], [ω]\}, including for example the Hadley circulation, are completely ageostrophic. The implication is that the important large-scale circulations are not necessarily quasigeostrophic. We note, however, that the strongest features in Fig. 3.12 do occur in the tropics, where geostrophy would be expected to lose its grip.

There are, of course, small systematic changes in the meridional distribution of atmospheric mass, associated in part with small systematic changes in the vertically integrated meridional wind. Trenberth et al. (1987) present an interesting discussion of these phenomena. Fig. 3.13, taken from their paper, shows the variation with season of the vertically integrated meridional velocity. The top panel shows variations with latitude, and the bottom panel shows the seasonal cycle at the Equator. The globally averaged surface pressure is of course very nearly invariant with time, apart from small changes associated with the seasonal cycle of atmospheric water vapor. Trenberth et al. (1987) discussed 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. 3.13.

Fig. 3.15 shows maps of the 850 mb meridional wind for January and July, respectively. Fig. 3.16 shows the corresponding 200 mb maps. 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. 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 mb than at 850 mb. In the Northern Hemisphere especially, there is a tendency for stronger features in winter than in summer.

As can be seen in Fig. 3.11, 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. 3.15 or Fig. 3.16, however. At 850 mb we can see a strong southerly flow 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. 3.17 shows the streamlines at 850 mb, for January and July, and Fig. 3.18 shows the corresponding streamlines at 200 mb. 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 mb winds. There are strong cross-equatorial flows at 850 mb in both the Pacific and Indian Oceans, in July. The 200 mb streamlines show the planetary wave patterns in the midlatitude winter, and also tropical phenomena including a strong monsoon-induced anticyclone over the Indian subcontinent in July.
3.4 Meridional wind

**Figure 3.13:** The variation with season of the vertically integrated meridional velocity. The top panel shows variations with latitude, and the bottom panel shows the seasonal cycle at the Equator. From Trenberth et al. (1987).
Figure 3.14: 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).
In the Northern Hemisphere, the height fields show pronounced seasonal changes; the gradients are tight in the winter and loose in the summer. The subtropical highs are easier to pick out in summer than in winter. At 850 mb, the winter fields in the Northern Hemisphere are decidedly “wavy,” while the summer fields are “lumpy.” The seasonal changes in the Southern Hemisphere are much weaker.

Fig. 3.19 shows maps of the geopotential height on the 850 mb surface, and Fig. 3.20 shows the corresponding maps for 200 mb. The geopotential thickness between two pressure surfaces measures the average temperature of the intervening layer. The meridional temperature gradients are clearly evident in the plots. We notice the weak gradients in the tropics, as discussed already in connection with Charney’s (1963) analysis. In addition, the geostrophic wind is proportional to the gradient of the height field. The relationships between
the zonal and meridional wind maps and the geopotential maps are clearly evident.

3.6 **Vertical velocity and the mean meridional circulation.**

A “stream function,” $\psi$, can be defined for the mean meridional circulation. Plots of $\psi$ effectively show the zonally averaged vertical velocity and the zonally averaged meridional velocity, together in one diagram. The definition of $\psi$ is:
3.6 Vertical velocity and the mean meridional circulation.

[Figure 3.17: Streamlines of the 850 mb wind, for January and July.]

\[
[v]2\pi a \cos \phi \equiv g \frac{\partial \psi}{\partial p},
\]

\[
[\omega]2\pi a^2 \cos \phi \equiv -g \frac{\partial \psi}{\partial \phi}.
\]

Here again the square brackets denote zonal averages. You should confirm that for, any given
distribution of $\psi$, the zonally averaged form of the continuity equation, (3.3), is automatically satisfied. Note that $\psi$ is, by virtue of its definition, independent of longitude.

It is easy to see that lines of constant $\psi$ cannot intersect the Earth’s surface; if they did, that would imply a flow of air across the Earth’s surface. Similarly, lines of constant $\psi$ cannot extend upward into space; if they did, that would imply that the Earth’s atmosphere was being accumulated from or lost into space. Note from (3.48) that

**Figure 3.18: Streamlines of the 200 mb wind, for January and July.**

An Introduction to the General Circulation of the Atmosphere
3.6 Vertical velocity and the mean meridional circulation.

Vertical velocity and the mean meridional circulation. 

\[
\psi \sim \left( \frac{\delta p}{g} \right) \frac{L}{t} \sim \frac{M}{L^2} \frac{L^2}{t} = \frac{M}{t},
\]

(3.49)

i.e. \( \psi \) has dimensions of mass per unit time. Typically \( \psi \) is expressed in units of \( 10^{12} \text{ g s}^{-1} \), which is the same as \( 10^9 \text{ kg s}^{-1} \). This unit is sometimes called a “Sverdrup,” especially in the oceanographic literature. The sign of \( \psi \) is arbitrary [i.e., the signs could be reversed in

Figure 3.19: Maps of the geopotential height at 850 mb. The contour interval is 25 m.
(3.48), and also an arbitrary constant can be added to $\psi$ without changing $[v]$ or $[\omega]$. In view of (3.48), we can compute $\psi$ from either $[v]$ or $[\omega]$, with the boundary condition

$$\psi = 0 \text{ at the Earth’s surface.} \quad (3.50)$$

The physical content of (3.50) is that $\psi$ is a constant along the Earth’s surface; the particular constant chosen (i.e., zero) has no physical significance.
Fig. 3.21 shows the latitude-height distribution of the stream function of the mean meridional circulation, for January and July, respectively. The observed meridional wind has been used to create these plots. A banded structure is clearly evident. Again, this is somewhat reminiscent of the banded circulation systems that are readily visible in images of Jupiter and Saturn. Deep rising motion occurs in the summer-hemisphere tropics, with sinking motion on either side. The strongest tropical rising motion is near 300 mb, but notice that weak rising motion continues into the tropical stratosphere. The strongest subsidence is in the winter hemisphere subtropics, again near 300 mb. Rising motion occurs in middle latitudes, and is
An overview of the observations

strongest in the winter. Maximum values tend to occur near 500 mb. Sinking motion is found near the poles, mainly in the lower troposphere.

The dominant cellular structures in the tropics are called “Hadley Cells.” There is a “large” Hadley circulation 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 Sverdups. A weaker Hadley circulation occurs in the summer hemisphere. Both Hadley cells are “direct” circulations, which means that their rising branches are warm and their sinking branches are cold. As discussed later, such 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 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° 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.

The correspondence between the zonally averaged vertical motion and the zonally averaged meridional motion is fairly obvious. The meridional currents can be interpreted as outflows from or inflows to the vertical currents.

It is interesting to examine the seasonal change of the mean meridional circulation, as shown in Fig. 3.22. Symmetry between the hemispheres is approximated about one month after the equinoxes. The Hadley cells are best developed (on the winter side, and least developed on the summer side) near the solstices. The reasons for this are discussed by Lindzen and Hou (1988) and Hack et al. (1989).

Fig. 3.23 shows maps of the 500 mb 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 millibars per day. Fig. 3.21 gives the impression that there are regular bands of rising and sinking motion, arranged along latitude circles. Zonal bands of rising and sinking motion are not easy to pick out in Fig. 3.23, however; at first glance, the patterns look almost random. There is some tendency, however, 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 regions of large-scale sinking motion in the middle troposphere.
3.6 Vertical velocity and the mean meridional circulation.

Figure 3.22: Seasonal change of the mean meridional circulation.
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.

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 hard to see a clear pattern of orographically forced vertical motions. Such motions do of course 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 850 mb 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, deserts, like the Sahara, are regions of sinking motion.

### 3.7 Angular momentum

As discussed in introductory physics textbooks (e.g. Feynman et al., 1962), the angular momentum per unit volume of a particle, \( \mathbf{L} \), with respect to some origin, is a vector, given by the cross product of the particle’s linear momentum (per unit volume), \( \rho \mathbf{V} \), and the displacement vector separating the particle from the origin, \( \mathbf{r} \):

\[
\mathbf{L} = \mathbf{r} \times \rho \mathbf{V}.
\]  \hspace{1cm} (3.51)

The angular momentum vector of the Earth’s atmosphere, with respect to an origin at the center of the Earth, could be computed by applying (3.51) to each air parcel, and integrating over the entire atmosphere.

Such a computation would reveal that the angular momentum of the air is mostly due to the rotation of the Earth. In addition, the motion of the air relative to the Earth includes strong jets that are oriented very nearly along latitude circles. For these reasons, the angular momentum vector for the atmosphere as a whole is very nearly parallel to the axis of the Earth’s rotation, and in practice when we discuss the angular momentum of the atmosphere we are almost always concerned with the component of the angular momentum vector that is parallel to the axis of the Earth’s rotation. This component of the angular momentum vector, per unit mass, is denoted by \( M \) in the discussion below, and is given by

\[
M \equiv (\Omega a \cos \varphi + u)a \cos \varphi.
\]  \hspace{1cm} (3.52)

The \( \Omega \) term of (3.52) represents the angular momentum due to the Earth’s rotation. The \( u \) term represents the relative angular momentum due to the rotation of the atmosphere relative to the Earth’s surface. Fig. 3.24 shows the zonally averaged total angular momentum and relative angular momentum, both per unit mass, for January and July, as analyzed by ECMWF. Note that the absolute angular momentum is generally much larger than the relative angular momentum. The absolute angular momentum varies mainly 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 per unit mass resemble those of the zonally averaged zonal wind (c.f. Fig. 3.8). In the tropical upper troposphere, the absolute angular momentum contours “tilt” slightly towards the winter pole, in both seasons. This is an indication that there is some tendency for the air flowing poleward in the main Hadley cell to conserve its absolute angular momentum as it goes.

The Lagrangian derivative of (3.52) can be written as
Figure 3.24: The absolute (left panels) and relative (right panels) atmospheric angular momentum per unit mass, for January and July, as analyzed by ECMWF. The units are $10^7$ m$^2$ s$^{-1}$.
Collecting terms, and using \( v = a \frac{D \phi}{D t} \), we can rewrite (3.53) as

\[
\frac{DM}{Dt} = a \cos \phi \frac{Du}{Dt} - (2 \Omega a \cos \phi + u) v \sin \phi .
\] (3.54)

To derive an expression for \( a \cos \phi \frac{Du}{Dt} \), which appears on the right-hand side of (3.54), we use the zonal component of (3.1), which can be written as

\[
\frac{Du}{Dt} = f v + \frac{u v \tan \varphi}{a} - \frac{1}{a \cos \phi} \frac{\partial \phi}{\partial \lambda} - \alpha \frac{\partial F_u}{\partial z} .
\] (3.55)

Multiply (3.55) by \( a \cos \phi \) to obtain

\[
a \cos \phi \frac{Du}{Dt} = f v a \cos \phi + u v \sin \phi - \frac{\partial \phi}{\partial \lambda} - a \cos \phi \left( \alpha \frac{\partial F_u}{\partial z} \right) = (2 \Omega a \cos \phi + u) v \sin \phi - \frac{\partial \phi}{\partial \lambda} - a \cos \phi \left( \alpha \frac{\partial F_u}{\partial z} \right) .
\] (3.56)

Substituting (3.56) into (3.54), we find that

\[
\frac{DM}{Dt} = - \frac{\partial \phi}{\partial \lambda} - a \cos \phi \left( \alpha \frac{\partial F_u}{\partial z} \right) .
\] (3.57)

This shows that \( M \) is conserved following a particle in the absence of (zonal) pressure torques and frictional torques. Note that (3.57) is considerably simpler than (3.55), indicating that angular momentum is more nearly conserved than linear momentum is.

Now use the continuity equation, (3.3), to convert (3.57) to flux form:

\[
\frac{\partial M}{\partial t} + \nabla \cdot (VM) + \frac{\partial}{\partial p} (\omega M) = - \frac{\partial \phi}{\partial \lambda} + g \frac{\partial F_u}{\partial p} (a \cos \phi) .
\] (3.58)

Integrate (3.58), term by term, through the entire column, and move the integrals inside the
derivatives, taking into account variations in the limits of integration. For example,

$$\int_{0}^{p_S} \frac{\partial M}{\partial t} dp = \frac{\partial}{\partial t} \left( \int_{0}^{p_S} M \, dp \right) - M_S \frac{\partial p_S}{\partial t}. \quad (3.59)$$

Here $p_S$ is the surface pressure. In this way, we find that

$$\frac{\partial}{\partial t} \left( \int_{0}^{p_S} M \, dp \right) + \nabla \cdot \left( \int_{0}^{p_S} \mathbf{V} M \, dp \right) - M_S \left( \frac{\partial p_S}{\partial t} + \mathbf{V}_S \cdot \nabla p_S - \omega_S \right)$$

$$= - \frac{\partial}{\partial \lambda} \int_{0}^{p_S} \phi dp + \phi S \frac{\partial \phi S}{\partial \lambda} + ga \cos \phi (F_u)_S$$

$$= - \frac{\partial}{\partial \lambda} \left( \int_{0}^{p_S} \phi dp - \phi S p_S \right) - p_S \frac{\partial \phi S}{\partial \lambda} + ga \cos \phi (F_u)_S. \quad (3.60)$$

The last term on the left-hand side of (3.60) drops out, because the condition that no mass crosses the Earth’s surface can be expressed as

$$\frac{\partial p_S}{\partial t} + \mathbf{V}_S \cdot \nabla p_S - \omega_S = 0. \quad (3.61)$$

Finally, integrate (3.62) around a latitude circle, to obtain:

$$\frac{\partial}{\partial t} \left( \int_{0}^{2\pi} \int_{0}^{p_S} M \, dp \, d\lambda \right) + \frac{1}{a \cos \phi \frac{\partial}{\partial \phi}} \int_{0}^{2\pi} \int_{0}^{p_S} \mathbf{V} \cos \phi \, M \, dp \, d\lambda$$

$$= - \int_{0}^{2\pi} p_S \frac{\partial \phi S}{\partial \lambda} d\lambda + ga \cos \phi \int_{0}^{2\pi} (F_u)_S d\lambda. \quad (3.62)$$

The first term on the right hand side of (3.62) represents the effects of “mountain torque.” It vanishes if either $p_S$ or $\phi_S$ is independent of $\lambda$. Near a mountain range, we expect higher pressure on the upstream side and lower pressure on the downstream side. This is called “form drag.” See Fig. 3.25. The spatial correlation shown in the figure need not occur at a particular moment with respect to a particular mountain range, but it does tend to occur in an average sense.

You should be able to prove that the mountain torque would vanish if $p_S$ were a function of $\phi_S$ only, i.e., if the data in Fig. 3.2 fell exactly onto a single curve. This would be expected in the absence of dynamical processes, e.g., if the atmosphere were at rest and in balance.
Angular momentum is conserved in the absence of mountain torque and surface friction. Note that both torques in (3.62) involve angular momentum exchange between the atmosphere and the underlying surface. 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 length of the day and the radius of the Moon's orbit are both increasing slowly. To the extent that we can neglect these small effects, the angular momentum of the Earth-atmosphere system must remain constant with time. Evidence that this is really true is shown in Fig. 3.26. 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.

Because the tendency of the angular momentum of the atmosphere is negligible in a long time average, the angular momentum exchanged between the Earth (i.e., the solid Earth and oceans together) and the atmosphere must average out to zero. If the angular momentum of the atmosphere increases, the Earth isolation must slow down and vice versa. This can be seen from the time average of (3.62):

\[
\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \int_0^{2\pi} \int_0^{p_S} \cos \phi \bar{v} M \, dp \, d\lambda = -\int_0^{2\pi} \left( p_S \frac{\partial \phi_S}{\partial \lambda} \right) d\lambda + ga \cos \phi \int_0^{2\pi} (\bar{F}_u) S d\lambda. \quad (3.63)
\]

The left-hand side of this equation represents the meridional transport of vertically and zonally integrated atmospheric angular momentum. The two terms on the right-hand side represent the sources and sinks due to mountain torque and surface friction torque, respectively.

The surface frictional stress transfers positive angular momentum from the atmosphere to the ocean-solid Earth system where the surface winds are westerly, and positive angular momentum flows back into the atmosphere where the surface winds are easterly. The mountain torques generally have the same sign as the frictional torques. It follows that for the
An Introduction to the General Circulation of the Atmosphere

Fig. 3.27 shows the observed zonally averaged zonal wind stress for the oceans only, as compiled by Han and Lee (1983). In this figure, positive zonal wind stresses indicate that positive (westerly) atmospheric momentum is being transferred to the ocean; such positive values naturally occur in regions where the surface wind is westerly, i.e. in the middle latitudes of both hemispheres. Negative zonal surface stresses, somewhat weaker in magnitude, occur in the tropics and subtropics, in association with the trade winds.

Fig. 3.28 shows estimates of the zonally averaged friction and mountain torques. The frictional torque dominates overall, but mountain torque is appreciable in the Northern Hemisphere.

Fig. 3.29 shows the flow of angular momentum from the tropics, where the surface easterlies extract it from the oceans and solid earth, to middle latitudes, where the surface westerlies deposit it in the oceans and the solid earth. Oort (1989) argues that the return flow occurs largely through the movement of the continents, rather than through the circulation of the oceans.

A tendency towards conservation of angular momentum in the air flowing poleward in the upper branch of the Hadley cell leads to the formation of westerlies, which take their maximum values in the jet streams near the latitudes where the Hadley circulations stop. The westerly jets can thus be interpreted as consequences of the poleward flow of air in the Hadley circulations. Further discussion is given later.
3.8 Temperature

Fig. 3.30 shows the (somewhat idealized) vertical distribution of temperature from the surface to the 100 km level. In the lowest 10 to 15 km, the temperature decreases monotonically upward; this is the troposphere. Above the upper boundary of the troposphere, which is called the tropopause, the temperature becomes uniform with height and then begins to increase upward. This region is the stratosphere. The upward increase of temperature 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 vertical distribution of ozone is also shown in Fig. 3.30. The stratopause occurs near the 1 mb (~50 km) level, which is above the ozone layer. Above the stratosphere is the mesosphere, within which the temperature decreases upwards again. The mesopause is near the 0.01 mb level. Above the mesopause is the "thermosphere," within which the temperature increases upwards again. Within the thermosphere, the composition of "dry air" begins to change significantly, the air loses the qualities of a perfect gas, electromagnetic forces become important for the dynamics of the atmosphere, and the molecular viscosity (per unit mass) becomes very large. In this course, we consider the troposphere and the stratosphere, but we will not discuss the mesosphere or the thermosphere.

Fig. 3.31 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 mb the coldest air is over the Equator. In fact, some of the lowest temperatures in the atmospheres 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 mid-latitudes, 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.

1. The 100 km level is approximately the height at which objects entering the Earth’s atmosphere begin to experience significant frictional heating.
The midlatitude low-level temperature gradients are quite strong in the winter hemisphere. Above 200 mb, 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 mb. This suggests easterlies in the summer stratosphere, which do in fact occur. In the winter hemisphere, the warmest 100 mb temperatures occur in middle latitudes. The strong decrease of temperature between midlatitudes and the poles is consistent with the polar night jets mentioned earlier.

In January, a temperature “inversion” (i.e. temperature increasing upward) appears over the North Pole. Generally speaking, the lapse rate, $\frac{\partial T}{\partial \phi}$, is largest in the tropics, and smaller (or even negative) near the poles.

Thermal wind balance between the meridional temperature gradient and the vertical shear of the zonal wind is well satisfied, as can be seen by comparison of Fig. 3.8 and Fig. 3.31.

It is interesting to compare the latitude-height section of zonally averaged temperature, shown in Fig. 3.31, with the corresponding latitude-height section of zonally

---

**Figure 3.28:** Meridional profiles of the total surface torque and the mountain torque for annual mean conditions. The horizontal axis is labeled in HUs, which are “Hadley units.” A Hadley unit is $10^{18}$ kg m$^2$ s$^{-2}$. Note that the scales in the two panels are different. Also note the difference in sign conventions relative to Fig. 3.27. (From Oort 1989, after Newton, 1971.)
An Introduction to the General Circulation of the Atmosphere

3.8 Temperature

71

\[ \theta \equiv T \left( \frac{p_0}{p} \right)^\kappa, \]

is a constant reference pressure, usually taken to be 1000 mb, and \( \kappa \equiv \frac{R}{c_p} \). The key physical facts about potential temperature are that it is conserved under dry adiabatic processes, and that it increases upward in a statically stable atmosphere. This upward increase is evident in the figure. 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 3.29: Cross sections of: a) the zonally averaged zonal flow (m s\(^{-1}\)), and b) the stream function of the zonally averaged relative angular momentum transport in units of \(10^{18} \text{ kg m}^2 \text{ s}^{-2}\). From Oort 1989.
Hoskins (1991) distinguishes the following three regimes: 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. The “Middleworld” has 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. 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. From the zonally averaged data, it appears that the largest potential temperatures in the Underworld are on the order of 300 K, although in reality much larger values do occur near the Earth’s surface, e.g., on a summer afternoon in the Sahara desert.

Fig. 3.33 shows maps of the 850 mb temperature for January and July, respectively. Fig. 3.34 shows the corresponding results for 200 mb. The expected winter-to-summer warming at 850 mb is very obvious in the Northern Hemisphere, but less so 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.
Naturally, the temperature gradient points mainly from the poles toward the tropics at 850 mb, but stationary eddies are plainly visible in the Northern Hemisphere in January. There are some regions in which the mean temperature actually increases poleward, at 850 mb, 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 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 mb are found in high latitudes, especially in winter. The eddy pattern is much stronger at 200 mb than at 850 mb. Particularly noticeable are the maxima over the North Pacific in January, over eastern North America in January, and over southern Asia in July. Note than in each of these regions the 200 mb temperature increases as we move from the tropics towards middle latitudes. This implies a tendency for the westerlies to weaken above this level.

3.9 A view in potential temperature coordinates

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). Any variable can be used as a vertical coordinate provided that it is monotonic with height; see the Appendix on vertical coordinates for a discussion. Potential temperature is almost but not quite always monotonic with height on the large scale. The exceptions are sufficiently minor that for most purposes they can be ignored, and we will ignore them in this course. We will use potential temperature coordinates at various places throughout the course.

The equations of motion in potential temperature coordinates can be written as

Figure 3.33: Maps of the 850 mb temperature for January and July. The contour interval is 5 K.
An overview of the observations follows.

### Physical Principle

Conservation of horizontal momentum

\[
\frac{DV}{Dt} + \left( f + \frac{u \tan \Theta}{a} \right) \mathbf{k} \times \mathbf{V} = - \nabla \theta s + g \frac{\partial F_V}{\partial p} \tag{3.64}
\]
3.9 A view in potential temperature coordinates

<table>
<thead>
<tr>
<th>Physical Principle</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrostatic equation</td>
<td>( \frac{\partial s}{\partial \theta} = \Pi ) (3.65)</td>
</tr>
<tr>
<td>Conservation of mass</td>
<td>( \left( \frac{\partial m}{\partial t} \right)<em>\theta + \nabla</em>\theta \cdot (m \mathbf{V}) + \frac{\partial}{\partial \theta} (m \dot{\theta}) = 0 ) (3.66)</td>
</tr>
<tr>
<td>Conservation of thermodynamic energy</td>
<td>( \dot{\theta} = g \frac{\partial F_\theta}{\partial p} + \left( \frac{\theta}{c_p T} \right) (Q_{rad} + Q_{lat}) ) (3.67)</td>
</tr>
<tr>
<td>Conservation of water vapor</td>
<td>( \frac{D q_v}{Dt} = g \frac{\partial F_v}{\partial p} - \frac{Q_{lat}}{L} ) (3.68)</td>
</tr>
</tbody>
</table>

Using (3.66) this can be rewritten as

\[
\frac{\partial}{\partial t} (m q_v) \bigg|_\theta + \nabla_\theta \cdot (m \mathbf{v} q_v) + \frac{\partial}{\partial \theta} (m \theta q_v) = m \left( \frac{\partial F_v}{\partial p} - \frac{Q_{lat}}{L} \right)
\] (3.69)

Equation of state

\( p \alpha = RT \) (3.70)

Compare (3.64) - (3.70) with (3.1)-(3.8). In (3.65),

\[ \Pi \equiv c_p \left( \frac{T}{\theta} \right) = c_p \left( \frac{P}{P_0} \right)^\kappa \] (3.71)

is the Exner function. In (3.64) and (3.65),

\( s \equiv c_p T + g z \) (3.72)

is the “Montgomery potential,” which is also called the “Montgomery stream function” and the “dry static energy.” With the use of hydrostatics, it can be shown that

\[ \frac{\partial s}{\partial z} = \Pi \frac{\partial \theta}{\partial z} \] (3.73)

so that \( s \), like the potential temperature, increases upward in a dry statically stable atmosphere. In (3.66) and (3.69)
An overview of the observations

\[ m \equiv -\frac{1}{g} \frac{\partial p}{\partial \theta} \]  

(3.74)

is the “pseudo-density,” which measures the amount of mass between two \( \Theta \)-surfaces. In the absence of heating, parcels remain on surfaces of constant \( \Theta \). As a result, the “vertical advection” terms of all equations vanish in the absence of heating. This is one of the major advantages of \( \Theta \)-coordinates.

Fig. 3.35 shows zonal and time-averages, plotted as functions of latitude and potential temperature of pressure, the pseudodensity, zonal wind, and meridional wind, according to Edouard et al. (1997). One of the advantages of \( \Theta \)-coordinates is that the potential vorticity (PV) equation can be derived very simply from the momentum equation and the continuity equation – much more simply than in \( p \)-coordinates, for example. The PV is a highly conservative variable that strongly controls the large-scale dynamical fields (e.g. Hoskins et al., 1985). Fig. 3.36 shows the zonally averaged PV, which is defined by

\[ \text{PV} = -(\zeta + f) \frac{\partial \theta}{\partial p} g, \]

(3.75)

where \( \zeta \) is the vertical component of the vorticity in which the horizontal derivatives are taken along isentropic surfaces. The figure shows that in middle latitudes the tropopause roughly coincides with a constant-PV surface, namely the \( \pm 2 \text{PVU} \) surface. A \( \text{PVU} \) or “potential vorticity unit” is \( 10^{-6} \text{m}^2 \text{Ks}^{-1} \text{kg}^{-1} \). Note from Fig. 3.35 that in the tropics the tropopause roughly coincides with a constant-\( \Theta \) surface, namely the 390 K surface. The stratosphere is a region of high PV, and intrusions of stratospheric air into the troposphere are characterized by anomalously large values of the PV. The troposphere has a relatively uniform PV, suggesting that PV is being mixed there. Stratospheric air entering the troposphere can be recognized by its large (absolute) PV.

3.10 The global distribution of water vapor

The total water mixing ratio, \( q_T \), can be separated into various subspecies by phase and/or particle size, as follows:

\[ q_T = q_v + q_c + q_i + q_r + q_s. \]

(3.76)

Here \( q_v \) is the mixing ratio of vapor; \( q_c \) is the mixing ratio of “cloud water,” which consists of droplets too small to precipitate significantly; \( q_i \) is the mixing ratio of “cloud ice,” which consists of crystals too small to precipitate significantly; \( q_r \) is the mixing ratio of rain water; and \( q_s \) is the mixing ratio of snow. In fact, much more elaborate break-downs are possible based on fine-grain distinctions among particles of various sizes. Such complexity is beyond the scope of this course, and in fact we will even ignore the ice phase for simplicity. We therefore replace (3.76) by
where $l$ represents the mixing ratio of liquid water, including drops of all sizes.

Conservation of total water is expressed by

$$
\frac{\partial}{\partial t} (\rho q_T) + \nabla \cdot (\rho \mathbf{V} q_T + \mathbf{F}_{q_T} - \mathbf{r}) = 0,
$$

(3.78)
where $q_T$ is the mixing ratio of water vapor, i.e.

$$ q_T = \frac{\rho_{\text{water}}}{\rho}, $$

(3.79)

$F_{q_T}$ is the vector flux of moisture due to molecular diffusion, and $r$ is the vector flux of total water due to precipitation. Note that if we put $q_T \equiv$ constant and drop the molecular diffusion term, then (3.78) reduces to the continuity equation. If we multiply the continuity
equation by \( q_T \), and subtract it from (3.78), we can re-write moisture conservation as

\[
\rho \frac{Dq_T}{Dt} = -\nabla \cdot (F_{q_T} - \mathbf{r}).
\] (3.80)

We refer to (3.78) as the “flux form” of the moisture conservation equation, and to (3.80) as the “advective form.”

Integration of (3.78) through the entire atmospheric column, and use of the lower boundary condition, gives

\[
\frac{\partial}{\partial t} \left( \int_{z_s}^{\infty} \rho q_T \, dz \right) + \nabla \cdot \left( \int_{z_s}^{\infty} \rho \mathbf{V}_T q_T \, dz \right) = \{ (F_{q_T})_{S} - \mathbf{r} \} \cdot \mathbf{n}_S.
\] (3.81)

where \( \mathbf{n}_S \) is a unit vector normal to the Earth’s surface.

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

\[
\frac{Dq_v}{Dt} = -\frac{1}{\rho} \nabla \cdot F_{q_v} - C,
\] (3.82)

\[
\frac{Dl}{Dt} = \frac{1}{\rho} \nabla \cdot \mathbf{r} + C.
\] (3.83)

Here \( C \) is the net condensation rate per unit mass, which converts vapor into liquid. Note that there is no precipitation term in (3.82), and no molecular diffusion term in (3.83). We have essentially assumed that \( F_{q_T} = F_{q_v} \) and \( F_{q_v} = 0 \).

Representative vertical distributions of relative humidity and mixing ratio are shown in Fig. 3.37. There is a nearly monotonic upward decrease of the relative humidity in the climatology, although of course this is not necessarily true at any particular time and place. The mixing ratio profile is representative of middle latitudes, in the annual mean. Note that the horizontal scale is logarithmic. The mixing ratio decreases by four orders of magnitude between the surface and the lower stratosphere. It actually increases upward in the stratosphere, because there is a chemical source of water vapor in the stratosphere due to the oxidation of methane.

Fig. 3.38 shows the latitude-height distributions of the zonally averaged water vapor mixing ratio for January and July, respectively. Because the data presented in these figures have been fed through the analysis/forecast system of a numerical weather prediction center (namely ECMWF), which can easily distort the distribution of water vapor, they should be taken with a grain of salt.

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
An overview of the observations

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. Despite the extreme dryness of the air aloft (a few parts per million by volume in the stratosphere), upper tropospheric water vapor is very important radiatively.

Since the mixing ratio is greatest near the surface, regions of low-level mass convergence, such as those apparent in the zonally averaged meridional wind (Fig. 3.12) 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 moist.

Fig. 3.39 shows maps of the 850 mb 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.

The distributions of temperature and moisture can be combined in a variable called the “moist static energy,” which is defined by

![Figure 3.37: Illustrative vertical profiles of a) the relative humidity (from Manabe and Wetherald, 1967), and b) the water vapor mixing ratio (from Dutton, 1976).](image)
The global distribution of water vapor

where $q_v$ is the water vapor mixing ratio. The moist static energy is of interest in part because it is conserved under both moist and dry adiabatic processes, and even under pseudoadiabatic processes in which precipitation occurs; these conservation properties are proven later. The
moist static energy is somewhat analogous to the equivalent potential temperature. The latitude-height distribution of moist static energy, as analyzed by ECMWF, is shown in Fig. 3.40. In the upper levels, the moist static energy increases upwards; in the tropics and in the midlatitude summer, the moist static energy decreases upwards in the lower troposphere; and especially in the tropics the moist static energy has a minimum in the middle troposphere. The reasons for this distribution are as follows: Above the middle troposphere, or at any altitude near the poles, the water vapor mixing ratio is negligible. In that case, the moist static energy reduces (approximately) to the dry static energy, mentioned earlier, which is given by

Figure 3.39: Maps of the 850 mb water vapor mixing ratio. The contour interval is 1 g kg\(^{-1}\) shaded values > 10 g kg\(^{-1}\).
The global distribution of water vapor. Recall that normally increases upwards. At lower levels, where water vapor is plentiful, especially in the tropics, the upward decrease of the water vapor mixing ratio overwhelms the upward increase of $s$, so that the moist static energy decreases upwards. It follows that the moist static energy has a minimum in the middle troposphere in the tropics.

$s \equiv c_p T + g z$. Recall that $s$ normally increases upwards. At lower levels, where water vapor is plentiful, especially in the tropics, the upward decrease of the water vapor mixing ratio overwhelms the upward increase of $s$, so that the moist static energy decreases upwards. It follows that the moist static energy has a minimum in the middle troposphere in the tropics.

Figure 3.40: The observed latitude–height distribution of the zonally averaged moist static energy, in kJ kg$^{-1}$, as analyzed by ECMWF.
3.11 Precipitation

The zonally averaged surface precipitation rates for January and July are shown in Fig. 3.41. Fig. 3.42 shows maps of the precipitation rate for January and July. Because precipitation tends to be very “spotty” in both space and time, average values are difficult to determine accurately even under ideal conditions. As a result, the values plotted are uncertain by at least 25%. Certainly the data for such remote regions as the South Pacific Ocean cannot be strongly defended.

The zonally averaged precipitation has its strongest maximum in the tropics, with secondary maxima in the middle latitudes. The tropical maximum is associated with the intertropical convergence zone and the monsoons, while the middle latitude maxima are associated with baroclinic wave activity in the winter, and monsoon circulations in the summer. There are minima in the subtropics, where the major deserts occur. The global mean of the precipitation rate is around 3 mm day\(^{-1}\); again, this value is uncertain by perhaps 25%.

Fig. 3.42 clearly shows that the rainiest regions of the world are in the tropics. There are major seasonal shifts in the locations of the tropical precipitation. 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 India and neighboring regions of Southeast Asia, and to the north of the maritime continent, off the east coast of tropical Asia. The seasonal shifts of tropical precipitation are quite spectacular, and are most clearly seen in the longitudes of South America, Africa, India, and Southeast Asia.
As is clear from Fig. 3.41, minima of the zonally averaged precipitation occur in the subtropics, and secondary maxima occur in the middle latitudes. Midlatitude precipitation is also highly variable. For example, the precipitation over northern Asia occurs mainly in the summer. The warm currents off the east coasts of North America and Asia receive heavy precipitation mainly in January. The northwestern portion of the United States receives heavy precipitation in January but not in July.

Regions that receive plentiful precipitation throughout the year include eastern North
As can be seen by comparing Fig. 3.23 and Fig. 3.42, there is a strong correlation between vertical motion and precipitation. Precipitation maxima correspond to regions of latent heat release inside the column. Regions of latent heat release also tend to be regions in which the net radiative heating of the atmosphere is positive, because of the preponderance of high cold clouds that block the emission of longwave radiation to space. The patterns of latent and radiative heating thus tend to reinforce each other. Now, rising motion tends to occur where there is heating, and this is particularly true in the tropics. We return to the analysis of Charney (1963). As already discussed, Charney concluded that horizontal temperature gradients tend to be weak in the tropics. Referring to the thermodynamic energy equation, (3.4), we re-write it in expanded form, as

\[ c_p \left( \frac{\partial T}{\partial t} + \mathbf{V} \cdot \nabla T + \omega \frac{\partial T}{\partial p} \right) = \omega \alpha + g \frac{\partial F}{\partial p} T + Q_{rad} + Q_{lat}. \]  

(3.85)

As already explained in Chapter 1, the time average of the time-rate-of-change term can be made as small as desired if the averaging interval is chosen to be long enough. For the case of atmospheric temperature, an averaging interval of one month is sufficient to make the time-rate-of-change term of (3.85) negligible. Then the time average of (3.85) can be written as

\[ \overline{\omega} \frac{\partial}{\partial p} (c_p T + \phi) = g \frac{\partial F}{\partial p} c_p T + \overline{Q_{rad}} + \overline{Q_{lat}}. \]  

(3.86)

Here the overbar denotes a time average, and the hydrostatic equation has been used. The terms on the left-hand side of (3.86) represent dry adiabatic processes, while those on the left-hand side represent “heating” of various kinds. The quantity \( c_p T + \phi \) is sometimes called the dry static energy. We show later that the dry static energy increases upwards when the atmosphere is stably stratified in the dry sense, i.e. \( \frac{\partial}{\partial p} (c_p T + \phi) < 0 \). The vertical motion term of (3.86) therefore represents cooling \( (-\omega \frac{\partial}{\partial p} (c_p T + \phi) < 0) \) when the air is rising \( (\omega < 0) \) and warming \( (-\omega \frac{\partial}{\partial p} (c_p T + \phi) > 0) \) when the air is sinking \( (\omega > 0) \). Eq. (3.86) simply says that in a time average, heating has to be balanced by a combination of horizontal and vertical advection.

Now recall Charney’s conclusion that horizontal temperature gradients are negligible in the tropics. This means that the first term on the left-hand side of (3.86) is relatively small in the tropics. It follows that the only way to balance tropical heating or cooling is through vertical motion, and so clearly there should be a very strong correspondence between the pattern of vertical motion and the pattern of heating. This is what we see when we compare Fig. 3.23 and Fig. 3.42. Tropical rising motion occurs almost exclusively where latent and radiative heating are active, i.e. where \( g \frac{\partial F}{\partial p} c_p T + Q_{rad} + Q_{lat} > 0 \), and tropical sinking motion occurs almost exclusively where radiative cooling is dominant, i.e. where
These conclusions apply fairly well even in the middle latitudes in summer, simply because horizontal temperature gradients are weak there as well. They definitely do not apply in the middle-latitude winter, where the horizontal advection term of (3.86) is critically important.

3.12 Surface fluxes due to turbulence

The atmosphere contains a variety of small-scale motions that significantly affect the global-scale circulation. These include cumulus convection, turbulence in the “planetary boundary layer” (PBL) near the Earth’s surface, turbulence above the PBL in clouds and in regions of strong shear, and gravity waves excited by a variety of mechanisms including flow over topography and convection. These phenomena can be lumped together under the generic heading “small-scale eddies.”

Small-scale eddies have important effects on the general circulation. The most important examples are:

- the vertical fluxes of energy, moisture, and momentum due to turbulence, primarily in the boundary layer and in clouds at any level;
- the vertical fluxes of energy, moisture, and momentum due to cumulus convection;
- the vertical flux of momentum due to small-scale gravity waves, which exert their effects mainly in the stratosphere and above.

Each of these phenomena will be discussed in more detail later in this course.

Fig. 3.43 shows maps of the “surface sensible heat flux” for January and July. The surface sensible heat flux is denoted by $F_{cpT}$ in Eq. (3.4). It is essentially a flux of temperature, multiplied by $c_p$, the specific heat of air at constant pressure, to give units of an energy flux. As can be seen in (3.4), the sensible heat flux directly modifies the temperature.

At this point, we digress to explain what the turbulent fluxes are. The detailed structures and life cycles of individual small-scale eddies are not crucial for the global circulation; it is only the average properties of many eddies working together that can exert significant effects on the large scale. For this reason, we introduce averaging operators, following the approach of “Reynolds Averaging,” which is explained in detail in an Appendix. Such averaging gives rise to terms that represent the statistical effects of the small-scale eddies on the large-scale circulation. Let

$$
\mathbf{V} = \bar{\mathbf{V}} + \mathbf{V}',
$$

$$
A = \bar{A} + A',
$$

(3.87)

where $(\bar{\cdot})$ denotes an averaged variable, and $(\cdot)'$ denotes a departure from or fluctuation.
Figure 3.43: Maps of the surface sensible heat flux, based on ECMWF reanalyses. These are not real observations, although they are based on observations. The contour interval is 10 W m$^{-2}$. Values higher than 50 W m$^{-2}$ are shaded.
about the averaged variable. In the study of the general circulation, the “large-scale” variables are the averaged quantities. The primed quantities represent fluctuations on smaller scales. In Eq. (3.87), \( \mathbf{V} \) is the vector wind and \( A \) is an arbitrary scalar variable. We assume that

\[
[(\bar{\cdot})]' = 0 \quad \text{and} \quad [\bar{\cdot}] = 0 \quad \text{and} \quad [(\bar{\cdot}')(\bar{\cdot})]' = 0 . \tag{3.88}
\]

The extent to which these assumptions are actually valid depends on the precise definition of the averaging operator; further discussion is given in the available handout on Reynolds Averaging.

We can show that the averaged temperature flux, \( \rho \bar{V} \bar{T} \), satisfies

\[
\rho \bar{V} \bar{T} = \rho \bar{V} \bar{T} + \rho \bar{V}' \bar{T} + \rho \bar{V} \bar{T}' + \rho \bar{V}' \bar{T}'
\]

\[
= \rho \bar{V} \bar{T} + \rho \bar{V}' \bar{T}' . \tag{3.89}
\]

The second line shows that the averaged temperature flux consists of two contributions. The first arises from the averaged wind and averaged temperature, and the second arises from the interactions of the fluctuating wind with the fluctuating temperature. The second contribution, due to the small-scale motions, can be quite important.

At this point you may be wondering whether the eddy-flux term of (3.89) should be expanded further in order to take into account contributions from \( \bar{\cdot}' \), the fluctuating part of the density. The answer is that such contributions are negligible, so that we just consider \( \rho \) in (3.89) to be the mean-state density, but omit the overbar for convenience. When we are working in potential temperature coordinates, however, the fluctuating part of the pseudo-density, \( m \equiv -\frac{1}{g} \frac{\partial \rho}{\partial \theta} \), can be important. This will be discussed later.

There are of course additional “molecular” fluxes due to the random molecular motions, which give rise to molecular conduction, diffusion, and viscosity. The turbulent fluxes are typically many orders of magnitude larger than the molecular fluxes, so we neglect the latter, for the most part, in atmospheric science.

It is important to remember, however, that molecular effects are ultimately responsible for the dissipation of kinetic energy, which is essentially a conversion of kinetic energy into internal energy (i.e. of macroscopic kinetic energy into microscopic kinetic energy). Similarly, the molecular thermal conductivity is ultimately responsible for the dissipation of thermal fluctuations. It is an amazing fact that even though the molecular processes act on scales of a few millimeters, they have profound effects on the global-scale circulation of the atmosphere! Further discussion of dissipation is given later.

When considering large-scale motions, we can assume that

\[
\nabla \cdot \rho \bar{V}' \bar{T} \equiv \frac{\partial}{\partial z} (\rho \bar{w}' \bar{T}') , \tag{3.90}
\]

because the depth of the atmosphere is shallow compared to horizontal extent of the large-
scale motions. Note that this is true even if the horizontal component of the flux vector is comparable in magnitude to the vertical component; it is the flux divergence, rather than the flux itself, that matters. Similarly, we have:

\[ \nabla \cdot \rho \nabla \theta' \equiv \frac{\partial F_\theta}{\partial z}, \quad (3.91)\]

\[ \nabla \cdot F_q \equiv \frac{\partial F_q}{\partial z}, \quad (3.92)\]

\[ \nabla \cdot F_V \equiv \frac{\partial F_V}{\partial z}. \quad (3.93)\]

Here we adopt, for convenience, a slight notational anomaly: From this point on, we use \( F_V \) to denote the upward turbulent flux of the (vector) horizontal momentum, \( F_\theta \) to denote the upward turbulent flux of potential temperature, and \( F_q \) to denote the upward turbulent flux of moisture. Note that \( F_\theta \neq |F_\theta| \), and \( F_q \neq |F_q| \).

The turbulent fluxes at the Earth’s surface are typically assumed to satisfy “bulk aerodynamic” formulae, e.g.

\[(F_\theta)_S = \rho_S c_T |V_S| (\theta_g - \theta_a), \quad (3.94)\]

\[(F_q)_S = \rho_S c_T |V_S| (q_g - q_a), \quad (3.95)\]

\[(F_V)_S = -\rho_S c_D |V_S| V_S. \quad (3.96)\]

Here the subscript \( S \) denotes a surface value, the subscript \( g \) denotes a value representative of the lower boundary, and the subscript \( a \) represents a value representative of a level inside the atmosphere but near the surface. The quantities \( c_T \) and \( c_D \) are the heat-and-moisture transfer coefficient and the drag coefficient, respectively. These coefficients are nondimensional, and (3.94)-(3.96) can be regarded as definitions of the coefficients. In practice, the coefficients \( c_T \) and \( c_D \) are determined as empirical functions of the surface roughness and the near-surface static stability and wind shear, and are used in (3.94)-(3.96) to compute the fluxes. For further discussion, see Stull (1988).

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. Over land, it is much more difficult to determine the appropriate value of \( q_g \), which 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, etc.).
Note from the minus sign in (3.96) that the near-surface momentum flux has a direction opposite to the surface wind. This means, for example, that if the near-surface wind is westerly, there will be a downward flux of westerly momentum into the Earth’s surface.

As can be seen from (3.94), 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. Similarly the latent heat flux is upward when the lower boundary is wetter than the air, and it dries the boundary while moistening the air. Finally, as mentioned above, the surface friction transfers momentum from the air to the surface, thus tending to reduce the momentum of the near-surface air. Because all three of these surface fluxes are self-destructive, they can continue over time only if some process acts to maintain them. For example, solar radiation absorbed by the ground can maintain an upward surface sensible heat flux, subsidence drying of the air near the surface can maintain an upward surface moisture flux, and the large-scale pressure gradient can maintain a surface momentum flux.

Having now finished our digression on the nature of the turbulent fluxes, we return to Fig. 3.43. 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 continents out over warm ocean currents. From (3.94), we see that strong sensible heat fluxes are to be expected under such conditions. Large values 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, because a downward sensible heat flux tends to damp the turbulence.

Maps of the surface moisture flux are shown in Fig. 3.44. Note the maxima over the subtropical oceans. The tradewinds diverge the evaporated moisture from the sub tropics and converge it into the deep tropics, where it is converted to precipitation. There are no large downward surface moisture fluxes.

Fig. 3.45 shows maps of the magnitude of the surface wind stress. This is the average of the magnitude, rather than the magnitude of the average vector. The surface stress is particularly strong in the storm-track regions. Note the maximum over the Arabian Sea in July, associated with the Somali jet.

Friction tends to make the wind near the surface depart from geostrophic balance, and flow down the pressure gradient, i.e., towards low pressure. To see how this works, look at Fig. 3.46. Because the air turns towards low pressure, surface lows tend to be regions of low-level convergence, and surface highs tend to be regions of low-level divergence. Of course, convergence near the surface must be balanced by divergence aloft, and vice versa. In this way, surface friction can influence the upper-tropospheric winds.

The effects of the surface wind stress tend 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. In particular, the east-to-west trade winds along the Equator 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), thus driving upwelling along the Equator. This is why the surface waters are colder along the Equator than they are on either side of the Equator. The equatorward winds along the west coasts of the continents drive equatorward currents, and also upwelling. Both the directions of the currents (from the poles) and the upwelling favor cold water, which is what is observed. See Fig. 3.47. Consider a surface wind blowing parallel to a coastline in the Northern Hemisphere, with the coastline on
Figure 3.44: Maps of the surface latent heat flux, based on ECMWF reanalyses. These are not real observations, although they are based on observations. The contour interval is 20 W m\(^{-2}\). Values higher than 150 W m\(^{-2}\) are shaded.
Figure 3.45: Maps of the magnitude of the surface wind stress based on satellite data. The contour interval is 0.02 Pa. Values higher than 0.2 Pa are shaded.
An overview of the observations

An Introduction to the General Circulation of the Atmosphere

its left side, as in the California high in July. The surface current moves to the right of the wind, i.e., is 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. Further discussion is given later.

3.13 A quick introduction to the effects of large-scale eddies on the zonally averaged f circulation

Just as small-scale eddies (turbulence, cumulus convection, and gravity waves) can produce vertical fluxes that affect the large-scale flow, “large-scale eddies” can produce both vertical and meridional fluxes that affect the zonally averaged flow. The large-scale eddies can be defined in terms of departures from the zonal mean.

As a first step, we adopt the following definitions, some of which have already been
In the following discussion, we consider various statistics derived from the three-dimensional fields of \( v \) and \( T \). Here \( v \) and \( T \) can stand for any variables. They could be the same variable.

The temporal covariance of \( v \) and \( T \) is

\[
\overline{vT} = (\overline{v} + v')(\overline{T} + T')
= \overline{v} \overline{T} + v'T' + vT' + v'T
= \overline{v} \overline{T} + vT'.
\]

(3.97)

The last line of (3.97) is only approximate, unless the averaging interval is infinite. Now decompose the first term of (3.97) into its zonal mean and eddy components:

\[
\overline{v} \overline{T} = ([v] + v^*) ([T] + T^*)
= ([v] + v^*) ([\overline{T}] + T^*)
= [\overline{v}]\overline{[T]} + v\overline{T} + [v]^*\overline{T} + v\overline{T}.\]

(3.98)

It follows that

\[
[\overline{v} \overline{T}] = [\overline{v}][T] + [v^{*} \overline{T}]. \quad \text{(stationary symmetric + stationary eddy)}
\]

(3.99)

Similarly,

\[
[v'T'] = [v'][T'] + [v'^{*} T'^{*}]. \quad \text{(transient symmetric + transient eddy)}
\]

(3.100)

Finally, substitution gives
An overview of the observations

\[
[\bar{v}T] = [\bar{v}][\bar{T}] + [\bar{v}' \bar{T}'] + [v'][\bar{T}'] + [v'\bar{T}'']
\]

stationary symmetric circulation  
stationary eddies  
transient symmetric circulation  
transient eddies

Some statistics of interest are:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>(v, T)</td>
<td>Time mean fields</td>
</tr>
<tr>
<td>2.</td>
<td>(\bar{vT})</td>
<td>Temporal covariance</td>
</tr>
<tr>
<td>3.</td>
<td>([v][T])</td>
<td>Total symmetric</td>
</tr>
<tr>
<td>4.</td>
<td>(\bar{v} \bar{T})</td>
<td>Total stationary</td>
</tr>
<tr>
<td></td>
<td>(1) \cdot (1)</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>(\bar{v'T'} = \bar{vT} - \bar{v} \bar{T})</td>
<td>Total transient</td>
</tr>
<tr>
<td></td>
<td>(2) - (4)</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>([\bar{v}] [\bar{T}])</td>
<td>Stationary symmetric</td>
</tr>
<tr>
<td></td>
<td>([1(1)] \cdot [1(1)])</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>([\bar{v}' \bar{T}'] = [\bar{v} \bar{T} - [\bar{v}][\bar{T}])</td>
<td>Stationary eddy</td>
</tr>
<tr>
<td></td>
<td>(4) - (6)</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>([v'][T'] = [v][T] - [\bar{v}][\bar{T}])</td>
<td>Transient symmetric</td>
</tr>
<tr>
<td></td>
<td>(3) - (6)</td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>([v'\bar{T}'] = [\bar{v'T'} - [v'][\bar{T}'])</td>
<td>Transient eddy</td>
</tr>
<tr>
<td></td>
<td>(5) - (8)</td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>([v'\bar{T}'] = [v'T'] + [v'\bar{T}''])</td>
<td>Total eddy</td>
</tr>
<tr>
<td></td>
<td>(7) + (9)</td>
<td></td>
</tr>
</tbody>
</table>

As an example, look at Fig. 3.48. The top map shows the eddy part of the zonally
An Introduction to the General Circulation of the Atmosphere

Figure 3.48: An example to illustrate 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 eddy parts of the January–mean meridional wind and temperature at 850 mb, respectively. The bottom panel shows the product, and on the bottom right the zonal mean of the product. See text for details.
averaged 850 mb meridional wind for January, i.e., $\bar{v}$; the corresponding map of the full field was given earlier. The middle panel shows a similar plot for the eddy temperature field at 850 mb, i.e., $\bar{T}$; 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. 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. 3.48 shows a map of the product, $\bar{v} \bar{T}$ and to its left is the corresponding zonal mean plot showing $[v \bar{T}]$, i.e., the zonally averaged meridional temperature flux due to stationary eddies. The flux is small except in middle latitudes of the Northern Hemisphere; as discussed later, strong stationary waves are produced by flow over topography in the Northern Hemisphere in winter.

To describe the mean meridional circulation, we use the zonally averaged forms of the primitive equations, in spherical coordinates, with pressure as the vertical coordinate, and including the “eddy” terms that arise through the analysis given above. The zonally averaged equations for the zonal component of the momentum, the potential temperature, and the water vapor mixing ratio are

$$\frac{\partial [u]}{\partial t} + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{ [v][u] \cos \phi + \{ v \ u \} \cos \phi \} + \frac{\partial}{\partial p} \{ [\omega][u] + \{ v \ u \} \}$$

$$= \{ [v][u] + \{ v \ u \} \} \tan \frac{\phi}{a} + f[v] + g \frac{\partial}{\partial p} [F_u], \quad (3.102)$$

$$\frac{\partial [v]}{\partial t} + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{ [v][v] \cos \phi + \{ v \ v \} \cos \phi \} + \frac{\partial}{\partial p} \{ [\omega][v] + \{ v \ v \} \}$$

$$= - \{ [u][u] + \{ u \ u \} \} \tan \frac{\phi}{a} - f[u] - \frac{1}{a} \frac{\partial [\phi]}{\partial \phi} + g \frac{\partial}{\partial p} [F_v], \quad (3.103)$$

$$\frac{\partial [\theta]}{\partial t} + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{ [v][\theta] \cos \phi + \{ v \ \theta \} \cos \phi \} + \frac{\partial}{\partial p} \{ [\omega][\theta] + \{ v \ \theta \} \}$$

$$= g \frac{\partial [F_\theta]}{\partial p} + \left[ \frac{\theta}{c_p T} \frac{\partial R}{\partial p} \right] + \left[ \frac{\theta}{c_p T} \frac{\partial [\omega]}{\partial p} \right] + \left[ \frac{\theta}{c_p T} \frac{\partial [\omega]}{\partial p} \right], \quad (3.104)$$

$$\frac{\partial [q_v]}{\partial t} + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{ [v][q_v] \cos \phi + \{ v \ q_v \} \cos \phi \} + \frac{\partial}{\partial p} \{ [\omega][q_v] + \{ v \ q_v \} \}$$

$$= g \frac{\partial [F_{q_v}]}{\partial p} - [C], \quad (3.105)$$

In (3.104), $R$ is the net upward energy flux due to radiation. It should be noted that in taking the zonal averages above we have assumed that the pressure surfaces along which the
averages are being taken do not intersect the ground. When the pressure surfaces do intersect the ground some additional complications arise, but those are ignored here for simplicity. Eqs. (3.102)- (3.105) will be used repeatedly throughout the course.

Fig. 3.49 shows the annually averaged meridional flux of zonal momentum, as

![Diagram of meridional flux of zonal momentum](image)

**Figure 3.49: Meridional flux of zonal momentum in the annual mean.** The top panel shows the total flux, the second panel shows the transient–eddy flux, the third panel shows the stationary eddy flux, and the bottom panel shows the flux due to the mean meridional circulation. From Peixoto and Oort (1992).

analyzed by Peixoto and Oort (1992). The total flux is shown, and this is then broken down into the contributions by transient eddies, stationary eddies, and the mean meridional circulation. Remember that it is the meridional change of the flux that drives the tendency, rather than the flux itself. The strongest momentum fluxes are found in the upper troposphere,
near the jet-stream level. The transient eddies dominate, especially in middle latitudes. The midlatitude eddy fluxes are associated with baroclinic eddies, i.e., eddies that arise through baroclinic instability and draw their energy from the available potential energy of the mean state. The available potential energy and the associated baroclinic energy conversions will be discussed in detail in a later chapter. The momentum fluxes associated with the baroclinic waves tend to drive westerlies in middle latitudes, and easterlies in the tropics.

Fig. 3.50 shows the meridional flux of temperature. In midlatitudes, the temperature flux is generally poleward. Transient eddies contribute strongly, especially in the lower troposphere. They tend to warm the higher latitudes and cool the subtropics. The eddy temperature field is extremely weak in the tropics, for reasons that have already been discussed.

The observations discussed in this chapter demand an encompassing explanation. Why does the global circulation of the atmosphere look as it does? We do not yet have a satisfactory theory. The main difficulties are:

- The heating processes at work in the atmosphere are very complicated, and motion-dependent. The mechanisms that govern the interactions between the heating and the motion must be understood as a prerequisite for understanding the general circulation. The heating is very closely related to moist processes,
including cloud formation and precipitation.

- Even if we pretend that the heating is “given,” the response of the atmospheric circulation to the heating is complicated because of the existence of eddies and their interactions with the zonal-mean flow. The eddies are neither purely random nor purely regular.

Despite our incomplete understanding of the overriding issues listed above, decades of research by many scientists have gradually built up a qualitative “view” of the general circulation, which can be summarized as follows: As can be seen from (3.102) to (3.105), a steady, zonally symmetric circulation is mathematically possible, or at least it would be if the Earth’s surface were zonally uniform. Such a symmetric circulation could be driven by tropical heating and polar cooling. The associated “thermal wind” shear of the zonal component would be very large, however. Correspondingly, a steady symmetric circulation would have a very large equator-to-pole temperature gradient, indicating that the poleward energy transport is inefficient in this symmetric regime. This inefficiency arises from the Earth’s rapid rotation, which inhibits meridional motions. If the energy transport could be made more efficient, the temperature gradient would be much weaker.

Nature chooses a more efficient way of transporting energy poleward, which results in a weaker meridional temperature gradient. This is possible because the symmetric circulation is unstable. The relevant mechanism is baroclinic instability, which causes the growth of quasi-horizontal, quasi-geostrophic eddies of cyclone scale. These eddies transport energy poleward, as shown in Fig. 3.50 and indicated schematically in Fig. 3.51. The flux is greatest in middle latitudes because that is where the vertical shear is strongest; the eddies are inactive in the tropics partly because is small there. Stationary eddies, forced by zonal inhomogeneities in the boundary conditions, e.g. mountains and land-sea contrast also transport some energy poleward, mainly in the Northern Hemisphere.

![Figure 3.51: A crude sketch of the variation of the eddy heat flux with latitude. Where the flux increases poleward, the air is cooled, and where it decreases poleward, the air is warmed.](image-url)
Although the meridional circulation is not independent of longitude, a mean meridional circulation (MMC) can, of course, be defined by zonal averaging; such an average is precisely what is plotted in Fig. 3.21. The MMC “feels” the eddy energy transport as a cooling in the subtropics and a warming in higher latitudes. In response, the MMC tends to produce sinking in the subtropics and rising motion in higher latitudes. This suggests a crude interpretation of why there is a three-cell circulation in the winter hemisphere (where baroclinic waves are particularly active), with a Ferrel Cell in middle latitudes, as shown from the data in Fig. 3.21 and indicated schematically in Fig. 3.52. In this interpretation the rising branch of the Ferrel Cell is the response of the MMC to eddy-induced warming. Note that the MMC fights against what the eddies are trying to do. That turns out to be true quite generally, as we will see later.

Similarly, the MMC “feels” the eddy momentum flux, which transports angular momentum from the subtropics to the middle latitudes, thus tending to reinforce the tropical easterlies and the midlatitude westerlies. Again, these eddy-induced tendencies have to be balanced by an adjustment of the MMC. In the midlatitudes, the positive (favoring westerlies) eddy momentum forcing is balanced primarily by the Coriolis acceleration associated with an equatorward component of the MMC, i.e., the upper branch of the Ferrel Cell. In the tropics, the negative (favoring easterlies) eddy momentum forcing is balanced primarily by the Coriolis acceleration associated with a poleward component of the MMC, i.e., it tends to strengthen the Hadley Cell.

The eddy warming and cooling are thus balanced by adjustments to the vertical branches of the MMC, while the eddy momentum flux divergences are balanced by adjustments to the meridional branches of the MMC. The vertical and horizontal branches of the MMC are of course coupled, and this is most easily seen in the zonally averaged

---

*Figure 3.52: A crude sketch of the mean meridional circulation. The subtropical sinking motion occurs where the eddies cool, and the midlatitude rising motion occurs where the eddies warm.*
continuity equation, i.e.,

\[
\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{ [v] \cos \phi \} + \frac{\partial}{\partial p} \{ \bar{\omega} \} = 0 .
\]  

(3.106)

Here we have discussed the effects of the momentum fluxes on the wind field and the
effects of the temperature fluxes on the temperature field. The two effects are closely linked,
however, because geostrophy and hydrostatics imply thermal wind balance, i.e., a tight
relationship between the temperature field and the wind field. Under the constraint of thermal
wind balance (or some similar balance), any process that tends to alter the wind field will
indirectly affect the temperature field, and vice versa. Examples are given later.

Would it be possible to construct a combination of temperature fluxes and momentum
fluxes that cancel each other out, so that there is no net effect on the zonally averaged zonal
wind and temperature? The answer, as shown later, is that this is not only possible, it is to be
expected under many conditions.

### 3.14 A view from theta coordinates

The time and zonally averaged thermodynamic energy equation in \( p \)-coordinates is

\[
\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{ [v\theta] \cos \phi \} + \frac{\partial}{\partial p} \{ \bar{\omega} \theta \} = [\bar{\theta}] .
\]  

(3.107)

Eq. (3.107) shows that heating will force (i.e. will have to be balanced by) some combination
of meridional and vertical motion. From (3.106), we can infer that the zonally averaged
vertical and meridional circulations are linked. This means that, even if the heating \( \dot{\theta} \) is
locally balanced by rising motion, there will have to be meridional motion somewhere in the
domain to satisfy continuity. In short, heating implies a meridional circulation.

The same analysis is simpler with \( \theta \)-coordinates. Applying both a time average and a
zonal average to the continuity equation in \( \theta \)-coordinates, i.e., (3.66), we find that

\[
\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{ [mv] \cos \phi \} = \frac{\partial}{\partial \theta} \{ \bar{m}\dot{\theta} \} ,
\]  

(3.108)

Eq. (3.108) shows directly that \( \dot{\theta} \) induces a meridional circulation, i.e. \( [\dot{\theta}] \neq 0 \) implies that
\( [v] \neq 0 \) somewhere. Note that this conclusion follows from mass conservation in \( \theta \)-space
(Townsend and Johnson, 1985; Hsu and Arakawa, 1990; Edouard et al., 1997). Corresponding
to (3.48), we can define a stream function by

\[
[mv]2\pi a \cos \phi \equiv \frac{\partial \Psi}{\partial \theta} ,
\]

\[
[m\dot{\theta}]2\pi a^2 \cos \phi \equiv -\frac{\partial \Psi}{\partial \phi} .
\]  

(3.109)
You should confirm that the stream function defined by (3.109) has the same dimensions (i.e., mass per unit time) as the stream function defined by (3.48).

Fig. 3.53 shows plots of the stream function of the seasonally varying mean meridional circulation as seen in $p$-coordinates (right panels) and $\theta$-coordinates (left panels). Units are $10^{10} \text{ kg s}^{-1}$. 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. From Townsend and Johnson (1985).

In $\theta$-coordinates, the stream function plots look very different. Hadley cells still appear, with roughly the same stream-function magnitudes, but these direct circulations extend all the way to the poles, primarily in the winter hemisphere. Ferrel cells do not appear at all.
To understand this, note that

$$[m v] = [m][v] + [m^* v^*].$$  \hfill (3.110)

The first term on the right-hand side of (3.110) 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. For example, if $m^*$ is large when $v^*$ is large, and vice versa, then $[m^* v^*]$ will be positive. The quantity $[m^* v^*]$ is called the "bolus mass flux," and $\frac{[m^* v^*]}{[m]}$ is called the "bolus velocity." According to (3.110), the mass transport by the mean flow is supplemented by a kind of "fluid dynamical peristalsis," in which velocity fluctuations are correlated with variations in the pseudo-density, leading to an eddy mass flux along isentropic surfaces. Think of $m$ as the width of a flexible pipe, and $v$ as the rate of flow through the pipe. If the width of the pipe and the flow both fluctuate, in a correlated way, then there can be a net mass transport even if the mean velocity, i.e., $[v]$, is zero. For example, suppose that the pipe is wide when the flow is toward the right and narrow when the flow is toward the left. In such a case there will be a net mass flux toward the right.

The horizontal mass flux in height coordinates, i.e. $\rho v$, where $\rho$ is the density, could be similarly decomposed to define a bolus mass flux in height coordinates, but variations of $\rho$ on height surfaces are so small that the implied bolus mass flux is negligible. It is the relatively strong variations of $m$ on isentropic surfaces that make $[m^* v^*]$ large enough to be important.

In the $\theta$-coordinate panels of Fig. 3.53, the transport by the zonally averaged flow, i.e. $[m][v]$, dominates in the tropics, where $m^*$ is small for the reasons that Charney gave. In mid-latitudes, however, $[v]$ is small due to geostrophy, but $m^*$ becomes large, and in the baroclinic waves of mid-latitudes $[m^* v^*]$ is large and essentially "takes over" from the tropical zonally averaged flow. The sum of $[m][v]$ and $[m^* v^*]$, i.e., $[mv]$, the total zonally averaged mass flux, is large both in the tropics and in middle latitudes. Since the stream function in $\theta$-coordinates is derived from $[mv]$, we see a single Hadley Cell that extends from the tropics to high latitudes.

Here is another way of thinking about it. In $p$-coordinates, the zonally averaged and time-averaged potential temperature equation is

$$\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \left( [v][\theta] \cos \phi + [v^* \theta^*] \cos \phi \right) + \frac{\partial}{\partial p} \left( [\omega][\theta] + [\omega^* \theta^*] \right) = [\tilde{\theta}].$$  \hfill (3.111)

In (3.111), the meridional derivative and zonal averages are taken along pressure surfaces. We
see that meridional fluxes of potential temperature are associated with both the mean meridional circulation and the eddies. In the tropics and in the subtropics of the winter hemisphere, the mean meridional circulation 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 as viewed in pressure coordinates, and the fluctuations of potential temperature on isobaric surfaces are large, especially in the winter hemisphere, due to baroclinic instability, so the eddy potential temperature flux along isobaric surfaces can become large enough, and does in fact become large enough, to “take over the job” of transporting potential temperature poleward.

We can derive an equation that corresponds to (3.111) in $\theta$-coordinates. Simply multiply (3.108) by $\theta$. Using the fact that the derivatives and zonal means in (3.108) are taken along isentropic surfaces, we can write the result as

$$\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \left\{ \overline{m v} \theta \cos \phi \right\} + \frac{\partial}{\partial \theta} \left\{ \overline{m \theta} \right\} = \overline{m \theta} \quad (3.112)$$

While (3.111) contains “eddy terms,” (3.112) does not. The reason is that, by definition, there are no fluctuations of potential temperature on $\theta$-surfaces, i.e. $\theta_e = 0$ on $\theta$-surfaces. The eddy flux of potential temperature along isentropic surfaces must, therefore, be exactly zero. Therefore the Hadley circulations must extend all the way to the poles when depicted in $\theta$-coordinates.

### 3.15 Lots of questions

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

- Why does the sea-level pressure tend to attain its maximum values in the subtropics?
- Why are the subtropical highs typically found on the eastern sides of the ocean basins?
- What determines the intensity and latitudes of the jet streams? Why do the winter jet maxima occur at particular longitudes?
- Why is the upper-level circulation “wavy”? Why is the lower-level circulation “lumpy”?
- What determines the number of “bands” seen, for example, in Fig. 3.21?
- What mechanisms generate the observed stationary waves?
- Why are there surface easterlies near the poles?
- Why are seasonal changes generally weaker in the Southern Hemisphere than in the Northern Hemisphere?
• What determines the magnitude of the pole-to-equator gradient of the surface temperature? 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 height of the tropopause, as a function of latitude? Why is the tropical tropopause so cold?

• 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 causes the summer and winter monsoons?

• Why is there a strong Siberian winter high?

• How much do individual Januaries and Julys, for particular years, differ from the “average” January and July conditions shown here? What causes such year-to-year variations?

• What determines the vertical distribution of water vapor?

• How are the observed patterns of large-scale rising and sinking motion produced and maintained?

• 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 general circulation appear “smooth,” rather than “noisy?”

• Why is the Intertropical Convergence Zone mainly north of the Equator?

These and many other questions will be discussed in the remainder of this course.

Problems

1. a) Estimate the total water vapor content of the atmosphere, in kg.

   b) Estimate the total mass of liquid and ice (combined) in the global atmosphere, in kg. This number is not actually known to better than an order of magnitude. You will have to be creative to come up with a credible estimate of your own. State any assumptions that you make.

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 much less than the rate at which the Earth absorbs solar radiation.

3. Derive (3.42) by starting from (3.3).
4. Prove that the mountain torque vanishes if \( p_S \) depends only on \( \phi_s \).

5. Derive (3.111) from (3.5) and (3.3).

6. When we use isentropic coordinates, the angular momentum equation, (3.58), becomes

\[
\frac{\partial (mM)}{\partial t} + \nabla \cdot (m \mathbf{V} M) + \frac{\partial}{\partial \theta} (m \dot{\theta} M) = -m \frac{\partial s}{\partial \lambda} - (a \cos \varphi) \frac{\partial F_u}{\partial \theta}. \tag{3.113}
\]

Derive (3.62) by starting from (3.113).
CHAPTER 4  Conservation of momentum and energy

Copyright 2005 David A. Randall

4.1 Introduction

In this chapter, we systematically explore the conservation of momentum and, especially energy in its various forms. We also introduce some approximations that are appropriate when considering large-scale motions in the Earth’s atmosphere, and derive some fundamental secondary relationships that will be used later in the course.

In the first part of the chapter, we write the equations as they apply at a point, and show molecular fluxes rather than turbulent fluxes. Near the end of the chapter we discuss the effects of the turbulent fluxes and also such processes as precipitation.

4.2 Conservation of momentum on a rotating sphere

The length of a day is 86400 s, so the Earth rotates about its axis with an angular velocity of \( \frac{2\pi}{(86400 \text{ s})} = 7.29 \times 10^{-5} \text{ s}^{-1} \). This angular velocity can be represented by a vector, \( \mathbf{\Omega} \), pointing towards the celestial North Pole. Consider a coordinate system that is rotating with the Earth, and refer to Fig. 4.1. Newton’s statement of momentum conservation, as applied in the rotating coordinate system, is

\[
\frac{D\mathbf{V}}{Dt} = -2\mathbf{\Omega} \times \mathbf{V} - \mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{r}) - \nabla \phi_a - \alpha \nabla p - \alpha \nabla \cdot \mathbf{F}. \tag{4.1}
\]

Here \( \mathbf{r} \) is a position vector extending from the center of the Earth to a particle of air whose position is generally changing with time. The gravitational potential is \( \phi_a \). The pressure-gradient term is \( -\alpha \nabla p \), where \( \alpha \) is the specific volume, and \( p \) is the pressure. The quantity \( \mathbf{F} \) is the stress tensor associated with molecular viscosity. The dimensions of \( \mathbf{F} \) are density times velocity squared, i.e., the units could be (kg m\(^{-3}\)) (m s\(^{-1}\))^2 = kg m\(^{-1}\) s\(^{-2}\). Note that \( \nabla \cdot \mathbf{F} \) is a vector.

The term \( -2\mathbf{\Omega} \times \mathbf{V} \) represents the Coriolis acceleration, whose direction is perpendicular to \( \mathbf{V} \). The term \( -\mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{r}) \) represents the centripetal acceleration. You should be able to show that

\[
\mathbf{V}_c = \mathbf{\Omega} \times \mathbf{r} = (\mathbf{\Omega} r \cos \phi)e_\lambda, \tag{4.2}
\]
where \( \mathbf{e}_\lambda \) is a unit vector pointing east, \( \varphi \) is latitude, and \( \mathbf{V}_e \) is the velocity (as seen in the inertial frame) that a particle at radius \( r \) and latitude \( \varphi \) experiences due to the Earth's rotation (refer to Fig. 4.1). With this notation, we find that

\[
-\Omega \times (\Omega \times \mathbf{r}) = (\Omega^2 r \cos \varphi) \mathbf{e}_\lambda \times \mathbf{e}_\Omega
= \Omega^2 \mathbf{r}_e,
\]

where \( \mathbf{r}_e \) is the vector shown in Fig. 4.1, and \( \mathbf{e}_\Omega \) is a unit vector pointing toward the celestial north pole. This shows that the centripetal acceleration points outward, in the direction of \( \mathbf{r}_e \), which is perpendicular to the axis of the Earth's rotation. It can be shown that

\[
\Omega^2 \mathbf{r}_e = \nabla \left[ \frac{1}{2} |\Omega \times \mathbf{r}|^2 \right].
\]

According to (4.4), the centripetal acceleration can be regarded as the gradient of a potential, called the “centrifugal potential.” The “apparent” gravity, \( \mathbf{g} \), due to the combined effects of true gravity and the centripetal acceleration, can be defined as

\[
\mathbf{g} = g_{\text{at}} - \Omega^2 \mathbf{r}_e,
\]
where $g_a \equiv \nabla \Phi_a$, and using (4.4) we see that the potential of $g$ is

$$\phi = \phi_a - \frac{1}{2} |\Omega \times r|^2.$$  \hspace{1cm} (4.6)

so that $g \equiv \nabla \phi$. We refer to $\phi$ as the “geopotential.”

As is well known, the Earth’s gravitational potential, $\phi_a$, decreases with distance from the center of the Earth. In addition, it varies with latitude and longitude, at a fixed distance from the center of the Earth, due to inhomogeneities in the distribution of mass within the solid Earth and the oceans. This means that a surface of constant $\phi_a$ is not spherical. Moreover, as can be seen from (4.4) and (4.6), the centripetal acceleration also varies geographically and with distance from the center of the Earth. As a result, surfaces of constant $\phi$ are only approximately spherical. In particular, the centripetal acceleration causes these surfaces to bulge outward at low latitudes, so that their shapes are well approximated by “oblate spheroids.” To the extent that we wish to consider geographical variations of $\phi$, the oblateness of the Earth’s surface should also be taken into account. For most purposes, however,

$$g \equiv g_a = -g k,$$  \hspace{1cm} (4.7)

because the centripetal acceleration is small compared to $g_a$. Here $k$ is a unit vector pointing upward, away from the center of the Earth. When we use (4.6) with a spatially constant value of $g$, as is conventional, we must also approximate the shape of the Earth as a sphere, with minor topographical bumps.

Using (4.6) we can now write the equation of motion (4.1) as

$$\frac{DV}{Dt} = -2\Omega \times V - \nabla \phi - \alpha \nabla p - \alpha \nabla \cdot F.$$  \hspace{1cm} (4.8)

Another useful form of this equation is

$$\frac{\partial V}{\partial t} + [2\Omega + (\nabla \times V)] \times V + \nabla \left( \frac{1}{2} V \cdot V \right) = -\nabla \phi - \alpha \nabla p - \alpha \nabla \cdot F.$$  \hspace{1cm} (4.9)

To obtain (4.9) from (4.8) we have used the vector identity

$$(V \cdot \nabla)V = (\nabla \times V) \times V + \nabla \left( \frac{1}{2} V \cdot V \right).$$  \hspace{1cm} (4.10)

We will have occasion to use both (4.8) and (4.9).

Now consider spherical coordinates, $(\lambda, \varphi, r)$. The unit vectors in the $(\lambda, \varphi, r)$ coordinates are $e_\lambda$, $e_\varphi$, and $e_r$, respectively. As shown in the Appendix on “Vectors and
coordinate systems,” the vector operators that will be used in this course, i.e. the gradient, divergence, curl, and Laplacian, can be expressed in spherical coordinates as follows:

\[
\nabla A = \left( \frac{1}{r \cos \phi} \frac{\partial A}{\partial \lambda}, \frac{1}{r} \frac{\partial A}{\partial \phi}, \frac{1}{r \sin \phi} \frac{\partial A}{\partial r} \right), \tag{4.11}
\]

\[
\nabla \cdot \mathbf{H} = \frac{1}{r \cos \phi} \frac{\partial H_\lambda}{\partial \lambda} + \frac{1}{r \cos \phi} \frac{\partial}{\partial \phi} (H_\phi \cos \phi) + \frac{1}{r^2} \frac{\partial}{\partial r} (H_r r^2), \tag{4.12}
\]

\[
\nabla \times \mathbf{H} = \left\{ \frac{1}{r} \left[ \frac{\partial H_r}{\partial \phi} - \frac{\partial}{\partial r} (r H_\phi) \right], \right. \\
\left. \frac{1}{r} \frac{\partial}{\partial r} (r H_\lambda) - \frac{1}{r \cos \phi} \frac{\partial H_\phi}{\partial \lambda}, \right. \\
\left. \frac{1}{r \cos \phi} \left[ \frac{\partial H_\phi}{\partial \lambda} - \frac{\partial}{\partial \phi} (H_\lambda \cos \phi) \right] \right\}, \tag{4.13}
\]

\[
\nabla^2 A = \frac{1}{r^2 \cos \phi} \left[ \frac{\partial}{\partial \lambda} \left( \frac{1}{\cos \phi} \frac{\partial A}{\partial \lambda} \right) + \frac{\partial}{\partial \phi} \left( \cos \phi \frac{\partial A}{\partial \phi} \right) + \frac{\partial}{\partial r} \left( \frac{2}{r} \cos \phi \frac{\partial A}{\partial r} \right) \right]. \tag{4.14}
\]

Here \( A \) is an arbitrary scalar, and \( \mathbf{H} = (H_\lambda, H_\phi, H_r) \) is an arbitrary vector.

Eq. (4.12) can be expanded as

\[
\nabla \cdot \mathbf{H} = \frac{1}{r \cos \phi} \frac{\partial H_\lambda}{\partial \lambda} + \frac{1}{r \cos \phi} \frac{\partial}{\partial \phi} (H_\phi \cos \phi) + \frac{1}{r} \frac{\partial}{\partial r} H_r + \frac{2 H_r}{r}. \tag{4.15}
\]

Because the Earth’s atmosphere is very thin compared to the radius of the Earth, the last term is negligible, and we can approximate the divergence operator by

\[
\nabla \cdot \mathbf{H} \approx \frac{1}{a \cos \phi} \frac{\partial}{\partial \lambda} H_\lambda + \frac{\partial}{\partial \phi} (H_\phi \cos \phi) + \frac{\partial}{\partial r} H_r. \tag{4.16}
\]

Note that \( r \) has been replaced by \( a \) in the first two terms. In this course, we normally use (4.16) rather than (4.12), largely because it is traditional to do so. It is not at all clear, however, that the approximation (4.16) actually makes our work simpler. Note that the approximation would not be applicable to a deep atmosphere, such as that of a star or of Jupiter.

We can represent the velocity vector in terms of zonal, meridional, and radial
components, as

\[ V = u e_\lambda + v e_\phi + w e_r, \]  \hspace{1cm} (4.17)

where

\[ u \equiv r \cos \phi \frac{D\lambda}{Dt}, \quad v \equiv r \frac{D\phi}{Dt}, \quad w \equiv \frac{Dr}{Dt}, \]  \hspace{1cm} (4.18)

and

\[ \frac{D}{Dt} = \frac{\partial}{\partial t} + \frac{D\lambda}{Dt} \frac{\partial}{\partial \lambda} + \frac{D\phi}{Dt} \frac{\partial}{\partial \phi} + \frac{Dr}{Dt} \frac{\partial}{\partial r} \]

\[ = \frac{\partial}{\partial t} + \frac{u}{r \cos \phi} \frac{\partial}{\partial \lambda} + \frac{v}{r} \frac{\partial}{\partial \phi} + \frac{w}{\partial r}. \]  \hspace{1cm} (4.19)

The directions in which the unit vectors \( e_\lambda, e_\phi \), and \( e_r \) actually point depend on where you are. Therefore, as an air particle moves from place to place, the directions of the unit vectors change. Because of this, we must expand (4.17), taking into account a total of six terms:

\[ \frac{DV}{Dt} = \left( \frac{Du}{Dt} e_\lambda + u \frac{De_\lambda}{Dt} \right) + \left( \frac{Dv}{Dt} e_\phi + v \frac{De_\phi}{Dt} \right) + \left( \frac{Dw}{Dt} e_r + w \frac{De_r}{Dt} \right). \]  \hspace{1cm} (4.20)

Simple geometrical reasoning leads to the following formulae:

\[ \frac{De_\lambda}{Dt} = \frac{D\lambda}{Dt} \sin \phi e_\phi - \cos \phi \frac{D\lambda}{Dt} e_r \]

\[ = \left( \frac{u \tan \phi}{a} \right) e_\phi - \frac{u}{a} e_r, \]  \hspace{1cm} (4.21)

\[ \frac{De_\phi}{Dt} = \frac{D\lambda}{Dt} \sin \phi e_\lambda - \frac{D\phi}{Dt} e_r \]

\[ = \left( \frac{v \tan \phi}{a} \right) e_\lambda - \frac{v}{a} e_r, \]  \hspace{1cm} (4.22)

\[ \frac{De_r}{Dt} = \cos \phi \frac{D\lambda}{Dt} e_\lambda + \frac{D\phi}{Dt} e_\phi \]

\[ = \frac{u}{a} e_\lambda + \frac{v}{a} e_\phi. \]  \hspace{1cm} (4.23)

When (4.21)-(4.23) are taken into account, (4.8) can be written as
Conservation of momentum and energy

\[ \frac{Du}{Dt} + \frac{uw}{r} - \frac{uv \tan \phi}{r} = f v - f w - \frac{\alpha}{r \cos \phi} \frac{\partial p}{\partial \lambda} - \alpha (\nabla \cdot \mathbf{F})_\lambda , \]

\[ \frac{Dv}{Dt} + \frac{vw}{r} + \frac{u^2 \tan \phi}{r} = -fu - \frac{\alpha}{r} \frac{\partial p}{\partial \phi} - \alpha (\nabla \cdot \mathbf{F})_\phi , \]

\[ \frac{ Dw}{Dt} - \left( \frac{u^2 + v^2}{r} \right) = f u - \frac{\alpha}{r} \frac{\partial p}{\partial r} - \alpha (\nabla \cdot \mathbf{F})_r - g . \]

Here

\[ f \equiv 2 \Omega \sin \phi \text{ and } \dot{f} \equiv 2 \Omega \cos \phi . \]

Eqs. (4.24) are the components of the equation of motion in spherical coordinates Further explanation is given in the Appendix.

By using the continuity equation in spherical coordinates, we can rewrite (4.24) in flux form:

\[ \frac{\partial}{\partial t} (\rho u) + \nabla \cdot (\rho \mathbf{V} u) + \rho \frac{uw}{r} - \rho \frac{uv \tan \phi}{r} = \rho f v - \rho f w - \frac{1}{r \cos \phi} \frac{\partial p}{\partial \lambda} - (\nabla \cdot \mathbf{F})_\lambda , \]

\[ \frac{\partial}{\partial t} (\rho v) + \nabla \cdot (\rho \mathbf{V} v) + \rho \frac{vw}{r} + \rho \frac{u^2 \tan \phi}{r} = -\rho f u - \frac{1}{r \sin \phi} \frac{\partial p}{\partial \phi} - (\nabla \cdot \mathbf{F})_\phi , \]

\[ \frac{\partial}{\partial t} (\rho w) + \nabla \cdot (\rho \mathbf{V} w) - \rho \left( \frac{u^2 + v^2}{r} \right) = \rho f u - \frac{\alpha}{r} \frac{\partial p}{\partial r} - \rho g - (\nabla \cdot \mathbf{F})_r . \]

These equations are fairly exact. Various approximations will be introduced later.

4.3 Conservation of kinetic energy and potential energy

As you probably already know, the kinetic energy equation can be derived from the equation of motion. To do this, start with the full three-dimensional equation of motion in the form

\[ \frac{\partial \mathbf{V}}{\partial t} + (\nabla \times \mathbf{V} + 2 \Omega) \times \mathbf{V} + \nabla (K + \phi) = -\alpha \nabla p - \alpha \nabla \cdot \mathbf{F} . \]

Here we are using height coordinates. Dotting (4.27) with \( \mathbf{V} \), we find that

\[ \frac{DK}{Dt} + \mathbf{V} \cdot \nabla \phi = -\alpha \mathbf{V} \cdot \nabla p - \alpha \mathbf{V} \cdot (\nabla \cdot \mathbf{F}) , \]

where
An Introduction to the General Circulation of the Atmosphere

4.3 Conservation of kinetic energy and potential energy

\[ K = \frac{1}{2} V \cdot V \]  

(4.29)

is the kinetic energy per unit mass.

Note that \( K \) and, therefore, the total energy depend on the choice of coordinate system. For example, if we compare a coordinate system that is rotating with the Earth to an inertial coordinate system, the actual value of the kinetic energy at a given place in the atmosphere will differ by a large amount. The Lagrangian time rate of change of the kinetic energy will be the same, however, and this is what matters, because it is \( \frac{DK}{Dt} \), rather than \( K \) itself, that appears in the energy conservation equation (4.28).

Since \( \phi \) is independent of time in height coordinates, we see that

\[ \frac{D\phi}{Dt} = V \cdot \nabla \phi = wg . \]  

(4.30)

This can be called “the potential energy equation.” Use of (4.30) allows us to rewrite (4.28) as

\[ \frac{D(K + \phi)}{Dt} = -\alpha V \cdot \nabla p - \alpha V \cdot (\nabla \cdot F) . \]  

(4.31)

We refer to \( K + \phi \) as the mechanical energy per unit mass. Eq. (4.31) is sometimes called the mechanical energy equation.

In (4.31), the rate at which work is done by the pressure force, per unit mass, is represented by \( -\alpha V \cdot \nabla p \). This expression can be rewritten as follows:

\[ -\alpha V \cdot \nabla p = -\alpha \nabla \cdot (pV) + \alpha(p \nabla \cdot V) \]

\[ = -\alpha \nabla \cdot (pV) + p \frac{D\alpha}{Dt} . \]

(4.32)

Here we have used the continuity equation to eliminate \( \alpha \nabla \cdot V \). The \( \nabla \cdot (pV) \) term on the second line of (4.32) has the form of a flux divergence, and so represents a spatial redistribution of energy by the pressure force. The \( p \frac{D\alpha}{Dt} \) term represents the work done by volume expansion (analogous to the work done in inflating a balloon). We refer to \( p \frac{D\alpha}{Dt} \) as the “expansion-work” term\(^1\).

\(^1\) Because \( \alpha \) is a constant for a fluid of constant density (e.g., “shallow water”), the internal energy of a constant-density fluid is nonconvertible (like the ruble), although it is not zero (unlike the ruble). Because it is nonconvertible, the internal energy of a constant-density fluid plays no role in the energy cycle; we can just ignore it.
Similarly, the friction term of (4.31) can be expanded to reveal two physically distinct parts, as follows:

$$-\alpha V \cdot (\nabla \cdot \mathbf{F}) = -\alpha \nabla \cdot (\mathbf{F} \cdot V) - \delta,$$  \hspace{1cm} (4.33)

where

$$\delta \equiv -\alpha (\mathbf{F} \cdot \nabla) \cdot V$$  \hspace{1cm} (4.34)

is the rate of kinetic energy dissipation per unit volume. The quantity $\nabla \cdot (\mathbf{F} \cdot V)$ in (4.33) has the form of a flux divergence, and so represents a spatial redistribution of kinetic energy as friction (represented by $\mathbf{F}$) causes air parcels to do work on each other. Because this is just 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 lower boundary.

In contrast, kinetic energy dissipation, represented in (4.33) by $\delta$, is a true sink of kinetic energy, i.e., it can be shown that

$$\delta \geq 0.$$  \hspace{1cm} (4.35)

As discussed later, the dissipation of kinetic energy appears as a source of thermodynamic energy, i.e., as “frictional heating.” It is a weak but persistent source of internal energy for the atmosphere.

At this point, I am conflicted. I want to explain enough about the effects of friction so that the preceding discussion is understandable, but I don’t think that it is worth a major digression. Here, therefore, is a small digression: For simplicity, consider a simple Cartesian coordinate system that is applied in some small volume of the atmosphere. The coordinates will be named $(x, y, z)$, and the corresponding velocity components will be named $(u, v, w)$. The stress tensor can be written schematically as a matrix:

$$\mathbf{F} = \begin{bmatrix} 0 & F_{v,x} & F_{w,x} \\ F_{u,y} & 0 & F_{w,y} \\ F_{u,z} & F_{v,z} & 0 \end{bmatrix}. $$  \hspace{1cm} (4.36)

As an example, $F_{u,y}$ 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 we want to consider only shearing stresses. 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 (4.36) as
The divergence can be written as a “row” vector, i.e.,
\[ \nabla \cdot \mathbf{F} = \begin{bmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \end{bmatrix}, \tag{4.38} \]
and so we find that
\[ \nabla \cdot \mathbf{F} = \left( \frac{\partial F_{u,y}}{\partial y} + \frac{\partial F_{u,z}}{\partial z} \right) \mathbf{i} + \left( \frac{\partial F_{u,y}}{\partial x} + \frac{\partial F_{u,z}}{\partial z} \right) \mathbf{j} + \left( \frac{\partial F_{u,y}}{\partial x} + \frac{\partial F_{v,z}}{\partial y} \right) \mathbf{k}, \tag{4.39} \]
where \( \mathbf{i} \), \( \mathbf{j} \), and \( \mathbf{k} \) are the unit vectors in the \( x \), \( y \), and \( z \) directions, respectively. Here you can see how it happens that the divergence of the tensor is a vector. Similarly,
\[ \mathbf{F} \cdot \mathbf{V} = (F_{u,y} v + F_{u,z} w) \mathbf{i} + (F_{u,y} u + F_{v,z} w) \mathbf{j} + (F_{u,z} u + F_{v,z} v) \mathbf{k} \tag{4.40} \]
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 energy exchange in the \( z \)-direction (hence, multiplied by \( \mathbf{k} \)) due to the work done as \( u \)-momentum is transferred in the \( z \)-direction by friction. The energy flux divergence is then
\[ \nabla \cdot (\mathbf{F} \cdot \mathbf{V}) = \frac{\partial}{\partial x} (F_{u,y} v + F_{u,z} w) + \frac{\partial}{\partial y} (F_{u,y} u + F_{v,z} w) + \frac{\partial}{\partial z} (F_{u,z} u + F_{v,z} v). \tag{4.41} \]

The dissipation rate can be constructed as follows: We can write
\[ \mathbf{F} \cdot \mathbf{\nabla} = \begin{bmatrix} 0 & F_{u,y} & F_{u,z} \\ F_{u,y} & 0 & F_{v,z} \\ F_{u,z} & F_{v,z} & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix} \]
\[ = \begin{bmatrix} (F_{u,y} \frac{\partial}{\partial y} + F_{u,z} \frac{\partial}{\partial z}) \mathbf{i} \\ (F_{u,y} \frac{\partial}{\partial x} + F_{v,z} \frac{\partial}{\partial z}) \mathbf{j} \\ (F_{u,z} \frac{\partial}{\partial x} + F_{v,z} \frac{\partial}{\partial y}) \mathbf{k} \end{bmatrix}. \tag{4.42} \]
Therefore
To prove (4.35), we must introduce actual expressions for the stresses. Air is an example of a “Newtonian fluid,” for which the stress is related to the spatial derivatives of the motion (in particular, the “strain”) by

\[
F_{u,y} = F_{u,y} = -\mu \left( \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right),
\]

\[
F_{u,z} = F_{u,z} = -\mu \left( \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \right),
\]

\[
F_{w,x} = F_{w,z} = -\mu \left( \frac{\partial w}{\partial y} + \frac{\partial v}{\partial z} \right),
\]

where \(\mu\) is the (positive, constant) molecular viscosity. Eq. (4.44)-(4.46) are called “stress-strain relationships.” For example, an upward increase of \(u\) will tend to favor a negative (downward) value of \(F_{u,z}\). The flow of momentum is from “fast” to “slow,” thus tending to homogenize the momentum over time. Such a flux is described as “down-gradient.” Substituting (4.44)-(4.46) into (4.43), we find that

\[
(F \cdot \nabla) \cdot \mathbf{V} = -\mu \left[ \left( \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right)^2 + \left( \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \right)^2 + \left( \frac{\partial w}{\partial y} + \frac{\partial v}{\partial z} \right)^2 \right],
\]

which establishes (4.35). This ends our digression.
All contributions to $\frac{\partial}{\partial t}(\rho K)$ on the left-hand side of (4.49) represent transport processes, which merely redistribute energy in space. In contrast, the expansion-work term, $p\frac{D\alpha}{Dt}$, need not integrate to zero. It 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 $p\frac{D\alpha}{Dt}$ term must be positive, i.e. it must act as a source of mechanical energy:

$$\int_V p\frac{D\alpha}{Dt} \rho dV = \int_V \delta\rho dV \geq 0 .$$  \hspace{1cm} (4.50)

Here the integral is taken over the entire mass of the atmosphere and the overbars represent a time average. Eq. (4.50) is a very fundamental result. It means that on the average the pressure force must do positive expansion work to compensate for the dissipation of kinetic energy. In order that $p\frac{D\alpha}{Dt}$ be positive, expansion must take place, in an average sense, at a higher pressure than compression. For example, we can have expansion in the lower troposphere and compression in the upper troposphere.

Given that in an average sense the expansion work term of (4.48) must act as a source of mechanical energy, we have to ask where this energy comes from. The answer is that it comes from the thermodynamic energy of the atmosphere. This will be demonstrated later. Expansion work represents an energy conversion process, which can have either sign locally but is positive when averaged over the whole atmosphere and over time.

Similarly, given that the dissipation term of (4.48) represents a sink of mechanical energy, we have to ask where the energy goes. The answer is that it appears as a source of thermodynamic energy. Dissipation is, therefore, another energy conversion process -- a conversion that runs in only one direction.

Eq. (4.50) simply means that the rate of kinetic energy dissipation must be equal, on the average, to the rate of kinetic energy generation. For the whole atmosphere, this rate has been estimated to be on the order of 5 W m$^{-2}$. For comparison, 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 kinetic energy.

The mechanical energy generation term can be written in another form. To see this, note that

$$p\frac{D\alpha}{Dt} = \frac{D}{Dt}(p\alpha) - \alpha\frac{Dp}{Dt}$$

$$= \frac{D}{Dt}(RT) - \omega\alpha .$$  \hspace{1cm} (4.51)
This shows that in a time average over the whole atmosphere the expansion-work term is closely related to the product $\omega \alpha$. We can interpret $\omega \alpha$ as a rate of conversion between mechanical and thermodynamic energy.

Substituting (4.51) into (4.48), we obtain

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

(4.52)

This is somewhat easier to interpret when we convert to flux form:

$$
\frac{\partial}{\partial t} [\rho (K + \phi)] + \nabla \cdot [\rho \mathbf{V} (K + \phi)] = -\nabla \cdot (\mathbf{F} \cdot \mathbf{V}) - \rho (\omega \alpha + \delta) + \frac{\partial p}{\partial t} .
$$

(4.53)

Note that the pressure-work term of (4.52), involving $\nabla \cdot (p \mathbf{V})$, has disappeared via a cancellation, but “in its place” we pick up a new term involving the local time-rate-of-change of the pressure, $\frac{\partial p}{\partial t}$. From (4.53) we see that in an average over the whole atmosphere, and over time, the $-\omega \alpha$ term of (4.53) must be positive, i.e. it must act as a source of mechanical energy:

$$
-\int_{V} \omega \alpha \rho dV = \int_{V} \delta \rho dV \geq 0 .
$$

(4.54)

Comparing (4.50) and (4.54), we see that $-\int_{V} \omega \alpha \rho dV = \int_{V} \left( \frac{\rho D \alpha}{Dt} \right) \rho dV$.

4.4 Conservation of thermodynamic energy

The internal energy of a perfect gas is given by

$$
e = c_v T ,
$$

(4.55)

where $c_v$, the heat capacity at constant volume, is a constant. For dry air, $c_v = \frac{5R}{2} \equiv 713$ J kg$^{-1}$ K$^{-1}$.

More generally, the internal energy also includes the latent heat associated with the potential condensation of water vapor$^2$, and we find that for moist air

$^2$ We could also add the latent heats of other atmospheric constituents, e.g., nitrogen, oxygen, and carbon dioxide, to represent the effects of their potential condensation. We do not bother to do so because those constituents do not condense under conditions realized in the Earth’s atmosphere.
This equation is approximate because we have neglected the heat capacity of the water vapor, as well as the heat capacity of any liquid (or ice) that might be present. In atmospheric science we frequently define the internal energy as the internal energy of dry air, and treat the latent heat as an "external" source or sink of internal energy. This is not strictly correct, but does no harm in most applications. We will follow this convention in this course, i.e., we will define the internal energy per unit mass by (4.55).

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:

$$\frac{De}{Dt} + p \frac{D\alpha}{Dt} = -\alpha \nabla \cdot (R + F_s) + LC + \delta.$$  \hspace{1cm} (4.57)

Here $e$ is given by (4.55); $F_s$ is the vector flux of internal energy due to molecular diffusion; $R$ is the vector flux of energy due to radiation\(^3\). Note that the dissipation rate appears here as a source of internal energy. Equation (4.57) is a statement of the conservation of thermodynamic energy\(^4\), applied to a moving particle.

An alternative statement of the conservation of thermodynamic energy, obtained using (4.51) in (4.57), is

$$c_p \frac{DT}{Dt} = \omega \alpha - \alpha \nabla \cdot (R + F_s) + LC + \delta,$$  \hspace{1cm} (4.58)

where

$$c_p = R + c_v \equiv 1000 \text{ J kg}^{-1} \text{ K}^{-1}.$$ \hspace{1cm} (4.59)

Eq. (4.58) shows that, for an adiabatic, isobaric process, the enthalpy, $\eta$, is a conserved variable. For an ideal gas, the enthalpy is given by

$$\eta = c_p T.$$  \hspace{1cm} (4.60)

More generally, the enthalpy can be written as

$$\eta = e + p\alpha.$$  \hspace{1cm} (4.61)

For saturated air containing liquid, it turns out that

\(^3\) This notation conflicts with that used for the gas constant, but there should be little chance of confusion.

\(^4\) Also called the “First Law of Thermodynamics,” although that terminology seems rather medieval.
Conservation of momentum and energy

where $L$ is the latent heat of condensation (e.g., Lorenz 1979; Emanuel, 1994) and $l$ is the liquid water mixing ratio.

It is also possible to express the conservation of thermodynamic energy in terms of the potential temperature. By using, we can show that

$$c_p \frac{D\theta}{Dt} = \left(\frac{\theta}{T}\right) \left[ -\alpha \nabla \cdot (\mathbf{R} + \mathbf{F}_s) + LC + \delta \right].$$

(4.63)

In the absence of heating and dissipation, $\frac{D\theta}{Dt} = 0$, i.e. $\theta$ is conserved following a particle in the absence of heating. This is one of the reasons that $\theta$ is a particularly useful quantity.

In summary, the thermodynamic energy equation can be expressed in the three equivalent forms (4.57), (4.58), and (4.63). Each of these forms will be used later.

4.5 Conservation of total energy

Adding (4.48) and (4.57) gives

$$\frac{D(K + \phi + e)}{Dt} = -\alpha \nabla \cdot (p \mathbf{V} + \mathbf{F} \cdot \mathbf{V}) - \alpha \nabla \cdot (\mathbf{R} + \mathbf{F}_s) + LC.$$  

(4.64)

Note that the $p \frac{D\alpha}{Dt}$ terms have cancelled, as have the dissipation terms. The cancellations occur because these terms represent conversions between thermodynamic and mechanical energy. Alternatively, we can add (4.52) and (4.58) to obtain

$$\frac{D(K + \phi + e)}{Dt} = -\alpha \nabla \cdot (p \mathbf{V} + \mathbf{F} \cdot \mathbf{V}) - \alpha \nabla \cdot (\mathbf{R} + \mathbf{F}_s) + LC.$$  

(4.65)

Here we have used $c_p T - p\alpha = e$, and the $\omega \alpha$ and dissipation terms have cancelled, as before, because they represent energy conversions. As should be expected, (4.64) and (4.65) are identical. You can think of (4.48) and (4.57) as a “matched set,” and (4.52) and (4.58) as a second, equivalent, matched set.

Although total energy conservation is expressed by (4.64) [and (4.65)], the energy conversions associated with latent heat release are not shown explicitly. To remedy this, we use the water vapor conservation equation in the form

$$\frac{Dq_v}{Dt} = -\alpha \nabla \cdot (\mathbf{F}_{q_v}) - C.$$  

(4.66)

where $\mathbf{F}_{q_v}$ is the flux of water vapor due to molecular diffusion, and $C$ is the rate of

An Introduction to the General Circulation of the Atmosphere
4.5 Conservation of total energy

condensation. Next, we multiply (4.66) by the latent heat of condensation, \( L \), neglect variations of \( L \) in time and space, and add the result to (4.64). This gives

\[
\frac{D}{Dt}(K + \phi + e + Lq_v) = -\alpha \nabla \cdot (pV + F \cdot V + R + F_h).
\]  (4.67)

Here \( F_h \equiv F_s + LF \cdot q \) is the sum of the molecular fluxes of sensible and latent heat; for reasons explained later, this sum will be called the molecular flux of moist static energy. From (4.67) we see that the total energy per unit mass is given by the sum of the kinetic, potential, internal, and latent energies:

\[
e_t = K + \phi + e + Lq_v.
\]  (4.68)

Every term on the right-hand side of (4.67) is the divergence of a flux, i.e., each term represents a spatial redistribution of energy. This shows that the total energy of the atmosphere is conserved apart from exchanges across its upper and lower boundaries. To see how this works, start by using the continuity equation to convert (4.67) to flux form:

\[
\frac{\partial}{\partial t}[\rho(K + \phi + e + Lq_v)] + \nabla \cdot [\rho V(K + \phi + e + Lq_v) + pV + F \cdot V + R + F_h] = 0.
\]  (4.69)

Next, distinguish between horizontal and vertical fluxes of energy, as follows:

\[
\frac{\partial}{\partial t}[\rho(K + \phi + e + Lq_v)]
+ \nabla_H \cdot [\rho V_H(K + \phi + e + Lq_v) + pV_H + (F \cdot V)_H + R_h + (F_h)_H]
+ \frac{\partial}{\partial z}[\rho w(K + \phi + e + Lq_v) + pw + (F \cdot V)_z + R_z + (F_h)_z] = 0.
\]  (4.70)

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. Next, we vertically integrate through the entire atmospheric column, using Leibniz’ Rule to take the integrals inside the derivatives. The result can be written as
Recall that the condition that no mass crosses the Earth’s surface can be written as

$$ \frac{\partial z}{\partial t} + \mathbf{V}_H \cdot \nabla z - w_S = 0. $$  

With the use of (4.72), we can rewrite (4.71) as

$$ \frac{\partial}{\partial t} \left[ \int_{z_S}^{\infty} \rho \mathbf{V} + \phi + e + Lq_v \, dz \right] + \nabla_H \cdot \left[ \int_{z_S}^{\infty} \left[ \rho \mathbf{V} + \phi + e + Lq_v \right] \, dz \right] $$

$$ + \frac{\partial z}{\partial t} \left[ \int_{z_s}^{\infty} \rho \mathbf{V} + \phi + e + Lq_v \, dz \right] $$

$$ + \nabla_H \cdot \left[ \int_{z_s}^{\infty} (\rho \mathbf{V} + (\mathbf{F} \cdot \mathbf{V})_H + \mathbf{R}_h + (\mathbf{F}_h)_H) \, dz \right] $$

$$ = -p_S [((\mathbf{V})_S \cdot \nabla z_s - w_S] - \{(\mathbf{F} \cdot \mathbf{V})_S \cdot \nabla z_s - [(\mathbf{F} \cdot \mathbf{V})_S] - (\mathbf{R}_z)_\infty \} - \{(\mathbf{R}_h)_S \cdot \nabla z_s - (\mathbf{R}_z)_S \}. $$

The terms on the right-hand side of (4.71) represent the effects of 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 by $-(\mathbf{R}_z)_\infty$. At the lower boundary there are energy fluxes due to pressure-work, frictional work, the molecular flux of moist static energy, and radiation.

When \( \frac{\partial z}{\partial t} = 0 \), the pressure-work term vanishes. Over the ocean, however, the surface height fluctuates due to the passage of waves, and so in the presence of such waves the atmosphere and ocean can exchange energy due to pressure work. It is interesting that one effect of such an energy exchange can be to add energy to the 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 can impart energy to the atmosphere through the pressure-work term.
The friction terms of (4.71) represent the work done by surface drag. This will be discussed further near the end of this Chapter.

As already discussed, the surface moist static energy flux is a very important energy source for the atmosphere.

Finally, the radiative energy flux is quite important at both the upper and lower boundaries of the atmosphere.

When we integrate (4.71) horizontally over the entire sphere, the horizontal flux divergence term integrates to zero, and we get

\[
\frac{d}{dt} \left[ \int_A \int_{z_s}^\infty \rho (K + \phi + e + L q_v) dz \right] dA = 0,
\]

(4.75)
i.e., the total energy of the atmosphere is invariant. Eq. (4.75) is a very important conclusion, which will be used later in the definition of available potential energy.

The combined effect of the pressure-work and frictional work terms of (4.74) is to remove energy from the global atmosphere. It follows that, in a time average, the radiative and molecular flux terms must act as an energy source for the atmosphere. As discussed later, the rate at which the atmosphere does the frictional work on the lower boundary is on the order of tenths of a W m\(^{-2}\). The pressure-work term is typically even smaller. As discussed in Chapter 1, the remaining individual terms on the right-hand side of (4.74) are typically larger by several orders of magnitude. It follows that these remaining terms must very nearly cancel in a time average, i.e.,

\[
0 \equiv - \int_A \left\{ \left[ (F_h)_H \right]_S \cdot \nabla z_S - \left[ (F_h)_z \right]_S \right\} dA
- \int_A (R)_{z,\infty} dA - \int_A \left\{ \left[ R_h \right]_S \cdot \nabla z_S - (R_z)_S \right\} dA.
\]

(4.76)

In this sense, the total “heating” of the atmosphere is very nearly zero. Eq. (4.76) can also be
written as
\[
\int \left[ - \alpha \nabla \cdot \left( \mathbf{R} + \mathbf{F}_h \right) + LC \right] \rho \, dV \equiv 0 , \tag{4.77}
\]
where \( \int \rho \, dV \) denotes a mass-weighted integral over the entire atmosphere.

### 4.6 Static energies

By using the equation of state and (4.59), we can rewrite the total energy equation, (4.69), as
\[
\frac{\partial}{\partial t} \left[ \rho (K + \phi + c_p T + L q_v) \right] + \nabla \cdot [\rho \mathbf{V} (K + \phi + c_p T + L q_v) + \mathbf{R} + \mathbf{F}_h] = - \nabla \cdot (\mathbf{F} \cdot \mathbf{V}) + \frac{\partial p}{\partial t} .
\tag{4.78}
\]

Here the enthalpy appears in place of the internal energy in the time derivative and flux divergence terms. A key difference between (4.69) and (4.78) is that there is no pressure-work term in the latter. A price that we pay for this simplification is the appearance of the \( \frac{\partial p}{\partial t} \) term on the right-hand side of (4.78). Note, however, that this term drops out in a time average.

The \( \frac{\partial p}{\partial t} \) term of (4.78) looks funny. As an aid in its interpretation, write
\[
\frac{\partial p}{\partial t} = \frac{\partial}{\partial z} \left( z \frac{\partial p}{\partial t} \right) - z \frac{\partial}{\partial z} \left( \frac{\partial p}{\partial t} \right) \\
= \frac{\partial}{\partial z} \left( z \frac{\partial p}{\partial t} \right) - \frac{\partial}{\partial t} \left( z \frac{\partial p}{\partial z} \right) \\
= \frac{\partial}{\partial z} \left( z \frac{\partial p}{\partial t} \right) - \frac{\partial}{\partial t} \left( z \frac{\partial p}{\partial z} \right) .
\tag{4.79}
\]

This can be substituted into (4.78) to obtain
\[
\frac{\partial}{\partial t} \left[ \rho (K + \phi + c_p T + L q_v) + z \frac{\partial p}{\partial z} \right] + \nabla \cdot [\rho \mathbf{V} (K + \phi + c_p T + L q_v) + \mathbf{R} + \mathbf{F}_h] = - \nabla \cdot (\mathbf{F} \cdot \mathbf{V}) + \frac{\partial}{\partial z} \left( z \frac{\partial p}{\partial t} \right) .
\tag{4.80}
\]
In the hydrostatic limit the $\rho \phi$ and $z \frac{d \rho}{d z}$ terms inside the time-rate-of-change operator cancel. The second term on the right-hand side of (4.80) represents a vertical flux of energy associated with the time-rate-of-change of the pressure.

The contribution of the kinetic energy to the total energy is typically quite negligible. For example, an air parcel zipping along at a rather extreme 100 m s$^{-1}$ has a kinetic energy per unit mass of $5 \times 10^3$ J kg$^{-1}$. (Keep in mind that the kinetic energy is proportional to the square of the wind speed, so that a parcel moving at a more typical 10 m s$^{-1}$ has a kinetic energy 100 times smaller.) If a parcel traveling at 100 m s$^{-1}$ resides on the 200 mb surface, its potential energy per unit mass (relative to sea level) is about $1.2 \times 10^5$ J kg$^{-1}$, or about 24 times greater than its kinetic energy. If the temperature of the fast parcel is a mere 200 K, which is if anything a little too cold for the 200 mb surface, its internal energy per unit mass is about $1.5 \times 10^5$ J kg$^{-1}$, about 30 times greater than its kinetic energy. For these reasons, we can usually neglect $K$ in (4.78).

In addition, the friction and pressure-tendency terms of (4.78) can often be neglected. With these simplifying approximations, (4.78) reduces to

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

(4.81)

where

$$
h \equiv c_p T + \phi + L q_v
$$

is the moist static energy, whose latitude-height distribution was discussed in Chapter 3. According to (4.81), 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, temperature, or geopotential height, the moist static energy is conserved even for pseudoadiabatic processes, in which condensed water is assumed to precipitate out immediately. For many practical purposes, conservation of total energy is (approximately) equivalent to conservation of moist static energy.

We did not have to use the hydrostatic approximation to derive (4.81); this is important, because it means that (4.81) can be used in the analysis of non-hydrostatic processes, e.g., cumulus convection. We will do just that in Chapter 5.

Note that conservation of moist static energy is an approximation to the total energy equation, rather than the thermodynamic energy equation. This is why there is no dissipation term in (4.81); 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,

$$
s \equiv c_p T + \phi ,
$$

(4.83)
is approximately conserved under dry adiabatic processes.

4.7 Entropy

For any gas or liquid, the entropy per unit mass, \( s \), satisfies

\[
\frac{Ds}{Dt} = \frac{De}{Dt} + p \frac{Da}{Dt}.
\] (4.84)

From (4.57), this implies that

\[
\frac{Ds}{Dt} = -\alpha \nabla \cdot (R + F_s) + LC + \delta.
\] (4.85)

Using the equation of state with (4.84), we can show that

\[
s = c_p \ln \left( \frac{T}{T_0} \right) - R \ln \left( \frac{p}{p_0} \right).
\] (4.86)

In (4.86), \( T_0 \) and \( p_0 \) are suitable reference values that arise here as “constants of integration.” There is a simple relationship between the entropy and the potential temperature, i.e.,

\[
s \equiv c_p \ln \left( \frac{\theta}{\theta_0} \right).
\] (4.87)

Here \( \theta_0 \) is the potential temperature corresponding to \( T_0 \) and \( p_0 \).

For saturated air containing liquid water, (4.87) must be replaced by

\[
s \equiv c_p \ln \left( \frac{\theta}{\theta_0} \right) - \frac{L_l}{T}
\] (4.88)

(e.g. Lorenz, 1979; Emanuel, 1994).

Eq. (4.85) can be rearranged as

\[
\frac{Ds}{Dt} = \frac{\left[ -\alpha \nabla \cdot (R + F_s) + LC + \delta \right]}{T}.
\] (4.89)

According to (4.89), dissipation and heating can change the entropy; because \( \delta \geq 0 \), dissipation never decreases and normally increases the entropy.
Now use continuity to rewrite (4.89) in flux form, then pass to an integral over a closed system to obtain:

\[
\frac{dS}{dt} = \int_{V} \left[ -\alpha \nabla \cdot (\mathbf{R} + \mathbf{F}_s) + LC + \delta \right] \frac{1}{T} \rho \, dV ,
\]

(4.90)

where

\[
S \equiv \int_{V} \rho \, sdV
\]

(4.91)

is the total entropy of the entire atmosphere. In an average over a sufficiently long time, (4.90) reduces to

\[
\int_{V} \left[ -\alpha \nabla \cdot (\mathbf{R} + \mathbf{F}_s) + LC + \delta \right] \frac{1}{T} \rho \, dV = 0 ,
\]

(4.92)

from which it follows that

\[
\int_{V} \left[ -\alpha \nabla \cdot (\mathbf{R} + \mathbf{F}_s) + LC \right] \frac{1}{T} \rho \, dV < 0 .
\]

(4.93)

This is an important result. It means that, for the atmosphere as a whole, heating acts to decrease the entropy. In order for (4.93) 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. This implies that heating and cooling try to make temperature contrasts increase with time. One way to arrange this 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 (4.93), let

\[
Q \equiv -\alpha \nabla \cdot (\mathbf{R} + \mathbf{F}_s) + LC
\]

(4.94)

denote the heating rate, and divide \(Q\) and \(T\) into averages, denoted by overbars, and the corresponding local departures from the means, denoted by primes:

\[
Q = \bar{Q} + Q' , T = \bar{T} + T'.
\]

(4.95)

Then we see that
Conservation of momentum and energy

An Introduction to the General Circulation of the Atmosphere

For \( \overline{Q} = 0 \) [see (4.77)], we get

\[
\frac{\overline{Q}}{T} = \frac{\overline{Q} + \overline{Q'}}{T + T'} = \overline{Q(1 + \frac{Q'/\overline{Q}}{T(1 + T'/T)}}
\]

\[
\equiv \frac{\overline{Q}}{T}(1 + \frac{Q'/\overline{Q}}{1 - T'/T})
\]

\[
= \frac{\overline{Q}}{T}\left[1 + \frac{Q'}{\overline{Q}} - \frac{T'}{T} - \frac{O'T'}{\overline{Q}T}\right]
\]

\[
= \frac{\overline{Q}}{T}\left(1 - \frac{O'T'}{\overline{Q}T}\right)
\]

\[
= \frac{\overline{Q}}{T} - \frac{O'T'}{T^2}.
\]  \hspace{1cm} (4.96)

If we add energy where the temperature is already warm, and remove energy where the temperature is already cool, then \( O'T' > 0 \), so \( (\overline{Q}/T) \) will be negative, and (4.93) will be satisfied. Such a process tends to increase the variability of temperature. We show later that this implies generation of available potential energy.

We have concluded that heating tends to increase the temperature contrasts within the atmosphere. In order for the system to achieve a steady state, some other process must oppose this effect of the heating. 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 general circulation tend to cool where the temperature is high (e.g. the tropical lower troposphere), and tend to warm where the temperature is low (e.g. the polar troposphere).

A final important point is that the global entropy budget is fundamentally different from the global energy budget. Energy is conserved, which means that in a time average the fluxes of energy into the system must be balanced by fluxes out. This is not true for entropy. Entropy is generated by a wide variety of dissipative processes; the Earth makes entropy. For this reason, there is a net entropy flux out of the Earth system, via radiation, and in a time average this flux has to be equal to the Earth’s entropy production rate. The Earth’s entropy production rate can, therefore, be measured using satellite data. This has been done. For further discussion, see Stephens and O’Brien (1992).
Introduction to the General Circulation of the Atmosphere

4.8 Approximations

Up to here the discussion has been fairly exact. In this and the following section, we introduce some approximations that are commonly used in the analysis of the large-scale circulation systems of the Earth’s atmosphere.

In the present section we focus on the equations of motion. The relevant approximations are:

(i) Replace \( r \) by \( a \) everywhere, where \( a \) is the 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, and so it is called the “thin atmosphere approximation.” It is a good approximation for Earth, but would not apply, e.g., to Jupiter.

(ii) Drop the terms containing \( \Omega \). This means that the horizontal component of \( \Omega \) disappears from the equations. This is often called “the traditional approximation.” There is an ongoing discussion as to whether or not this is a good idea.

(iii) Neglect \( \frac{u w}{r} \) and \( \frac{v w}{r} \), the curvature terms involving \( w \), in the equations for \( u \) and \( v \), respectively, neglect \( \frac{u^2 + v^2}{r} \) in the equation of vertical motion.

Next, we introduce a fourth, very familiar approximation, called the quasi-static approximation. For resting air, the vertical component of (4.26) reduces to the “hydrostatic equation:

\[
\frac{\partial p}{\partial z} = -\rho g .
\] (4.98)

With an appropriate boundary condition, (4.98) allows us to compute \( p(z) \) from \( \rho(z) \). Even when the air is moving, (4.98) gives a good approximation to \( p(z) \), simply because \( \frac{Dw}{Dt} \) and the vertical component of the friction force are small compared to \( g \). Eq. (4.98) as applied to moving air is called the hydrostatic approximation, and it is applicable to virtually all meteorological phenomena, including violent thunderstorms.

For large-scale circulations, the approximate \( p(z) \) determined through the use of (4.98) can be used to compute the pressure gradient force in the equation of horizontal motion. To do so is to use the quasi-static approximation.

The quasi-static approximation applies very well to large-scale motions, but it is not
applicable to many small-scale motions, such as thunderstorms. When the quasi-static approximation is made, the effective kinetic energy is due entirely to the horizontal motion; the contribution of the vertical component, \( w \), is neglected. For large-scale motions, \( w \ll (u, v) \), so that this quasistatic kinetic energy is very close to the true kinetic energy. Further discussion is given in a hand-out on the quasi-static approximation, available from the instructor.

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 gravity waves can be accurately approximated by the vertical derivative of the vertical component of the flux.

With these approximations, (4.24) is replaced by

\[
\begin{align*}
\frac{Du}{Dt} - \frac{uv \tan \varphi}{a} &= fv - \alpha \frac{\partial p}{a \cos \varphi} \frac{\partial}{\partial \lambda} - \alpha \frac{\partial}{\partial z} (F_w), \\
\frac{Dv}{Dt} + \frac{u^2 \tan \varphi}{a} &= -fu - \alpha \frac{\partial p}{a \cos \varphi} - \alpha \frac{\partial}{\partial z} (F_v), \\
0 &= -g - \alpha \frac{\partial p}{\partial z}.
\end{align*}
\]

The vector equation of horizontal motion is given by

\[
\frac{\partial \mathbf{V}}{\partial t} + (\zeta + f) \mathbf{k} \times \mathbf{V} + \nabla_z K + w \frac{\partial \mathbf{V}}{\partial z} = -\alpha \nabla_z p - \alpha \frac{\partial}{\partial z} (F_v),
\]

where the symbol \( \mathbf{V} \) now represents the horizontal velocity vector,

\[
\zeta_z \equiv \mathbf{k} \cdot (\nabla_z \times \mathbf{V})
\]

is the vertical component of the relative vorticity as seen in height coordinates,

\[
K \equiv \frac{1}{2} \mathbf{V} \bullet \mathbf{V},
\]

and \( F_v \) is the vertical flux of horizontal momentum due to turbulence. Note that (4.100) is analogous to (4.27).

**4.9 The mechanical energy equation in other vertical coordinate systems**

Before leaving this Chapter, we return briefly to the mechanical energy equation. This time we derive it in pressure coordinates, using the quasi-static approximation for the pressure-gradient term. We start from the vector equation of horizontal motion in pressure coordinates, which is

\[
\frac{\partial \mathbf{V}}{\partial t} + (\zeta + f) \mathbf{k} \times \mathbf{V} + \nabla_p K + w \frac{\partial \mathbf{V}}{\partial z} = -\alpha \nabla_p p - \alpha \frac{\partial}{\partial z} (F_v),
\]
where
\[
\zeta_p = \mathbf{k} \cdot (\nabla_p \times \mathbf{V})
\] (4.104)
is the vertical component of the horizontal vorticity computed along isobaric surfaces. Again, the vector \( \mathbf{V} \) represents the horizontal motion, and the kinetic energy per unit mass, \( K \), is that part associated with the horizontal motion only. Taking the dot product of (4.103) with \( \mathbf{V} \), we obtain the kinetic energy equation in the form
\[
\frac{DK}{Dt} + \mathbf{V} \cdot \nabla_p \phi + g \frac{\partial}{\partial p} (\mathbf{V} \cdot \mathbf{F}_V) = -\delta ,
\] (4.105)
where
\[
\frac{D}{Dt} = \left( \frac{\partial}{\partial t} \right)_p + \mathbf{V} \cdot \nabla_p + \omega \frac{\partial}{\partial p} ,
\] (4.106)
and the dissipation rate is approximated by
\[
\delta \equiv g \mathbf{F}_V \cdot \frac{\partial \mathbf{V}}{\partial p} .
\] (4.107)

Using the continuity and hydrostatic equations in pressure coordinates, this can be rewritten as
\[
\frac{\partial K}{\partial t} + \nabla_p \cdot [\mathbf{V}(K + \phi)] + \frac{\partial}{\partial p} [\omega(K + \phi)] + g \frac{\partial}{\partial p} (\mathbf{V} \cdot \mathbf{F}_V) = -\omega \alpha - \delta .
\] (4.108)

Compare (4.108) with (4.48) and (4.53). The terms \( \nabla_p \cdot (\mathbf{V}\phi) + \frac{\partial}{\partial p} (\omega \phi) \) in (4.108) are analogous to \( \nabla_z \cdot (p \mathbf{V}) \) in (4.48). We can also write (4.108) as
\[
\frac{\partial}{\partial t} (K + \phi) + \nabla_p \cdot [\mathbf{V}(K + \phi)] + \frac{\partial}{\partial p} [\omega(K + \phi)] + g \frac{\partial}{\partial p} (\mathbf{V} \cdot \mathbf{F}_V) = -\omega \alpha - \delta + \frac{\partial \phi}{\partial t} ,
\] (4.109)
which is more directly analogous to (4.53). The manipulation
Conservation of momentum and energy

(4.110) allows us to rewrite (4.109) as

\[
\frac{\partial \phi}{\partial t} = \frac{\partial}{\partial p} \left( p \frac{\partial \phi}{\partial t} \right) - p \frac{\partial}{\partial p} \left( \frac{\partial \phi}{\partial t} \right)
\]

\[
= \frac{\partial}{\partial p} \left( p \frac{\partial \phi}{\partial t} \right) - \frac{\partial}{\partial t} \left( p \frac{\partial \phi}{\partial p} \right)
\]

\[
= \frac{\partial}{\partial p} \left( p \frac{\partial \phi}{\partial t} \right) - \frac{\partial}{\partial t} \left( \frac{\partial \phi}{\partial p} \right)
\]

\[
= \frac{\partial}{\partial p} \left( p \frac{\partial \phi}{\partial t} \right) + \frac{\partial}{\partial t} (p \alpha)
\]

\[
= \frac{\partial}{\partial p} \left( p \frac{\partial \phi}{\partial t} \right) + \frac{\partial}{\partial t} (RT)
\]

The total energy equation corresponding to (4.111) is

\[
\frac{\partial}{\partial t} \left( K + \phi - RT \right) + \nabla_p \cdot [V(K + \phi)] + \frac{\partial}{\partial p} \left( \omega(K + \phi) \right) + g \frac{\partial}{\partial p} (V \cdot F_V)
\]

\[
= - \omega \alpha - \delta + \frac{\partial}{\partial p} \left( p \frac{\partial \phi}{\partial t} \right)
\]

(4.111)

(4.112)

Compare with (4.80).

Finally, by starting from the equation of motion in isentropic coordinates, i.e.,

\[
\frac{\partial V}{\partial t} + (\zeta_\theta + f) k \times V + \nabla_\theta (K + M) + \hat{\theta} \frac{\partial V}{\partial \theta} = \frac{1}{m \hat{\theta}} (F_V)
\]

(4.113)

where

\[
\zeta_\theta \equiv k \cdot (\nabla_\theta \times V)
\]

(4.114)

we can derive the mechanical energy equation in isentropic coordinates:
4.10 The effects of turbulence

In practice, the molecular fluxes of momentum, temperature, and moisture are quite negligible compared with the corresponding fluxes due to turbulence. The large scales of atmospheric motion feel the turbulent fluxes directly, and in effect contribute some of their large-scale energy to maintain the energy of the turbulent scales. This will be discussed in detail in a later chapter. For now, it suffices to say that there is a “flow” of energy from large scales to smaller scales, and from the smaller scales to molecular dissipation.

When $\mathbf{F}$ is dominated by the turbulent momentum flux, we can use the approximation

$$\nabla \cdot (\mathbf{F} \cdot \mathbf{V}) = -\frac{\partial}{\partial z} (\mathbf{V} \cdot \rho w' \mathbf{V}') .$$

This frictional work term is small throughout most of the atmosphere; it matters most of all in the turbulent boundary layer near the surface. In particular, when suitably integrated this term expresses the energy loss by the atmosphere due to work done on the ocean by the surface wind stress.\(^5\) The rate at which the atmosphere does work on the oceans can be roughly estimated as follows: As discussed later, 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 current speed, which, using the values given above, is roughly $10^{-2}$ W m\(^{-2}\). The energy that the atmosphere imparts to the wind-driven ocean circulation 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, the rate at which energy is lost through work done on the ocean is utterly negligible, compared for example with the net surface radiation.

In a similar way, we can approximate the dissipation rate by

$$\delta \equiv -\alpha (\rho w' \mathbf{V}') \cdot \frac{\partial \mathbf{V}}{\partial z} .$$

\(^5\) It can also represent the rate at which work is done by the atmosphere as the wind disturbs the vegetation on the land surface, or as dust is lofted into the air.
As will be discussed later, this is an example of what is called a “gradient production” term. Specifically it represents the rate of production of turbulence kinetic energy (TKE) by conversion from the kinetic energy of the mean flow. The physical picture is that the kinetic energy of the mean flow is converted to TKE, which is then dissipated, i.e. the actual dissipation occurs on small, “turbulent” scales.

We can apply (4.118) to estimate the rate of dissipation in the planetary boundary layer (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 (4.118) through the depth of the PBL by

\[ \int_{\text{PBL}} \rho \delta dz \equiv - (\rho w'V')_S V_M, \]  

(4.119)

where \( V_M \) is the horizontal wind in the upper part of the PBL, near the top of the shear layer. The bulk aerodynamic formula tells us that

\[ (\rho w'V')_S = - \rho_s C_D |V_M| V_M. \]  

(4.120)

Substitution of (4.120) into (4.119) gives

\[ \int_{\text{PBL}} \rho \delta dz \equiv \rho_s C_D |V_M|^3. \]  

(4.121)

This shows that the rate of dissipation in the PBL increases very strongly as the wind speed increases. For \( V_M = 10 \text{ m s}^{-1} \), we find that \( \int_{\text{PBL}} \rho \delta dz \equiv 1 \text{ W m}^{-2} \). The rate of dissipation of atmospheric kinetic energy in the PBL is thus considerably larger than the rate at which the atmosphere does work on the ocean through the surface wind stress.

Bister and Emanuel (1998) have pointed out that very large dissipation rates must 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 (4.121). They show that dissipative heating acts to increase storm intensity, 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\(^{-2}\).

4.11 Summary

In this chapter we have discussed the conservation principles for momentum, kinetic energy, potential energy, thermodynamic energy, total energy, and entropy. We have also explored the conversion processes that connect these conservation principles with one another. Useful approximations have been introduced. The results obtained will be used extensively in later chapters.
Problems

1. The velocity associated with the Earth’s rotation is

\[ \mathbf{v}_e = \Omega r \cos \phi \mathbf{e}_\lambda, \]  \hspace{1cm} (4.122)

where \( \mathbf{e}_\lambda \) is a unit vector pointing east. Show that

\[ \mathbf{k} \cdot \nabla \times \mathbf{v}_e = 2\Omega \sin \phi \]  \hspace{1cm} (4.123)

2. Prove the following about the unit vectors in spherical coordinates:

\[ \nabla \cdot \mathbf{e}_\lambda = 0, \quad \nabla \cdot \mathbf{e}_\phi = -\frac{\tan \phi}{r}, \quad \nabla \cdot \mathbf{e}_r = \frac{2}{r}, \]  \hspace{1cm} (4.124)

\[ \nabla \times \mathbf{e}_\lambda = \frac{\mathbf{e}_\phi}{r} + \frac{\tan \phi}{r} \mathbf{e}_r, \quad \nabla \times \mathbf{e}_\phi = \frac{\mathbf{e}_\lambda}{r}, \quad \nabla \times \mathbf{e}_r = 0. \]  \hspace{1cm} (4.125)

3. Using (4.5), calculate the radius of a geostationary orbit. You will have to take into account the variation of \( g_a \) with distance from the center of the Earth.

4. Prove that

\[ \Omega^2 r_e = \nabla \left( \frac{1}{2} |\Omega \times r|^2 \right). \]  \hspace{1cm} (4.126)

You may adopt a coordinate system for this purpose if you feel that you need one.

5. Suppose that you can bench press 100 kg at the North Pole. Assume that the Earth is a perfect sphere and that the acceleration due to the Earth’s gravity is horizontally uniform. How much can you lift at the Equator?

6. For a two-dimensional spherical coordinate system, \((\lambda, \phi)\), prove by direct calculation that

\[ \frac{D \mathbf{v}}{Dt} + 2\Omega \times \mathbf{v} + \Omega \times (\Omega \times \mathbf{r}) = \]  \hspace{1cm} (4.127)

\[ \frac{\partial \mathbf{v}}{\partial t} + (\nabla \times \mathbf{v} + 2\Omega) \times \mathbf{v} + \nabla \left( \frac{1}{2} \mathbf{v} \cdot \mathbf{v} \right) = \nabla \left[ \frac{1}{2} |\Omega \times r|^2 \right]. \]

7. Derive (4.24) from (4.9).
8. Consider a spherical coordinate system \((\lambda', \varphi', r')\), which differs from \((\lambda, \varphi, r)\) in that the “pole” of the \((\lambda', \varphi', r')\) system is \textit{tilted away from the pole of the Earth’s rotation axis}, by an angle \(\alpha\), as illustrated in Fig. 4.2. Rewrite the horizontal momentum equation in a the \(\lambda', \varphi', r'\) coordinate system.

\[
\frac{\partial \mathbf{V}_H}{\partial t} + (2 \Omega + \nabla \times \mathbf{V}_H) \times \mathbf{V}_H + \nabla \left(\frac{1}{2} \mathbf{V}_H \cdot \mathbf{V}_H\right) + \frac{\partial \mathbf{V}_H}{\partial z} = -\alpha \nabla p - \alpha \frac{\partial \mathbf{F}_V}{\partial z} \\
(4.128)
\]

9. Consider a planet that is shaped like a doughnut, as illustrated in the sketch given in Fig. 4.3. The planet is spinning about an axis through the hole in the doughnut. It has an “outside equator” and an “inside equator.” Define coordinates \(\lambda\) and \(\varphi\) as shown in Fig. 4.3. These are orthogonal curvilinear coordinates. Let \(\varphi = 0\) on the outside equator. Note that both \(\varphi\) and \(\lambda\) range between 0 and \(2\pi\). Assume that gravity acts perpendicular to the surface of the planet locally, everywhere. (This would not really be true.)

As can be deduced from Fig. 4.3, an increment of distance in the \(\varphi\) direction is

\[
dy = rd\varphi, \\
(4.129)
\]

and an increment of distance in the \(\lambda\) direction is
Figure 4.3: A planet shaped like a doughnut. The inner radius is $a$ and the outer radius is $b$. Latitude is measured from the outer Equator. See Problem 9.

\[
dx = [b - r(1 - \cos \varphi)]d\lambda \\
\equiv c(\varphi)d\lambda .
\]  

(4.130)

a) Work out the two-dimensional gradient and curl operators in the $(\lambda, \varphi)$ system. Do not worry about the third coordinate, which measures height above the surface of the planet.
b) Define velocity components in the \( \lambda \) and \( \varphi \) directions. Call these \( u \) and \( v \), respectively. They should have dimensions of length per unit time (e.g. m s\(^{-1}\)). Give formulae for \( u \) and \( v \) in terms of \( \frac{D\lambda}{Dt} \) and \( \frac{D\varphi}{Dt} \). Write down the two component equations of motion as they apply to \( u \) and \( v \). Use Eq. (4.9) as your starting point. Ignore the "vertical" component of the motion. Ignore friction.

c) Derive a form of angular momentum conservation for this planet. Consider only the component of the angular momentum with respect to the planet’s axis of rotation. Hint: Start by writing down a suitable definition for the angular momentum per unit mass.

d) Push a particle due “north” (i.e. toward larger \( \varphi \)) starting from the outer equator at \( (\lambda = 0, \varphi = 0) \). Note that the initial value of \( u \) is zero. Ignore the pressure gradient force and friction. What is the minimum initial \( v \) to make the particle travel “all the way around” to \( \varphi = 2\pi \), with \( \varphi \) increasing monotonically en route?

e) When this minimum initial \( v \) is used, what will be the longitude of the particle when it arrives at \( \varphi = 2\pi \)? Note: This longitude can be expressed in terms of an integral. You are not required to evaluate the integral; just set it up.

10. Prove that for a closed volume

\[
\int \frac{DA}{V} \rho dV = \frac{d}{dt} \int A \rho dV .
\] 

(4.131)

11. Show that for an isentropic process

\[
\frac{p}{p_0} = \left( \frac{p}{p_0} \right)^\gamma = \left( \frac{T}{T_0} \right)^{\gamma - 1} ,
\] 

(4.132)

where subscript “0” denotes a reference state.

12. The “Exner function” is defined by

\[
\pi \equiv \frac{c_p}{p_0} T = c_p \left( \frac{p}{p_0} \right)^{\gamma - 1} .
\] 

(4.133)

Show that, for an arbitrary process,
\[ \alpha dp = \theta d\pi, \]
\[ Tds = \pi d\theta, \]
\[ h = \pi \theta + \text{constant}. \] (4.134)

Here \( s \) is the entropy per unit mass.

13. A process that mixes (i.e., homogenizes) potential temperature increases entropy. The following two exercises illustrate this:

a) Consider two parcels of equal mass. Parcel number one has potential temperature

\[ \theta_1 = \theta_0 + \Delta \theta, \] (4.135)

and parcel number two has potential temperature

\[ \theta_2 = \theta_0 - \Delta \theta, \] (4.136)

where

\[ 0 \leq \Delta \theta < \theta_0. \] (4.137)

Define the entropy as

\[ s = c_p \ln \left( \frac{\theta}{\theta_{ref}} \right), \] (4.138)

so that the state of zero entropy has \( \theta = \theta_{ref} \). Show that mixing the two parcels in such a way that potential temperature is homogenized results in an increase of the total entropy.

b) Consider a process that mixes potential temperature, i.e.,

\[ \frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x} \left( K \frac{\partial \theta}{\partial x} \right), \] (4.139)

where \( K \geq 0 \), acting over a domain that is either periodic or else closed in the sense that \( K \frac{\partial \theta}{\partial x} = 0 \) on the boundaries. Prove that the process described by (4.139) causes the average entropy, over the whole domain, to increase with time.
14. Starting from

\[
\int_{V} \rho \frac{D\alpha}{Dt}\rho dV > 0 ,
\]

(4.140)

show that

\[
\int_{V} T \frac{Ds}{Dt}\rho dV > 0 ,
\]

(4.141)

and

\[
\int_{V} \pi \frac{D\theta}{Dt}\rho dV > 0 .
\]

(4.142)

Discuss the physical meaning of these inequalities.

15. Prove that in a hydrostatic atmosphere

\[
\frac{\partial s}{\partial z} = c_{p} \frac{T \partial \theta}{\partial z} .
\]

(4.143)

Here \( s \) is the dry static energy.

CHAPTER 5  

The mean meridional circulation

5.1  The observed meridional transports of energy and moisture

As discussed in Chapter 3, moist static energy conservation is expressed by

\[
\frac{Dh}{Dt} = g \frac{\partial}{\partial p} (R + F_h)
\]

(5.1)

where

\[
h \equiv c_p T + \phi + Lq
\]

(5.2)

is the moist static energy. Using the methods discussed in Chapter 2, we can derive

\[
\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \int_0^{p_s} \left\{ [v][h] \cos \phi + [v^* h^*] \cos \phi \right\} \frac{dp}{g} = \left[ F_{h} + \bar{R} \right]_S - \left[ \bar{R}_T \right].
\]

(5.3)

Here the square brackets denote zonal means, \( R_T \) is the net radiation at the top of the atmosphere, and a time average has been used to eliminate the tendency term. Eq. (5.3) expresses the requirement of total energy balance for a “ring” of air along a latitude circle. By further integration of (5.3) with respect to latitude, we can obtain the meridional energy transport by the atmosphere, with dimensions of energy per unit time, much as we worked out the meridional energy transport by the atmosphere and ocean combined in Chapter 1, based on the net top-of-the-atmosphere radiation. Observations of this atmospheric total meridional energy transport are discussed below.

We can derive a corresponding equation for the “precipitable water,” or vertically averaged water vapor amount, using essentially the same procedure again. The result is

\[
\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \int_0^{p_s} \left\{ [v][q_v] \cos \phi + [v^* q_v^*] \cos \phi \right\} \frac{dp}{g} = \left[ F_{q_v} \right]_S - \int_0^{p_s} \left[ \bar{C} \right] \frac{dp}{g},
\]

(5.4)

and the corresponding result for liquid water is
The globally averaged evaporation and precipitation rates, which must balance in a time average, are not accurately known, but are roughly 3 kg m\(^{-2}\) day\(^{-1}\). It follows that an average water molecule thus spends a “residence time” of about 8 days in the atmosphere between its introduction by surface evaporation, and its removal by precipitation. This means that if we average the vertically integrated moisture budget of the atmosphere over a time interval at least several times longer than 8 days, e.g. over one month or more, the local time rate of change term of (5.4) becomes negligible, and we obtain a balance between evaporation, precipitation, and the lateral flux of moisture into or out of the region. Strictly speaking, (5.4) applies to the time rate of change of the vertical integral of the total atmospheric moisture, including atmospheric liquid and ice, but recall that liquid and ice make only a tiny contribution to the total moisture content of the atmosphere. According to (5.9), when precipitation exceeds evaporation, moisture must be laterally “imported” by the atmosphere. When evaporation exceeds precipitation, the opposite occurs. The observed meridional transport of moisture is discussed below. For obvious reasons, it is convenient to discuss energy and moisture together.
Recall that the dry static energy satisfies $s = h - Lq$. Then (5.3) and (5.4) can be combined, with the use of (5.7), to derive a corresponding equation for the dry static energy:

\[
\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \int_{-\infty}^{\infty} \left[ \frac{1}{g} \frac{\partial F_s}{\partial \phi} + \frac{1}{s} \frac{\partial}{\partial \phi} \left( \frac{\partial F_s}{\partial \phi} \right) \right] dp = \left[ \bar{F}_s + \bar{R}_T \right] - \left[ \bar{R}_T \right] + L[\bar{P}] .
\]

We now present some observations of the energy transports of the atmosphere and ocean as reported by Trenberth and Caron (2001) and Masuda (1988). Fig. 5.1 shows the zonally integrated poleward transport of dry static energy for December, January, and February (DJF) and June, July, and August (JJA). The dry static energy transport can be

---

**Figure 5.1:**
- **a)** Northward transport of dry static energy during DJF (December–January–February) in units of PW. (a) Total transport; (b) transport by the mean meridional circulation; (c) transport by stationary eddies; (d) transport by transient eddies. For all panels, the solid lines show the results obtained from ECMWF data, the dashed lines from GFDL data, and the dotted lines from Oort’s (1983) statistics. Figure from Masuda (1988).
computed from the expression \( p_s \frac{2\pi a \cos \phi}{g} \int_{\theta} \{(v s) + (v^* s^*)\} \, dp \); compare with (5.10).

The dimensions of the dry static energy transport are energy per unit time. The total dry static energy transport can be broken down into components associated with the mean meridional circulation (hereafter MMC), the stationary eddies, and the transient eddies. The MMC dominates in the tropics. It gives weak equatorward fluxes in mid-latitudes of the winter hemispheres, due to the Ferrell Cells. The stationary eddies are significant in the Northern Hemisphere midlatitudes, particularly in winter, but not in the tropics or the Southern Hemisphere midlatitudes. The transient eddies are important in the middle latitudes of both hemispheres. In the Northern Hemisphere, they are noticeably more vigorous in DJF than in JJA, but there is relatively little seasonal change in the Southern Hemisphere.

Fig. 5.2 shows the corresponding results for “latent energy transport;” this quantity is

![Figure 5.2: a) Northward transport of latent energy during DJF (December-January-February). b) Northward transport of latent energy during JJA (June-July-August). From Masuda (1988).](image)

closely related to the zonally averaged difference between precipitation and evaporation, as shown by (5.9). Again the mean meridional circulation dominates in the tropics. Note also that in the tropics the latent energy transport is away from the poles, as discussed earlier.

---

1. Masuda refers to the stationary eddies as “steady eddies.” As Dave Barry would say, “Steady Eddie and the Transients” would be a great name for a band.
Stationary eddy contributions are noticeable in the tropics, due to monsoon circulations there, but are not so important in middle latitudes. Keep in mind here that in the middle latitudes, particularly in winter, the lower temperatures imply that the atmosphere contains little water vapor, so that there cannot be much latent energy transport.

The sum of the dry static energy transport and the latent energy transport gives the moist static energy transport, which is, practically speaking, nearly the same as the total energy transport. The moist static energy transport is shown in Fig. 5.3, for each of the four seasons. These curves are relatively simple, and shaped roughly like $\sin(2\varphi)$. The northward flux of $h$ increases northward from about 40°S to 40°N, implying that the atmosphere receives a net input of energy at its upper and/or lower boundaries, through this range of latitudes. The poleward transport shows a noticeable seasonal cycle in the Northern Hemisphere, with the strongest transport in DJF, but the Southern Hemisphere transport shows relatively little

Figure 5.3: Northward transport of moist static energy during the four seasons from the analyses of (a) ECMWF, (b) GFDL and (c) Oort (1983). The solid lines are for DJF, dashed lines for MAM, dash-dotted lines for JJA, and dotted lines for SON seasons. From Masuda (1988).
seasonal change. Fig. 5.3 shows results from three different analyses, and there are significant differences among them. This illustrates that the observations are subject to serious uncertainties.

Fig. 5.4 shows the annual mean moist static energy transport, this time from six different analyses. The disagreements among the analyses suggest that the peak values in middle latitudes are uncertain by at least 25%. This is a very troubling discrepancy for such an important quantity as the annually averaged total energy transport by the atmosphere. Of course, the disagreements among the various analyses cannot necessarily be interpreted directly as uncertainties, because the different methods and/or data used may a priori be of different degrees of merit.

Fig. 5.5 breaks the annual mean moist static energy transport down into contributions by the MMC, the stationary eddies, and the transient eddies. Viewed in this annual mean sense, the MMC appears to play a small role, because its contributions in different seasons cancel, while the transient eddies and to a lesser extent the stationary eddies contribute consistently throughout the year, and so dominate the annual mean moist static energy transport. As the previous figures make clear, this picture is misleading. In reality, the MMC plays a very strong role in the total energy transport in the individual seasons.

We have already shown, in Chapter 1, the annual mean northward energy transport by the ocean and atmosphere combined; this quantity was first diagnosed by Vonder Haar and Oort (1973). The curve of the annually averaged total energy transport has a pleasingly simple shape. The maximum absolute values in middle latitudes of both hemispheres are on the order of 6 PW. This total energy transport by the climate system can be inferred directly from satellite measurements of the Earth’s radiation budget, and so is known with relatively good accuracy now.

By subtracting the annually averaged atmospheric energy transport from the annually averaged total (atmosphere plus ocean) energy transport, it is possible to infer the annually averaged energy transport by the oceans, as a residual. This also was done by Vonder Haar and Oort (1973). The ocean energy transport is very difficult to determine directly from measurements of the ocean’s temperature and currents, because of the lack of suitable data. In other words, we lack the data to work out the energy transport by the oceans using...
Some results for the poleward transport of energy by the oceans are shown in Fig. 5.6. These curves were obtained by the “residual” method described above. One direct, “hydrographic” estimate (due to Bennett, 1978) is also shown in the figure, however. The peak ocean energy transports are on the order of 30 to 40% of the total transport by the atmosphere and ocean combined; in other words, the oceans appear to make a very significant contribution to the total poleward energy transport by the climate system. Fig. 5.7 summarizes the contributions of the ocean and atmosphere to the poleward energy transport.

5.2 A simple theory of the Hadley circulation

We now discuss a theory of an idealized mean meridional circulation, without eddies. Much of the discussion is based on the work of Held and Hou (1980), as summarized by Lindzen (1990). For this purpose, we temporarily adopt a simplified set of equations to describe the mean meridional circulation, as follows:

\[
\nabla \cdot \mathbf{V} = 0, \quad \quad (5.11)
\]

\[
\nabla \cdot (\mathbf{V} u) - \left( f + \frac{u \tan \phi}{a} \right) v = \frac{\partial}{\partial z} \left( \sqrt{\frac{\partial u}{\partial z}} \right), \quad \quad (5.12)
\]
Figure 5.6: The northward ocean heat transports from the NCEP-derived and ECMWF-derived products are compared (top) for the Atlantic Ocean with direct ocean estimates from sections, as identified in the key. The dashed curves show estimates of the uncertainty for the derived transports. Where given in the original source, error bars are also plotted and the symbol is solid. Slight offsets in latitude are introduced where overlap would otherwise occur. Several sections are not exactly along a latitude circle, notably those for Bacon (1997) at 55°N and the Saunders and King (1995) section along 45°S (South America to 10°E) to 35°S (Africa), plotted at 40°S. (middle) Comparison of the derived results with transports from the simulations with a coupled ocean–atmosphere model called HADCM3 and another model called CSM for the Atlantic. (bottom) Results for the global ocean along with those from Macdonald and Wunsch (1996) at 24°N and 30°S, and at 24°N the combined Lavin et al. (1998) and Bryden et al. (1991) and for Ganachaud and Wunsch (2000). From Trenberth and Caron (2001).
Here $\mathbf{V} = (v, \omega)$ is a two-dimensional vector in the latitude-height plane, $\phi = p/\rho_0$, and $\theta_0$ and $\rho_0$ are constant “reference” values of the potential temperature and density, respectively. These equations are idealized, and require some explanation:

- We have assumed a steady state.
- We have assumed that there are no longitudinal variations whatsoever. One effect of this assumption is to eliminate the pressure gradient term of the equation of zonal motion.
- We are using the Bousinesq approximation for simplicity, so that the continuity equation reduces to non-divergence of the velocity field, and the hydrostatic equation reduces to the form given by (5.15).
- We have assumed that “friction” is due to downgradient mixing, with a non-negative and spatially constant mixing coefficient $\nu$.

\begin{align}
\nabla \cdot (\mathbf{V} v) + \left( f + \frac{u \tan \phi}{a} \right) u &= -\frac{1}{a \phi} \frac{\partial \phi}{\partial \phi} + \frac{\partial}{\partial z} \left( \mathbf{V} \frac{\partial v}{\partial z} \right), \quad (5.13) \\
\n\nabla \cdot (\mathbf{V} \theta) &= \frac{\partial}{\partial z} \left( \mathbf{V} \frac{\partial \theta}{\partial z} \right) - \frac{(\theta - \theta_E)}{\tau}, \quad (5.14)
\end{align}
The mean meridional circulation

• We have assumed that there is a vertical mixing of potential temperature, with the same mixing coefficient as for momentum.

• We have assumed that the heating can be represented by “relaxation,” with constant (positive) time scale $\tau$, to an “equilibrium” potential temperature, $\theta_E$, which can be thought of as the distribution of $\theta$ that would occur in radiative-convective equilibrium. Note that the symbol $\theta_E$ does not denote the equivalent potential temperature.

We simply specify $\theta_E$, as follows:

$$\theta_E(\varphi, z) = \theta_0 \left[ 1 - \Delta_H \sin^2 \varphi + \Delta_v \left( \frac{z}{H} - \frac{1}{2} \right) \right]. \quad (5.16)$$

Here $\theta_0 \equiv 400$ K is a reference potential temperature, $\Delta_H \equiv 0.3$ is the fractional potential temperature drop (of the “radiative-convective equilibrium state”) from the Equator to the pole, and $\Delta_v \equiv 0.3$ is the potential temperature drop (again, of the “radiative-convective equilibrium state”) from $z = H$ to the ground. We have to chose $\Delta_H \leq 1$ and $\Delta_v \leq 1$. For $\Delta_H > 0$, $\theta_E$ decreases towards the poles. For $\Delta_v > 0$, $\theta_E$ increases upward. A plot of $\theta_E$ is given in Fig. 5.8.

![Figure 5.8: A plot of $\theta_E$ (vertical axis) as a function of latitude in radians (front axis) and normalized height (right axis).](image)

As boundary conditions, we use
A simple theory of the Hadley circulation

(5.17)

\[ \frac{\partial u}{\partial z} = \frac{\partial v}{\partial z} = 0 \text{ at } z = H, \]

(5.18)

\[ \frac{\partial \theta}{\partial z} = 0 \text{ at } z = H \text{ and } z = 0, \]

(5.19)

\[ w = 0 \text{ at } z = H \text{ and } z = 0, \]

(5.20)

\[ v\frac{\partial u}{\partial z} = Cu \text{ at } z = 0, \]

(5.21)

\[ v\frac{\partial v}{\partial z} = Cv \text{ at } z = 0, \]

(5.22)

\[ v = 0 \text{ at } \varphi = 0. \]

Here \( C \) is interpreted as a kind of “effective drag coefficient,” which is actually the true drag coefficient times an average wind speed. Eqs. (5.20) and (5.21) thus represent linearizations of the bulk aerodynamic drag law discussed earlier; this is a simplifying assumption. Eq. (5.22) is a symmetry assumption, also made for simplicity, that allows us to restrict our analysis to a single hemisphere.

The model presented above allows a balanced, purely zonal flow if \( v = 0 \), i.e. if there is no friction. In this simple solution, \( \theta = \Theta_E \) everywhere, and the zonal wind is given by \( u = u_E \), where, by definition, \( u_E \) satisfies

(5.23)

\[ \frac{\partial}{\partial z} \left( fu_E + \frac{u_E^2 \tan \varphi}{a} \right) = -\frac{g}{a \theta_0} \frac{\partial \Theta_E}{\partial \varphi}. \]

The boundary conditions (5.19) and (5.20) are not relevant in the absence of friction. If we use the lower boundary condition

(5.24)

\[ u_E = 0 \text{ at } z = 0 \text{ for all } \varphi, \]

we find\(^2\) that

(5.25)

\[ \frac{u_E}{\Omega a} = \left[ \left(1 + 2RoTz_H \right)^{1/2} - 1 \right] \cos \varphi, \]

where

\(^2\) The use of (5.24) is somewhat awkward, because we have assumed no friction. Ordinarily we would attribute the weakness of the surface winds to the effects of surface drag.
The mean meridional circulation

An Introduction to the General Circulation of the Atmosphere

\[ Ro_T \equiv \frac{g H \Delta_H}{(\Omega a)^2} \]  (5.26)

is called the “thermal Rossby number.” Eq. (5.25) describes a zonal velocity that becomes increasingly westerly with height at all latitudes, even over the Equator. A plot is given in Fig. 5.9.

![Figure 5.9: A plot of \( u_E \) (vertical axis) as a function of latitude in radians (front axis) and normalized height (right axis).](image)

A physical interpretation of the thermal Rossby number is as follows. The thermal wind relation can be written as

\[ \frac{\partial u}{\partial z} = -\frac{1}{f} \frac{g}{T} \frac{\partial T}{\partial \phi}, \]  (5.27)

which is analogous to

\[ \frac{u_T}{H} \sim \frac{g \Delta_H}{\Omega a}, \]  (5.28)

where \( u_T \) is the wind at the tropopause. A “Rossby number” can then be defined as

\[ \frac{u_T}{\Omega a} = \frac{g H \Delta_H}{(\Omega a)^2}, \]  (5.29)
which agrees with (5.26). In short, the thermal Rossby number is a “regular” Rossby number (i.e. a wind scale divided by the product of the Coriolis parameter and the horizontal length scale), constructed using the wind at the tropopause (as the wind scale), divided by the product of the Earth’s rotation rate (in the place of the Coriolis parameter) times the radius of the Earth (as the horizontal length scale). For the Earth’s atmosphere,

\[ Ro_T \equiv 0.226. \]  

(5.30)

Let \( M_E \) be the angular momentum per unit mass associated with \( u_E \):

\[ M_E = a \cos \varphi (u_E + \Omega a \cos \varphi). \]  

(5.31)

From (5.25), we see that

\[ \frac{M_E}{\Omega a^2} = \cos^2 \varphi \left( 1 + 2 Ro_T \frac{z}{H} \right)^{1/2}. \]  

(5.32)

Inspection of (5.32) shows that

\[ \frac{M_E}{\Omega a^2} > 1 \text{ at the Equator.} \]  

(5.33)

In fact, \( M_E \) has a maximum on the Equator, at \( z = H \).

We now prove that with the “downgradient” momentum diffusion assumed here, a purely zonal circulation is impossible. This is called “Hide’s theorem,” after work by R. Hide. As a first step, we rewrite (5.12) as a conservation law for angular momentum, i.e.

\[ \nabla \cdot (\nabla M) = \nabla \cdot (v \nabla M), \]  

(5.34)

where, as before, \( M \) is the angular momentum per unit mass. We have introduced two-dimensional eddy mixing for generality. Suppose that \( M \) has a maximum somewhere in the interior of the atmosphere. We can draw a contour of constant \( M \) around this maximum, in the \((\varphi, z)\) plane. If we integrate over the region enclosed by this contour, the advection term on the left-hand side of (5.34) must integrate to zero, because of (5.11). The friction term on the right-hand side will represent a sink of \( M \), however, because we have assumed that \( M \) is a maximum inside the contour. This means that (5.28) cannot be satisfied in this case, and we conclude that our assumption of a maximum of \( M \) inside the atmosphere is not tenable. For a similar reason, \( M \) cannot have a minimum inside the atmosphere.

Suppose now that \( M \) has a maximum at the Earth’s surface. Again we can draw a contour of constant \( M \) around this maximum and close it off along the Earth’s surface. As before, we can conclude, through the use of (5.11), that the advective term of (5.34) must vanish when integrated over the area enclosed by this contour. Friction with the air outside the
contour will still represent a sink of angular momentum, so the only chance for balance is if there is a source of angular momentum through drag with the Earth’s surface. Such a source can only occur where the surface winds are easterly.

Similarly, using (5.17), we can show that \( M \) cannot have a maximum at \( z = H \).

We thus conclude that, in the absence of eddies and with downgradient momentum transfer, any maximum of \( M \) must occur at the Earth’s surface in a region of easterlies. This is observed, as can be seen in the plots of \( M \) presented in Chapter 2.

We have already worked out that in the absence of a mean meridional circulation, \( u \) will satisfy (5.25), which means that \( M \) will increase upward at all latitudes including the Equator. This violates our conclusion above that \( M \) can have a maximum only at the Earth’s surface in a region of easterlies -- a conclusion that was reached based on the assumption of non-zero friction. We can conclude, therefore, that the solution given by (5.25) cannot apply if there is any friction in the system. In other words, the existence of friction guarantees that there will be a mean meridional circulation.

In particular there will have to be a Hadley circulation. When friction is present, the zonal wind given by \( u = u_E \) is unrealistically strong, because the meridional rate of change of potential temperature given by \( \Theta = \Theta_E \) is unrealistically rapid. A Hadley circulation makes the solution more realistic, because it transports heat poleward, thus reducing \( \Theta \) near the Equator, and increasing it at higher latitudes. The zonal wind decreases, accordingly.

Held and Hou further simplified their model in order to obtain approximate solutions analytically. Assuming conservation of angular momentum at the tropopause gives

\[
\begin{align*}
    u(\phi, H) &= \frac{\Omega a \sin^2 \phi}{\cos \phi} . \\
    \text{(5.35)}
\end{align*}
\]

Gradient balance tells us that

\[
\begin{align*}
    f u(H) + \tan \phi \frac{u^2}{a} (H) &= \frac{1}{a} \frac{d}{d \phi} \phi(H) . \\
    \text{(5.36)}
\end{align*}
\]

The geopotential on the right-hand side of (5.36) can be evaluated by using the hydrostatic equation, giving

\[
\frac{\phi(H)}{H} = g \frac{\hat{\Theta}}{\Theta_0} ,
\]

where the hat denotes a vertical mean. We can eliminate \( u(H) \) in (5.36) by using (5.35), leading to
This is a first-order ordinary differential equation for \( \hat{\theta} \). It can be integrated to yield

\[
\hat{\theta}(\varphi) = \hat{\theta}(0) - \frac{\theta_0}{2gH} \left( \frac{\Omega a \sin^2 \varphi}{\cos \varphi} \right)^2.
\]  
(5.39)

Now suppose that beyond the poleward edge of the Hadley circulation, i.e. for latitudes \( \varphi > \varphi^* \), the temperature is in radiative-convective equilibrium. We assume that the temperature is continuous at \( \varphi = \varphi^* \). Then

\[
\hat{\theta}(\varphi^*) = \hat{\theta}_E(\varphi^*).
\]  
(5.40)

Finally, we note that the Hadley circulation is an advective process, and so merely redistributes the potential temperature, without changing its average value, so that

\[
\int_0^{\varphi^*} \hat{\theta} \cos \varphi d\varphi = \int_0^{\varphi^*} \hat{\theta}_E \cos \varphi d\varphi.
\]  
(5.41)

Substituting into (5.40) and (5.41) from (5.39) and (5.16), we get

\[
\hat{\theta}(0) - \frac{\theta_0}{2gH} \left( \frac{\Omega a \sin^2 \varphi^*}{\cos \varphi^*} \right)^2 = \theta_0(1 - \Delta_H \sin^2 \varphi^*),
\]  
(5.42)

and

\[
\int_0^{\varphi^*} \left[ \hat{\theta}(0) - \frac{\theta_0}{2gH} \left( \frac{\Omega a \sin^2 \varphi}{\cos \varphi} \right)^2 \right] \cos \varphi d\varphi = \int_0^{\varphi^*} \theta_0(1 - \Delta_H \sin^2 \varphi) \cos \varphi d\varphi.
\]  
(5.43)

Eqs. (5.42) and (5.43) can be solved as two equations for the two unknowns \( \varphi^* \) and \( \hat{\theta}(0) \). For \( \varphi^* \ll 1 \), we find that

\[
\varphi^* \approx \frac{5}{3} \frac{R_o T}{\ell^3}.
\]  
(5.44)
This gives $\varphi^* \approx 35^\circ$, which is close to the right answer. Recall that $Ro_T$ varies as the inverse square of the rotation rate. The theory predicts, therefore, that strongly rotating planets will have Hadley cells that are tightly confined near the Equator, while slowly rotating planets will have Hadley cells that extend out further towards the poles. This prediction is discussed further in the next section.

5.3 Extension to other planetary atmospheres

Williams (1988) explored the sensitivity of the general circulation to the planetary rotation rate, using a numerical model originally developed to simulate the general circulation of the Earth’s atmosphere. The model is based on equations similar to those discussed in Chapters 2 and 3, with parameterizations of radiation, moist convection, and surface fluxes due to turbulence. Williams modified the model so that the lower boundary is a global ocean of zero heat capacity, and he ignored the possibility that the ocean could freeze. The insolation was prescribed to be the observed annual mean, and the distribution of cloudiness was prescribed to crudely mimic the observed cloudiness of the Earth. The model does not include a diurnal cycle, so that the sun is essentially a bright “torus” encircling the planet, rather than a point in the sky on the day-side of the planet; this idealization may be acceptable for sufficiently rapid rotation rates, but leads to obvious problems of interpretation for very slow rotation rates.

Williams performed a suite of extended numerical simulations in which the model adjusted to the rotation rate specified in each case. He measured the rotation rate in terms of

$$\Omega^* = \frac{\Omega}{\Omega_E}, \quad (5.45)$$

where $\Omega$ is the rotation rate of the hypothetical planet being simulated, and $\Omega_E$ is the rotation rate of the Earth. A real planet that rotates much more slowly than Earth is Venus. A real planet that rotates more rapidly than Earth is Jupiter.

Fig. 5.10 shows the latitude-height distribution of the zonally averaged zonal wind, for values of $\Omega^*$ ranging from 0 to 8. Look first at the panel corresponding to $\Omega^* = 1$, i.e. Earth-like conditions. We see a westerly jet at a latitude of 30°, and with peak strength on the order of 50 m s$^{-1}$, which is comparable to what is observed on Earth. We also see easterlies in the tropics and at high latitudes. The solution is in fact reasonably Earth-like, even though the effects of mountains, etc., have been completely omitted from the model.

As $\Omega^*$ decreases towards zero, the jet moves poleward, ultimately disappearing in the limit of no rotation. As $\Omega^*$ increases to 8, the westerly jet moves in towards the Equator.

Fig. 5.11 shows the stream function of the mean meridional circulation, for various values of $\Omega^*$. Not surprisingly, the case of $\Omega^* = 1$ produces an Earth-like Hadley circulation, consistent with the westerly jet noted above. As the rotation rate decreases, the Hadley cell broadens. In the limit of no rotation it extends all the way to the pole. As the rotation rate increases, the Hadley cell contracts towards the Equator, and additional cells appear in middle and higher latitudes. The dependence of the latitudinal extent of the Hadley cell on the rotation rate is broadly consistent with the theory of Held and Hou (1980), as discussed in the previous subsection. The additional cells that appear at higher rotation rates are reminiscent of the many zonal bands that appear in pictures of Jupiter.
an introduction to the general circulation of the atmosphere

5.3 Extension to other planetary atmospheres

161

Fig. 5.12 shows the latitude–height cross section of the temperature, for various values of $\Omega^*$. The meridional temperature gradient is strong on rapidly rotating planets, and weak on slowly rotating planets. Rotation evidently interferes with efficient poleward transport of energy. It does so by restricting the latitudinal excursions of particles, as discussed in Chapter 3.

Finally, Fig. 5.13 shows the latitude–height cross section of the eddy kinetic energy,
The mean meridional circulation

An Introduction to the General Circulation of the Atmosphere

for various values of \( \Omega^* \). The exact definition of this quantity and the nature of its sources and sinks will be discussed in detail later; for now we simply note that it is a measure of the vigor of circulations that have longitudinal structures and arise, in these simulations, primarily from baroclinic instability. An interesting point is that the maximum eddy kinetic energy occurs for \( \Omega^* = 1 \); planets that rotate either more rapidly or less rapidly than Earth have less vigorous eddies, by this measure. Apparently the Earth’s rotation rate is just what is

\[ \Omega^* = \frac{\Omega}{\Omega_E} \]

Figure 5.11: The stream function of the mean meridional circulation, in \( 10^{10} \) kg s\(^{-1} \), for various values of the rotation parameter. From Williams (1988).
needed to maximize the storminess of the middle latitudes! The Earth is the stormiest of all possible planets — a meteorologist’s paradise.

5.4 Particle trajectories on the sphere: A partial explanation of “bandedness”

The bandedness of the circulation can be partially interpreted as a simple consequence of kinetic energy conservation in combination with angular momentum conservation. In
Combination, angular momentum conservation and kinetic energy conservation imply some very strong constraints on the motion of particles on a spherical surface (Cushman-Roisin, 1982; Paldor and Killworth, 1988; Pennell and Seitter, 1990). To see this, consider the motion of a particle in the absence of pressure forces and friction. The equations of motion for the particle are simply:

Figure 5.13: The latitude–height distribution of the eddy kinetic energy, per unit mass, in m$^2$ s$^{-2}$, for various values of the rotation parameter. From Williams (1988).
### 5.4 Particle trajectories on the sphere: A partial explanation of “bandedness”

It follows from (5.46) that the kinetic energy $K$ and the angular momentum $M$ of the particle do not vary as it moves:

$$K = \frac{1}{2}(u^2 + v^2) = \frac{1}{2}s^2 = \text{constant}, \quad (5.47)$$

$$M = (\Omega a \cos \varphi + u)a \cos \varphi = \text{constant}. \quad (5.48)$$

In (5.47) $s$ is the speed of the particle. Solving (5.48) for $u$, we find that

$$u = \frac{M - \Omega a^2 \cos^2 \varphi}{a \cos \varphi}. \quad (5.49)$$

You should be able to see from (5.46) that if a particle starts on the Equator with $v = 0$, it will remain on the Equator for all time, simply because $f + \frac{u \tan \varphi}{a} = 0$ there.

As a more interesting example, suppose that $u = 0$ at the Equator. Then (5.48) leads to

$$u = \Omega a \left(\frac{1 - \cos^2 \varphi}{\cos \varphi}\right) \geq 0. \quad (5.50)$$

From this result, it appears that $u \to \infty$ as $\varphi \to \pm \frac{\pi}{2}$. Recall, however, that $K$ is also conserved. It follows that $|u| \leq s$. This implies that there is a maximum value of $|\varphi|$ beyond which the particle cannot go; the particle is confined within a ring centered on the Equator. At the north and south edges of the ring, $u = s$, $v = 0$, and $|\varphi| = \varphi_{\text{max}}$. See Fig. 5.14.

We now demonstrate that in general a particle’s motion is confined within a range of latitudes. We allow an arbitrary choice of the initial latitude and velocity. Without loss of generality, assume that

$$\Omega > 0. \quad (5.51)$$

Let

$$y \equiv \cos \varphi, \quad (5.52)$$
and note that

\[ 0 \leq y \leq 1. \quad (5.53) \]

As the particle moves, its latitude changes so long as \( v \neq 0 \); its meridional motion is blocked where \( v = 0 \), i.e., it cannot cross a latitude where \( v = 0 \). In view of (5.47), at a latitude where \( v = 0 \) we have either \( u = s \) or \( u = -s \). Consider these two possibilities one at a time. In the first case, (5.48) reduces to

\[ y_1^2 + xy_1 - \mu = 0, \quad (5.54) \]

while in the second case we get

\[ y_2^2 - xy_2 - \mu = 0. \quad (5.55) \]

Here

\[ x \equiv \frac{s}{\Omega a} > 0 \quad \text{and} \quad \mu \equiv \frac{M}{\Omega a^2}. \quad (5.56) \]

Note that \( x \) and \( \mu \) do not change as the particle moves around; they are “invariants of the motion.” For the Earth’s atmosphere, \( \mu > 0 \) in virtually every conceivable situation. In principle, however, it would be possible to have \( \mu < 0 \).

The solutions of (5.54) and (5.55) are
5.4 Particle trajectories on the sphere: A partial explanation of “bandedness”

\[ y_1 = -\frac{1}{2}x + \sqrt{\frac{1}{4}x^2 + \mu} , \]  

(5.57)

and

\[ y_2 = \frac{1}{2}x + \sqrt{\frac{1}{4}x^2 + \mu} , \]  

(5.58)

respectively. In both cases, we have chosen the plus sign before the discriminant in order to satisfy \( y > 0 \). Note that

\[ \frac{1}{4}x^2 + \mu = \frac{\frac{1}{4}v^2 + \left(\frac{1}{2}u + \Omega a \cos \varphi\right)^2}{(\Omega a)^2} \geq 0 , \]  

(5.59)

which implies that \( y_1 \) and \( y_2 \) are real numbers.

By subtracting (5.57) from (5.58), we obtain

\[ y_2 - y_1 = x > 0 . \]  

(5.60)

This is a measure of the width of the latitude band within which the particle is confined. Rotation inhibits latitudinal excursions larger than \( \cos^{-1}\left(\frac{s}{\Omega a}\right) \). As the rotation rate increases, \( y_2 - y_1 \), decreases. From (5.60) we see that \( y_2 \) is always greater than \( y_1 \). This means that \( y_2 \) corresponds to the latitude closer to the Equator (in either hemisphere), and \( y_1 \) corresponds to the latitude further from the Equator.

From (5.57) we can show that the condition for \( y_1 = 0 \) is simply \( \mu = 0 \), i.e. the particle has no angular momentum. A particle without angular momentum can reach the poles. A particle with \( \mu \neq 0 \) will never visit the poles.

From (5.58), we can show that \( y_2 > 1 \) is equivalent to

\[ x + \mu > 1 . \]  

(5.61)

When \( y_2 > 1 \), i.e., when (5.61) is satisfied, a particle moving towards the Equator will not encounter a latitude where \( v = 0 \); it can therefore cross the Equator, and in fact it will keep moving until it reaches the latitude where \( y = y_1 \) in the opposite hemisphere. In such a case, the particle is confined in the vicinity of the Equator. We can say that the particle is “equatorially trapped.”
When (5.61) is not satisfied, a particle moving toward the Equator will encounter a latitude \( y = y_2 \) where \( \nu = 0 \) before it gets to the Equator. The particle will, therefore, stop short of the Equator, i.e., it will spend eternity in one hemisphere.

Since \( x \) decreases as the rotation rate increases, (5.60) suggests that rotation inhibits the latitudinal excursions of a particle undergoing inertial motion, i.e., rotation effectively confines the particle to a limited range of latitudes. To explore this further, consider first the case in which the particle’s motion is confined to one hemisphere. In terms of meridional distance,

\[
a(\varphi_1 - \varphi_2) = a(\cos^{-1}y_1 - \cos^{-1}y_2)
\]

\[
\equiv \frac{a(y_1 - y_2)}{\sqrt{1 - \left(\frac{y_1 + y_2}{2}\right)^2}}
\]

\[
= \frac{ax}{\sqrt{1 - \left(1 - \frac{x^2}{4} + \mu\right)}}.
\]

The approximation on the second line of (5.62) is valid when the band is sufficiently narrow. From (5.59), we can show that, for meteorologically relevant values of \( u \) and \( \nu \),

\[
\frac{1}{4}x^2 + \mu \equiv \cos^2 \varphi.
\]

Substituting (5.63) into (5.62), we find that

\[
a(\varphi_1 - \varphi_2) \equiv \frac{ax}{\sin \varphi} = \frac{s}{\Omega \sin \varphi} = \frac{2s}{f}.
\]

Here \( \varphi \) should be interpreted as a latitude in the middle of the particle’s “band.” We do not need to worry about division by zero in (5.64), because we are considering the case in which the particle’s motion is confined to one hemisphere.

Now consider the case in which the particle is equatorially trapped. Then the width of the particle’s band is given by

\[
2a\varphi_1 = 2a\cos^{-1}y_1
\]

\[
= 2a\cos^{-1}\left(-\frac{1}{2}x + \frac{1}{\sqrt{4x^2 + \mu}}\right).
\]

Using (5.63), this can be approximated as
Here we have used (5.56) and the approximation \( \cos \phi \equiv 1 \), which is appropriate for equatorially trapped particles. Since \( \cos x \equiv 1 - \frac{x^2}{2} \), we can simplify (5.66) to

\[
2a\phi_1 \equiv 2a \cos^{-1}\left(1 - \frac{s}{2\Omega a}\right). \tag{5.67}
\]

This is similar to the “Rhines length,” \( \frac{s}{\sqrt{\Omega a}} \), where \( \beta \equiv \frac{1}{\sqrt{\frac{d\phi}{a d\varphi}}} \), which will be discussed in a later chapter.

It is clear from (5.64) and (5.67) that faster rotation confines the motion of a particle to a narrower latitudinal band. We can interpret this as a partial explanation for the results of Williams’ numerical experiments, which show that more “bands” occur as the planetary rotation rate increases. When there are more bands, each band is narrower, and particles moving within a band are confined to a narrower range of latitudes.

The simple analysis given above is useful for interpretation of the observed mean meridional circulation. Nevertheless it is important to keep in mind that this analysis is highly idealized and has only limited relevance to the atmospheric circulation. A key simplifying assumption made above is that no pressure-gradient forces act on the moving particles. Notice that it would be possible for two of our inertially moving particles to collide. In the atmosphere, such collisions are prevented by the pressure-gradient force, which acts as a kind of “air-traffic control.”

### 5.5 Summary

We discussed the energy balances of the atmosphere and ocean, which imply meridional transports of energy by the circulations of the atmosphere and ocean. The transports were dissected to show the contributions of the mean meridional circulation, the transient eddies, and the stationary eddies. We also separated out the transports of dry static energy and latent energy.

This was followed by a discussion of the theory of a zonally symmetric circulation, following Held and Hou (1980). Finally, we examined the numerical simulations of Williams (1988), which show how the mean meridional circulation varies as the planetary rotation rate is changed.

### Problems

1. To estimate the width of the Hadley Cell, we used Eq. (5.40), which states that the temperature is continuous at the edge of the cell. Repeat the analysis, assuming
instead that the zonal wind is continuous. Compare the results obtained with these two assumptions.

2. Consider an air parcel at rest at the sea surface on the Equator. If the parcel rises from the surface to an altitude of 15 km, conserving its angular momentum, what is its zonal velocity? For purposes of this problem, define the axial component of the angular momentum by

\[ M \equiv r \cos \phi (\Omega r \cos \phi + u) , \quad (5.68) \]

where \( r \) is the radial distance of the parcel from the center of the Earth.

3. Look up the information that you need to determine the thermal Rossby number for Mars. Tabulate the information that you use and the source of the information. Estimate the widths of the Hadley circulation, in degrees of latitude, for Mars.

4. Derive (5.44) from (5.42) and (5.43).
CHAPTER 6  An overview of the effects of radiation and convection

6.1 Convective energy transports

Riehl and Malkus (1958; Fig. 6.1) 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 intertropical convergence zone (ITCZ). Recall that the “body” of the main solstitial Hadley Cell lies in the winter hemisphere. They neglected the mass transport across the boundary of the ITCZ on the summer side. They then attempted to evaluate the lateral energy transports across into and out of the ITCZ, as functions of height. They considered transports of internal energy, potential energy, and latent energy. 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 next export of energy by the meridional flow implies that there is a compensating net input of energy at the top and bottom of the column. Their estimates of the various quantities are summarized in Table 6.1.

Because energy flows into the ITCZ at low levels, and out at high levels, Riehl and Malkus concluded that there must be a net upward transport of energy inside the ITCZ. They argued, however, that this upward energy flux cannot be due to the mean flow, because the observed profile of moist static energy has a minimum at mid levels, as discussed in Chapter

Figure 6.1: Joanne Malkus (now Joanne Simpson), and the late Herbert Riehl.
2. If the mean flow was acting alone, then since $h$ is conserved following parcels $h$ would be uniform with height throughout the ascending column. Similar reasoning shows that diffusive energy transport cannot explain the observed upward energy flux. Riehl and Malkus concluded that the upward energy transport must occur in deep convective clouds that penetrate through the troposphere.

**Table 6.1: Lateral energy transports on the poleward side of the ITCZ. Adapted from Riehl and Malkus (1958).**

<table>
<thead>
<tr>
<th>$\delta p$</th>
<th>$v$</th>
<th>$M_0$</th>
<th>$s$</th>
<th>$sM_0$</th>
<th>$Lq$</th>
<th>$hM_0$</th>
<th>Eddy moisture transport,</th>
<th>Total energy transport out of the ITCZ,</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-9</td>
<td>-1.3</td>
<td>-5.2</td>
<td>301.5</td>
<td>-1.56</td>
<td>37.6</td>
<td>-0.20</td>
<td>10$^{16}$ J s$^{-1}$</td>
<td>10$^{16}$ J s$^{-1}$</td>
</tr>
<tr>
<td>9-8</td>
<td>-1.1</td>
<td>-4.4</td>
<td>305.7</td>
<td>-1.34</td>
<td>27.6</td>
<td>-0.12</td>
<td>10$^{16}$ J s$^{-1}$</td>
<td>10$^{16}$ J s$^{-1}$</td>
</tr>
<tr>
<td>8-7</td>
<td>-0.4</td>
<td>-1.6</td>
<td>311.6</td>
<td>-0.49</td>
<td>18.4</td>
<td>-0.03</td>
<td>10$^{16}$ J s$^{-1}$</td>
<td>10$^{16}$ J s$^{-1}$</td>
</tr>
<tr>
<td>7-6</td>
<td>0</td>
<td>0</td>
<td>317.8</td>
<td>0</td>
<td>11.3</td>
<td>0</td>
<td>10$^{16}$ J s$^{-1}$</td>
<td>10$^{16}$ J s$^{-1}$</td>
</tr>
<tr>
<td>6-5</td>
<td>0</td>
<td>0</td>
<td>323.3</td>
<td>0</td>
<td>7.1</td>
<td>0</td>
<td>10$^{16}$ J s$^{-1}$</td>
<td>10$^{16}$ J s$^{-1}$</td>
</tr>
<tr>
<td>5-4</td>
<td>0.3</td>
<td>1.2</td>
<td>329.1</td>
<td>0.39</td>
<td>4.2</td>
<td>0.00</td>
<td>10$^{16}$ J s$^{-1}$</td>
<td>10$^{16}$ J s$^{-1}$</td>
</tr>
<tr>
<td>4-3</td>
<td>0.6</td>
<td>2.4</td>
<td>335.4</td>
<td>0.80</td>
<td>2.1</td>
<td>0.00</td>
<td>10$^{16}$ J s$^{-1}$</td>
<td>10$^{16}$ J s$^{-1}$</td>
</tr>
<tr>
<td>3-2</td>
<td>1.3</td>
<td>5.2</td>
<td>340.8</td>
<td>1.77</td>
<td>0.8</td>
<td>0.00</td>
<td>10$^{16}$ J s$^{-1}$</td>
<td>10$^{16}$ J s$^{-1}$</td>
</tr>
<tr>
<td>2-1.25</td>
<td>0.8</td>
<td>2.4</td>
<td>348.4</td>
<td>0.83</td>
<td>0</td>
<td>0</td>
<td>10$^{16}$ J s$^{-1}$</td>
<td>10$^{16}$ J s$^{-1}$</td>
</tr>
<tr>
<td>10--5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-1.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-1.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Neelin and Held (1987) considered the moist static energy budget of the ITCZ from a similar perspective. In a time average, the vertically integrated moist static energy budget is expressed by

$$g^{-1} \nabla \cdot \left( \int_0^{h^*} \mathbf{v} dh \right) = -(N_S - N_T) ,$$

(6.1)
where $N$ is the net downward flux of energy due to turbulence, convection and radiation, and subscripts $T$ and $S$ denote the top of the atmosphere and the surface, respectively. Similarly, mass continuity gives

$$\nabla \cdot \left( \int_0^{p_S} \mathbf{V} dp \right) = 0 . \quad (6.2)$$

Divide the column into upper and lower portions, and write

$$p_S^{-1} \nabla \cdot \left( \int_0^{p_S} \mathbf{V} h dp \right) = \nabla \cdot \left( \mathbf{V} h \right)_u + \nabla \cdot \left( \mathbf{V} h \right)_l = -g \left( \frac{N_S - N_T}{p_S} \right), \quad (6.3)$$

$$p_S^{-1} \nabla \cdot \left( \int_0^{p_S} \mathbf{V} dp \right) = \nabla \cdot \mathbf{V}_u + \nabla \cdot \mathbf{V}_l = 0 . \quad (6.4)$$

For the tropics, horizontal variations of $h$ are weak, so that it is useful to define $h_u$ and $h_l$ by

$$\nabla \cdot (\mathbf{V} h)_u \equiv h_u (\nabla \cdot \mathbf{V}_u), \quad (6.5)$$

$$\nabla \cdot (\mathbf{V} h)_l \equiv h_l (\nabla \cdot \mathbf{V}_l) . \quad (6.6)$$

Combining (6.3) - (6.6) gives

$$\nabla \cdot \mathbf{V}_l = \frac{g}{p_S} \left( \frac{N_S - N_T}{h_u - h_l} \right) . \quad (6.7)$$

This shows 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

$$h_u - h_l > 0 . \quad (6.8)$$

We note that (6.8) must be satisfied if the upper branch carries energy away faster than the lower branch carries energy in. Neelin and Held call $h_u - h_l$ the “gross moist stability.” According to (6.7), the pattern of the gross moist stability is closely linked to the pattern of low-level convergence, for a given distribution of $N_T - N_S$. As discussed later [see (6.26)], we expect the gross moist stability to be small where cumulus convection is active.

### 6.2 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.
An overview of the effects of radiation and convection

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.

The simplest model that can represent this process is called a radiative-convective model. The basic idea is very simple. First, assemble physical parameterizations that suffice to determine the time rates of change of the temperature within the atmospheric column and at the Earth’s surface. Combine them into a model, and integrate using a time step on the order of an hour or so. Repeat until a steady state is approached to sufficient accuracy. Depending on the initial conditions, convergence can take on the order of 500 simulated days. No significance is ascribed to the time evolution itself; only the steady state is of interest. It is not obvious a priori that a radiative-convective model will actually approach a steady state, but the models discussed below do.

The physical ingredients of a radiative-convective model include parameterizations of radiation, convection, turbulence, and the processes that determine the change of the surface temperature. We write

\[ \rho c_p \frac{\partial T}{\partial t} = L C - \frac{\partial F_S}{\partial z} + Q_R. \]  

(6.9)

Horizontal and vertical advection are deliberately omitted in (6.9), 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 (6.9), we must determine the condensation rate and the convective fluxes. Obviously 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

\[ Q_R = \frac{\partial}{\partial z} (S - R). \]  

(6.10)

Here \( S \) is the net solar radiation (positive down), and \( R \) is the net terrestrial radiation (positive up).

We also impose an energy budget for the Earth’s surface:

\[ C_g T_S \frac{\partial T_S}{\partial t} = N_S(T_S). \]  

(6.11)

Here \( C_g \) is the effective “heat capacity” of the surface, \( T_S \) is the surface temperature, and \( N_S(T_S) \) denotes the net downward vertical flux of energy due to turbulence, convection, and radiation. Eq. (6.11) can be used to determine \( T_S \), provided that the functional form of \( N_S(T_S) \) is specified. The value of \( C_g \) determines how rapidly the surface temperature
changes in response to a given value of $N_\text{S}$. When $C_g$ is large $T_\text{S}$ changes slowly. When $C_g \to 0$, $T_\text{S}$ adjusts instantaneously so as to keep $N_\text{S}(T_\text{S}) = 0$.

The radiative-convective balance requirements that must be satisfied in equilibrium can be stated as follows:

- There can be no net radiative energy flux at the top of the atmosphere:
  \[ N_\text{T} = S_\text{T} - R_\text{T} = 0. \]  
  (6.12)

- This was discussed in Chapter 1. The atmospheric column must be in energy balance:
  \[ N_\text{S} = N_\text{T}, \]  
  (6.13)
  where
  \[ N_\text{S} \equiv S_\text{S} - R_\text{S} - (F_h)_\text{S}. \]  
  (6.14)

- From (6.12) - (6.13), it follows that in equilibrium the Earth’s surface must also be in energy balance:
  \[ N_\text{S} = 0. \]  
  (6.15)

We now discuss some results from radiative-convective models. In a series of studies during the 1960s (see bibliography), Manabe and his colleagues investigated the degree to which pure radiative equilibrium and/or radiative-convective equilibrium can explain the observed vertical distribution of temperature. These studies made major advances in our understanding of the vertical structure of the atmosphere. Because little was known about moist physics during the 1960s\(^1\), Manabe et al. did not explicitly represent moist processes in their model; instead, they made alternative assumptions for 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.

They used a time-marching method to find equilibrium solutions. Suppose that initial conditions are specified for the temperatures of the atmosphere and surface. From this information, together with a specification of the composition of the atmosphere, it is possible to determine the net radiative cooling of the atmosphere at each level, due to both terrestrial and solar radiation. By integration, we can then determine the net atmospheric radiative cooling, $ARC$, which is given by

\[ ARC = (R_\text{T} - R_\text{S}) - (S_\text{T} - S_\text{S}). \]  
(6.16)

---

\(^1\) And only a little more is known now...
Manabe et al. assumed that on each time step the net radiative cooling of the atmosphere is equal to the net radiative warming of the surface, i.e.

\[ ARC = S_S - R_S. \quad (6.17) \]

From (6.12) and (6.16) it is clear that (6.17) must hold in equilibrium, but it is not logically required during the approach to equilibrium.

Manabe et al. also assumed that on each time step the excess of net downward surface solar radiation over net upward surface longwave radiation equals the net integrated radiative cooling of the atmospheric column. Again, this must really be true only in equilibrium. From (6.14), (6.15), and (6.17), it follows that

\[ (F_h)_S = ARC. \quad (6.18) \]

An updated surface temperature can then be determined by time-stepping (6.11) with (6.14).

Manabe et al. imposed (6.15) even during the approach to equilibrium. The physical interpretation of this assumption is that the heat capacity of Earth’s surface is zero; in that case, (6.15) follows immediately from (6.11). Realistically, of course, the Earth’s surface does have a finite heat capacity (primarily in the ocean), and so (6.15) does not apply at any given instant. Because Manabe et al. were mainly interested in the equilibrium solution anyway, however, their assumption is fairly harmless.

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 Wetherald did not attempt to compute the vertical distribution of ozone; instead they prescribed it according to observations.

They also had to determine the distribution of water vapor. Although they did not attempt to compute it from a moisture budget (a very serious omission), they did consider two alternative assumptions. The first is fixed specific humidity, and the second is fixed relative humidity. These were both prescribed from observations similar to those shown in Chapter 2. Although Manabe and Wetherald’s model does not include a water budget, more modern radiative-convective models do include explicit water budgets. In these models, water vapor is introduced by evaporation from the sea surface; convection and to a much smaller degree diffusion carry the moisture upward; and precipitation removes it. An example will be given later.

A radiative-convective equilibria can be found through the procedure summarized in Fig. 6.2. 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 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\(^{-1}\). The vertical distribution of water vapor is then corrected, using the assumption of either fixed absolute humidity or fixed relative humidity.
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, which is close to 6.5 K km\(^{-1}\) in the tropical lower troposphere. The physical meaning of this hypothesis will be discussed further later in this chapter.

Fig. 6.3 shows Manabe and Wetherald’s results for three cases:

- pure radiative equilibrium of the clear atmosphere with a given distribution of relative humidity,
- 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 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 radiative-convective calculations, this instability is assumed to be removed by convection. The radiative-convective equilibrium with fixed relative humidity is amazingly realistic, considering that the model ignores all large-scale dynamical processes.

We now present the results of more modern radiative-convective equilibrium calculations performed using the physical parameterizations of the Colorado State University general circulation model. In these simulations, the surface temperature was prescribed according to the observed zonal annual means, and annually averaged insolation was used at each latitude. The model incorporates an elaborate theory of the interactions of cumulus convection with the large-scale circulation (see the next subsection), based on the work of...
An overview of the effects of radiation and convection

Arakawa and Schubert (1974, discussed in the next sub-section), Randall and Pan (1993), and Pan and Randall (1998) as well as representations of stratiform cloud processes (Fowler et al., 1996) and boundary-layer turbulence (Suarez et al. 1983). A surface wind speed is needed to determine the surface fluxes of sensible and latent heat using bulk aerodynamic formulae similar to those discussed earlier; a value of 5 m s\(^{-1}\) was assumed for this purpose. The boundary layer, cumulus, and stratiform cloud parameterizations together determine the equilibrium distribution of moisture. The boundary layer, cumulus, stratiform cloud, and radiation parameterizations determine the distribution of temperature. Clouds in the radiative sense were neglected. The top of the model was placed at 1 mb, and 29 levels were used. Stratospheric ozone amounts were prescribed from observations. Note that in this model the distribution of moisture is predicted; this is a key difference from the work of Manabe and colleagues, discussed above.

Fig. 6.4 shows the results with and without prescribed stratospheric ozone. When ozone is present, the model produces a very obvious stratosphere in which the temperature increases upward; when ozone is neglected, on the other hand, the upper regions of the model atmosphere become more or less isothermal. The structure of the tropospheric temperature sounding is only slightly altered by the effects of stratospheric ozone.

Fels (1985) reported the results of similar radiative-convective equilibria on the sphere, but his model included a representation of photochemistry, and he emphasized the
Figure 6.4: Temperature profiles for radiative-convective equilibrium, as simulated by the physical parameterizations of the CSU GCM. The surface temperature was prescribed as a more or less realistic function of latitude, and for each latitude the annual mean insolation was used.
An overview of the effects of radiation and convection

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. The results obtained by Fels are shown in Fig. 6.5. The observed state of the

Figure 6.5: (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).

atmosphere (see panel b of Fig. 6.5) resembles that predicted by the model in the summer stratosphere, but the model is much too cold in the winter polar stratosphere. 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 large-scale motions, 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. It appears that large-scale motions have little effect on the thermal structure of the middle atmosphere in summer, however. This indicates that in the summer the middle atmosphere is close to a state of radiative equilibrium.

6.3 The observed vertical structure of the atmosphere, and the mechanisms of vertical energy transport

Fig. 6.6 shows the observed vertical structure of the atmosphere in the tropics, the subtropical tradewind regime, and the subtropical marine stratocumulus regime. The quantities plotted are the dry static energy, defined by

\[ s = c_p T + gz \]  

(6.19)

the moist static energy, defined by
Figure 6.6: Representative observed soundings for Darwin, Australia, 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 curves plotted show the dry static energy, the moist static energy, and the saturation moist static energy. The panels on the right cover both the troposphere and lower stratosphere, while those on the right zoom in on the lower troposphere to show more detail. Values are divided by $c_p$ to give units in K.
and the saturation moist static energy, defined by

\[ h^* \equiv s + Lq^*. \] (6.21)

Here \( q^* \) is the saturation mixing ratio. At cold temperatures, these three quantities are nearly equal, because both the water vapor mixing ratio and the saturation mixing ratio are small. Quite generally, we have

\[ s \leq h \leq h^*. \] (6.22)

In cloudy air, \( h = h^* \). In very dry air, \( h \equiv s \). In very cold air, \( h^* \equiv h \equiv s \). It can be shown that the dry static energy increases upward in a statically stable atmosphere. The dry static energy is approximately conserved under dry adiabatic processes, while the moist static energy is approximately conserved under dry adiabatic, moist adiabatic, and pseudo-adiabatic processes. The saturation moist static energy is not a conservative variable, and despite its name it does not actually carry any information about the moisture field; the sounding of the saturation moist static energy is essentially determined by the temperature sounding.

If a parcel of air containing vapor 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. Once the lifting condensation level has been exceeded, the moist static energy of the now-cloudy parcel will continue to remain constant, while its dry static energy 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 be conserved.

Note that conservation of \( h \) and \( q_v + l \) implies that

\[ s_t \equiv h - L(q_v + l) \] (6.23)

is also conserved. We refer to \( s_t \) 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 \( s_t \), but it does not change the value of \( h \). This means that \( h \) is “more conservative” than \( s_t \). Nevertheless both variables are useful.

The temperature difference between the cloudy air and its environment at the same level is simply proportional to the saturation moist static energy difference, i.e.

\[ c_p(T_c - \bar{T}) \sim h^* - \bar{h}^* \] (6.24)
where a subscript \( c \) denotes the cloudy air, and an overbar denotes the environment. Because the cloudy air is saturated, however, we can write

\[
c_p(T_c - \bar{T}) \sim h_c - \bar{h}^*.
\]

(6.25)

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. The cloudy updraft will stop when it encounters a level where \( h_c = \bar{h}^* \); if this level is high and cold, then \( \bar{h} \equiv \bar{h} \), and so we expect to find

\[
\bar{h}_{\text{tropopause}} \equiv \bar{h}_{\text{boundary layer}} \text{ in regions where deep convection is active.}
\]

(6.26)

In the terminology of Neelin and Held (1987), Eq. (6.29) means that the gross moist stability is small. See (6.8).

We can apply these ideas to the tropical (i.e. Darwin) sounding shown in Fig. 6.6. Near the surface, \( h^* > h \), indicating that the air is unsaturated. If we lift a parcel adiabatically from near the surface, its moist static energy will follow 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 \( h_c > \bar{h}^* \). According to (6.25), the parcel will then be positively buoyant, \( \text{if it is saturated} \). We can thus tell, simply by looking at Fig. 6.6, 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.

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. The sounding 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. 6.7.

The physical picture represented by the three soundings shown in Fig. 6.6 is
summarized in Fig. 6.8. 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 region of deep convection, characterized by warm sea surface temperatures and large-scale rising motion.

As the air descends in the subtropical branches of the Hadley cells, it is gradually
An Introduction to the General Circulation of the Atmosphere

6.3 The observed vertical structure of the atmosphere, and the mechanisms of verti--

The observed vertical structure of the atmosphere, and the mechanisms of vertical overturn is cooled by radiation. As a result, the potential temperature of the air in the subtropical free atmosphere decreases downward, or in other words it increases upward.

The lapse rate of the deep convective zones is essentially determined by convection. As already noted, 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. We can write a rough thermodynamic balance for the descending branch of the Hadley cell as follows:

\[ \omega \frac{\partial s}{\partial p} = Q_R. \]  (6.27)

Here \( \omega \) is the positive large-scale pressure velocity, corresponding to sinking motion; \( \frac{\partial s}{\partial p} \) is the rate of change of the dry static energy with height, which we have just argued is essentially imposed, above the trade inversion, by the moist convective processes of the deep tropics; and \( Q_R < 0 \) is the radiative cooling. We see from (6.27) that the speed of the large-scale sinking motion in the subtropics is essentially determined by the requirement of thermodynamic balance. Typical clear-sky tropospheric radiative cooling rates in the tropics are on the order of 2 K per day (Fig. 6.9).

The sinking air passes through the trade inversion. How does this happen? It is remarkable, for example, that the average mixing ratio of the air suddenly increases from perhaps 1 g kg\(^{-1}\) above the trade inversion to 6 or 7 g kg\(^{-1}\) below the trade inversion. This is the

Figure 6.8: Schematic diagram summarizing the relationships among the cloud regimes depicted in Fig. 6.6, and how they fit into the mean meridional circulation. From Schubert et al. (1995).
same air, after all; how does it suddenly become so moist? The answer is that the convective vertical motions associated with the shallow stratocumulus and/or trade cumulus clouds transport moisture upwards, and deposit it at the base of the inversion, where it is used to moisten the sinking air. The moisture used for this purpose is carried up from the sea surface by a combination of turbulence and shallow cloudy convection.

The air also cools as it descends through the inversion. This cooling is produced by a combination of concentrated radiative cooling near cloud tops, evaporative cooling due to the evaporation of liquid deposited at the trade inversion level by the shallow clouds, and a downward flux of sensible heat that cools the air as it crosses the inversion.

A macroscopic view of this entrainment process is as follows. Let $A$ be an arbitrary scalar, satisfying a conservation equation that can be written in “flux form” as

$$\frac{\partial}{\partial t}(\rho A) + \nabla \cdot (\rho \nabla A) + \frac{\partial}{\partial z} (\rho w A) = -\frac{\partial F_A}{\partial z} + S_A,$$

(6.28)

where $F_A \equiv \rho w^T A'$ is the upward turbulent flux of $A$, bars are omitted on the mean.

Figure 6.9: Estimates of the clear-sky radiative cooling rate in the eastern and central Pacific, based on ECMWF data.

An overview of the effects of radiation and convection
quantities, and \( S_A \) is a source or sink of \( A \), per unit volume. Integrating (6.28) from just below to just above the inversion, and using Leibniz’ rule, we get

\[
\frac{\partial}{\partial t} \int_{z_B^-}^{z_B^+} \rho A dz - \Delta(\rho A) \int_{z_B^-}^{z_B^+} \rho V_A dz - \nabla \cdot \int_{z_B^-}^{z_B^+} \rho V A dz - \Delta(\rho V A) \cdot \nabla z_B
\]

\[
+ \Delta(\rho w A) = - (F_A)_B^+ + (F_A)_B^- + \int_{z_B^-}^{z_B^+} S_A dz ,
\]

(6.29)

where the indicated terms drop out as the domain of integration shrinks to zero and/or because all of the turbulence variables go to zero above the inversion. Here we have used the notation 
\( \Delta(\ ) \equiv (\ )_{z = z_B^+} - (\ )_{z = z_B^-} \) and henceforth subscripts \( B \) and \( B^+ \) denote levels just above and just below the inversion, respectively. For \( A \equiv 1 \), (6.29) reduces to mass conservation in the form

\[
\rho_{B^+} \left( \frac{\partial z_B}{\partial t} + V_{B^+} \cdot \nabla z_B - w_{B^+} \right) = \rho_B \left( \frac{\partial z_B}{\partial t} + V_B \cdot \nabla z_B - w_B \right) = E ,
\]

(6.30)

where \( E \) is the downward mass flux across the inversion. In essence, (6.30) simply says that the mass flux is continuous across the PBL 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 (6.30), we can rewrite (6.29) as

\[
-\Delta A E = (F_A)_B^- + \int_{z_B^-}^{z_B^+} S_A dz ,
\]

(6.31)

For \( S_A \equiv 0 \), (6.31) 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 \( A_B \). Lilly (1968; Fig. 6.10) was the first to derive (6.31) using the approach followed here.

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

\[
-\Delta q_T E = (F_{q_T})_B^- ,
\]

(6.32)
An overview of the effects of radiation and convection

which is a special case of (6.31). Here $q_T \equiv q_r + l$ is the total water mixing ratio.

After sinking through the trade inversion, the air is subjected to friction. This causes its angular momentum to decrease. As it flows back equatorward, near-surface easterlies result.

Turning now to middle and high latitudes, Fig. 6.6 shows representative summer and winter soundings for Denver, Colorado and Barrow, Alaska. The Denver sounding is conditionally unstable in summer, but the dry near-surface air has to be lifted quite a long way before it can become positively buoyant; we expect to see high cloud bases. The winter sounding for Denver is strongly stable near the surface. The Barrow sounding is quite stable all year, but especially so in winter. Note that the tropopause is much lower at Barrow that at Darwin. In summer, there are low clouds at Barrow.

6.4 More on moist convection

The preceding discussion suggests that moist convection is important in two ways:

- As shown observationally by Riehl and Malkus (1958), moist convection is the primary mechanism to transport energy upward in the deep tropics. 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. Even if we accept the validity of this hypothesis, we are faced with the problem of determining when and where convection is active.

In addition, of course, moist convection is important because:

- Convection produces a large fraction of the Earth’s precipitation.

- Convection generates radiatively important stratiform clouds, especially in the upper troposphere in regions of deep convection.

For the four reasons summarized in the four bullets above, moist convection is crucially
Figure 6.11: Representative observed soundings for Denver, Colorado and Barrow, Alaska, for July and January. The curves plotted show the dry static energy, the moist static energy, and the saturation moist static energy. The panels on the right cover both the troposphere and lower stratosphere, while those on the right zoom in on the lower troposphere to show more detail. Values are divided by $c_p$ to give units in K.
important for the general circulation of the Earth’s atmosphere. It is no exaggeration to say that we cannot understand the general circulation unless we understand the interactions between the general circulation and moist convection. In fact there is an enormous literature on this subject of “cumulus parameterization,” and many contentious issues remain unresolved, while many additional issues have not even been confronted yet. Current ideas are that cumulus convection exerts its effects on the large-scale stratification by transporting mass vertically (the so-called “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, including moisture convergence from neighboring columns.

The effects of convection on the large-scale state can be analyzed following Arakawa and Schubert (1974; hereafter AS; see Fig. 6.12). Let an overbar denote a suitable average.

Figure 6.12: Prof. Akio Arakawa, cruising along in mid-lecture.

The averaged budget equations for mass, dry static energy, water vapor mixing ratio, and liquid water mixing ratio are:

\[ 0 = -\nabla \cdot (\overline{\rho \overline{V}}) - \frac{\partial}{\partial z}(\overline{\rho \overline{w}}), \quad (6.33) \]

\[ \frac{\partial}{\partial t} \overline{s} = -\rho \overline{V} \cdot \nabla \overline{s} - \rho \overline{w} \frac{\partial}{\partial z} \overline{s} + \overline{Q_R} + \overline{LC} - \frac{\partial}{\partial z} \overline{F_s}, \quad (6.34) \]

\[ \frac{\partial}{\partial t} \overline{q_v} = -\rho \overline{V} \cdot \nabla \overline{q_v} - \rho \overline{w} \frac{\partial}{\partial z} \overline{q_v} - \overline{C} - \frac{\partial}{\partial z} \overline{F_{q_v}}, \quad (6.35) \]
\[ \rho \frac{\partial \bar{\rho}}{\partial t} = -\rho \nabla \cdot \bar{V} - \rho w \frac{\partial \bar{\rho}}{\partial z} + \bar{C} - \frac{\partial F_{\chi}}{\partial z} - \bar{\chi}. \]  

(6.36)

Here \( \rho \) is the density of the air, which is presumed to be quasi-constant at each height\(^2\); 
\( s \equiv c_p T + g z \) is the dry static energy; \( T \) is temperature; \( c_p \) is the specific heat at constant pressure; \( q \) is the water vapor mixing ratio; \( w \) is the vertical velocity; \( \nabla \) is the horizontal velocity; \( L \) is the latent heat of evaporation; \( C \) is the net rate of condensation, \( Q_R \) is the radiative heating rate, and \( \chi \) is the rate at which liquid water is being converted into precipitation, which then falls out and acts as a sink of \( l \). The vertical “eddy fluxes,” \( F_s \equiv \rho ws - \rho \bar{ws} \) and \( F_{q_v} \equiv \rho wq_v - \rho \bar{q}_v \), represent quite a variety of physical processes, but here we assume for simplicity that above the boundary layer these fluxes are due to the vertical currents associated with cumulus convection. Compare (6.34) with (6.9).

AS used a very simple cumulus cloud model to formulate the eddy fluxes that appear in (6.34) and (6.35) 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, using a single cloud type for simplicity.

As a first step, we divide the domain into an arbitrary number \( N \) of subdomains, each having a characteristic fractional area \( \sigma_i \), a characteristic vertical velocity \( w_i \), 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 fractional areas must sum to unity:

\[ \sum_{i=1}^{N} \sigma_i = 1. \]  

(6.37)

The area-averaged vertical velocity and moist static energy satisfy:

\[ \sum_{i=1}^{N} \sigma_i w_i = \bar{w}, \]  

(6.38)

\(^2\) This is the “anelastic” approximation. A handout on this topic is available from the instructor. Because the density is assumed to be quasi-constant at each height, we do not bother to put an overbar on it.
An overview of the effects of radiation and convection

\[ \sum_{i=1}^{N} \sigma_i h_i = \bar{h}, \]  \hspace{1cm} (6.39)

and other area-averages are constructed in a similar way. It then follows algebraically that

\[ F_h \equiv \rho \bar{w} \bar{h} - \rho \bar{w} \bar{h} = \sum_{i=1}^{N} \rho \sigma_i (w_i - \bar{w}) (h_i - \bar{h}). \]  \hspace{1cm} (6.40)

You should work through the derivation to satisfy yourself that this is true.

We now assume for simplicity that mesoscale organization and convective-scale downdrafts are not significant, 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 tilde to denote an environmental value, and a subscript \( c \) to denote the collective properties of the cloudy updrafts. Then (6.37) can be rewritten as

\[ \sigma_c + \tilde{\sigma} = 1. \]  \hspace{1cm} (6.41)

Here

\[ \sigma_c \equiv \sum_{\text{all clouds}} \sigma_i \]  \hspace{1cm} (6.42)

is the total fractional area covered by all of the convective updrafts, and \( \tilde{\sigma} \) is the fractional area of the environment. Similarly, (6.38) and (6.39) become

\[ \sigma_c w_c + \tilde{\sigma} \bar{w} = \bar{w}, \]  \hspace{1cm} (6.43)

\[ \sigma_c h_c + \tilde{\sigma} \bar{h} = \bar{h}, \]  \hspace{1cm} (6.44)

where we define

\[ w_c \equiv \frac{\sum_{\text{all clouds}} \sigma_i w_i}{\sigma_c}, \]  \hspace{1cm} (6.45)

\[ h_c \equiv \frac{\sum_{\text{all clouds}} \sigma_i h_i}{\sigma_c}. \]  \hspace{1cm} (6.46)
It is observed that

$$\sigma_c \ll 1 \text{, and } \bar{\sigma} \equiv 1.$$  \hspace{1cm} (6.47)

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 Bjerknes (1938). Suppose that at a certain
time the temperature is horizontally uniform, with lapse rate

$$\Gamma \equiv -\frac{\partial T}{\partial z}.$$  \hspace{1cm} (6.48)

Consider temperature changes due to adiabatic vertical motion only. In a cloudy region, the
temperature satisfies

$$\frac{\partial T_c}{\partial t} = w_c (\Gamma - \Gamma_m).$$  \hspace{1cm} (6.49)

In a neighboring clear region,

$$\frac{\partial \bar{T}}{\partial t} = \bar{w} (\Gamma - \Gamma_d).$$  \hspace{1cm} (6.50)

The mean vertical motion satisfies (6.43), from which it follows that

$$w_c = \bar{w} + (1 - \sigma_c) (w_c - \bar{w}),$$  \hspace{1cm} (6.51)

$$\bar{w} = \bar{w} - \sigma_c (w_c - \bar{w}).$$  \hspace{1cm} (6.52)

Substituting, we find that

$$\frac{\partial}{\partial t} (T_c - \bar{T}) = w_c (\Gamma - \Gamma_m) - \bar{w} (\Gamma - \Gamma_d)$$

$$= \bar{w} (\Gamma_d - \Gamma_m) + (w_c - \bar{w}) [(1 - \sigma_c) (\Gamma - \Gamma_m) + \sigma_c (\Gamma - \Gamma_d)].$$  \hspace{1cm} (6.53)

If the sounding is conditionally unstable, then

$$\Gamma - \Gamma_m > 0 \text{ and } \Gamma - \Gamma_d < 0.$$  \hspace{1cm} (6.54)

Then, if the cloudy air is rising and the environmental air is sinking, both $T_c$ and $\bar{T}$ will tend
to increase with time. We can therefore conclude from (6.53) that $T_c - \bar{T}$ will increase most
rapidly if $\sigma_c \to 0$. 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 the rate of temperature increase in the downdraft is proportional to the downdraft speed. Therefore, the 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 $w_c - w$, by making the updraft narrow, and the downdraft broad.

The smallness of fractional area covered by the convective towers is crucially important for the general circulation because it means that even in the regions where deep convection is most active there is a lot of clear sky. The convective clouds “tunnel through” deep layers of (mostly) unsaturated air. Cloud processes would be relatively simple if they involved only uniform cloudiness over large regions. The importance of narrow saturated updrafts in clear environments makes the interaction of moist convection with the general circulation a much more subtle (and interesting) problem than it would otherwise be.

In view of (6.47), we can write (6.44) as

$$\tilde{h} \equiv \bar{h},$$

where

$$\tilde{h} \equiv \bar{h},$$

It is not true, however, that $\tilde{w} \equiv \bar{w}$, because the cumulus updrafts are typically several orders of magnitude stronger than the large-scale vertical motions, i.e.,

$$w_i \gg \bar{w}.$$  \hspace{1cm} (6.56)

Similarly, it is not true in general that $\tilde{l} \equiv \bar{l}$, because there may be no liquid water at all in the environment of the convective clouds.

Using (6.56), we can approximate (6.40) by

$$F_h \equiv \sum_{i=1}^{N} M_i (h_i - \tilde{h})$$  \hspace{1cm} (6.57)

where

$$M_i \equiv \rho \sigma_i \bar{w}_i$$  \hspace{1cm} (6.58)

is the “convective mass flux” associated with cloud type $i$. The convective mass flux is a key concept. It represents the rate at which mass is pumped through the convective circulations -- through the updraft, through the compensating sinking motion outside the updraft, and through the horizontal branches of the convective circulation that connect the updraft and the sinking motion.
From this point we simplify the discussion by considering only one type of convective cloud, whose properties are denoted by subscript $c$. We write

$$F_s \equiv M_c (s_c - \tilde{s}), \quad (6.59)$$

$$F_{q_v} \equiv M_c [(q_{v,c} - \overline{q_v})], \quad (6.60)$$

$$F_l \equiv M_c (l_c - \tilde{l}). \quad (6.61)$$

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

$$M_c \equiv \rho \sigma_c w_c. \quad (6.62)$$

Note that we allow the possibility of liquid water in the environment of the cumulus clouds. We also use

$$\tilde{s} = (1 - \sigma_c) s + \sigma_c s_c, \quad (6.63)$$

$$\tilde{q}_v = (1 - \sigma_c) q_v + \sigma_c (q_{v,c}), \quad (6.64)$$

$$\tilde{l} = (1 - \sigma_c) l + \sigma_c l_c, \quad (6.65)$$

$$\tilde{C} = (1 - \sigma_c) C + \sigma_c C_c, \quad (6.66)$$

$$\tilde{\chi} = (1 - \sigma_c) \chi + \sigma_c \chi_c. \quad (6.67)$$

We can now rewrite (6.34) - (6.36) as

$$\rho \frac{\partial \tilde{s}}{\partial t} = - \rho \vec{V} \cdot \nabla \tilde{s} - \rho_w \frac{\partial \tilde{s}}{\partial z} + R + L (\tilde{C} + \sigma_c C_c) - \frac{\partial}{\partial z} [M_c (s_c - \tilde{s})], \quad (6.68)$$

$$\rho \frac{\partial \tilde{q}_v}{\partial t} = - \rho \vec{V} \cdot \nabla \tilde{q}_v - \rho_w \frac{\partial \tilde{q}_v}{\partial z} + \tilde{C} + \sigma_c C_c) - \frac{\partial}{\partial z} \{M_c [(q_{v,c} - \overline{q_v})], \quad (6.69)$$

$$\rho \frac{\partial \tilde{l}}{\partial t} = - \rho \vec{V} \cdot \nabla \tilde{l} - \rho_w \frac{\partial \tilde{l}}{\partial z} + (\tilde{C} + \sigma_c C_c) - \frac{\partial}{\partial z} [M_c (l_c - \tilde{l})] - [(1 - \sigma_c) \tilde{\chi} + \sigma_c \chi_c]. \quad (6.70)$$
Eq. (6.47) has been used here. Note that the convective condensation rate, $C_c$, appears in all three of these equations, as would be expected.

To go further we need to describe what is going on inside the convective updrafts; for example, we need to know the soundings in side the updrafts. We assume that all cumulus clouds originate from the top of the PBL, carrying the mixed-layer properties upward. The mass flux changes with height according to

$$\frac{\partial}{\partial z} M_c(z) = E(z) - D(z). \quad (6.71)$$

Here $E$ is the entrainment rate, and $D$ is the detrainment rate. The in-cloud profile of moist static energy, $h_c(z)$, is governed by

$$\frac{\partial}{\partial z} \left[ M_c(z) h_c(\lambda, z) \right] = E(z) \tilde{h}(z) - D(z) h_c(z). \quad (6.72)$$

There are no source or sink terms in this equation because the moist static energy is unaffected by phase changes and/or precipitation processes.\(^3\) By combining (6.71) and (6.72), we can obtain

$$\frac{\partial}{\partial z} h_c(z) = \frac{E(z)}{M_c} \left[ \tilde{h}(z) - h_c(z) \right]. \quad (6.73)$$

This shows that $h_c(z)$ is affected by entrainment, which dilutes the cloud with environmental air, but not by detrainment, which has been assumed to expel from the cloud air that has the average moist static energy of the cloud at each level. Similarly, we can write

$$\frac{\partial}{\partial z} (M_c s_c) = \tilde{E} s - D s_c + \sigma_c L C_c, \quad (6.74)$$

$$\frac{\partial}{\partial z} \left[ M_c (q_v)_c \right] = \tilde{E} q_v - D (q_v)_c - \sigma_c C_c, \quad (6.75)$$

$$\frac{\partial}{\partial z} (M_c l_c) = \tilde{E} l - D l_c + \sigma_c C_c - \chi_c. \quad (6.76)$$

A simple microphysical model is used to determine $\chi_c$, 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

\(^3\) Radiation effects on the in-cloud moist static energy are neglected here.
lower troposphere by evaporating as it falls, is actually an important and currently active topic in the arena of cumulus parameterization, but it will not be discussed here.

By using these expressions for the convective fluxes, the large-scale budget equations can be rewritten in a very interesting form, as follows. First, consider the dry static energy. We write

\[
-\frac{\partial}{\partial z} [M_c(s_c - \tilde{s})] = -\frac{\partial}{\partial z} (M_c s_c) + M_c \frac{\partial}{\partial z} s - s \frac{\partial M_c}{\partial z} .
\]  

(6.77)

Now substitute from (6.74) and (6.77) into (6.80) to obtain

\[
-\frac{\partial}{\partial z} [M_c(s_c - \tilde{s})] = -(E\tilde{s} - Ds_c + L\sigma_c C_c) + M_c \frac{\partial}{\partial z} s + s(E - D)
\]

\[
= M_c \frac{\partial}{\partial z} s + D(s_c - \tilde{s}) - L\sigma_c C_c .
\]

(6.78)

This allows us to rewrite (6.68) as

\[
\rho \frac{\partial}{\partial t} \tilde{s} = -\rho \nabla \cdot \tilde{V} s - \rho\frac{\partial}{\partial z} \tilde{s} + \frac{\tilde{Q}_R}{c_p} + L\tilde{C} + M_c \frac{\partial}{\partial z} s + D(s_c - \tilde{s}) .
\]

(6.79)

The last two terms on the right-hand side of (6.79) 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 called “compensating subsidence.” The effect can be seen more explicitly by combining the two “vertical advection” terms of (6.79), and using (6.81), to obtain

\[
\rho \frac{\partial}{\partial t} \tilde{s} = -\rho \nabla \cdot \tilde{V} s - M \frac{\partial}{\partial z} \tilde{s} + \frac{\tilde{Q}_R}{c_p} + L\tilde{C} + M_c \frac{\partial}{\partial z} s + D(s_c - \tilde{s}) .
\]

(6.80)

\[
\tilde{M} \equiv \rho\bar{w} - M_c ,
\]

(6.81)

The reason that \(M\) appears in (6.81) is that (6.55) applies, i.e. \(s = \tilde{s}\). The last term on the right-hand side of (6.79) represents the effects of detrainment. Notice that the cumulus condensation rate does not appear in (6.79). An interpretation is that condensation inside the updraft cannot directly warm the environment, and since almost the entire area is the environment, this means that condensation does not, to any significant degree, directly affect the area-averaged value of the dry static energy. Instead, the effects of condensation are felt indirectly, through the compensating subsidence term, which we have already interpreted. The physical role of condensation, then, is to make possible 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 indirect condensation heating rate due to
compensating subsidence is in general different from the vertical profile of the convective condensation rate itself.

In a similar way, we find by analogy with (6.79) that the water vapor and liquid water budget equations can be rewritten as

\[ \rho \frac{\partial}{\partial t} \bar{q}_v = -\rho \bar{V} \cdot \nabla \bar{q}_v - \rho w \frac{\partial}{\partial z} \bar{q}_v - \bar{C} + M_c \frac{\partial}{\partial z} \bar{q}_v + D[(q_v)_c - \bar{q}_v], \quad (6.82) \]

\[ \rho \frac{\partial}{\partial t} \bar{l} = -\rho \bar{V} \cdot \nabla \bar{l} - \rho w \frac{\partial}{\partial z} \bar{l} + \bar{C} + M_c \frac{\partial}{\partial z} \bar{l} + D(l_c - \bar{l}) - [(1 - \sigma_e) \bar{\chi} + \sigma_c \chi_c]. \quad (6.83) \]

Eq. (6.82) describes the convective drying in terms of convectively-induced subsidence in the environment, which brings down drier air from aloft. Detrainment of water vapor from the convective clouds can moisten the environment. Detrained liquid (or ice) can persist in the form of stratiform “anvil” and cirrus clouds.

Within the limits of applicability of the assumptions discussed above, (6.79) and (6.82) - (6.82) are equivalent to (6.34) - (6.36).

As discussed earlier [see (6.25)], the buoyancy of the cloudy air at height \( z \), is approximately

\[ B(z) \equiv T_c - T - \frac{\bar{h}(z) - \bar{h}^*(z)}{c_p}, \quad (6.84) \]

where \( \bar{h}^* \) 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 \( \hat{p} \), where

\[ B(\hat{p}) = 0. \quad (6.85) \]

This equation can be used to find \( p \), after the in-cloud sounding has been determined using (6.71) - (6.76).

To determine the entrainment rate, AS assumed that

\[ E = \lambda M_c, \quad (6.86) \]

where \( \lambda \), which is called the fractional entrainment rate and has the units of inverse length, is

---

4: Virtual temperature and ice effects can easily be included in this analysis but have been omitted here for simplicity.
assumed to be a constant for each cloud type. Larger values of $\lambda$ mean stronger entrainment; $\lambda = 0$ means no entrainment; $\lambda < 0$ has no physical meaning and so is not allowed. For a given sounding, clouds with smaller values of $\lambda$ (weaker mixing) will have higher tops. Because $\lambda$ is assumed to be a constant, the solution of (6.71) gives an exponential profile for $\eta(\lambda, z)$, from the cloud-base level to the cloud top level.

For simplicity, AS assumed that detrainment occurs only at cloud top.\footnote{Lord (1978) investigated the effects of detrainment distributed continuously along the sides of the cloud, and found that they are negligible.} 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 (6.71) reduces to

$$\frac{\partial}{\partial z} M_c(z) = E(z).$$  \hspace{1cm} (6.87)

By combining (6.71) and (6.71), and using the assumption that $\lambda$ is a constant for each cloud type, we see that

$$M_c(z, \lambda) = M_B(\lambda) \exp(\lambda z),$$  \hspace{1cm} (6.88)

where $M_B(\lambda)$ is the “cloud-base mass flux,” which can vary among cloud types but is not a function of height. We define a normalized mass flux, denoted by $\eta(\lambda, z)$; the normalization is in terms of the cloud-base mass flux:

$$M_c(z, \lambda) \equiv M_B(\lambda) \eta(\lambda, z).$$  \hspace{1cm} (6.89)

Note that by virtue of its definition, $\eta(\lambda, z_B) = 1$; here $z_B$ is the cloud-base height.

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:

$$A(\lambda) = \int_{z_B}^{z_D(\lambda)} \frac{g}{c_p T(z)} \eta(z, \lambda)[s_v(z, \lambda) - \bar{s}_v(z)] dz.$$

\hspace{1cm} (6.90)
Here $z_D(\lambda)$ is the height of the detrainment level for cloud type $\lambda$; and $s_v$ denotes the virtual static energy. From (6.90) we see that the function $A(\lambda)$ is a property of the large-scale environment. A positive value of $A(\lambda)$ means that a cloud with fractional entrainment rate $\lambda$ can convert the potential energy of the mean state into convective kinetic energy. For $\lambda = 0$, $A(\lambda)$ 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 $T(z)$ and $q(z)$, from which $A(\lambda)$ can be determined; therefore, these models indirectly predict $A(\lambda)$. By taking the time derivative of (6.90), and using the conservation equations for thermodynamic energy and moisture, AS showed that

$$\frac{d}{dt} A(\lambda) = J(\lambda) M_B(\lambda) + F(\lambda). \quad (6.91)$$

The $J M_B$ term of (6.91) represents all of the terms involving convective processes, each of which turns out to be proportional to $M_B$. Note, however, that Eq. (6.91) is written in simplified form. The $J M_B$ 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 $J M_B$ term of (6.91) 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 (6.91) holds for each cumulus subensemble.

The $F(\lambda)$ term of (6.91) 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 $F$ as “large-scale;” a better term would be “non-convective.”

AS assumed quasi-equilibrium (QE) of the cloud work function, i.e.
\[
\frac{d}{dt} A(\lambda) = JM_B(\lambda) + F(\lambda) \equiv 0 \quad \text{when } F(\lambda) > 0 .
\] (6.92)

Eq. (6.92) 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 (6.92) approximately balance each other. In a steady-state situation, this balance is of course trivially satisfied, by definition. The physical content of (6.92) is, therefore, the assertion that near-balance is maintained even when \( F(\lambda) \) is varying with time, provided that the variations of \( F(\lambda) \) 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.

The QE approximation is expected to hold if \( \tau_{LS} \), the time scale for changes in \( F(\lambda) \), is much longer than the “adjustment time,” \( \tau_{adj} \), 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 \( \tau_{adj} \) by describing 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 \( \tau_{adj} \) and estimated to be on the order of a couple of 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 (6.92) is justified, as an approximation, for the simulation of “weather” whose time scale is on the order of one day or longer, i.e. at least one order of magnitude longer than \( \tau_{adj} \).

By using (6.92), together with

\[
|JM_B| \sim \frac{A}{\tau_{adj}} ,
\] (6.93)

AS found that

\[
A \sim \tau_{adj} F \ll \tau_{LS} F ,
\] (6.94)

where \( \tau_{LS} \) is the time-scale on which the forcing itself is varying. This means that the cloud work function is “small” compared to \( \tau_{LS} F \), which is the value that the cloud work function would take if the forcing acted without opposition over a time scale \( \tau_{LS} \). Although we should
expect to see day-to-day variations of \( A \), we should not expect to see values as large as \( \tau_{LS} F \). This means that \( A \) is trapped in the range of values between zero (since by definition \( A \) cannot be negative) and \( \tau_{adj} F \). In this sense, \( A \) is “close to zero” (see also Xu and Emanuel, 1989)\(^6\).

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

\[
\frac{dA}{dt} \equiv 0 ,
\]

which is a short-hand form of Eq. (6.92); and that “the CAPE is small” in convectively active regimes, i.e.

\[
A \equiv 0 ,
\]

which is a short-hand form of Eq. (6.94). A sounding for which \( A \equiv 0 \) tends to follow a saturated moist adiabat, throughout the depth of the convective layer. This provides a rationalization for Manabe et al.’s assumption that the lapse rate cannot exceed 6.5 K km\(^{-1}\).

Unfortunately, because (6.95) and (6.96) are short-hand forms, they are subject to misinterpretation. For example, data like those shown in Fig. 6.13 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. (6.92) does not imply that \( A \) is invariant from day to day, and the observed day-to-day changes in the CAPE, such as those shown in Fig. 6.13, are not necessarily in conflict with QE. What QE does imply is that the changes in the CAPE that we actually see, from day to day, are much smaller than they would be if the negative convective term of (6.92) could somehow be suppressed, so that the positive (under disturbed conditions) forcing term could have its way with the sounding.

A practical application of (6.92) is to solve it for the convective mass flux as a function of cloud type, \( \lambda \). After discretization this leads to a system of linear equations (Lord

---

\(^6\) Here an earthy analogy, suggested by the pastoral life in Fort Collins, may help. Think of \( A \) as a measure of the length of the grass in a field, the convective clouds as a herd of sheep, and the forcing as an irrigation system controlled by a large-scale dynamicist. The sheep eat the grass as quickly as it grows, so that the grass is always short, no matter how generous the supply of water. The sheep are white and fluffy, like convective clouds, and one can even find analogs, in this parable, to the precipitation process.
et al. 1982). Although the system is linear, the mass flux distribution function, \( M_\beta(\lambda) \), is required to be non-negative for all \( \lambda \). 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 (1987), Grell et al. (1991), Wang and Randall (1994), and Cripe (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).

For further discussions of cumulus parameterization see the collections of essays edited by Emanuel and Raymond (1993) and Smith (1998).

6.5 Summary

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 models, which completely ignore the effects of large-scale dynamics. Obviously this does not mean that dynamics is unimportant, but it does suggest that dynamics 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 general circulation. It is impossible to understand the circulation without understanding the heating, and vice versa. Further discussion is given in later chapters.

Problems

1. Derive (6.40).
An overview of the effects of radiation and convection

2. Refer to Table 6.1.

   a) Consider the number in the bottom right corner of the table, i.e., \(1.30 \times 10^{15} \, \text{J s}^{-1}\). Assuming that there is no net radiative source or sink of moist static energy at any level in the Equatorial Trough Zone, what does this number represent? Make a simple sketch to explain your answer.

   b) Using the numbers given in the table, estimate the total upward transport of moist static energy across the 500 mb surface in the Equatorial Trough Zone, in J s\(^{-1}\).

   c) Using the numbers given in the table, estimate the upward transport of moist static energy across the 500 mb surface due to the large-scale rising motion in the Equatorial Trough Zone, in J s\(^{-1}\).

   d) Assume that the area covered by the Equatorial Trough Zone is \(4 \times 10^{13} \, \text{m}^2\). Using your answers from above, work out the numerical value of \(F_h\), the upward flux of moist static energy due to convection, in W m\(^{-2}\).

   e) Estimate a rough numerical value (in kg m\(^{-2}\) s\(^{-1}\)) of the convective mass flux, \(M_c\), at the 500 mb level in the ITCZ. You will need to use

   \[
   F_h = M_c(h_c - \bar{h}),
   \]  

   where \(h_c\) is the in-cloud moist static energy, and \(\bar{h}\) is the large-scale mean moist static energy. State your assumptions.

3. Suppose that moist static energy is simply conserved, i.e.

   \[
   \frac{\partial h}{\partial t} = -\nabla \cdot (Vh) - \frac{\partial}{\partial z}(wh).
   \]  

   The density has been omitted here and throughout the rest of this problem for simplicity. The corresponding continuity equation is

   \[
   0 = -\nabla \cdot V - \frac{\partial w}{\partial z}.
   \]  

   a) By using Reynolds averaging, show that

   \[
   \frac{\partial \bar{h}}{\partial t} = -\nabla \cdot (\bar{V}h + \bar{V}'h') - \frac{\partial}{\partial z}(\bar{w}h + \bar{w}'h').
   \]
For large-scale averages this can be approximated by

\[
\frac{\partial \tilde{h}}{\partial t} \equiv - \nabla \cdot (\overline{V \hat{h}}) - \frac{\partial}{\partial z} (\overline{w h + w' h'}). \tag{6.101}
\]

b) Show that the moist static energy variance, \( h'^2 \), satisfies

\[
\frac{\partial}{\partial t} \overline{h'^2} = - \nabla \cdot \overline{\left( V h'^2 + V' h'h' \right)} - \frac{\partial}{\partial z} \left( \overline{w h'^2 + w' h'h'} \right) - 2 \overline{V' h'} \cdot \nabla \tilde{h} - 2 w' h' \frac{\partial \tilde{h}}{\partial z}. \tag{6.102}
\]

For large-scale averages, the time-rate-of-change term of (6.102) is negligible, as are the terms representing horizontal and vertical advection of \( h'^2 \) by the mean flow, as are the other terms involving horizontal derivatives, so that (6.102) can be drastically simplified to

\[
0 \equiv - \frac{\partial}{\partial z} (w' h' h') - 2 h' \frac{\partial \tilde{h}}{\partial z}. \tag{6.103}
\]

c) Now suppose that the vertical velocity fluctuations represented by \( w' \) are associated with a single family of cumulus updrafts covering fractional area \( \sigma \). Show that

\[
\overline{w' h'} = \sigma (1 - \sigma) (w_u - w_d) (h_u - h_d), \tag{6.104}
\]

\[
\overline{w' h'h'} = \sigma (1 - \sigma) (1 - 2 \sigma) (w_u - w_d) (h_u - h_d)^2, \tag{6.105}
\]

where \( w_u \) and \( w_d \) are the updraft and downdraft velocities, respectively, and \( h_u \) and \( h_d \) are the corresponding values of the moist static energy.

d) Define a “convective mass flux” by

\[
M \equiv \sigma w_u. \tag{6.106}
\]

You may assume

\[
\sigma \ll 1, \tag{6.107}
\]

and correspondingly that
Using (6.107)-(6.106), show that if the convective mass flux is independent of height, then (6.103) can be approximated by

$$\frac{\partial}{\partial z}(w'h') = M\frac{\partial}{\partial z} \bar{h}. \quad (6.109)$$

Discuss (6.109) as it relates to the ideas on cumulus parameterization presented in the notes.
CHAPTER 7  The Energy Cycle

7.1 Available potential energy

We have shown that, under dry adiabatic and frictionless processes,

$$\frac{d}{dt} \int (c_v T + \phi + K) \rho \, dV = 0 .$$

(7.1)

Here the integral is over the mass of the whole atmosphere. For each vertical column, the vertically integrated sum of the mass-weighted internal and potential energies is equal to the vertical integral of the mass-weighted enthalpy, i.e.,

$$\int_0^{p_s} (c_v T + \phi) \, dp = \int_0^{p_s} c_p T \, dp ,$$

(7.2)

provided that we have hydrostatic balance. This can be demonstrated as follows: Using hydrostatics in the form

$$\frac{\partial p}{\partial z} = -\rho g ,$$

(7.3)

we find that the vertically integrated potential energy, $P$, satisfies

$$P \equiv \int_0^\infty g z \rho \, dz = -\int_0^\infty \left( \frac{\partial p}{\partial z} \right) dz = -\int_0^\infty \left[ \frac{\partial}{\partial z} (pz) - p \right] dz$$

$$= -[ (pz) \bigg|_{z=0}^{z=\infty} - \int_0^\infty \rho RT \, dz ] = \int_0^\infty \rho RT \, dz .$$

(7.4)

Eq. (7.4) says that the total potential energy of the column is proportional to the average temperature of the column. The explanation is that warmer air occupies a larger volume, so that a warmer column is “taller.” It follows from (7.4) that
Here $I$ is the vertically integrated internal energy. Note that $c_v T + \phi = c_p T$ \textit{is not true}; this would imply that $\phi = RT$, which is obviously nonsense.

Let $H$ and $K$ be the total (mass integrated) enthalpy and kinetic energy of the entire atmosphere. It follows from (7.1) and (7.2) that $H + K$ is invariant, under dry-adiabatic frictionless processes, i.e.,

$$\frac{d}{dt} (H + K) = 0.$$  
(7.6)

Imagine that we have the power to spatially rearrange the mass of the atmosphere at will, adiabatically and reversibly. 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.”

You should prove for yourself 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.

In passing from the given state to the A-state we have

$$H_{gs} \rightarrow H_{min},$$

$$K_{gs} \rightarrow K_{gs} + (H_{gs} - H_{min}) = K_{max},$$

(7.7)

where $H_{min}$ is the value of $H$ in the A-state, and subscript $gs$ denotes the given state. The non-negative quantity

$$A \equiv H_{gs} - H_{min} \geq 0$$

(7.8)

is called the “available potential energy,” or APE. The APE was first defined by Lorenz (1955; see Fig. 7.1). Eq. (7.8) gives the fundamental definition of the APE.
Notice that the APE is a property of the entire atmosphere; it cannot be rigorously defined for a portion of the atmosphere, although the literature does contain studies in which the APE is computed, without rigorous justification, for a portion of the atmosphere, e.g., the Northern Hemisphere.

The A-state is invariant under adiabatic processes, because it depends only on the probability distribution of $\theta$ over the mass, rather than on any particular spatial arrangement of the air. Therefore

$$
\frac{dA}{dt} = \frac{d}{dt}(H_{gs} - H_{rs}) = \frac{dH_{gs}}{dt}.
$$

(7.9)

Then (7.6) implies that

$$
\frac{d}{dt}(A + K) = 0.
$$

(7.10)

The sum of the available potential and kinetic energies is invariant under adiabatic frictionless processes. This means that such processes only convert between $A$ and $K$.

In order to compute $A$, we have to find the reference state and its (minimum) enthalpy. We can deduce the properties of the A-state as follows: There can be no 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 $\theta$ surfaces. This means that both the variance of $\theta$ on

---

**Figure 7.1:** Prof. Edward N. Lorenz, who proposed the concept of available potential energy, and has published a great deal of additional very fundamental research in the atmospheric sciences and the relatively new science of nonlinear dynamical systems.

Notice that the APE is a property of the *entire* atmosphere; it cannot be rigorously defined for a portion of the atmosphere, although the literature does contain studies in which the APE is computed, without rigorous justification, for a portion of the atmosphere, e.g., the Northern Hemisphere.

The A-state is invariant under adiabatic processes, because it depends only on the probability distribution of $\theta$ over the mass, rather than on any particular spatial arrangement of the air. Therefore

$$
\frac{dA}{dt} = \frac{d}{dt}(H_{gs} - H_{rs}) = \frac{dH_{gs}}{dt}.
$$

(7.9)

Then (7.6) implies that

$$
\frac{d}{dt}(A + K) = 0.
$$

(7.10)

The sum of the available potential and kinetic energies is invariant under adiabatic frictionless processes. This means that such processes only convert between $A$ and $K$.

In order to compute $A$, we have to find the reference state and its (minimum) enthalpy. We can deduce the properties of the A-state as follows: There can be no 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 $\theta$ surfaces. This means that both the variance of $\theta$ on
isobaric surfaces and the variance of $p$ on isentropic surfaces are measures of the APE. There can be no static instability ($\partial \theta / \partial z < 0$) in the A-state for a similar reason. From these considerations we conclude that, in the A-state, $\theta$ and $T$ are uniform on each pressure-surface, or equivalently that $p$ is uniform on each $\theta$-surface, and also that $\theta$ 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 $\bar{\theta}$, 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.

A complication arises: What about $\theta$-surfaces that intersect the ground in the given state? These are treated as though they continue along the ground, as shown in Fig. 7.2.

![Figure 7.2: Sketch illustrating the concept of “massless layers.”](image)

Because the “layers” between the isentropic surfaces that are following the ground contain no mass, they have no effect on the physics. They are called “massless layers.” Where a $\theta$-surface intersects the Earth’s surface, the pressure is $p = p_S$.

The concept of massless layers allows us to write

$$\int_0^{p_S} (\cdot) dp = \int_0^{\infty} (\cdot) \frac{\partial p}{\partial \theta} d\theta, \quad (7.11)$$

where $(\cdot)$ can be anything. Note that the lower limit of integration on the right-hand side of (7.11) is zero. Eq. (7.11) will be used below and it is important for you to understand why it is true.

A useful expression for the APE can be derived as follows. The total enthalpy is given by
Integration by parts gives

\[
H = \frac{c_p a^2}{(1 + \kappa) g p_0} \int \int \left[ \int p^1 + \kappa \cos \phi \, dp \, d\lambda \, d\phi \right].
\]  
(7.13)

Note that vertical integration is now with respect to \( \theta \) rather than \( p \), and that the lower limit of integration is \( \theta = 0 \). Let \( \bar{p}^{\theta} \) be the average pressure on an isentropic surface (taking into account intersections with the ground). Recall that, provided that there are no regions of dry static instability, \( \bar{p}^{\theta} \) is the same in the A-state as in the given state. Then use of (7.13) in (7.8) gives

\[
A = \frac{c_p a^2}{(1 + \kappa) g p_0} \int \int \left[ \int (p^1 + \kappa) \left( \frac{\bar{p}^{\theta} - \theta}{p_0} \right) \cos \phi \, d\theta \, d\lambda \, d\phi \right].
\]  
(7.14)

Note that (7.14) is valid only if \( \partial \theta / \partial z \geq 0 \) everywhere in the given state, because we have assumed that \( \bar{p}^{\theta} \) is the same in the A-state as in the given state. So long as this assumption is satisfied, (7.14) is exact. The most general expression for the available potential energy is the definition \( A \equiv H_{gs} - H_{min} \).

Let \( p' \) be the departure of \( p \) from its average on an isentropic surface, so that \( p = \bar{p}^{\theta} + p' \), where \( \bar{p}^{\theta} = 0 \). The binomial theorem tells us that

\[
p^1 + \kappa = (p^1 + \kappa)^{1 + \kappa} \left( \frac{p'}{\bar{p}^{\theta}} \right)^{1 + \kappa} = \left( \frac{p'}{\bar{p}^{\theta}} \right)^{1 + \kappa} \left[ 1 + (1 + \kappa) \frac{p'}{\bar{p}^{\theta}} + \frac{\kappa(1 + \kappa)}{2!} \left( \frac{p'}{\bar{p}^{\theta}} \right)^2 + \ldots \right].
\]  
(7.15)
and he showed that this is actually a rather good approximation. Substitution of (7.16) into (7.14) gives

$$A \equiv \frac{Ra^2}{2g} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{0}^{2\pi} \frac{\Theta^2}{2} \frac{\Theta^2}{\Theta} \cos \varphi d\varphi d\lambda d\varphi \ . \ (7.17)$$

Because he wanted to express his results in terms of $\Theta$ perturbations on isobaric surfaces, rather than pressure perturbations on isentropic surfaces, Lorenz also used

$$p' \equiv \theta \frac{\partial p}{\partial\Theta}$$

$$= \theta \frac{\partial p}{\partial\Theta} , \quad (7.18)$$

where, as before, $p'$ represents the departure of $p$ from its global average on an isentropic surface, and $\Theta'$ represents the departure of $\Theta$ from $\bar{\Theta}$, its global average on a $p$-surface. Substitution of (7.18) into (7.17) gives

$$A \equiv \frac{Ra^2}{2g} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{0}^{2\pi} \left(\frac{\partial \Theta}{\partial p}\right)^2 \cos \varphi d\varphi d\lambda d\varphi \ . \ (7.19)$$

Note that in (7.19) the independent variable for vertical integration has been changed from $\Theta$ to $p$. Eq. (7.19) involves a weighted average of the square of the departure of $\Theta$ from its mean on the pressure surface. The average of the square of the departure from the mean is called the “variance about the mean,” or just the variance. The variance is a measure of how variable a quantity is; if the quantity is constant, and so everywhere equal to its mean, then its variance must be zero. If the quantity is not constant, its variance is positive. Because we are interested in variability, variances are quite important in the study of the general circulation.

Finally, Lorenz used the hydrostatic equation in the form

$$\frac{\partial \Theta}{\partial p} = \frac{\kappa \Theta}{p} \left(\frac{\Gamma_d - \Gamma}{\Gamma_d}\right) , \quad (7.20)$$
where \( \Gamma = -\frac{\partial T}{\partial z} \) is the lapse rate of temperature and \( \Gamma_d = \frac{g}{c_p} \) is the dry adiabatic lapse rate, as well as

\[
\frac{\theta'}{\theta} = \frac{T'}{T},
\]

(7.21)
to rewrite (7.19) as

\[
A = \frac{a^2}{2} \int_{-\pi/2}^{\pi/2} \int_{0}^{2\pi} \int_{0}^{\infty} \frac{T}{\Gamma_d - \Gamma} \left( \frac{T'}{T} \right)^2 \cos \phi dp d\lambda d\phi.
\]

(7.22)

This result shows that the available potential energy is closely related to the variance of temperature on isobaric surfaces. It also increases as the lapse rate of temperature increases, i.e., as the atmosphere becomes less stable in the dry static sense.

Observations show that the APE is only about half a percent of \( P + I \). The APE is comparable in magnitude to the total kinetic energy. Both are on the order of \( 10^6 - 10^7 \) J m\(^{-2}\).

As shown by the heavy arrows in Fig. 7.3, 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; the APE of the A-state itself is obviously zero.

### 7.2 The gross static stability

The A-state used in the definition of \( A \) defines a correspondence or mapping between \( \theta \) and \( p \). For each \( p \) there is one possible value of \( \theta \) (the converse is not necessarily true). We can say that in the A-state, \( p \) and \( \theta \) are perfectly correlated, (or, more correctly, perfectly anti-correlated).

Consider the opposite limit, in which \( \theta \) and \( p \) are completely uncorrelated. In this “S-state,” all possible values of \( \theta \) occur, with equal probability, for any given \( p \). This will be the case if the \( \theta \) surfaces are vertical, so that \( \partial \theta / \partial p = 0 \), and if the surface pressure is globally uniform. See Fig. 7.4.

To see why the surface pressure must be globally uniform, suppose that there are variations in the surface pressure in the S-state (independent of height at each location). If the surface pressure varies geographically, we can say that it varies with \( \theta \). As an example, suppose that, in the S-state, for \( \theta = \theta_1 \) the surface pressure is 900 mb. Then if I tell you that the pressure where I am is 1000 mb, you know that my \( \theta \) cannot be \( \theta_1 \), i.e., you have a clue that will help you (at least a little bit) if you try to guess my \( \theta \). This shows that variations of the surface pressure in the S-state would violate the rule that \( \theta \) and \( p \) are completely uncorrelated in the S-state. Therefore, the surface pressure must be uniform in the S-state.
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 specification of the S-state, in which the surface pressure is allowed to vary in a simple way with the surface height, but the potential temperature is spatially distributed so that it is uncorrelated with the surface height (and, therefore, uncorrelated with the surface pressure).

The total enthalpy of the S-state is given by

\[ H = H_{gs} - H_{min} = APE \]

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 specification of the S-state, in which the surface pressure is allowed to vary in a simple way with the surface height, but the potential temperature is spatially distributed so that it is uncorrelated with the surface height (and, therefore, uncorrelated with the surface pressure).

The total enthalpy of the S-state is given by

\[ S \equiv H_{S\text{-state}} - H_{gs} \]
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 write
where $\bar{p}_S$ is the globally averaged surface pressure, which is the same in the given state and the S-state, and $\bar{\Theta}$ 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 we assume that the S-state is such that $\Theta$ is uncorrelated with $p_S^{1+\kappa}$, then it follows immediately that

$$\frac{c_p a^2}{g P_0^{\kappa}} \int_{-\pi/2}^{\pi/2} \int_{0}^{2\pi} \left[ \frac{(\bar{p}_S)^{1+\kappa}}{1+\kappa} \right] d\lambda \cos \varphi d\varphi = \frac{c_p 4\pi a^4}{g P_0^{\kappa}} \left[ \frac{(\bar{p}_S)^{1+\kappa} \bar{\Theta}}{1+\kappa} \right].$$

We conclude that the two definitions of the S-state give exactly the same result for $H_{\text{S-state}}$.

Now we show why $S$ is called the “gross static stability.” Eq. (7.24) can also be written as

$$H_{\text{S-state}} = \frac{c_p a^2}{g \bar{P}_0^{\kappa}} \int_{-\pi/2}^{\pi/2} \int_{0}^{2\pi} \left[ \frac{(\bar{p}_S)^{1+\kappa}}{1+\kappa} \right] \cos \varphi d\varphi d\lambda d\varphi,$$

where

$$\bar{p}_S^{\kappa} = \frac{(\bar{p}_S)^{1+\kappa}}{1+\kappa}.$$
This allows us to rewrite (7.27) as

\[
H_{S\text{-state}} = \frac{c_p a^2}{g} \left\{ \int_{-\pi/2}^{\pi/2} \int_{0}^{2\pi} \left[ \frac{\pi}{2} \int_{0}^{p_S} \theta dp \right] \cos \varphi d\lambda d\varphi \right\} \quad \text{given state}
\] (7.30)

Now substitute (7.30) into (7.23), to obtain

\[
S = \frac{c_p a^2}{g} \left\{ \int_{-\pi/2}^{\pi/2} \int_{0}^{2\pi} \left[ \frac{\pi}{2} \int_{0}^{p_S} \theta dp \right] \cos \varphi d\lambda d\varphi \right\} \quad \text{given state}
\] (7.31)

This can be rearranged to

\[
S \equiv \frac{c_p a^2}{g p_0 \kappa} \left\{ \int_{-\pi/2}^{\pi/2} \int_{0}^{2\pi} \left[ \int_{0}^{p_S} (p - p_S)^\kappa \theta dp \right] \cos \varphi d\lambda d\varphi \right\} \quad \text{given state}
\] (7.32)

Integration by parts gives

\[
S \equiv \frac{c_p a^2}{g p_0^\kappa (1 + \kappa)} \left\{ \int_{-\pi/2}^{\pi/2} \int_{0}^{2\pi} \left[ \int_{0}^{p_S} \left( \frac{\partial \theta}{\partial p} \right) dp \right] \cos \varphi d\lambda d\varphi \right\} \quad \text{given state}
\] (7.33)

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 \( \theta \) are invariant under adiabatic processes. This means 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

---

An Introduction to the General Circulation of the Atmosphere
Since \( \frac{d}{dt}(H + K) = 0 \), it follows from (7.34) that

\[
\frac{dS}{dt} = \frac{dH}{dt}.
\]

This shows that when \( K \) is produced by conversion from APE, the gross static stability increases. Isn’t that what you would expect?

7.3 **Examples: The available potential energies of three simple systems**

7.3.1 **The APE associated with static instability**

As an example, consider a simple system containing two parcels of equal mass. In the given state, parcels with potential temperature \( \theta_1 \) and \( \theta_2 \) 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 \( c_p \theta_i \left( \frac{p_i}{p_0} \right)^K \). If the parcels are interchange (“swapped”) so that parcel number 2 goes to pressure \( p_1 \) and vice versa, the change in the total enthalpy per unit mass is

\[
\Delta H = c_p (\theta_1 - \theta_2) \left[ \left( \frac{p_2}{p_0} \right)^K - \left( \frac{p_1}{p_0} \right)^K \right],
\]

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 (7.36) is the available potential energy of the system per unit 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.

7.3.2 **The APE associated with meridional temperature gradients**

As a second example 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:

\[
\theta_{gs}(\mu) = \theta_0 (1 - \Delta_H \mu^2),
\]

Here \( \Delta_H \) is a constant, and \( \mu \equiv \sin \varphi \). The subscript “gs” denotes the given state. Eq. (7.37)
is the same meridional distribution of $\theta$ as used for $\theta_E$ in Chapter 4. Recall that for realistic states $0 < \Delta_H < 1$. For realistic states, $\Delta_H > 0$. What is the available potential energy of this idealized atmospheric state? To answer this question, we need to complete the total enthalpies of the given state and the A-state, and subtract them. The first step is to find the A-state.

The mass in a latitude belt of width $d\varphi$ is:

$$dm = 2(2\pi a \cos \varphi) \left(\frac{P_S}{g}\right)(a d\varphi)$$

$$= \frac{4\pi a^2}{g} P_S d\mu .$$

(7.38)

where $\mu \equiv \sin \varphi$, and $d\mu = \cos \varphi d\varphi$. In (7.38) the leading factor of two is included because we have symmetry across the equator, so when we increment latitude in one hemisphere we actually pick up mass from two “rings” of air, one in each hemisphere. The rate of change of $\theta$ as we add mass is:

$$\frac{d\theta}{dm} = \left(\frac{d\theta}{d\mu}\right)_{gs} \left(\frac{dm}{d\mu}\right)^{-1} .$$

(7.39)

The subscript “gs” denotes the given state. Combining (7.38) through (7.40), we get

$$\frac{d\theta}{dm} = \frac{g}{4\pi a^2 P_S} \left(\frac{d\theta}{d\mu}\right)_{gs} .$$

(7.40)

In the A-state, the $\theta$-surfaces are flat, and the increment of mass between two $\theta$-surfaces is

$$dm = \frac{4\pi a^2}{g} dp ,$$

(7.41)

or

$$dm = \frac{4\pi a^2}{g} \left(\frac{dp}{d\theta}\right)_{rs} d\theta ,$$

(7.42)

so that

$$\left(\frac{d\theta}{dp}\right)_{rs} = \frac{d\theta}{dm} \frac{4\pi a^2}{g} .$$

(7.43)

The subscript “rs” denotes the A-state.
Now substitute $\frac{d\theta}{dm}$ from (7.40) into (7.43). We can do this because the distribution of $\theta$ over the mass must be the same in the A-state as in the given state. The result is

$$\left(\frac{d\theta}{dp_*}\right)_{rs} = \left(\frac{d\theta}{d\mu}\right)_{gs}.$$  \hspace{1cm} (7.44)

where

$$p_* \equiv \frac{p}{p_S}.$$  \hspace{1cm} (7.45)

We have to be careful when we look at this equation. The left-hand-side refers to the distribution of $\theta$ with pressure in the A-state. The right-hand-side refers to the distribution of $\theta$ with $\mu$ in the given state. Keep in mind that, in this idealized example, $\theta$ does not vary with pressure in the given state, and it does not vary with $\mu$ in the A-state.

From (7.44), we see that $p_*$ plays the same role in the A-state as $\mu$ plays in the given state. Also recall that the potential temperature must increase upward in the A-state. Referring back to (7.37), we conclude that

$$\theta_{rs} = \theta_0(1 - \Delta_H p_*^2),$$  \hspace{1cm} (7.46)

which is the desired formula for the distribution of potential temperature in the A-state. It can be verified that (7.46) gives the “right” values of $\theta$ 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 (7.37) gives

$$H_{gs} = \int_{-1}^{1} \int_0^{p_S} 2\pi a^2 c_p \theta_0(1 - \Delta_H \mu^2) \left(\frac{p}{p_0}\right)^\kappa \frac{dp}{g} d\mu$$

$$= 4\pi a^2 c_p \theta_0 \left(\frac{p_S}{g}\right) \left(\frac{p_S}{p_0}\right)^\kappa \left(\frac{1 - \frac{1}{3}\Delta_H}{1 + \kappa}\right).$$  \hspace{1cm} (7.47)

Similarly, the total enthalpy of the A-state is

$$H_{min} = \int_{-1}^{1} \int_0^{p_S} 2\pi a^2 c_p \theta_0 \left[1 - \Delta_H \left(\frac{p}{p_S}\right)^2\right] \left(\frac{p}{p_0}\right)^\kappa \frac{dp}{g} d\mu$$

$$= 4\pi a^2 c_p \theta_0 \left(\frac{p_S}{g}\right) \left(\frac{p_S}{p_0}\right)^\kappa \left[\left(\frac{1}{1 + \kappa}\right) - \left(\frac{\Delta_H}{3 + \kappa}\right)\right].$$  \hspace{1cm} (7.48)
Finally, we obtain

\[ A = H_{gs} - H_{min} = 4\pi a^2 c_p \theta_0 \Delta_H \left( \frac{P_S}{g} \right) \left( \frac{P_S}{P_0} \right)^K \left[ \frac{2\kappa}{3(3 + \kappa)(1 + \kappa)} \right] \] . \quad (7.49)

Note that \( A \) is proportional to \( \Delta_H \), as might be expected.

Now consider a meridional transport process:

\[ \frac{\partial \theta_{gs}}{\partial t} = \frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} \left( F_\theta \cos \varphi \right) \]
\[ = \frac{1}{a \mu} (F_\theta \sqrt{1 - \mu^2}). \quad (7.50) \]

Here \( F_\theta \) is a meridional flux of potential temperature, which we regard as given, and assume to be independent of height and longitude and symmetrical about the Equator. Note that \( F_\theta = 0 \) at both poles. \quad (7.51)

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:

\[ \frac{\partial H_{gs}}{\partial t} = \int_{-1}^{1} \int_0^{P_S} 2\pi a^2 c_p \left( \frac{\partial \theta}{\partial t} \right) \left( \frac{P}{P_0} \right)^K \frac{dp}{g} \, d\mu \]
\[ = \frac{2\pi a^2 c_p (P_S)}{(1 + \kappa) \frac{P_S}{P_0}} \int_{-1}^{1} \frac{\partial \theta}{\partial t} \, d\mu \]
\[ = \frac{2\pi a^2 c_p (P_S)}{(1 + \kappa) \frac{P_S}{P_0}} \int_{-1}^{1} \frac{\pi}{2} \frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} \left( F_\theta \cos \varphi \right) \, d\varphi \]
\[ = 0. \quad (7.52) \]

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 (7.50). According to (7.52), the specified transport
process has no effect on the total enthalpy of the given state. The reason is that the average $\theta$ 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 $\theta$ 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 $\ln(\theta)$, however, and the average value of $\ln(\theta)$ is altered by the transport process. Another point of view is that generally speaking the specified transport process is not reversible. For example, if $F_\theta$ is a downgradient flux due to diffusive mixing, then it could, in principle, homogenize $\theta$ 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 (7.52), it follows that

$$\frac{dA}{dt} = -\frac{dH_{\text{min}}}{dt}.$$  \hspace{1cm} (7.53)

Now recall that, based on comparison of (7.37) and (7.46), $p_*$ plays the same role in the A-state as $\mu$ plays in the given state. In particular, $p_* = 1$ (the surface) corresponds to $\mu = 1$ (the pole), and $p_* = 0$ (the “top of the atmosphere”) corresponds to $\mu = 0$ (the Equator). Our goal is to determine the time-rate of change of $\theta$ in the A-state at a particular instant, namely, the time when the distribution of $\theta$ satisfies (7.37) [and (7.46)], and the time-rate-of-change of $\theta$ in the given state satisfies (7.50). We can find the time-rate-of-change of $\theta$ in the A-state at this instant by going to our expression for the time-rate-of-change of $\theta$ in the given state, and simply replacing $\mu$ by $p_*$, everywhere. The time-rate-of-change of $\theta$ in the A-state thus satisfies

\[
\frac{\partial \theta}{\partial t} = -\frac{1}{a\partial p_*}[(F_\theta \sqrt{1 - p_*^2})] .
\]  \hspace{1cm} (7.54)

Again there is a possibility of confusion. We have specified that $F_\theta$ is not a function of height, although of course it does depend on latitude. It would thus appear that we can pull $F_\theta$ out of the derivative in (7.54), but this is not correct. The reason is that, when $\mu$ was replaced by $p_*$ we also replaced the $\mu$-dependence of $F_\theta$ by a corresponding $p_*$-dependence. Thus, in (7.54), $F_\theta$ should be regarded as a function of $p_*$, but not as a function of latitude! This is understandable, because $F_\theta$ is acting to change $\theta_{rs}(p_*)$. As an example, suppose that $F_\theta$ 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 \( F_\theta \) tends to warm the poles and cool the tropics, reducing \( \Delta H \) 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 \( \theta_{rs} \) 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

\[
\frac{\partial H_{\text{min}}}{\partial t} = \int p_s \left( \frac{\partial}{\partial t} \right) c_p \left( \frac{\partial \theta}{\partial t} \right) \frac{\kappa}{g} dp \, d\mu
\]

\[
= \frac{4 \pi a^2 c_p p_s}{g} \left[ \int_0^1 \left( \frac{\partial\theta}{\partial t} \right) p_s^n dp \right] \kappa \, dp_s
\]

\[
= \frac{4 \pi a^2 c_p p_s}{g a} \left[ \int_0^1 \left( -\frac{\partial}{\partial p_s} (F_\theta \sqrt{1-p_s^2}) \right) p_s^n dp_s \right] \kappa \, dp_s .
\]

We cannot do the integral on the last line of (7.55), because the form of \( F_\theta \) has not been specified. The last step would be to substitute (7.55) into (7.53).

**7.3.3 The APE associated with surface pressure variations**

APE can occur in the presence of surface pressure gradients, even in the absence of 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, \( \theta_0 \). The surface pressure, \( p_s(\lambda, \varphi) \), is given as a function of longitude, \( \lambda \) and latitude \( \varphi \). 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 (7.8), the basic definition of the APE. The total enthalpy satisfies

\[
H = \int \int \frac{2 \pi p_s}{2 \pi} c_p \left( \frac{\partial \theta}{\partial t} \right) \frac{\kappa}{g} dp \, d\mu \cos \varphi d\lambda d\varphi .
\]

Because \( \theta = \theta_0 \) everywhere in this problem, we can simplify (7.56) considerably:

\[
H = \frac{c_p \theta_0 a^2}{g (1 + \kappa) p_0} \pi \int \frac{p_s^{1+\kappa} \cos \varphi d\lambda d\varphi}{\pi \frac{2}{2}}
\]

Eq. (7.57) can be applied to both the given state and the A-state, so that the APE is given by.
Before we can evaluate the double integrals in (7.58), we must substitute for $[p_s(\lambda, \varphi)]_{gs}$ and $[p_s(\lambda, \varphi)]_{rs}$. The former is assumed to be known. Our problem thus reduces to finding $[p_s(\lambda, \varphi)]_{rs}$. This is very simple, 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 (7.58) as

$$A = \frac{c_p \theta_0 a^2}{g (1 + \kappa) p_0} \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} \left[ \left( p_s \right)_{gs}^{1 + \kappa} - \left( p_s \right)_{rs}^{1 + \kappa} \right] \cos \varphi d\lambda d\varphi. \tag{7.59}$$

As an exercise, prove that (7.59) gives $A \geq 0$.

### 7.4 Variance budgets

From (7.22) we see that the available potential energy is closely related to the variance of temperature or 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, $A_E$. The eddy potential temperature variance interacts with the meridional gradient of the zonally averaged potential temperature through

$$\frac{\partial}{\partial t} [\theta^2] \sim \frac{-[\theta^* v^*]}{a} \frac{\partial}{\partial \phi} [\theta]. \tag{7.60}$$

The term shown in the right-hand side of (7.60) is called the “meridional gradient-production term.” There are actually several additional terms; the others are discussed below. To gain an intuitive understanding of the gradient production terms, consider the simple example illustrated in Fig. 7.5. State A consists of two latitude belts of equal mass, each with uniform $\theta$. State B is obtained by homogenizing State A. Consider the average of the square of $\theta$, for each state. For State A,

$$\overline{\theta^2} = \frac{1}{2} (\theta_1^2 + \theta_2^2). \tag{7.61}$$
Here the overbar denotes an average over both latitude belts. For State B,

\[
\overline{\theta^2} = \left[ \frac{1}{2} (\theta_1 + \theta_2) \right]^2 = \frac{1}{4} (\theta_1^2 + 2\theta_1 \theta_2 + \theta_2^2)
\]

\[
= \frac{1}{2} (\theta_1^2 + \theta_2^2) - \frac{1}{4} (\theta_1^2 - 2\theta_1 \theta_2 + \theta_2^2)
\]

\[
= \frac{1}{2} (\theta_1^2 + \theta_2^2) - \frac{1}{4} (\theta_1 - \theta_2)^2
\]

\[
\leq \frac{1}{2} (\theta_1^2 + \theta_2^2).
\]

(7.62)

This shows that down-gradient transport (in this case, mixing) reduces the square of the mean state. It correspondingly increases the square of the fluctuations. In other words, it leads to a conversion of mean-state variance to eddy variance. This appears as a “dissipation” of the

\[\theta_1\]

\[\theta_2\]

\[\theta = \frac{1}{2} (\theta_1 + \theta_2)\]

State A

State B

Figure 7.5: A simple example used to explain the idea of gradient production. State B is obtained by homogenizing State A. In both State A and State B, the average of \(\theta\) is \(\overline{\theta} = \frac{1}{2} (\theta_1 + \theta_2)\). Variance seems to “disappear” in passing from State A to State B. In reality it is converted from the variance of the mean to an eddy variance.

mean-state variance. Further discussion is given later.

To see where (7.60) comes from, start from the conservation equation for potential temperature, i.e.,
\[
\frac{\partial \theta}{\partial t} + \frac{1}{a \cos \phi} \frac{\partial}{\partial \lambda} (u \theta) + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} (v \cos \phi) + \frac{\partial}{\partial p} (\omega \theta) = \theta, \quad (7.63)
\]

and also mass continuity:
\[
\frac{1}{a \cos \phi} \frac{\partial u}{\partial \lambda} + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} (v \cos \phi) + \frac{\partial \omega}{\partial p} = 0. \quad (7.64)
\]

Zonally average (7.63) and (7.64), to obtain
\[
\frac{\partial}{\partial t} ([\theta] + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} ([v] \cos \phi) + \frac{\partial}{\partial p} [\omega \theta]) = [\theta], \quad (7.65)
\]
\[
\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} ([v] \cos \phi) + \frac{\partial}{\partial p} [\omega] = 0. \quad (7.66)
\]

Subtracting (7.65) and (7.66) from (7.63) and (7.64), respectively, we find that
\[
\frac{\partial \theta^*}{\partial t} + \frac{1}{a \cos \phi} \frac{\partial}{\partial \lambda} (u^* \theta) + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{(v \cos \phi - [v]) \cos \phi \} + \frac{\partial}{\partial p} ((\omega \theta - [\omega \theta])
\]
\[
= \frac{\partial \theta^*}{\partial t} + \frac{1}{a \cos \phi} \frac{\partial}{\partial \lambda} (u^* [\theta] + [u] \theta^* + u^* \theta^*)
\]
\[
+ \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{(v^* [\theta] + [v] \theta^* + v^* \theta^* - [v^* \theta^*]) \cos \phi \}
\]
\[
+ \frac{\partial}{\partial p} ((\omega^* [\theta] + [\omega] \theta^* + \omega^* \theta^* - [\omega^* \theta^*])
\]
\[
= \frac{\partial \theta^*}{\partial t} + \frac{1}{a \cos \phi} \frac{\partial}{\partial \lambda} ([u] \theta^*) + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{([v] \theta^*) \cos \phi \} + \frac{\partial}{\partial p} ([\omega \theta^*]
\]
\[
+ \frac{1}{a \cos \phi} \frac{\partial}{\partial \lambda} (u^* \theta^*) + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} (v \cos \phi) + \frac{\partial}{\partial p} (\omega \theta^*)
\]
\[
+ \frac{1}{a \cos \phi} \frac{\partial}{\partial \lambda} (u^* [\theta]) + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{([v] [\theta]) \cos \phi \} + \frac{\partial}{\partial p} (\omega [\theta])
\]
\[
- \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} ([v^* \theta^*] \cos \phi) - \frac{\partial}{\partial p} ([\omega^* \theta^*])
\]
\[
= \dot{\theta}^*,
\]
\[
\frac{1}{a \cos \phi} \frac{\partial u^*}{\partial \lambda} + \frac{1}{a \cos \phi} \frac{\partial (v^* \cos \phi)}{\partial \phi} + \frac{\partial \omega^*}{\partial p} = 0. \quad (7.68)
\]
To obtain the first equality of (7.67), we have used
\[
\nu\theta = ([\nu] + \nu_*)([\theta] + \theta_*) = [\nu][\theta] + \nu_*[\theta] + [\nu]\theta_* + \nu_*\theta_*,
\]  
(7.69)

\[
[v\theta] = [\nu][\theta] + [\nu_*\theta_*],
\]  
(7.70)

\[
\nu\theta - [\nu\theta] = \nu_*[\theta] + [\nu]\theta_* + \nu_*\theta_* - [\nu_*\theta_*],
\]  
(7.71)

and so on. We can use (7.66) and (7.68) to rewrite (7.67) as follows:
\[
\left(\frac{\partial}{\partial t} + \frac{[u]}{a\cos\phi}\frac{\partial}{\partial \lambda} + \frac{[v]}{a}\frac{\partial}{\partial \phi} + [\omega]\frac{\partial}{\partial p}\right)\theta_* + \left(\frac{u_*}{a\cos\phi}\frac{\partial}{\partial \lambda} + \frac{v_*}{a}\frac{\partial}{\partial \phi} + \omega_*\frac{\partial}{\partial p}\right)\theta_*
\]  
(7.72)

\[
+ \frac{\nu_*}{a}\frac{\partial}{\partial \phi}[\theta] + \omega_*\frac{\partial}{\partial p}[\theta] = \frac{1}{a\cos\phi}\left([\nu_*\theta_*]\cos\phi + \frac{\partial}{\partial p}[\omega_*\theta_*] + \theta_*. \right.
\]

Multiplying (7.72) by \(\theta_*\), and using (7.68) again, we obtain:
\[
\left(\frac{\partial}{\partial t} + \frac{[u]}{a\cos\phi}\frac{\partial}{\partial \lambda} + \frac{[v]}{a}\frac{\partial}{\partial \phi} + [\omega]\frac{\partial}{\partial p}\right)\left(\frac{1}{2}\theta_*^2\right)
\]  
(7.73)

Zonally averaging gives
\[
\left(\frac{\partial}{\partial t} + \frac{[v]}{a}\frac{\partial}{\partial \phi} + [\omega]\frac{\partial}{\partial p}\right)\left(\frac{1}{2}[\theta_*^2]\right)
\]  
(7.74)

\[
+ \frac{1}{a\cos\phi}\frac{\partial}{\partial \phi}\left([\nu_*\theta_*\frac{1}{2}\theta_*^2]\cos\phi\right) + \frac{\partial}{\partial p}\left([\omega_*\theta_*\frac{1}{2}\theta_*^2]\right)
\]

\[
= - \frac{[\nu_*\theta_*]}{a}\frac{\partial}{\partial \phi}[\theta] - [\omega_*\theta_*]\frac{\partial}{\partial p}[\theta] + [\dot{\theta}\theta_*].
\]

Finally, we can use (7.66) to rewrite (7.74) in flux form:
According to (7.75), $\frac{1}{2}[\theta_*^2]$ can change due to advection by the mean meridional circulation, or due to transport by the eddies themselves, or due to “gradient production.”

Eq. (7.75) governs the “eddy variance.” (7.65) There is also a contribution to the global variance of $\theta$ that comes from the meridional and vertical gradients of $[\theta]$. To derive an equation for this part of the $\theta$ global variance of $\theta$, start by rewriting (7.65) as

$$
\left( \frac{\partial}{\partial t} + \frac{[v]}{a} \frac{\partial}{\partial \varphi} + [\omega] \frac{\partial}{\partial p} \right) [\theta]
= - \frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} ([v_* \theta_*] \cos \varphi) - \frac{\partial}{\partial p} [\omega_* \theta_*] + [\theta].
$$

(7.76)

Multiplication by $[\theta]$ gives

$$
\left( \frac{\partial}{\partial t} + \frac{[v]}{a} \frac{\partial}{\partial \varphi} + [\omega] \frac{\partial}{\partial p} \right) \frac{1}{2}[\theta]^2
= - \frac{[\theta]}{a \cos \varphi} ([v_* \theta_*] \cos \varphi) - [\theta] \frac{\partial}{\partial p} ([\omega_* \theta_*]) + [\theta][\dot{\theta}].
$$

(7.77)

This can be rearranged to
Converting back to flux form, we find that

\[
\begin{align*}
\left( \frac{\partial}{\partial t} + \frac{v}{a} \frac{\partial}{\partial \phi} + [\omega] \frac{\partial}{\partial p} \right) \frac{1}{2} [\theta]^2 \\
= - \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} ([\theta][v_\ast \theta_\ast \cos \phi]) - \frac{\partial}{\partial p} ([\theta][\omega_\ast \theta_\ast]) \\
+ \frac{[v_\ast \theta_\ast]}{a} [\theta] + [\omega_\ast \theta_\ast] \frac{\partial}{\partial \phi} [\theta] + [\dot{\theta}][\theta] .
\end{align*}
\]  

(7.78)

When we add (7.79) and (7.75), the gradient production terms cancel. This shows that the gradient production terms represent a “conversion” between \( \frac{1}{2} [\theta]^2 \) and \( \frac{1}{2} \theta_\ast^2 \). We obtain:

\[
\begin{align*}
\left( \frac{\partial}{\partial t} + \frac{v}{a} \frac{\partial}{\partial \phi} + [\omega] \frac{\partial}{\partial p} \right) \left( \frac{1}{2} [\theta]^2 + \frac{1}{2} \theta_\ast^2 \right) \\
= - \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} ([\theta][v_\ast \theta_\ast \cos \phi]) - \frac{\partial}{\partial p} ([\theta][\omega_\ast \theta_\ast]) \\
+ \frac{[v_\ast \theta_\ast]}{a} [\theta] + [\omega_\ast \theta_\ast] \frac{\partial}{\partial \phi} [\theta] + [\dot{\theta}][\theta] .
\end{align*}
\]  

(7.79)

Finally, integration of (7.80) over the entire atmosphere gives

\[
\frac{d}{dt} \int_M \left( \frac{1}{2} [\theta]^2 + \frac{1}{2} \theta_\ast^2 \right) dM = \int_M \left( [\theta][\dot{\theta}] + [\dot{\theta}][\theta] \right) dM .
\]  

(7.81)

This shows that, in the absence of heating, the sum of the eddy variance and the variance of the zonal mean is a constant. Note also that the potential temperature variance increases with time if we “heat where it’s hot and cool where it’s cold.”
7.5 **Generation of available potential energy, and its conversion into kinetic energy**

Earlier we derived (7.22), Lorenz’s approximate expression for the available potential energy of a statically stable atmosphere, which is repeated here for convenience:

$$ A = \frac{a^2}{2} \int_{-\pi/2}^{\pi/2} \int_{0}^{2\pi} \frac{\bar{T}}{T} \left( \frac{T'}{T} \right)^2 \cos \phi d\phi d\lambda. $$  \hspace{1cm} (7.82)

Recall that in this equation an overbar represents a global mean on a pressure surface, and a prime denotes a departure from the global mean. The APE is an integral of the variance of the temperature about its global mean on pressure 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 and conversion to or from kinetic energy.

Let the subscript “GM” denote a global mean on an isobaric surface, i.e.,

$$ (\ )_{\text{GM}} \equiv \frac{1}{4\pi a^2} \int_{-\pi/2}^{\pi/2} \int_{0}^{2\pi} (\ ) a^2 \cos \phi d\phi d\lambda. $$ \hspace{1cm} (7.83)

We can show that, for any two quantities $A$ and $B$,

$$ (AB)_{\text{GM}} = ([A][B] + [A*B])_{\text{GM}}, $$ \hspace{1cm} (7.84)

and

$$ \{ (A - A_{\text{GM}})(B - B_{\text{GM}}) \}_{\text{GM}} = (AB)_{\text{GM}} - A_{\text{GM}}B_{\text{GM}} $$

$$ = ([A][B] + [A*B])_{\text{GM}} - A_{\text{GM}}B_{\text{GM}}. $$ \hspace{1cm} (7.85)

As a special case of (7.85), the variance of an arbitrary quantity $A$ about its global mean is given by

$$ A_{\text{Var}} \equiv (A - A_{\text{GM}})^2_{\text{GM}} $$

$$ = ([A]^2 + [A^2])_{\text{GM}} - A_{\text{GM}}^2. $$ \hspace{1cm} (7.86)

Using the notation introduced above, we can approximate (7.78) by
We now work out an equation for $\frac{dA}{dt}$, based on (7.83). The global means of (7.63) and (7.64) are

\[
\frac{\partial \theta_{GM}}{\partial t} + \frac{\partial (\omega \theta)_{GM}}{\partial p} = \dot{\theta}_{GM},
\]

(7.88)

and

\[
\frac{\partial \omega_{GM}}{\partial p} = 0,
\]

(7.89)

respectively. Since $\omega = 0$ at $p = 0$, it follows from (7.89) that

\[
\omega_{GM} = 0 \text{ for all } p.
\]

(7.90)

This allows us to write

\[
(\omega \theta)_{GM} = \omega_{GM} \omega_{GM} + \{(\omega - \omega_{GM})(\theta - \theta_{GM})\}_{GM}
\]

\[
= \{\omega(\theta - \theta_{GM})\}_{GM}.
\]

(7.91)

Area-weighted integration of (7.80) over all latitudes, at a given pressure-level, leads to

\[
\frac{\partial}{\partial t} \left( \frac{1}{2} [\theta]^2 + \frac{1}{2} [\theta^*]^2 \right)_{GM} + \frac{\partial}{\partial p} \left[ \omega \left( \frac{1}{2} [\theta]^2 + \frac{1}{2} [\theta^*]^2 \right) \right]_{\text{GM}}
\]

\[
= - \frac{\partial}{\partial p} \left( \{\theta\} [\omega \theta^*]_{GM} + (\{\theta\} \dot{\theta} + [\dot{\theta} \theta^*]_{GM} \right).
\]

(7.92)

Here we ignore, as usual, the complications arising from the fact that some pressure surfaces intersect the Earth’s surface. From (7.82), we see that
Substituting into (7.89) from (7.88) and (7.92), we find that

\[
\frac{\partial}{\partial t} \left( \frac{\Theta_{\text{Var}}}{2} \right) = \frac{\partial}{\partial p} \left\{ \omega \left( \frac{1}{2} \Theta^2 + \frac{1}{2} \Theta_*^2 \right) + \left( \omega_* \left( \frac{1}{2} \Theta_*^2 \right) \right) + ([\Theta] \Theta_* \Theta)_{\text{GM}} \right\}_{\text{GM}}
\]

\[+ \Theta_{\text{GM}} \frac{\partial (\omega \Theta)}{\partial p} + ([\Theta] \dot{\Theta} + \dot{\Theta}_* \Theta)_{\text{GM}} \Theta_{\text{GM}} \hat{\Theta}_{\text{GM}} \]

\[= \frac{\partial}{\partial p} \left\{ \left[ \omega \left( \frac{1}{2} \Theta^2 + \frac{1}{2} \Theta_*^2 \right) + \left( \omega_* \left( \frac{1}{2} \Theta_*^2 \right) \right) + ([\Theta] \Theta_* \Theta)_{\text{GM}} - \Theta_{\text{GM}} \omega \Theta \right\}_{\text{GM}} \right\}
\]

\[- \{ \omega (\Theta - \Theta_{\text{GM}}) \} \frac{\partial}{\partial p} \theta_{\text{GM}} + \{ (\Theta - \Theta_{\text{GM}})(\dot{\Theta} - \dot{\Theta}_{\text{GM}}) \} \theta_{\text{GM}} .
\]

Here we can recognize what is going on. Vertical transport is clearly visible, as are gradient production and “Heat where it’s hot, cool where it’s cold.” Now we apply (7.20) in the form

\[
\frac{\partial \Theta_{\text{GM}}}{\partial p} = \frac{\kappa \theta_{\text{GM}} (\Gamma_d - \Gamma_{\text{GM}})}{p}.
\]

This leads to

\[
\frac{\partial}{\partial t} \left( \frac{\Theta_{\text{Var}}}{2} \right) =
\]

\[\frac{\partial}{\partial p} \left\{ \omega \left( \frac{1}{2} \Theta^2 + \frac{1}{2} \Theta_*^2 \right) + \left( \omega_* \left( \frac{1}{2} \Theta_*^2 \right) \right) + ([\Theta] \Theta_* \Theta)_{\text{GM}} - \Theta_{\text{GM}} \omega \Theta \right\}_{\text{GM}}
\]

\[+ \{ \omega (\Theta - \Theta_{\text{GM}}) \} \frac{\kappa \theta_{\text{GM}} (\Gamma_d - \Gamma_{\text{GM}})}{p} + \{ (\Theta - \Theta_{\text{GM}})(\dot{\Theta} - \dot{\Theta}_{\text{GM}}) \} \theta_{\text{GM}} ,
\]

or

An Introduction to the General Circulation of the Atmosphere
7.5 Generation of available potential energy, and its conversion into kinetic energy

\[
\frac{\partial}{\partial t} (\theta_{\text{Var}}) =
\]
\[
-\frac{\partial}{\partial p} \left\{ \left[ \omega \left( \left[ \theta \right]^2 + \left[ \theta^2 \right] \right) + \left[ \omega \left( \theta^2 \right) \right] + 2 \left( \left[ \theta \right] \left[ \omega, \theta \right] \right) \right] - 2 \theta_{\text{GM}} \left( \omega \theta \right)_{\text{GM}} \right\}_{\text{GM}} \quad (7.96)
\]
\[
+ 2 \{ \omega (\theta - \theta_{\text{GM}}) \} \left[ \frac{\kappa \theta_{\text{GM}} \left( \Gamma_d - \Gamma_{\text{GM}} \right)}{p} \right] + 2 \{ (\theta - \theta_{\text{GM}}) (\dot{\theta} - \dot{\theta}_{\text{GM}}) \} \left[ \frac{\Gamma_d - \Gamma_{\text{GM}}}{\Gamma_d} \right] .
\]

Vertical integration of (7.96) gives
\[
\frac{d}{dt} \left[ \int_{0}^{p_s} \left( \theta_{\text{Var}} dp \right) \right] = \int_{0}^{p_s} \left\{ 2 \{ \omega (\theta - \theta_{\text{GM}}) \} \left[ \frac{\kappa \theta_{\text{GM}} \left( \Gamma_d - \Gamma_{\text{GM}} \right)}{p} \right] dp 
\]
\[
+ 2 \{ (\theta - \theta_{\text{GM}}) (\dot{\theta} - \dot{\theta}_{\text{GM}}) \} \right\}_{\text{GM}}. \quad (7.97)
\]

This can be approximated by
\[
\frac{d}{dt} \left[ \int_{0}^{p_s} \left( \frac{T_{\text{GM}}}{\Gamma_d - \Gamma_{\text{GM}}} \left( \frac{\theta_{\text{Var}}}{\theta_{\text{GM}}} \right) \right) dp \right] = \int_{0}^{p_s} \left\{ 2 \{ \omega (\theta - \theta_{\text{GM}}) \} \left[ \frac{\kappa T_{\text{GM}}}{p \Gamma_d} \right] dp 
\]
\[
+ 2 \int_{0}^{p_s} \left\{ (\theta - \theta_{\text{GM}}) (\dot{\theta} - \dot{\theta}_{\text{GM}}) \right\}_{\text{GM}} \frac{T_{\text{GM}}}{\theta_{\text{GM}} \left( \Gamma_d - \Gamma_{\text{GM}} \right)} dp .
\]

Finally, note that
\[
\frac{\theta - \theta_{\text{GM}}}{\theta_{\text{GM}}} = \frac{\alpha - \alpha_{\text{GM}}}{\alpha_{\text{GM}}} = \frac{T - T_{\text{GM}}}{T_{\text{GM}}} . \quad (7.99)
\]

so that
Recall that \( \omega \alpha \) represents conversion between kinetic and non-kinetic energy, so by inspection of (7.100) we can identify rate of conversion of KE into APE as

\[
\frac{dA}{dt} = 4\pi a^2 \int_0^{p_S} \left\{ \omega \left( \frac{\alpha - \alpha_{GM}}{\alpha_{GM}} \right) \right\}_{GM} \frac{kT_{GM}}{p\Gamma_d} dp \\
+ 4\pi a^2 \int_0^{p_S} \left\{ \left( \theta - \theta_{GM} \right) \left( \dot{\theta} - \dot{\theta}_{GM} \right) \right\}_{GM} \frac{T_{GM}}{\theta_{GM}^2 (\Gamma_d - \Gamma_{GM})} dp
\]

\[
= 4\pi a^2 \int_0^{p_S} \omega \left( \alpha - \alpha_{GM} \right) \{ GM dp \\n+ 4\pi a^2 \int_0^{p_S} \left\{ \left( \frac{\theta - \theta_{GM}}{\theta_{GM}} \right) \left( \frac{T_{GM}}{\theta_{GM}} \right) \left( \dot{\theta} - \dot{\theta}_{GM} \right) \right\}_{GM} \frac{T_{GM}}{(\Gamma_d - \Gamma_{GM})} dp. \tag{7.100}
\]

Recall that \( \omega \alpha \) represents conversion between kinetic and non-kinetic energy, so by inspection of (7.100) we can identify rate of conversion of KE into APE as

\[
C \equiv 4\pi a^2 \int_0^{p_S} \omega \left( \alpha - \alpha_{GM} \right) \{ GM dp \} \tag{7.101}
\]

The rate of generation or destruction of APE by heating is

\[
G \equiv \pi a^2 \int_0^{p_S} \left\{ \left( \frac{\theta - \theta_{GM}}{\theta_{GM}} \right) \left( \frac{T_{GM}}{\theta_{GM}} \right) \left( \dot{\theta} - \dot{\theta}_{GM} \right) \right\}_{GM} \frac{T_{GM}}{(\Gamma_d - \Gamma_{GM})} dp. \tag{7.102}
\]

Note that \( G > 0 \) if we “heat where it’s hot and cool where it’s cold.” This should sound familiar. A heating field that generates APE destroys entropy.

We can summarize (7.100) as

\[
\frac{dA}{dt} = C + G. \tag{7.103}
\]

Similarly, we can show that

\[
\frac{dK}{dt} = -C - D, \tag{7.104}
\]

Where \( D \) 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 KE, in order to supply the KE that is destroyed by dissipation. Because this conversion depletes the APE, the generation of APE must be positive. In other words, energy must flow as shown in Fig. 7.6.

**Figure 7.6:** Sketch illustrating the flow of energy through the atmospheric general circulation. Generation produces APE, which is converted to KE, which in turn is dissipated.

### 7.6 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, and so will be omitted here for brevity. A handout showing the details of the derivations is available on request.

We define the eddy kinetic energy per unit mass by

\[
KE \equiv \frac{1}{2} [(u^*)^2 + (v^*)^2].
\]

It satisfies the following equation:

\[
\frac{\partial KE}{\partial t} + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \left\{ \left[ v \right] KE + \frac{1}{2} \left[ v^* u^* u^* \right] + \frac{1}{2} \left[ v^* v^* v^* \right] \right\} \cos \phi
\]

\[
+ \frac{\partial}{\partial p} \left\{ \left[ \omega \right] KE + \frac{1}{2} \left[ \omega^* u^* u^* \right] + \frac{1}{2} \left[ \omega^* v^* v^* \right] \right\} - \left\{ \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \left[ \left( v^* \phi^* \right) \cos \phi \right] + \frac{\partial}{\partial p} \left[ \omega^* \phi^* \right] \right\}
\]

\[
= -\frac{v^* u^*}{a} \frac{\partial}{\partial \phi} [u] - \frac{v^* v^*}{a} \frac{\partial}{\partial \phi} [v] - \left[ \omega^* u^* \right] \frac{\partial}{\partial p} [u] - \left[ \omega^* v^* \right] \frac{\partial}{\partial p} [v]
\]

\[
- \left[ \omega^* \alpha^* \right] + \left[ \left( u^* u^* \right) [v] - \left( u^* v^* \right) [u] \right] \tan \phi \frac{\partial}{\partial \phi} \left[ \tau_{\phi} \right] + \left[ v^* g \frac{\partial}{\partial p} \tau_{\phi} \right] + \left[ v^* g \frac{\partial}{\partial p} \tau_{\phi} \right].
\]
The terms on the first line of the right-hand side represent gradient production, i.e. 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 $\omega^*\alpha^*$ term represents eddy kinetic energy generation from eddy available potential energy, while the terms involving $\phi^*$ represent the effects of “pressure work.”

The appearance of the metric terms in (7.106) 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

$$KZ \equiv \frac{1}{2}([u]^2 + [v]^2),$$

and note that

$$[K] = KZ + KE.$$  

(7.108)

All three quantities in Eq. (7.108) are independent of longitude. The zonal kinetic energy satisfies

$$\begin{align*}
\frac{\partial}{\partial t}KZ + \frac{1}{a \cos \phi \partial \phi}([v]KZ \cos \phi) + \frac{\partial}{\partial p}([\omega]KZ) \\
= -\frac{1}{a \cos \phi \partial \phi}([v][\phi] \cos \phi) - \frac{\partial}{\partial p}([\phi][\omega]) - [\omega][\alpha] \\
- \frac{[v]}{a \cos \phi \partial \phi}([v^* u^*] \cos \phi) - [u] \frac{\partial}{\partial p}([\omega^* u^*]) \\
- \frac{[v]}{a \cos \phi \partial \phi}([v^* v^*] \cos \phi) - [v] \frac{\partial}{\partial p}([\omega^* v^*]) \\
+ ([u][v^* u^*] - [v][u^* u^*]) \tan \phi \\
+ [u] g \frac{\partial}{\partial p} [\tau_\lambda] + [v] g \frac{\partial}{\partial p} [\tau_\phi].
\end{align*}$$

(7.109)

The terms on the third and fourth lines of the right-hand side of (7.109) can be interpreted as representing the work done by the mean flow against the “forces” exerted on the mean flow by the eddies, through eddy momentum transport.

Comparison with the eddy kinetic energy equation shows that, as expected, the metric terms do not affect the zonally averaged total kinetic energy. Adding the equations for
7.6  The governing equations for the eddy kinetic energy, zonal kinetic energy, and \( K \)

The governing equations for the eddy kinetic energy, \( K \), gives the equation for the zonally averaged total kinetic energy, \([K]\):

\[
\begin{align*}
\frac{\partial}{\partial t}[K] &+ \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \left\{ \left[ v \right][K] + \frac{1}{2} \left[ v^* u^* u^* \right] + \frac{1}{2} \left[ v^* v^* v^* \right] \right\} \cos \phi \\
&+ \frac{\partial}{\partial p} \left( \left[ \omega \right][K] + \frac{1}{2} \left[ \omega^* u^* u^* \right] + \frac{1}{2} \left[ \omega^* v^* v^* \right] \right) \\
= &- \frac{[v^* u^*]}{a} \frac{\partial}{\partial \phi} \left[ u \right] \frac{[v^* v^*]}{a} \frac{\partial}{\partial \phi} \left[ v \right] - \left[ \omega^* u^* \right] \frac{\partial}{\partial p} \left[ u \right] - \left[ \omega^* v^* \right] \frac{\partial}{\partial p} \left[ v \right] \\
&- \left[ a \cos \phi \frac{\partial}{\partial \phi} \left[ v^* u^* \right] \cos \phi \right] - \left[ u \right] \frac{\partial}{\partial p} \left( \left[ \omega^* u^* \right] \right) \\
&- \left[ a \cos \phi \frac{\partial}{\partial \phi} \left[ v^* v^* \right] \cos \phi \right] - \left[ v \right] \frac{\partial}{\partial p} \left( \left[ \omega^* v^* \right] \right) \\
&\left\{ \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \left( [v] [\phi] \cos \phi \right) + \frac{\partial}{\partial p} \left( [\phi] [\omega] \right) \right\} - \left[ \omega [\alpha] \right] \\
&- \left\{ \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \left( [v^* \phi^*] \cos \phi \right) + \frac{\partial}{\partial p} \left( [\omega^* \phi^*] \right) \right\} - \left[ \omega^* \alpha^* \right] \\
&+ \left[ u^* g \frac{\partial}{\partial p} \tau^*_\lambda \right] + \left[ v^* g \frac{\partial}{\partial p} \tau^*_\phi \right] + \left[ u \right] g \frac{\partial}{\partial p} \left[ \tau^*_\lambda \right] + \left[ v \right] g \frac{\partial}{\partial p} \left[ \tau^*_\phi \right].
\end{align*}
\]

The first three lines on the right-hand side of (7.110) come from the “gradient production” terms of the eddy kinetic energy equation and the terms of the zonal kinetic energy equation that represent the work done by the mean flow against the forces exerted on the mean flow by the eddies. Inspection shows that these terms can be combined, as shown in Eq. (7.111):
Here the terms mentioned above have been combined on the first two lines of the right-hand side, and it is apparent that they take the form of a divergence. This implies, of course, that the two terms together integrate to zero over the whole atmosphere. The interpretation of this result is that the gradient production terms represent conversion between the kinetic energy of the mean flow and the kinetic energy of the eddies. We do not see simple cancellation between the corresponding terms of the $KE$ and $KZ$ equations, because the conversion process is not local; it occurs over a region that is extended in the meridional and vertical directions. Cancellation occurs only when we integrate over the whole atmosphere that.

7.7 Observations of the energy cycle

Arpé et al. (1986) discussed the observed energy cycle of the atmosphere, based on ECMWF analyses. They wrote the following equations for the energy cycle:

$$\frac{d}{dt} KZ = -\sum_m CK(m) + CZ - DZ ,$$  \hspace{1cm} (7.112)

$$\frac{d}{dt} AZ = -\sum_m CA(m) - CZ + GZ ,$$  \hspace{1cm} (7.113)

$$\frac{\partial}{\partial t} KE(m) = CK(m) + LK(m) + CE(m) - DE(m) ,$$  \hspace{1cm} (7.114)

$$\frac{\partial}{\partial t} AE(m) = CA(m) + LA(m) - CE(m) + GE(m) .$$  \hspace{1cm} (7.115)
Here $m$ is the zonal wave number. The eddy kinetic energy and eddy available potential energy are defined as functions of the zonal wave number, and the contributions for the individual waves have been worked out. 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. If (7.112)-(7.115) are added together, all terms on the right-hand side of the result cancel, except for \( GZ, GE, DZ \), and \( DE \).

Baroclinic instability of the zonal flow would be 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 \( KE \).

A tendency for eddies to pump momentum into the jet, increasing its strength, would be represented by negative values of \( CK \).

A direct mean meridional circulation, such as the Hadley Circulation, converts \( AZ \) into \( KZ \), and so would be associated with positive values of \( CZ \).

It is important to understand the meaning of each term of (7.112) through (7.115), as outlined above. It is also important, however, to notice that certain terms are not present, i.e. certain processes do not exist. For example, there is no process that directly converts \( AZ \) into \( KE(m) \). Such conversion can occur only indirectly, via a two-step process, e.g. first \( AZ \rightarrow AE \) and then \( AE \rightarrow KE \).

Fig. 7.7 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 Arpé et al. have defined the APE for the Northern and Southern Hemispheres separately; as discussed earlier, this is not strictly kosher.

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, however, so that 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.

It is obvious from Fig. 7.7 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 pumping angular momentum into the jet stream, and \( KZ \) is converted back to \( AZ \). 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. 7.8 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.
7.8 The role of heating

Fig. 7.9 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. 7.10 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.

7.8 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 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.

7.9 Summary

We defined the available potential energy and the gross static stability, and studied the generation and conversion of available potential energy. Finally we presented observations of the atmospheric energy cycle.

The observations of Arpé et al. remind us again of the wide range of eddy scales that are simultaneously at work in the general circulation the 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 a later chapter, in which we view the general circulation as a kind of large-scale turbulence.

Problems

1. Prove that the mass-integrated potential energy of the entire atmosphere is lower in the A-state than in the given state.

2. Derive (7.20).

3. Prove that
Figure 7.10: Mean annual cycles of energy amounts and conversions in both hemispheres and contributions from wavenumber groups calculated from 12 hour forecasts. From Arpé et al. (1986).
Note that on the right-hand side the lower limit of integration is zero. This result is used to derive (7.13) from (7.12).

4. Prove that the mass-integrated gravitational potential energy of the entire atmosphere is smaller in the A-state than in the given state.

5. Prove that \( S \geq 0 \), where \( S \) is given by (7.33). State any assumptions.

6. The shallow water equations are:

\[
\frac{\partial h}{\partial t} + \nabla \cdot (h \mathbf{V}) = 0 ,
\]

\[
\frac{\partial \mathbf{V}}{\partial t} + (\zeta + f) \mathbf{k} \times \mathbf{V} + \nabla [ K + g(h + h_S) ] = 0 ,
\]

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 \( h_S \) is the height of the lower boundary (see the sketch below).

Show that the available potential energy of the system, per unit area, is

\[
A = \frac{1}{2}g(\overline{H^2} - \overline{H}^2) ,
\]

where \( H \equiv h + h_S \), and the bar represents an average over the whole domain. Prove that \( \frac{d}{dt}(A + \overline{hK}) = 0 \).

7. a) For the example given in the discussion beginning with Eq. (7.37), calculate the variance of \( \Theta \) for both the given state and the A-state, and demonstrate that the two
variances are equal.

b) Continuing the example, assume that

\[ F_\theta = -\frac{D\partial \theta}{\partial \varphi}, \tag{7.119} \]

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. (7.37).

c) Suppose that we add a “heating” term to (7.50), so that it becomes

\[ \frac{\partial \theta}{\partial t} = -\frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} (F_\theta \cos \varphi) + Q(\varphi). \tag{7.120} \]

Continue to use the form of \( F_\theta \) given in part b) above. Find the form of \( Q(\varphi) \) needed to maintain a steady state. Plot \( Q(\varphi) \). Find the global mean of \( Q(\varphi) \). Find the covariance of \( Q(\varphi) \) and \( \theta \). Discuss.

8. Suppose that

\[ \frac{d}{dt} AZ = GZ - \frac{AZ}{\tau_Z} \equiv 0, \]
\[ \frac{d}{dt} KE = CE - \frac{KE}{\tau_E} \equiv 0, \]

Estimate the values of the time scales \( \tau_Z \) and \( \tau_E \), by using the numerical values given in Fig. 7.7. Compare \( GZ \) and \( GE \) with the actual rates of change of \( AZ \) and \( KE \) as shown in Fig. 7.8.
CHAPTER 8

Planetary-scale waves and other eddies

Copyright 2005 David A. Randall

8.1 Introduction

Earlier we introduced the concept of the zonally averaged circulation, and defined “eddies” as zonal inhomogeneities. In atmospheric science, the term “eddy” is often used synonymously with “wave,” but actually waves are special kinds of eddies. A “wave” can be defined as a process that transports energy without transporting mass. Waves can also transport momentum. Familiar examples are sound waves, and water waves seen at the beach; the latter are examples of what meteorologists call “gravity waves.”

If a wave does not transport mass, then it cannot transport any property that is “attached” to the mass, i.e., any property that is conserved under adiabatic frictionless processes. Examples include the mixing ratios of water vapor and other minor constituents, the potential temperature (for dry adiabatic processes), and the potential vorticity. Waves cannot produce fluxes of any of these quantities.

Waves can be produced by a variety of mechanisms (Fig. 8.1). Orographic forcing, heating that is localized at particular longitudes, and baroclinic and barotropic instability are among the mechanisms that can generate large-scale waves in the atmosphere. In the case of a wave forced by heating, the wave itself may or may not alter the heating.

Waves arise from a variety of physical mechanisms, and appear on many spatial and temporal scales. The atmosphere supports Rossby waves, Kelvin waves, inertia-gravity waves, mixed Rossby-gravity waves and 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 to weeks. The distributions of energy and fluxes over these various scales are complicated and interesting.

From elementary physics, we know that a particle does not undergo any net displacement as a small-amplitude wave passes by; it moves in a more or less circular path, and returns to its starting point. This is consistent with the definition of “wave” given above. We also know, however, that it is possible for a surfer to ride on a large-amplitude water wave. We can think of the surfer as a particle of mass that is, in fact, transported by the wave. This illustrates that large-amplitude (often called “finite-amplitude”) waves can transport mass. A wave that can transport mass can also transport any intensive property that is “attached” to the mass, such as potential temperature or water vapor mixing ratio. If a wave is defined, as above, as a process that transports energy but not mass, then beyond some limit a large-amplitude

1. This definition was given to me by Jule Charney. I have never seen it written down anywhere.
Planetary-scale waves and other eddies

An Introduction to the General Circulation of the Atmosphere

wave is not a wave at all. Note, however, that finite-amplitude waves tend to “break,” i.e., they disintegrate into turbulence.

We can use the world “eddy” as a generic term to denote zonal inhomogeneities that may or may not be waves.

As discussed earlier, it is useful to 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 hidden in time averages.

Eddies appear in the winds, temperatures, geopotential heights, surface pressure, water vapor mixing ratio, and all other fields that characterize the circulation. They are important aspects of the circulation in their own right, and in addition they are of interest because they can affect the zonally averaged flow by producing fluxes and flux convergences. Such feedbacks of the eddies on the zonally averaged flow are the subject of the next chapter.

As mentioned earlier, atmospheric waves come in a variety of physically distinct types. Rossby waves, which arise from conservation of potential vorticity on the sphere, are the type most properly referred to as “planetary waves,” because they depend for their existence on rotation in the presence of spherical geometry, and so are characteristic of

Figure 8.1: 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.
planetary atmospheres, although of course they also occur in oceans, and in stars. Rossby waves propagate westward relative to the mean flow, so that it is possible for them to 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 stratospheric sudden warmings, as discussed in the next chapter.

Gravity waves arise through the action of the buoyancy force under stable stratification. They occur on many scales. When their scale is larger than the radius of deformation, they are strongly influenced by rotation, and are called “gravity-inertia waves.” They can be produced by many mechanisms, including topographic forcing and convection. Vertically propagating gravity waves are thought to produce important vertical momentum transports that strongly affect the large-scale circulation, especially in the stratosphere and above. Today, the role of gravity waves in the general circulation is a “hot” topic.

The tropical atmosphere is home to two special classes of equatorially trapped waves. These are the “mixed Rossby-gravity waves,” also called Yanai waves, which propagate westward; and Kelvin waves, which propagate eastward. As described later, these two types of waves have been implicated in the physical mechanism that drives the Quasi-Biennial Oscillation, although gravity waves are now believed to be quite important. Kelvin waves are also believed to play important roles in El Niño and in the Madden-Julian Oscillation.

The purpose of the present 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.

8.2 Free and forced small-amplitude oscillations of a thin spherical atmosphere

8.2.1 Perturbation equations

Laplace (originally published in French in 1799; English translation in 1832) was the first to investigate the free and forced oscillations of a thin atmosphere on a spherical planet. His 200-year-old paper is still very relevant today. Here we briefly outline his work, omitting the mathematical details. A handout giving those details and discussing the applications of Laplace’s work to the problem of atmospheric tides is available on request.

Laplace considered a spherical planet without mountains, and with a highly idealized basic state:

\[
\bar{\nabla} = 0, \bar{\omega} = 0, \frac{\partial \bar{\phi}}{\partial p} = -\bar{\alpha}, p\bar{\alpha} = R\bar{T}(p), p_S = p_0 = \text{constant} .
\]  

(8.1)

Here \( T(p) \) is an arbitrary function of \( p \). Note that \( T \) does not depend on latitude. This basic state has no meridional temperature gradient and no mean flow. It is, of course, in balance.

The linearized governing equations are

\[
\frac{\partial u'}{\partial t} = (2\Omega \sin \phi)v' - \frac{1}{a \cos \phi} \frac{\partial \phi'}{\partial \lambda} ,
\]

(8.2)
where

\[ S_p = -\frac{\alpha \partial \Phi}{\Phi \partial p} \]  

is the static stability, which depends only on \( p \), and \( Q \) is the heating. Friction has been neglected in the momentum equations. Also,

\[ \phi' = g z' + \Phi(\lambda, \phi, t) \]  

where \( \Phi(\lambda, \phi, t) \) is the external gravitational tidal potential, due to the moon and/or sun. In (8.6), we recognize that the atmosphere experiences gravitational accelerations due to the pulls of the moon and sun, in addition to that of the Earth. The variation of \( \Phi \) with \( p \) is negligible, because the atmosphere is thin compared to the distances to the sun and moon. Note that these equations are valid only for atmospheres that are shallow compared to the planetary radius, \( a \).

We look for separable solutions of the form

\[
\begin{bmatrix}
u' \\
v' \\
\omega' \\
\phi' \\
Q \\
\Phi
\end{bmatrix}
= \sum_n \begin{bmatrix}
U_n^{\sigma, s}(p) \\
V_n^{\sigma, s}(p) \\
W_n^{\sigma, s}(p) \\
Z_n^{\sigma, s}(p) \\
J_n^{\sigma, s}(p) \\
G_n^{\sigma, s}(p)
\end{bmatrix} \Theta_n^{\sigma, s}(\phi) \exp[i(s \lambda + \sigma t)] \text{phase},
\]

where the \( \Theta_n^{\sigma, s}(\phi) \) are as-yet-undetermined functions of latitude only, and
8.2 Free and forced small-amplitude oscillations of a thin spherical atmosphere

\( s = \text{zonal wave number} = 0, 1, 2 \ldots \)
\( \sigma = \text{frequency}, \sigma < 0 \rightarrow \text{eastward moving} \)
\( \sigma > 0 \rightarrow \text{westward moving} \).

The superscripts \((\sigma, s)\) simply denote the particular frequency and zonal wave number associated with each mode. The subscript \(n\) is introduced to recognize the possibility of multiple solutions, and the summation over \(n\) represents a superposition of these solutions. It can be shown that the set \(\{\Theta_n^{\sigma, s}(\varphi)\}\) for all \(n\) is complete for \(\frac{-\pi}{2} \leq \varphi \leq \frac{\pi}{2}\). At this point, we do not know what meridional structures are represented by the \(\Theta_n^{\sigma, s}(\varphi)\).

After several pages of manipulation, we can derive the following two equations:

\[
F(\Theta_n^{\sigma, s}) = -\varepsilon_n \Theta_n^{\sigma, s}, \quad (8.9)
\]

\[
\frac{d^2 W_n^{\sigma, s}}{dp^2} + \frac{S_p}{gh_n} W_n^{\sigma, s} = -\frac{R}{gh_n c_p} \left( \frac{f_n^{\sigma, s}}{p} \right). \quad (8.10)
\]

Here \(F\) is a linear operator:

\[
F \equiv \frac{d}{d\mu} \left( \frac{1 - \mu^2}{\sqrt{\nu^2 - \mu^2}} \frac{d}{d\mu} \right) - \frac{1}{\nu (\nu^2 - \mu^2)} \left( \frac{\mu}{\nu (\nu^2 - \mu^2)} + \frac{\mu^2}{\nu} \right); \quad (8.11)
\]

\(\nu \equiv \frac{\sigma}{2\Omega}\) is the normalized frequency; and \(\mu \equiv \sin \varphi\), so that \(d\mu \equiv \cos \varphi \, d\varphi\). We have introduced the nondimensional quantity

\[
\varepsilon_n \equiv \frac{4\Omega^2 a^2}{gh_n}. \quad (8.12)
\]

The quantity \(h_n\), which appears in (8.10) and (8.12), is a “separation constant,” because it arises during the separation of variables. It is called the “equivalent depth,” for reasons that will become clear later.

Not surprisingly, Eq. (8.9) is called the meridional structure equation, and Eq. (8.10) is called the vertical structure equation. Recall that we have derived these equations using the assumptions that the basic state is at rest, and that the temperature depends on pressure (i.e. height) only. Separation of variables is not possible if the basic state is made more realistic, e.g., if the observed zonally averaged temperature and winds are used.
Eq. (8.9) was derived by Laplace about 200 years ago. It is often called the Laplace Tidal Equation, or LTE. It is a second-order ordinary differential equation, and so two boundary conditions are needed. It suffices to assume that the $\Theta_n^{\sigma,s}$ are bounded at the poles, i.e. at $\mu = -1$ and 1. Note that (8.9) and its boundary conditions are satisfied quite nicely by the trivial solution $\Theta_n^{\sigma,s} \equiv 0$. Non-trivial solutions do exist, but only for particular choices of the parameters $\nu$ and/or $h_n$ (or $\varepsilon_n$). If these parameters are chosen “at random,” the only solution of (8.9) that satisfies the boundary conditions is the trivial solution $\Theta_n^{\sigma,s} \equiv 0$. A problem of this type is called an “eigenvalue” problem. The frequencies and/or equivalent depths that allow non-trivial solutions are the eigenvalues, and the $\Theta_n^{\sigma,s}$ are the eigenfunctions or eigenvectors, which (for this particular problem) are called Hough functions.

All information about the planetary radius, rotation rate, and gravity is “buried” in the parameters $\varepsilon_n$ and $\nu$. The parameter $\varepsilon_n$ is sometimes called by the imposing name “the terrestrial constant.” Because it contains only two non-dimensional parameters characterizing the planet, the LTE does not “know” or “care” very much about the particular planet to which it is being applied. For given $\nu$ and $\varepsilon_n$, the eigenvalues and eigenfunctions of (8.9) are the same for all planets, provided that the atmosphere in question is shallow compared to the planetary radius. This means that the solutions of (8.9) have a very broad applicability.

The vertical structure equation, (8.10), is a second-order ordinary differential equation for $W_n^{\sigma,s}(p)$. For now we regard $J_n^{\sigma,s}$ as known, so that (8.10) contains the single unknown $W_n^{\sigma,s}$. The assumption that $J_n^{\sigma,s}$ is known means that $J_n^{\sigma,s}$ is at least approximately independent of the motion. This assumption would be reasonable for heating due to absorption of solar radiation by ozone, but it would be completely inappropriate for cumulus heating, for example. At the top of the atmosphere we apply the boundary condition

$$W_n^{\sigma,s} = 0 \text{ at } p = 0.$$  \hspace{1cm} (8.13)

This is exact. The exact lower boundary condition (in the absence of mountains) is $$w \equiv \frac{Dz}{Dt} = 0 \text{ at } p = p_S(\lambda, \varphi, t).$$ We apply the linearized boundary condition

$$\frac{Dz}{Dt} \equiv \left( \frac{\partial z'}{\partial t} \right)_p + \omega \frac{\partial z'}{\partial p} = 0 \text{ at } p = p_0. \hspace{1cm} (8.14)$$

Here $p_0$ is the spatially and temporally constant value of $p_S$ in the basic state. Because

$$gz' = \phi' - \Phi,$$  \hspace{1cm} (8.15)

where $\Phi(\lambda, \varphi, t)$ is known, and using the hydrostaticity of the basic state, as expressed by
we can rewrite the linearized lower boundary condition (8.14) as
\[
\frac{\partial \phi'}{\partial t} - \omega' g \frac{H_0}{p_0} = \frac{\partial \Phi}{\partial t} \text{ at } p = p_0. \tag{8.17}
\]
This involves both $\phi'$ and $\omega'$. After some additional algebra to eliminate $\phi'$ in (8.17), we can finally express the lower boundary condition entirely in terms of $W_n^{\sigma, s}$, as
\[
\frac{dW_n^{\sigma, s}}{dp} - \frac{H_0 W_n^{\sigma, s}}{h_n p_0} = \frac{i\sigma}{gh_n} G_n^{\sigma, s} \text{ at } p = p_0. \tag{8.18}
\]
Note that the gravitational forcing enters the problem through the lower boundary condition on the vertical structure equation. The thermal forcing enters through the vertical structure equation itself. The gravitational and thermal forcings do not appear in the LTE.

**8.2.2 Free oscillations of the first and second kinds**

A free oscillation is one for which there is no thermal or gravitational forcing. When there is no thermal forcing the vertical structure equation (8.10) reduces to
\[
\frac{d^2 W}{dp^2} + \frac{S_p}{gh} W = 0. \tag{8.19}
\]
Here the superscripts $(\sigma, s)$ and the subscript $n$ have been dropped for simplicity. When there is no gravitation forcing the surface boundary condition (8.18) can be simplified to
\[
\frac{dW}{dp} - \frac{H_0 W}{h p_0} = 0 \text{ at } p = p_0. \tag{8.20}
\]
We also have
\[
W = 0 \text{ at } p = 0. \tag{8.21}
\]
The system (8.19) - (8.21) has non-trivial solutions only for special values of $h$. These eigenvalues are denoted by $\hat{h}$. For $h \neq \hat{h}$, only the trivial solution [i.e. $W(p) \equiv 0$] exists. In order to find the $\hat{h}$ and the corresponding solutions for $W(p)$, we have to specify the static stability $S_p$ as a function of height. Different choices for $S_p$ will give different $\hat{h}$ and $W(p)$.
Planetary-scale waves and other eddies

An Introduction to the General Circulation of the Atmosphere

For the case of free oscillations, the solution procedure is summarized in Fig. 8.2. Note that in this case we have two eigenvalue problems: One from the vertical structure equation, and a second from the LTE. In the vertical structure problem, the eigenvalues are the equivalent depths. In the meridional structure problem, the eigenvalues are the frequencies.

As a very simple example, suppose that \( S_p = 0 \). Then we find from (8.19) that

\[
\frac{d^2 W}{dp^2} = 0. \tag{8.22}
\]

A solution of (8.22) that is consistent with the upper boundary condition (8.21) is

\[
W = Ap, \tag{8.23}
\]

where \( A \) is an arbitrary constant. Use of (8.23) in the lower boundary condition (8.20) gives

\[
\hat{h} = H_0. \tag{8.24}
\]

This is the only possible equivalent depth for free oscillations of an isentropic atmosphere. With more general stratifications there can be many (infinitely many) equivalent depths. As mentioned earlier, this is why we need the subscript \( n \), i.e. we use \( n \) to denote a particular solution. The procedure used to find the equivalent depth in this simple example can also be used for other stratifications.

With \( \hat{h} \) given by (8.24), nontrivial solutions of (8.9) exist only when there is a special relation (called the dispersion relation) among \( v \), \( s \), and \( n \). We refer to \( n \) as the “wave type.” The Hough functions (i.e., the solutions of the LTE) have been tabulated by Longuet-Higgins (1968) and others. Here we consider only some limiting cases. First suppose that there is no rotation, so that \( v = \frac{\sigma}{2\Omega} \to \infty \). We continue to assume that \( S_p = 0 \) so that (8.24) applies. For this case, find that

---

An Introduction to the General Circulation of the Atmosphere
8.2 Free and forced small-amplitude oscillations of a thin spherical atmosphere

\[ \nabla^2 F \rightarrow \frac{d}{d \mu} \left[ (1 - \mu^2) \frac{d}{d \mu} \right] - \frac{s^2}{1 - \mu^2}, \quad (8.25) \]

and

\[ \nabla^2 \varepsilon \rightarrow \frac{\sigma^2 a^2}{g \hat{h}} = \frac{\sigma^2 a^2}{g H_0}. \quad (8.26) \]

Then the LTE reduces to

\[ \frac{d}{d \mu} \left[ (1 - \mu^2) \frac{d \Theta}{d \mu} \right] + \left( \frac{\sigma^2 a^2}{g H_0} - \frac{s^2}{1 - \mu^2} \right) \Theta = 0. \quad (8.27) \]

It can be shown that (8.27) has solutions that are bounded as \( \mu \rightarrow \pm 1 \) only for

\[ \frac{\sigma^2 a^2}{g H_0} = n(n + 1), \quad n = 1, 2, 3 \ldots \quad (8.28) \]

These frequencies that satisfy (8.28) are eigenvalues. The eigenfunctions, i.e., the nontrivial solutions of (8.27), are called associated Legendre functions of order \( n \) and rank \( s \), denoted by

\[ \Theta_n = P_n^s(\mu), \quad n \geq s. \quad (8.29) \]

Note that \( n \) and \( s \) are both integers such that \( n \geq s \). The \( P_n^s \) are discussed in the Appendix. When (8.29) is combined with the longitudinal structure shown in (8.7), we find that the horizontal structure of the waves is given by

\[ Y_n^s(\mu, \lambda) = P_n^s(\mu) \exp(is \lambda). \quad (8.30) \]

These are the spherical harmonics (see the Appendix on this topic). Here \( n \) is the total number of nodal circles, \( s \) is the zonal wave number, and \( n - s \) is the number of nodes in the meridional direction, also known as the “meridional nodal number.”

The solutions found here are external gravity waves. They are called “external” because they have no nodes in the vertical and they do not propagate vertically. A stratified (i.e., non-isentropic) atmosphere can support both external and internal gravity waves.

The frequencies can be written as

\[ \sigma = \pm \sqrt{\frac{n(n + 1)gH_0}{a}}; \quad (8.31) \]

An Introduction to the General Circulation of the Atmosphere
they depend on the wave’s horizontal scale through the two-dimensional index, \( n \), but they are independent of \( s \). For example, when \( n = 1 \), \( s \) can be either 0 or 1 (because \( n \geq s \) and \( s \geq 0 \), but both modes have the same frequency. This is not true when rotation is present, because with rotation the zonal direction (in which scale is measured by \( s \)) becomes physically “different” from the meridional direction. A non-isentropic atmosphere can support both external and internal gravity waves.

Now consider \( \Omega \neq 0 \), still for an isentropic atmosphere, and neglect all details. Define a stream function \( \psi \) and a velocity potential \( \chi \) so that

\[
\begin{align*}
    u' &= -\frac{1}{a} \left( \frac{\partial \psi}{\partial \phi} \right)_p + \frac{1}{a \cos \phi} \left( \frac{\partial \chi}{\partial \lambda} \right)_p, \\
    v' &= \frac{1}{a \cos \phi} \left( \frac{\partial \psi}{\partial \lambda} \right)_p + \frac{1}{a \partial \phi}.
\end{align*}
\]

(8.32)

The vorticity is then \( \xi_p = k \cdot (\nabla_p \times V_h) = \nabla^2_p \psi \), and the divergence is \( \delta_p = \nabla_p \cdot V_h = \nabla^2_p \chi \). The equation of horizontal motion leads to

\[
\frac{\partial}{\partial t} \nabla^2_p \psi + \left( \frac{2 \Omega \cos \phi}{a} \right) v' + 2 \Omega \sin \phi (\nabla^2_p \chi) = 0
\]

(8.33)

(the vorticity equation) and

\[
\frac{\partial}{\partial t} \nabla^2_p \chi + \left( \frac{2 \Omega \cos \phi}{a} \right) u' - 2 \Omega \sin \phi (\nabla^2_p \psi) = -g \nabla^2_p z'
\]

(8.34)

(the divergence equation). We can also show (see the problems at the end of this chapter) that for an isentropic atmosphere

\[
\frac{\partial z'}{\partial t} + H_0 \nabla^2_p \chi = 0.
\]

(8.35)

Equations (8.33) through (8.35) form a closed set that can be solved for \( \psi, \chi, \) and \( z' \).

From (8.33) and (8.35) we see that when \( \Omega \neq 0 \) stationary motion cannot exist unless \( v' = 0 \). For nontrivial stationary motion with \( v' = 0 \) it follows from (8.2) that \( s = 0 \), i.e. the motion must be both purely zonal and zonally uniform.

Margules (1893) and Hough (1898) showed that the LTE has two classes of solutions, which they named Free Oscillations of the First and Second Classes. For the case of \( \epsilon_n \equiv \frac{4 \Omega^2 a^2}{g h_n} \) small (weak rotation), we can obtain approximate solutions of (8.33) and (8.34)
by expanding in spherical harmonics (see Longuet-Higgins, 1968). These Free Oscillations of
the First Class (FOFC) are essentially gravity waves, satisfying
\[
\begin{align*}
\chi &\equiv A_n \psi_n(\mu) e^{i(\lambda + \sigma t)} \\
\psi &\equiv 0 \quad \text{(irrotational)} \\
\sigma^2 &\equiv \frac{gH_0}{a^2} n(n+1)
\end{align*}
\tag{8.36}
\]

Compare with (8.31), obtained for \( \Omega = 0 \). Haurwitz (1937) obtained a more accurate
expression for the frequency of the FOFC:
\[
\sigma \equiv \frac{\Omega s}{n(n+1)} \pm \sqrt{\frac{\Omega^2 s^2}{n^2(n+1)^2} + n(n+1) \frac{gH_0}{a^2}}.
\tag{8.37}
\]

This should be compared with (8.31). For large \( n \), (8.37) reduces to (8.31). The additional
terms in (8.37) involve \( \Omega \). For \( n \geq 4 \) the error in (8.37) is less than 1%. From (8.37) we see
that eastward propagating inertia gravity waves have frequencies slightly different from those
of westward propagating inertia gravity waves. The difference is due to rotation.

The Free Oscillations of the Second Kind (FOSC) are the so-called Rossby-Haurwitz
waves, which satisfy
\[
\begin{align*}
\psi &\equiv i B_n^s \psi_n(\mu) e^{i(\lambda + \sigma t)} \\
\chi &\equiv 0 \quad \text{(nondivergent)} \\
\sigma &\equiv -\frac{2 \Omega s}{n(n+1)} > 0
\end{align*}
\tag{8.38}
\]

Note that, since \( \sigma > 0 \), the FOSC always move westward. They are nearly nondivergent. They
can be found by assuming \( \chi = 0 \) from the beginning, as follows: For \( x = 0 \), Eq. (8.33)
reduces to
\[
\frac{\partial}{\partial t} \nabla^2 \psi + \frac{2 \Omega}{a^2} \frac{\partial \psi}{\partial \lambda} = 0.
\tag{8.39}
\]

Therefore
\[
i s \left\{ \frac{1}{a^2 \cos^2 \phi} \left[ \cos \phi \frac{d}{d\phi} \left( \cos \phi \frac{d\psi}{d\phi} - s^2 \psi \right) \right] + \frac{2 \Omega i s}{a^2} \psi \right\} = 0,
\tag{8.40}
\]
or

\[
\frac{d}{d\mu} \left[ (1 - \mu^2) \frac{d\psi}{d\mu} \right] + \left( \frac{2\Omega s}{\sigma} - \frac{s^2}{1 - \mu^2} \right) \psi = 0 .
\]  

(8.41)

This is another eigenvalue problem. The solution of (8.41) is \( \psi = B_n^p P_n^p(\mu) \). Note that for Rossby-Haurwitz waves, in contrast to pure gravity waves, \( \sigma \) does depend explicitly on \( s \).

The westward propagation of Rossby-Haurwitz waves is due to the Earth’s sphericity. To see this, rewrite (8.39) as

\[
\frac{\partial \xi}{\partial t} < 0 \quad \frac{\partial \xi}{\partial t} > 0
\]

Figure 8.3: Chain of vortices along a latitude circle, illustrating the westward propagation of Rossby waves.

\[
\frac{\partial \xi}{\partial t} + \beta v = 0 ,
\]  

(8.42)

where

\[
\beta \equiv \frac{2\Omega \cos \varphi}{a} = \frac{1}{a\partial \varphi} \frac{df}{d\phi} .
\]  

(8.43)

Note that \( \beta \geq 0 \). Consider a chain of vortices along a latitude circle, as shown in Fig. 8.3. In places where \( v > 0 \), \( \beta v > 0 \), so \( \frac{\partial \xi}{\partial t} < 0 \). This occurs to the west of the place where \( \xi < 0 \).

Similarly, where \( v < 0 \), \( \beta v < 0 \), so \( \frac{\partial \xi}{\partial t} > 0 \). This occurs to the west of the place where \( \xi > 0 \).

A direct test of the theory of non-divergent Rossby waves was made by Eliassen and Machenhauer (1965) and Deland (1965). They performed a spherical-harmonic analysis of the 500 mb stream function, isolating transient waves by taking the difference in 24 hours. Their results, illustrated in Fig. 8.4, show westward propagation. Table 8.1 compares the computed and observed phase speeds, in degrees of longitude per day. The model overpredicts the westward phase speeds. This error is due to our neglect of the effects of divergence. Table 8.2 gives examples of the periods of the FOFC and FOSC.
8.3 Observations of stationary and transient eddies in middle latitudes

Blackmon (1976) discussed the observed eddy activity in the Northern Hemisphere, as

An Introduction to the General Circulation of the Atmosphere
Planetary-scale waves and other eddies

Blackmon filtered the data in both space and time, in order to isolate particular space-time scales. He expanded the height fields into spherical harmonics $Y_n^m$ (see the Appendix on spherical harmonics), where superscript $m$ denotes the zonal wave number, and subscript $n$ seen in the 500 mb 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.

<table>
<thead>
<tr>
<th>(n)</th>
<th>(s)</th>
<th>First Class To West, hours</th>
<th>First Class To East, hours</th>
<th>Second Class Only to West, days</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>13.76</td>
<td>39.34</td>
<td>1.21</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>10.92</td>
<td>14.50</td>
<td>5.37</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11.80</td>
<td>18.10</td>
<td>1.64</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>9.11</td>
<td>10.09</td>
<td>8.75</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9.23</td>
<td>10.96</td>
<td>3.87</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>9.63</td>
<td>12.17</td>
<td>2.10</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>7.65</td>
<td>7.99</td>
<td>12.78</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7.65</td>
<td>8.28</td>
<td>6.11</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7.75</td>
<td>8.66</td>
<td>3.79</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>7.94</td>
<td>9.18</td>
<td>2.56</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>6.52</td>
<td>6.67</td>
<td>17.75</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6.52</td>
<td>6.79</td>
<td>8.70</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6.52</td>
<td>6.95</td>
<td>5.62</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6.58</td>
<td>7.24</td>
<td>3.92</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6.69</td>
<td>7.49</td>
<td>3.05</td>
</tr>
</tbody>
</table>

Table 8.2: Periods of the free oscillations on the sphere, as computed from theory. Note that the periods of the gravity waves are given in hours, while those of the Rossby waves are given in days. From Phillips (1963).
denotes the “two dimensional index.” The number of nodes in the meridional direction, i.e. the “meridional nodal number” (try saying that five times, fast) is \( n - m \geq 0 \); note that \( m \leq n \) 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.

The spherical harmonics form a complete orthonormal basis that can be used to represent an arbitrary function on the sphere (see the Appendix). Using the spherical harmonics, the data can be expanded as follows:

\[
Z(\lambda, \varphi) = \sum_{n=-\infty}^{\infty} \sum_{m=-n}^{n} C_n^m Y_n^m .
\]  

(8.44)

Here the \( C_n^m \) are the expansion coefficients. Blackmon defined three groups 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 can be determined for each observation time. 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 was used in Blackmon’s study.

The lower panel of Fig. 8.5 shows the time-averaged 500 mb 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 2, of course.
The upper panel of Fig. 8.5 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. 8.6 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
Observations of stationary and transient eddies in middle latitudes

where blocking commonly occurs. Blocking will be discussed later.

Fig. 8.7 is similar to Fig. 8.6, but for medium-pass waves, 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. 8.6, so that the total amount of activity indicated in Fig. 8.7 is less.

Figure 8.6: Maps of the low-pass filtered rms fields (winter): (a) all waves, contour interval 10 m; (b) waves in Regime I, contour interval 5 m; (c) waves in Regime II, contour interval 5 m; (d) waves in Regime III, contour interval 5 m. From Blackmon (1976).
Even less power resides in high-pass waves (not shown).

Table 8.3 shows that in winter the greatest power is found along a “ridge” running from lower left (small $n$, low frequency) to upper right (large $n$, high frequency). By far the largest power occurs in the low frequencies and for two-dimensional index $n$ on the order of
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>51.7</td>
<td>49.8</td>
<td>45.8</td>
<td>39.3</td>
<td>37.0</td>
<td>26.7</td>
<td>21.2</td>
<td>17.3</td>
<td>13.7</td>
<td>10.1</td>
<td>7.2</td>
<td>5.3</td>
<td>4.1</td>
<td>3.4</td>
<td>3.2</td>
<td>3.4</td>
</tr>
<tr>
<td>17</td>
<td>68.7</td>
<td>66.5</td>
<td>58.6</td>
<td>48.1</td>
<td>40.1</td>
<td>34.0</td>
<td>27.9</td>
<td>22.1</td>
<td>15.8</td>
<td>11.0</td>
<td>7.6</td>
<td>5.4</td>
<td>4.3</td>
<td>3.1</td>
<td>3.2</td>
<td>3.3</td>
</tr>
<tr>
<td>16</td>
<td>98.5</td>
<td>97.5</td>
<td>82.4</td>
<td>69.6</td>
<td>56.5</td>
<td>45.0</td>
<td>34.8</td>
<td>26.0</td>
<td>18.4</td>
<td>12.2</td>
<td>7.9</td>
<td>5.5</td>
<td>4.1</td>
<td>3.3</td>
<td>3.3</td>
<td>3.5</td>
</tr>
<tr>
<td>15</td>
<td>141.6</td>
<td>131.0</td>
<td>108.1</td>
<td>85.9</td>
<td>70.3</td>
<td>58.6</td>
<td>75.6</td>
<td>32.3</td>
<td>19.7</td>
<td>11.6</td>
<td>7.3</td>
<td>5.1</td>
<td>4.1</td>
<td>3.3</td>
<td>3.3</td>
<td>3.4</td>
</tr>
<tr>
<td>14</td>
<td>210.1</td>
<td>195.5</td>
<td>158.9</td>
<td>122.3</td>
<td>99.5</td>
<td>79.2</td>
<td>53.3</td>
<td>32.3</td>
<td>18.2</td>
<td>10.8</td>
<td>6.9</td>
<td>4.9</td>
<td>3.5</td>
<td>3.3</td>
<td>3.3</td>
<td>3.4</td>
</tr>
<tr>
<td>13</td>
<td>302.8</td>
<td>272.5</td>
<td>213.2</td>
<td>164.7</td>
<td>129.0</td>
<td>92.6</td>
<td>56.1</td>
<td>31.6</td>
<td>16.8</td>
<td>9.4</td>
<td>5.7</td>
<td>4.4</td>
<td>3.2</td>
<td>2.9</td>
<td>2.9</td>
<td>3.3</td>
</tr>
<tr>
<td>12</td>
<td>502.6</td>
<td>435.2</td>
<td>311.1</td>
<td>225.5</td>
<td>159.9</td>
<td>99.8</td>
<td>53.5</td>
<td>26.8</td>
<td>13.6</td>
<td>8.0</td>
<td>5.3</td>
<td>3.9</td>
<td>3.2</td>
<td>3.0</td>
<td>3.0</td>
<td>3.4</td>
</tr>
<tr>
<td>11</td>
<td>746.3</td>
<td>639.7</td>
<td>443.3</td>
<td>290.6</td>
<td>176.5</td>
<td>94.4</td>
<td>43.8</td>
<td>21.1</td>
<td>10.6</td>
<td>6.0</td>
<td>4.1</td>
<td>3.6</td>
<td>3.0</td>
<td>2.8</td>
<td>3.0</td>
<td>3.4</td>
</tr>
<tr>
<td>10</td>
<td>1190.1</td>
<td>943.3</td>
<td>545.9</td>
<td>322.8</td>
<td>181.0</td>
<td>80.5</td>
<td>33.6</td>
<td>15.6</td>
<td>8.3</td>
<td>5.4</td>
<td>3.9</td>
<td>3.3</td>
<td>3.0</td>
<td>2.9</td>
<td>3.1</td>
<td>3.6</td>
</tr>
<tr>
<td>9</td>
<td>1815.2</td>
<td>1417.6</td>
<td>739.3</td>
<td>323.8</td>
<td>135.8</td>
<td>55.0</td>
<td>23.2</td>
<td>11.4</td>
<td>6.1</td>
<td>4.1</td>
<td>3.1</td>
<td>2.8</td>
<td>2.4</td>
<td>2.7</td>
<td>2.9</td>
<td>3.6</td>
</tr>
<tr>
<td>8</td>
<td>3305.7</td>
<td>2175.9</td>
<td>732.7</td>
<td>245.2</td>
<td>104.5</td>
<td>42.0</td>
<td>17.3</td>
<td>9.1</td>
<td>4.3</td>
<td>3.6</td>
<td>2.3</td>
<td>2.6</td>
<td>2.2</td>
<td>2.7</td>
<td>2.5</td>
<td>3.4</td>
</tr>
<tr>
<td>7</td>
<td>3732.6</td>
<td>2370.4</td>
<td>698.2</td>
<td>191.2</td>
<td>71.2</td>
<td>29.3</td>
<td>12.9</td>
<td>7.7</td>
<td>3.8</td>
<td>3.6</td>
<td>2.3</td>
<td>2.9</td>
<td>1.9</td>
<td>2.6</td>
<td>2.5</td>
<td>4.3</td>
</tr>
<tr>
<td>6</td>
<td>3731.5</td>
<td>2374.2</td>
<td>657.5</td>
<td>149.3</td>
<td>51.0</td>
<td>20.6</td>
<td>9.0</td>
<td>5.8</td>
<td>2.4</td>
<td>2.8</td>
<td>1.3</td>
<td>2.3</td>
<td>1.5</td>
<td>2.4</td>
<td>2.3</td>
<td>4.2</td>
</tr>
<tr>
<td>5</td>
<td>3563.5</td>
<td>2192.6</td>
<td>522.5</td>
<td>114.0</td>
<td>39.1</td>
<td>16.2</td>
<td>7.4</td>
<td>4.9</td>
<td>3.4</td>
<td>2.7</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.6</td>
<td>2.6</td>
<td>4.8</td>
</tr>
<tr>
<td>4</td>
<td>2553.1</td>
<td>1600.8</td>
<td>420.2</td>
<td>95.1</td>
<td>345.9</td>
<td>13.8</td>
<td>5.7</td>
<td>4.2</td>
<td>2.1</td>
<td>2.6</td>
<td>1.5</td>
<td>2.1</td>
<td>1.3</td>
<td>2.0</td>
<td>2.1</td>
<td>3.8</td>
</tr>
<tr>
<td>3</td>
<td>1448.2</td>
<td>992.6</td>
<td>329.2</td>
<td>79.7</td>
<td>27.0</td>
<td>9.4</td>
<td>4.6</td>
<td>3.3</td>
<td>2.1</td>
<td>1.8</td>
<td>1.4</td>
<td>1.5</td>
<td>1.2</td>
<td>1.5</td>
<td>1.5</td>
<td>4.6</td>
</tr>
<tr>
<td>2</td>
<td>1115.3</td>
<td>642.0</td>
<td>146.3</td>
<td>41.8</td>
<td>18.5</td>
<td>8.3</td>
<td>3.3</td>
<td>2.3</td>
<td>1.6</td>
<td>1.6</td>
<td>1.1</td>
<td>1.3</td>
<td>1.2</td>
<td>1.4</td>
<td>2.2</td>
<td>3.7</td>
</tr>
<tr>
<td>1</td>
<td>294.5</td>
<td>221.7</td>
<td>105.9</td>
<td>45.1</td>
<td>15.8</td>
<td>4.1</td>
<td>2.0</td>
<td>1.7</td>
<td>1.2</td>
<td>1.0</td>
<td>0.9</td>
<td>1.0</td>
<td>1.0</td>
<td>1.1</td>
<td>6.0</td>
<td>12.4</td>
</tr>
<tr>
<td>0</td>
<td>168.6</td>
<td>92.6</td>
<td>17.1</td>
<td>4.7</td>
<td>2.5</td>
<td>1.4</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Table 8.3: Power as a function of wave number and frequency for the winter season. Units: m² rad⁻¹ (15 days)⁻¹
6. The diurnal tide is also apparent near the bottom right, with \( n = 2 \).

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 the centers of action are shifted towards the pole. As in winter, the long waves and low frequencies dominate.

One conclusion from 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 total heat and momentum transport across a latitude circle as a function of wave number for selected latitudes. As shown in Fig. 8.8, by far the strongest contributions come from zonal wave numbers less than 10. For middle latitudes, this corresponds to wave lengths longer than 2000 km. The results show that the waves of largest spatial scale do most of the work of transporting energy and momentum poleward; the “synoptic-scale” and mesoscale eddies that we think of as “weather” play only a relatively minor role. They are fleas on the back of the general circulation.

8.4 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 orographic forcing (i.e., by blocking the flow) or by acting as elevated heat sources. Thermal forcing is also associated with land-sea contrasts, sea surface temperature gradients, and the like.

Held (1983) summarized the work of Charney and Eliassen (1949), who tried to understand 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 (see the Appendix on the shallow water equations), with the quasigeostrophic approximation, i.e.

\[
\left( \frac{\partial}{\partial t} + \mathbf{V}_g \cdot \nabla \right) q = -\left( \frac{r \zeta}{h - h_T} \right),
\]  

(8.45)

where

\[
q \equiv \frac{\zeta + f}{h - h_T}.
\]  

(8.46)

Here \( h \) is the height of the free surface, \([h]\) is the zonally averaged value of \( h \), \( h_T(x) \) is the height of the topography, assumed to vary in the \( x \)-direction only, \( h^* = h - [h] \), \( \mathbf{V}_g \) is the geostrophic wind, and \( r \) is a Rayleigh friction coefficient. See Fig. 8.9. We use the quasigeostrophic approximation, with

\[
\mathbf{V}_g = \frac{g}{f_0} \mathbf{k} \times \nabla h \quad \text{and} \quad \zeta = \frac{g}{f_0} \nabla^2 h.
\]  

(8.47)

We linearize about the zonally averaged state, which is assumed to be geostrophically balanced. Using (8.46) and (8.48), we can show that the zonal wind of the basic state satisfies
We assume that \([u]\) is independent of \(y\) as well as \(x\) and \(t\). The potential vorticity gradient of the zonally averaged flow is given by

\[
[u] = \frac{g}{f_0} \frac{\partial}{\partial y} [h].
\] (8.48)

In linearizing about the basic state given by (8.48) and (8.49), we assume not only that the perturbation vorticity and height are small, but also that \(h_T\) is small compared to \([h]\), so that we can neglect any products of \(h_T\) with a perturbation quantity. The linearized version of

---

**Figure 8.8:** Upper panel: The total eddy heat transport across a latitude circle as a function of wave number for selected latitudes, for January 1962. Lower panel: The eddy momentum transport averaged with respect to pressure, as a function of zonal wave number, for selected latitudes, based on observations for January 1962. In both panels the gray scale at the bottom denotes the zonal wave number. From Wiin-Nielsen et al. (1963).
Planetary-scale waves and other eddies

An Introduction to the General Circulation of the Atmosphere

(8.45) is then

\[
\left( \frac{\partial}{\partial t} + (u) \frac{\partial}{\partial x} \right)q^* + v^* \frac{\partial}{\partial y}[q] = \frac{r\zeta^*}{h}. \tag{8.50}
\]

where all perturbation quantities can be expressed in terms of \( h^* \), as follows:

\[
v^* = \frac{g}{f_0} \frac{\partial h^*}{\partial x}, \tag{8.51}
\]

\[
\zeta^* = \frac{g}{f_0} \nabla^2 h^*. \tag{8.52}
\]

\[
[h]q^* = \zeta^* - f_0 \frac{(h^* - h_T)}{[h]} \tag{8.53}
\]

By substitution from (8.49) and (8.51)-(8.53), we can rewrite the perturbation potential vorticity equation, (8.50), as
An Introduction to the General Circulation of the Atmosphere

8.4 Theory of orographically forced stationary waves

\[ \left( \frac{\partial}{\partial t} + [u] \frac{\partial}{\partial x} \right) \left\{ \frac{g}{f_0^2} \nabla^2 h^* - \frac{f_0}{h} \left( \frac{h^*}{h} - \frac{h_T}{h} \right) \right\} + \frac{g}{f_0^2} \frac{\partial h^*}{\partial x} \left\{ \beta + \frac{f_0}{g} \frac{\partial h_T}{h} \right\} = -r \frac{g}{f_0^2} \nabla^2 h^* , \quad (8.54) \]

or, after rearranging,

\[ \frac{\partial}{\partial t} \left( \frac{g}{f_0^2} \nabla^2 h^* - \frac{f_0}{h} h^* \right) + \frac{g}{f_0^2} [u] \frac{\partial}{\partial x} \left( \nabla^2 h^* + \frac{\beta g}{f_0^2} \frac{\partial h^*}{\partial x} + r \frac{g}{f_0^2} \nabla^2 h^* \right) = - \left[ \frac{u}{h} \frac{f_0}{\partial h_T}{\partial x} \right] - \beta \frac{g}{f_0^2} \frac{\partial h_T}{h} \quad (8.55) \]

Here some nice cancellation has occurred, and we have placed the “topographic forcing” term, involving \( \frac{\partial h_T}{\partial x} \), on the right-hand side, to set it apart. The waves described by (8.55) are “forced” by topography, which enters mathematically through the inhomogeneous term on the right-hand side, i.e. \(- \left[ \frac{u}{h} \frac{f_0}{\partial h_T}{\partial x} \right] - \beta \frac{g}{f_0^2} \frac{\partial h_T}{h} \). In the presence of such forcing, a non-zero \( h^* \) is demanded by (8.55).

Assume that the perturbations have the form

\[ h^* = \text{Re} \{ \hat{h} \exp[i(kx + ly - \omega t)] \} , \quad (8.56) \]

where \( \hat{h} \) is a constant, and also that the topography satisfies

\[ h_T = \text{Re} \{ \hat{h}_T \exp[i(kx + ly)] \} . \quad (8.57) \]

With (8.56), a positive value of \( \omega \) indicates eastward propagation. Note that we could choose \( l = 0 \), which would give us the special case of no meridional variations. Substitution of (8.56) and (8.57) into (8.55) gives:

\[ \{- \omega (K^2 + \lambda^2) + \beta [ |u| K^2 - \beta ] - \beta r K^2 \} \hat{h} e^{-i\omega t} = [u] \lambda^2 \hat{k} h_T . \quad (8.58) \]

Here \( K \) is the total wave number, which is defined by

\[ K^2 = k^2 + l^2 , \quad (8.59) \]

and \( \lambda \) is the radius of deformation, which is defined by

\[ \lambda^2 \equiv g[h]/f_0^2 . \quad (8.60) \]

Note that \( \lambda \) is defined in terms of the basic state. We see directly from (8.58) that the
amplitude of the waves, as measured by \( \hat{h} \), is proportional to the amplitude of the forcing, as measured by \( \hat{h}_T \). 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 (8.58) reduces to a dispersion formula, which can be written as

\[
-\omega (K^2 + \lambda^{-2}) + k([u]K^2 - \beta) - irK^2 = 0 ,
\]

or

\[
\omega = \frac{k([u]K^2 - \beta) - irK^2}{K^2 + \lambda^{-2}} .
\]

Eq. (8.62) describes 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

\[
\omega = \omega_0 - i(\tau_f)^{-1} ,
\]

where

\[
\omega_0 = \frac{k([u]K^2 - \beta)}{K^2 + \lambda^{-2}} ,
\]

and

\[
(\tau_f)^{-1} = \frac{rK^2}{K^2 + \lambda^{-2}} .
\]

Then (8.56) can be rewritten as

\[
h^* = e^{-t/\tau_f} \text{Re}\{\hat{h}\exp[i(kx + ly - \omega_0 t)]\} .
\]

The wave-like solution, with period \( \omega_0 \), decays with \( e \)-folding time \( \tau_f \).

A stationary wave is one for which \( \text{Re}\{\omega\} = 0 \). Under what conditions can a free Rossby wave be stationary? It is clear from (8.64) that a stationary free wave is possible only for \( [u] > 0 \). The reason is that the Rossby wave propagates toward the west relative to the mean flow, so the mean flow must be from the west to hold the wave steady relative to the Earth’s surface. In such a case, the total wave number of the stationary wave is
Here the subscript $S$ or $K_S$ stand for “stationary.”

With this preparation, we return now to the topographically forced case, and assume a stationary, neutral wave, i.e., $\omega = 0$. For the case of no friction ($r = 0$), we find from (8.58) that

$$\hat{h} = \frac{\hat{h}_T}{\lambda^2 (K^2 - K_S^2)}.$$  \hspace{1cm} (8.68)

Note that 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, the amplitude of the stationary neutral wave is

$$\hat{h} = \frac{\hat{h}_T}{\lambda^2 (K^2 - K_S^2 - i r K^2 / k[u])}.$$ \hspace{1cm} (8.69)

Eq. (8.69) shows that, mathematically, friction makes the amplitude complex. Clearly the denominator of (8.69) cannot become zero, so long as $r > 0$. For $K^2 = K_S^2$, (8.69) simplifies to

$$\hat{h} = \hat{h}_T \left( \frac{ik\beta}{\lambda^2 r} \right).$$  \hspace{1cm} (8.70)

Fig. 8.10 shows how the squared amplitude of the steady-state wave response varies as the zonal wind speed changes, for different values of $r$. As shown in Fig. 8.10, the Charney-Eliassen model, despite its extreme simplicity, can explain reasonably well the observed zonal structure of the 500 mb height field at $45^\circ$ 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 general circulation model.

Fig. 8.12 shows the observed time-latitude sections of the zonal and meridional contributions to the stationary wave kinetic energy, i.e. $[u^*]^2$ and $[v^*]^2$, 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 will be discussed later.
8.5 Tropical waves

Matsuno (1966) studied the linearized shallow water equations (see the Appendix on the shallow water equations) applied to an equatorial $\beta$-plane. When he began this work, his motivation was to investigate to what extent near-equatorial motions are geostrophic. In the end, however, he discovered two new classes of tropical waves, which soon after were recognized in the observations. These same waves are actually among the solutions found by Laplace, but this was not recognized until later. The model studied 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 is amazing, in view of the model’s extreme simplicity.

The shallow-water equations on an equatorial $\beta$-plane, linearized about a state of rest, are:

$$
\frac{\partial u}{\partial t} - fv + g \frac{\partial h}{\partial x} = 0,
$$

$$
\frac{\partial v}{\partial t} + fu + g \frac{\partial h}{\partial y} = 0,
$$

$$
\frac{\partial h}{\partial t} + H \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = 0.
$$

Here $f \equiv \beta v$, where $y$ is distance in the meridional direction, measured from $y = 0$ at the
Equator (i.e., $y = a\varphi$), and $\beta \equiv \frac{df}{dy}$ is approximated by a constant value. Of course, $f = 0$ at the Equator, so that $f = \beta y$. Matsuno defined a time scale, $T \equiv \frac{1}{\sqrt{c\beta}}$, and a length scale, $L \equiv \frac{c}{\sqrt{\beta}}$. Here $c \equiv \sqrt{gH}$ is the phase speed of a pure gravity wave. These definitions would have to be altered slightly if the Earth’s rotation were reversed. Why? See Fig. 8.14 for a sketch defining the other quantities. Fig. 8.15 shows how $T$ and $L$ vary as functions of $c$. For $c = 10$ m s$^{-1}$, we find that $L \equiv 1000$ km and $T \equiv 1$ day. Nondimensionalizing the governing equations by $T$ and $L$, we obtain

$$
\frac{\partial u}{\partial t} - yv + \frac{\partial \phi}{\partial x} = 0, \\
\frac{\partial v}{\partial t} + yu + \frac{\partial \phi}{\partial y} = 0, \\
\frac{\partial \phi}{\partial t} + \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0.
$$

Equation (8.72)
Here \( \phi \) 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), and so are more readily applicable to the real atmosphere than one might guess. For example, consider a two-level model:
Figure 8.13: A photograph of Prof. T. Matsuno, taken as he gave a lecture at UCLA in January 1998.

Figure 8.14: Model and coordinate system. From Matsuno (1966).

\[
\begin{align*}
\frac{\partial V_1}{\partial t} + f k \times V_1 + \nabla \phi_1 &= 0, \\
\frac{\partial V_3}{\partial t} + f k \times V_3 + \nabla \phi_3 &= 0, \\
\frac{\partial}{\partial t} (\phi_3 - \phi_1) + S \Delta p \omega_2 &= 0.
\end{align*}
\]

(8.73)
As shown in Fig. 8.16, subscript 1 denotes the upper level and subscript 3 denotes the lower level. The vertical velocity is defined in between, at level 2. Here $\Delta p = p_3 - p_1$ is the pressure thickness between the two layers, and $S \equiv \frac{\alpha \partial \theta}{\partial p}$ is the static stability of the basic state. Let

$$V_d = V_3 - V_1, \quad (8.74)$$

$$\phi_d = \phi_3 - \phi_1, \quad (8.75)$$

be the vertical shear (actually, difference) of the horizontal wind between the two layers, and
the thickness between the two layers, respectively. Then (8.73) implies that

\[
\frac{\partial V_d}{\partial t} + f k \times V_d + \nabla \phi_d = 0 ,
\]

(8.76)

and

\[
\frac{\partial \phi_d}{\partial t} + \frac{S \Delta p}{2} \nabla \cdot V_d = 0 ,
\]

(8.77)

which are identical to the shallow water equations, and we can identify

\[
c_i = \Delta p \sqrt{\frac{S}{2}}
\]

(8.78)

as the phase-speed of the internal gravity waves.

We return now to our discussion of (8.72). Assume solutions of the form \( e^{i(kx + \omega t)} \), with \( k \geq 0 \), so that \( \omega > 0 \) implies westward propagation. We can derive a single equation that governs the meridional structure of \( v \):

\[
\frac{d^2 v}{dy^2} + \left( \omega^2 - k^2 + \frac{k}{\omega} - y^2 \right)v = 0 .
\]

(8.79)

Here

\[
v \rightarrow v(y) ,
\]

(8.80)

where the “original \( v \)” is \( \hat{v} e^{i(kx + \omega t)} \). As boundary conditions, we use

\[
v \rightarrow 0 \text{ as } y \rightarrow \pm\infty .
\]

(8.81)

Nontrivial solutions satisfying these boundary conditions exist when

\[
\omega^2 - k^2 + \frac{k}{\omega} = 2n + 1 \text{ for } n = 0, 1, 2, \ldots .
\]

(8.82)

Note that the expression on the right-hand side of (8.82) generates all positive odd integers, so that Eq. (8.82) is equivalent to the statement that \( \omega^2 - k^2 + \frac{k}{\omega} \) is an odd positive integer. The solutions of (8.79) are given by

\[
v(y) = C e^{-\frac{1}{2}y^2} H_n(y) ,
\]

(8.83)
where \( H_n(y) \) is the \( n \)th Hermite polynomial, which is given by

\[
H_n(y) \equiv (-1)^n e^{y^2} \frac{d^n}{dy^n}(e^{-y^2})
\] (8.84)

(see the Appendix on Hermite polynomials). Because of the factor \( e^{-\frac{1}{2}y^2} \) in (8.83), these modes decay rapidly as we move away from the Equator. The e-folding distance is about 1000 km.

The dispersion equation, (8.82), is cubic in \( \omega \), and so there are three \( \omega \)'s for each \((k, n)\). Two of these correspond to inertia-gravity waves. For large \( k \), they can be approximated by

\[
\omega_{1,2} \equiv \pm \sqrt{k^2 + 2n + 1}.
\] (8.85)

These expressions can be compared with (8.37). The third root corresponds to a Rossby wave. For large \( k \), it can be approximated by

\[
\omega_3 \equiv \frac{k}{\sqrt{k^2 + 2n + 1}}.
\] (8.86)

For the special case \( n = 0 \), the dispersion equation (8.82) can be factored:

\[
(\omega - k)(\omega^2 + k\omega - 1) = 0.
\] (8.87)

Matsuno shows that for \( n = 0 \) the three roots can be interpreted as follows:

**Eastward gravity wave:** \( \omega_1 = -\frac{k}{2} - \frac{(\sqrt{k^2/2})^2}{\sqrt{k^2/2} + 1} \),

\[
(8.88)
\]

**Westward gravity wave:** \( \omega_2 = \begin{cases} 
\sqrt{\frac{k^2}{2}} + 1 - \frac{k}{2} & \text{for } k \leq \frac{1}{\sqrt{2}} \\
k & \text{for } k \geq \frac{1}{\sqrt{2}}
\end{cases} \),

\[
(8.89)
\]
Notice that for $n = 0$ the westward gravity wave and the Rossby wave are not really distinct. They coincide for $k = 1/\sqrt{2}$. Matsuno pointed out that the root $\omega = k$ has to be thrown out in (8.89) and (8.90). The reason is that in order to derive (8.79) it is necessary to assume that

\[
\omega = \frac{\omega y v + k \frac{d v}{d y}}{i(\omega - k)(\omega + k)} .
\]  

In view of (8.91), $\omega = -k$ is not permitted unless the numerator is identically zero for all $y$. Matsuno concluded, therefore, that for $n = 0$ we have only two waves: an eastward moving gravity wave, and a “mixed Rossby-gravity wave,” which is also known as the “Yanai wave.” 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

\[
\omega = \sqrt{\frac{(k/2)^2}{\sqrt{\frac{3}{2}}} + 1 - \frac{k}{2}}.
\]  

Because $H_0(y) = 1$, (8.83) reduces to

\[
v(y) = C e^{-\frac{1}{2}y^2},\]

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.

Another special case is the Kelvin wave. To find this mode, begin by putting $v \equiv 0$ everywhere in (8.72). Then (8.72) reduces to

\[
i\omega u + i k \phi = 0 ,
\]

\[
y u + \frac{\partial \phi}{\partial y} = 0 ,
\]  

\[
i\omega \phi + i k u = 0 .
\]

Nontrivial solutions exist only for

\[
(\omega - k)(\omega + k) = 0 ,
\]  

\[
\omega_3 = \begin{cases} 
 k & \text{for } k \leq \frac{1}{\sqrt{2}} \\
 \sqrt{\frac{(k/2)^2}{\sqrt{\frac{3}{2}}} + 1 - \frac{k}{2}} & \text{for } k \geq \frac{1}{\sqrt{2}}.
\end{cases}
\]
and are given by

\[
\begin{align*}
u &= \phi = Ce^{\frac{1}{2}y^2}, \quad \omega = -k \\
-u &= \phi = Ce^{\frac{1}{2}y^2}, \quad \omega = k.
\end{align*}
\]

(8.96)

The second solution does not satisfy our boundary condition as \( y \to \pm \infty \). The first solution is the Kelvin wave. The Kelvin wave can be interpreted as corresponding to \( n = -1 \), in the sense that for \( n = -1 \) (8.82) has \( \omega = -k \) as a root. We can accept \( \omega = -k \), because we have not divided by \( \omega + k \) in the course of deriving (8.96).

The various wave-solutions of Matsuno’s model are summarized in Fig. 8.17, which shows the roots of the dispersion equation. Recall that positive values of the frequency

![Figure 8.17: Frequencies as functions of wave number. Positive frequencies correspond to westward propagation. Thin solid line: eastward propagating inertia–gravity waves. Thin dashed line: westward propagating inertia–gravity waves. Thick solid line: Rossby (quasi–geostrophic) waves. Thick dashed line: The Kelvin wave. The westward moving wave with \( n=0 \) is the mixed Rossby–gravity or Yanai wave. It is denoted by a dashed line for \( k < 1/\sqrt{2} \), and by a solid line for \( k > 1/\sqrt{2} \). From Matsuno (1966)](image)

correspond to westward propagating waves, and negative values (lower part of the figure) to eastward propagating waves. The thick 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 thin 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 mixed Rossby-gravity wave, or 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.

The thick dashed line proceeding downwards towards the right from the origin represents the Kelvin wave.

For \( n = 0 \), the eastward moving inertia-gravity wave and westward moving Yanai wave have the structures shown in the upper and middle panels of Fig. 8.18. 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 \( n = 1 \) are shown on the left side of Fig. 8.19. 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 \( n \)th Hermite polynomial. As can be seen in the figure, higher values of \( n \) correspond to more nodes in the meridional direction.

The structure of the Kelvin wave is shown in Fig. 8.20. Note that the velocity vectors are purely zonal, and that the tendency of the zonal wind is in phase with the pressure, as in a
Planetary-scale waves and other eddies

There have been many observational studies of tropical waves. Maruyama and Yanai (1966) observed the mixed-Rossby-gravity wave shortly after Matsuno had predicted its existence\(^2\), and Wallace and Kousky (1968) found the Kelvin wave shortly thereafter.

\(^2\) Amazingly, Matsuno and Yanai were office mates at about this time.
8.6 The response of the tropical atmosphere to stationary heat sources and sinks

Wheeler and Kiladis (1999) examined the space-time variability of the tropical outgoing long wave radiation. 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 (Fig. 8.21 through Fig. 8.23).

8.6 The response of the tropical atmosphere to stationary heat sources and sinks

Fig. 8.24, which is taken from 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, as 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, as in the eastern equatorial Pacific.
Planetary-scale waves and other eddies

Figure 8.22: Longitudinal propagation of the Madden–Julian Oscillation (MJO; discussed later), 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 wave numbers and frequencies that fall in the corresponding boxes in Fig. 8.21. This is what is meant by “filtering.” From Wheeler and Kiladis (1999).
The response of the tropical atmosphere to stationary heat sources and sinks

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. The westerlies can be interpreted as a “monsoon-like” westerly inflow to a region of heating. Further discussion is given later.

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. Since much of the convective heating in the tropics is confined over three relatively small land regions (Africa, South America, and the Indonesian region), 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, 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.
Planetary-scale waves and other eddies

An Introduction to the General Circulation of the Atmosphere

Corresponding to (8.72), we have

\[
\begin{align*}
\epsilon u - y v + \frac{\partial \phi}{\partial x} &= 0, \\
\epsilon v + y u + \frac{\partial \phi}{\partial y} &= 0, \\
\epsilon \phi + \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= -Q;
\end{align*}
\]

and, as a purely diagnostic relation,

\[
w = \epsilon \phi + Q.
\]

The winds \( u \) and \( v \) represent the lower-tropospheric variables. Here \( \epsilon^{-1} \) is a dissipation time scale, and \( Q \) is a “heating rate” that must be specified. The variables \( \phi \), \( w \), and \( Q \) are defined in the middle troposphere. Gill included dissipation in the form of Rayleigh friction and Newtonian cooling, and for simplicity assumed that the time scales, given by \( \epsilon^{-1} \), are equal. Rayleigh friction is a simple parameterization of friction in which the velocity is divided by a frictional time scale.

Gill focused primarily on cases for which the heating is symmetric or anti-symmetric.

Figure 8.24: Stationary circulation pattern (lower panel) forced by the mass source and sink shown in the upper panel. From Matsuno (1966).
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 forming a vorticity equation for the case of no damping, and then substituting from the continuity equation, Gill found that

\[ v = yQ. \tag{8.99} \]

This is closely related to what is sometimes called “Sverdrup balance,” in which the “meridional advection of the Coriolis parameter,” i.e. the so-called \( \beta \) term of the vorticity equation, is balanced by the divergence term, which is represented on the right-hand side of (8.99). For a incompressible atmosphere with a rigid lid at \( z = D \) and a constant lapse rate, the gravest mode (the mode with the largest vertical scale) has horizontal velocity components that vary as \( \cos(\pi z/D) \), i.e., they pass through zero in the middle troposphere. This is similar to the observed vertical structure of the Hadley-Walker circulation.

For \( Q > 0 \), (8.99) 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 symmetric case as a simulation of the Walker circulation, and the asymmetric case as a simulation of the Hadley circulation.

For heating centered on the Equator, as in Fig. 8.25, 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 a Walker circulation, as indicated in Fig. 8.25, and a surface pressure field with a minimum pressure slightly to the west of the heating.

When the heating is antisymmetric across the Equator, as in Fig. 8.26, 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. 8.27, the model produces a circulation that looks remarkably similar to that of the Asian summer monsoon, as discussed in the next subsection.
Although Gill plausibly 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, 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. Second, the model includes neither a moisture budget nor cloud radiative effects. Last and most important, the model was linearized about a resting atmosphere. Although linearization is useful, Gill’s results must be interpreted as the response of the tropical atmosphere to a perturbation of a basic state, not as a prediction of the basic-state climate. Gill demonstrated the sensitivity of the tropical troposphere to the spatial distribution of heating, but his study does not really address the basic-state climate.

### 8.7 Monsoons

Monsoons occur in many parts of the world. They can be viewed as thermally forced stationary planetary waves associated with land-sea contrast. The most spectacular monsoon on Earth is the one associated with Asia, our largest continent.

The Asian monsoon can be defined in a number of ways. By definition, a monsoon is a dramatic seasonal reversal of the low-level prevailing winds (Lighthill and Pearce, 1981). 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. 8.28. 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
Monsoons 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. 8.29), which extends upward to about the 500 mb 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 mb temperature for the monsoon region is shown in Fig. 8.30. An island of warm air is centered over the Tibetan plateau.

After the non-permanent 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

\[\text{Figure 8.26: The response to antisymmetric heating. On the left side, the top panel shows contours of the mid-level vertical velocity superimposed on the horizontal wind vectors for the lower layer. The lower panel shows contours of the perturbation surface pressure, again with the lower-layer horizontal wind field superimposed. The right-hand panels show the zonally integrated solution corresponding to the results in the left-hand panels. The upper panel shows the latitude–height distributions of the zonal velocity and the stream function of the mean meridional circulation, as well as the meridional profile of the surface pressure. From Gill (1980).}\]
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. In some cases, the progression of the monsoon trough is expedited by the passage of a tropical or extratropical cyclone to the north (Mooley and Shukla, 1987). Cloudiness and precipitation begin to increase at the southern tip of India in late May. Nearly the entire country has begun to receive monsoon precipitation by the end of June.
Figure 8.28: Observed 850 mb wind vectors for a) January, and b) July.

Figure 8.30: Observed JJA climatological 500 mb temperatures. Contour interval is 2 K.
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. The data shows, as one might expect, that 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.

From an agricultural standpoint, the beginning of the precipitation associated with the Asian summer monsoon is probably one of the most anticipated events in the world. The onset of the monsoon is generally defined as the beginning of consistent rainfall of the monsoon season. Fig. 8.31 shows a plot of the dates of onset of the Asian monsoon from Krishnamurti et al. (1990). 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. 8.32 a. Many areas within the monsoon region also receive significant amounts of precipitation from tropical cyclones. Fig. 8.32 b 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 (Johnson and Houze, 1987).

Fig. 8.33 shows JJA means of the precipitation across the monsoon region, according to the climatological data set of Legates and Wilmott (1990). There are two major precipitation maxima. One is west of the southwest coast of India, and the other is 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 and western parts of the monsoon region are quite dry, receiving less than 2 mm day$^{-1}$ of rain. The observed 850 mb 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.
The precipitation in much of the monsoon region varies on many time scales. 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. 8.34. 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 by some (Webster, 1987) to the 40-50 day oscillation.
discovered by Madden and Julian (1972), which is discussed in detail later.

One of the most prominent signatures of the Asian monsoon is the monsoon trough, which extends 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.

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 mb vertical velocity for the monsoon region is shown in Fig. 8.35. 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 2 mm day$^{-1}$ of rain. The area mean vertical velocity is upward, at -10 mb day$^{-1}$.

The upper-level divergence is associated with a broad, strong anticyclonic circulation over the Plateau at 200 mb, as shown in Fig. 8.36. The strong, upper-level easterlies (15 - 20 m s$^{-1}$) 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, 1987; Yanai and Li, 1993). There is a slight northerly component to the 200 mb 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 30 m s$^{-1}$ at about 35°N. Fig. 8.37 shows a latitude-pressure cross-section of the observed zonal wind at 77.5° E.

From the above discussion, it is clear that we can interpret the monsoon as a zonally localized 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.

The left-hand column of Fig. 8.38 shows the hourly record of surface pressure and its Fourier spectrum at 10°N, 77.5°E, near the southern tip of India, at 26°N, 77.5°E, and at 26°N, 77.5°E in northern India. All three locations show a large peak at 12 hours, which corresponds to the semidiurnal tide. There is also a noticeable diurnal tide. The right-hand column of Fig. 8.38 shows similar observations for the same three locations, but for longer periods. Beyond 24 hours, all three locations show oscillations with time scales of a few days, which is the time scale of synoptic weather systems. There is also a spectral peak at 10-20 days, which may be associated with the period of time it takes for the monsoon trough to move from the southern tip of India to the Himalayan foothills and back again (Webster, 1987). The 10-20 day peak for 26° N is at about 18 days, while the peak at 10° N is only 13 days. The difference between the two periods may be due to the monsoon trough not making it as far north as 26° N during some of its north-south oscillations, which would cause the surface pressure dip to be less frequent there. There is also an oscillation with a period of 90 days that might be associated with the onset and departure of the monsoon itself.
Precipitation in the monsoon region has also been observed to vary systematically on time scales of 2-5 days, 10-20 days, and 40-50 days (Webster, 1987). Due to the convective nature of much of the precipitation that takes place in the Asian monsoon, there are also significant variations on time scales of only a few hours. For periods on the order of 10-20 days, the spectral analyses of precipitation are similar to those of surface pressure. This is understandable, because the heaviest precipitation is generally associated with the monsoon trough. There do not appear to be any semidiurnal or diurnal variations in precipitation, although there is a large amount of high-frequency spectral energy, as might be expected with convective rainfall. There are some indications of spectral peaks of precipitation at 25-35 days. These might be associated with the Madden-Julian oscillation.

There have been many efforts to understand the interannual variability that occurs with the Asian monsoon. The 1987 summer monsoon was remarkably dry over most of India (Krishnamurti et al., 1989) whereas the 1988 summer monsoon was quite wet in the same region (Krishnamurti et al., 1990). Charney and Shukla (1981) argued on the basis of modeling studies that the largest variations in monsoon rainfall and sea level pressure are due to changes in the boundary conditions. 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. 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 (La Niña) by some researchers (Shukla and Paolino, 1983; Rasmusson and Carpenter, 1983), but there is some dissension on this point (Webster and Yang, 1992). The interannual variability of monsoons has also been linked to variations in snow cover (Yanai and Li, 1994; Barnett et al., 1989) and the thermal contrast between the equatorial Pacific and the Tibetan Plateau (Fu and Fletcher, 1985).
8.8 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 mechanisms to be discussed later. For this reason, the Walker Circulation is best

Figure 8.38: Observed Fourier power spectra of surface pressure. a) at 10°N, 77.5°E, near the southern tip of India, for oscillations with a period of 0–5 days; b) at 18°N, 77.5°E, for oscillations with a period of 0–5 days; c) at 26°N, 77.5°E, for oscillations with a period of 0–5 days; d) at 10°N, 77.5°E, for oscillations with a period of 0–90 days; e) at 18°N, 77.5°E, for oscillations with a period of 0–90 days; and f) at 26°N, 77.5°E, for oscillations with a period of 0–90 days.
understood as a coupled ocean-atmosphere phenomenon. It undergoes strong interannual variability. Fig. 8.39 is a schematic illustration of the walker circulation and its relation to the surface wind field in the southern Hemisphere. 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. Fig. 8.40 shows the observed longitude-height cross sections of the zonal wind and vertical velocity, for January.

The Hadley Circulation is defined in terms of zonal averages, and so a particle participating in the Hadley Circulation through motions in the latitude-height plane cannot “escape” by moving to a different longitude. In contrast, the Walker Circulation is restricted to a narrow band of tropical latitudes, so that a particle participating in the Walker Circulation can escape by moving off to a different latitude; in fact, such meridional escapes are to be expected in view of the strong meridional motions associated with the Hadley Circulation. For this reason, we should not think of the Walker Circulation as a closed “race track;” it is better to view the Hadley and Walker Circulations as closely linked. For instance, 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. 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 which depends on the strength of the angular momentum flux divergence.

Fig. 8.41 shows that the 1000-mb winds above the tropical Pacific (between 10° N and 10° S) have an easterly component in both solstitial seasons. For both seasons, easterly flow near the equator occurs 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, a notable characteristic is the strong cross-equatorial flow in the eastern Pacific. The zone of convergence at 1000 mb during the NH summer has moved north of 10° N over the eastern Pacific. At the latitude of the ITCZ, the easterly fetch originates to the east of Central America during both seasons. If we consider the Walker circulation to occur at near-equatorial latitudes, then easterly flow at 1000 mb cannot originate over the continents because the mountains of Peru act as a vertical barrier on the eastern boundary of the ocean. However, we note that a strong southerly component is evident during both seasons.

Lindzen and Nigam (1987) used a simple model to show that SST gradients are capable of forcing low-level winds and convergence in the tropics. They assumed that near the surface the Coriolis acceleration is balanced by the sum of the horizontal pressure-gradient force and wind stress; this is called an Ekman balance. 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; hereafter N96) compared water-vapor data from the Upper Atmosphere Research Satellite (UARS) with upper-air wind data from the ECMWF
Planetary-scale waves and other eddies

An Introduction to the General Circulation of the Atmosphere

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. 8.42 shows the vertical velocity fields at 300-mb from the ECMWF reanalysis dataset for the solstitial months. During January 1989, centers of ascending motion were located near 145° E at latitudes 5° N and 5° S. The SPCZ is clearly evident in the January 1989 data, with a large region of ascending motion that extends southeastward from 145° E to 160° W. A region of strong sinking motion straddles the equator and extends eastward from 160° E. During July 1989, the ascending region remains fixed at 145° E, but the NH and SH centers of ascending motion have merged on the equator. During the NH summer, the ITCZ is well developed at 5° N, and so the zone of sinking motion has slipped southward from its January position, particularly the zone over the central Pacific. 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 and 160° E. West of 160° E, the low-level equatorial winds are very weak. However, easterlies span the equatorial Pacific at 5° S and 5° N. Fig. 8.43 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 (NH), weak westerly flow appears between 170° W and 140° W poleward of 15° N. In the southern hemisphere (SH), 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 SH. A reexamination of Fig. 8.42b indicates that sinking motion is confined mainly to the SH, and occurs as far west as 165° W.

The Walker Circulation is an atmospheric phenomenon, but it is closely tied to east-west sea surface temperature gradients that are produced by atmospheric phenomena including aspects of the Walker Circulation itself. The Walker Circulation can thus be viewed as a phenomenon of the coupled atmosphere-ocean system. Fig. 8.44 shows that a sea-surface temperature (SST) maximum occurs over the tropical region centered on 120° E, and for this reason the region is known as the tropical Warm Pool. The “cold tongue” is a band of relatively cold waters along the equator that stretches from South America westward to near 160° E. Although a noticeable SST gradient exists along and across the cold tongue, the temperature variation is still much smaller than that which is generally observed in extratropical or polar regions of the globe. The tropical climate is characterized by sea surface and horizontal air temperature gradients that are weak compared to the corresponding mid-latitude gradients. As explained by Charney (1963), for cloud-free regions of the tropics, pressure and temperature gradients must be small compared to those of midlatitudes.

The distribution of tropical convection is strongly related to both the local SST and the SST gradient. The tropical-Pacific Warm Pool is a region of intense deep convection. In
Planetary-scale waves and other eddies

An Introduction to the General Circulation of the Atmosphere

Fig. 8.45, 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 convection. From the figure, we see that convection occurs throughout the Warm Pool, and in the South Pacific convergence zone (SPCZ). On the other hand, the OLR is generally larger.

Figure 8.43: Streamlines and horizontal wind vectors for the tropical Pacific at 200 mb for a) January 1989, and b) July 1989.

Figure 8.44: Tropical skin temperature for January 1989 from the ECMWF reanalysis dataset obtained from NCAR. Resolution for this dataset is 2.5° and the contour interval is 2 K.
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 (1991) 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.

In the eastern tropical Pacific, stratus clouds in the boundary layer 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 cold-water upwelling in the eastern Pacific, and along the equator in the eastern and central Pacific. As discussed in Chapter 2, the equatorial cold tongue is due to upwelling driven by the trade winds.

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 exists, as discussed in Chapter 2.
Figure 8.46: Observed annual-mean low-cloud amount (upper panel) and the net effects of clouds on the Earth’s radiation budget (lower panel). Negative values in the lower panel indicate a cooling, i.e. shortwave reflection dominates longwave trapping.
Pierrehumbert (1995; hereafter P95) presented a two-box model of the Hadley/Walker circulation that has strongly influenced recent studies of the tropical climate. Fig. 8.47 presents a schematic of his “furnace/radiator-fin” model. The model has separate energy budgets for its Cold-Pool and Warm-Pool regions. 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 at $z = z_T/2$, where $z_T$ is the assumed height of the tropopause. 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 net radiative flux at the Warm-Pool TOA was assumed to be balanced by a horizontal energy transport to the Cold Pool. The Cold-Pool SST and radiating temperature were then computed under the constraint that the net diabatic cooling must balance the energy imported laterally from the Warm Pool.

![Hadley-Walker Circulation Diagram](image)

**Figure 8.47: Reproduction of Pierrehumbert’s schematic representation of the ‘furnace/radiator-fin’ model of the tropical circulation. The symbols E and TS represent the evaporation rate and SST, respectively. The subscripts 1 and 2 denote the Warm Pool or furnace and the Cold Pool or radiator fin.**

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 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 decreases...
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 the model is that it fails to account for cloud-radiative effects. Miller (1997) extended Pierrehumbert’s model by showing that low clouds in the Cold Pool act as a thermostat for the SST throughout the tropics.

8.9 The Madden–Julian Oscillation

The tropical atmosphere undergoes a powerful oscillation with a period in the range 30 - 60 days (Fig. 8.48). The oscillation appears most clearly over the Indian and western Pacific Oceans, and involves many meteorological variables, including the zonal wind, surface pressure, temperature, and humidity (Madden and Julian, 1971, 1972). It is referred to as the Madden-Julian Oscillation (MJO), or sometimes the Tropical Intraseasonal Oscillation.

Observations have revealed 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. 8.49. The observed phase propagation is shown in Fig. 8.50. The MJO does not produce 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.
Embedded within the MJO are fluctuations on smaller space and time scales (Nakazawa, 1988; Fig. 8.51).

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 (Gray, 1979), and the initiation of El Niño events (Lau and Chan, 1985). 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).

The MJO is a first baroclinic mode, equatorially-trapped, convectively-coupled disturbance that propagates at a phase speed of about 5 m s\(^{-1}\) as it travels from the Indian

---

*Figure 8.49: Schematic depiction of the time and space (zonal plan) variations of the disturbance associated with the 40–50–day 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 is 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). Taken from Madden and Julian (1972 a).*
An understanding of the MJO has been elusive. Most theories fall into one of the following three categories:

1) wave-CISK theories (e.g., Hayashi, 1970; Lindzen, 1974; Hayashi and Sumi, 1986; Lau and Peng, 1987; Salby et al., 1994),

2) surface evaporation feedback theories (Emanuel, 1987; Neelin and Yu, 1994; Yu and Neelin, 1994), called WISHE (wind-induced surface heat exchange) theories, and

3) local forcing or discharge-recharge theories (Blade and Hartmann, 1993; Hu and Randall, 1994; Salby and Garcia, 1997; Flatau et al., 1997).

Each category of theories has strengths and weaknesses.

The basic premise of wave-CISK theories is that the waves and the cumulus heating
An Introduction to the General Circulation of the Atmosphere

309

8.9 The Madden–Julian Oscillation

The Madden-Julian Oscillation sustain each other and propagate eastward together via their mutual interactions and feedbacks. In order to maintain a continuous progression of waves at any given longitude, the waves have to complete a global circuit without dissipating. A major problem with these theories is that they have difficulty explaining the observed phase speeds and vertical structures of the waves. The wave-CISK theory (Hayashi, 1970; Lindzen, 1974) predicts a vertical wavelength of 8 - 9 km for waves with a Doppler-shifted phase speed of 10-15 m s\(^{-1}\) (consistent with a global circuit in 40-50 days). This vertical wavelength is less than half that of the observed 30 - 60 day waves in the tropics (15 - 30 km, see Madden and Julian, 1972). In other words, a direct application of the classic wave-CISK theory predicts a phase speed twice as fast as observed, for waves of the observed depth.

Chang (1977) produced a new, “viscous” mode by introducing a damping term that raised the order of the dispersion equation by one. Although the viscous mode’s eastward phase speed is slow compared to the that of the classic wave-CISK modes, it has significant amplitude only near the heat source. Chang did not explain the mechanism that produces the oscillating heat source. Along similar lines, Lau and Peng (1987) proposed a theory in which negative heating is not permitted. Although their results show modes that realistically combine deep vertical scales with slow phase speeds, these are a consequence of their specified heating profile that has a maximum near the 700 mb level (see their Table 2). Chang and Lim (1988) argued that such a low-level heating maximum is necessary in order to generate slowly propagating modes. Wang (1988) suggested that interactions between internal and external modes can generate slow, eastward-propagating waves. Wang and Chen (1989) showed that the frictionally forced boundary-layer convergence of mass and moisture can induce such

Figure 8.51: 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).
interactions. When frictional moisture convergence is strong enough, the growth of low-frequency planetary waves is favored. It remains to be seen whether this mechanism can explain how a low-frequency oscillation is excited.

Pursuing a rather different idea, Neelin et al. (1987) and Emanuel (1987) hypothesized a feedback based on “wind-induced surface heat exchange” (WISHE) to explain the Madden-Julian oscillation. According to their theory, the heat source is maintained by both the waves through their interaction with the low-level flow. The maintenance and eastward propagation of convection in this theory depends on the existence of a mean easterly flow in the Indian and western Pacific Oceans. Such winds are not observed, however (e.g., Oort, 1983). In addition, observations show that the strongest evaporation in an MJO disturbance occurs on the west side of the convection, rather than the east side. WISHE does appear in some numerical model results, however. It is possible that WISHE accounts for the initial (i.e., small-amplitude) growth of the MJO, even though it does not appear to describe fully developed events.

The discharge-recharge idea has received some support in recent years. Hartmann and Gross (1988) analyzed a 22-year time series of tropical precipitation, and showed that the 30 - 60 day oscillation of precipitation occurs only at stations within or near the areas of intense convection in the Indian and western Pacific Oceans, e.g., the SPCZ. They found that the oscillations of precipitation lead those of the zonal wind by 5 - 6 days. This phase difference suggests that the observed propagating low-frequency waves in the zonal winds are forced by the convection, but not the reverse.

Hsu et al. (1990) reported a stationary, fluctuating, low-frequency latent heat source in the tropical western Pacific. Waves appear to emanate from the region of the oscillating heat source, and to propagate both eastward and poleward. The heating itself is relatively localized and does not appear to be symbiotically coupled with the propagating waves. Interestingly, Hsu et al. (1990) found that the eastward-propagating 30 - 60 day waves typically complete only about a half cycle before they dissipate. This appears to rule out the possibility that episodes of enhanced convection are triggered by the arrival, from the west, of eastward-propagating waves that were forced by their predecessors.

Holton (1972) showed that a specified oscillating tropospheric heat source excites waves in both the troposphere and the stratosphere; naturally, these waves have the same period as the heating. Their vertical scale is on the order of 20-30 km. For Kelvin waves, the zonal wavelength is linearly proportional to the vertical scale of the waves; deep modes have zonal wave number one. Studies by Yamagata and Hayashi (1984), Hayashi and Miyahara(1987), Salby and Garcia (1987), and Garcia and Salby (1987), among others, provide examples of the application of this forcing – response concept to the Madden-Julian oscillation. Salby and Garcia showed that the response of the tropospheric winds to a stationary 30-day heating oscillation in the tropical western Pacific Ocean closely resembles that of the observed Madden-Julian oscillation. They found that the period of the oscillation increases with the period of the heating. They further demonstrated that eastward-propagating waves excited in the lower troposphere in the Western Pacific are likely to succumb to dissipation before reaching the Americas. As shown by Salby and Garcia (1987) and Garcia and Salby (1987), a low-frequency tropical heat source excites waves that propagate slowly and have structures similar to the observed Madden-Julian waves. In summary, several studies have shown that slow eastward- and poleward-propagating waves with vertical scales comparable to that of the Madden-Julian oscillation can be explained quite naturally as responses to localized, low-frequency tropical heating.

Recharge of atmospheric moisture on MJO time scales has been reported in the
An Introduction to the General Circulation of the Atmosphere

8.10 Summary

Some authors have suggested that fluctuations of the SST are important for the MJO. Flatau et al. (1997) and Walliser et al. (1999) reported that simulated 30-60-day tropical variability increased when their atmospheric general circulation models (GCMs) were coupled to slab ocean models. They concluded that the increased variability at MJO times scales was linked to the response time of sea surface temperatures (SSTs) to cooling associated with the convective phase of the MJO. Wang and Xie (1998) simulated the coupled tropical system within a linear model framework and concluded that the ocean mixed layer sustains the MJO by destabilizing the moist Kelvin wave, providing a longwave selection mechanism, and slowing down the phase propagation to the expected 40-50-day period, although a physical mechanism by which this occurs was not clearly stated. Shinoda et al. (1998), through an analysis of observational data, also proposed that the ocean mixed layer controls (slows) the period of oscillation by reducing surface latent and sensible heat fluxes at MJO timescales below those that would be expected in the absence of MJO time scale variability in the upper layers of the ocean. In summary, discharge-recharge theories are attractive in that they offer a possible explanation for the observed seasonality of the MJO, and invoke coupling to the ocean as a mechanism to explain the period of the oscillation.

It has also been suggested that the MJO can be triggered by low-frequency variability in middle latitudes. In concluding that some sort of atmospheric discharge-recharge process controls the MJO, Blade and Hartmann (1993) mention east Asian pressure surges as one possible “trigger” for setting off the MJO after the atmosphere has recharged from the previous MJO episode. Meehl et al. (1996) present composites of 6-30-day OLR variability over the Indian and Pacific Ocean regions and examine the accompanying upper-level wave trains and surface pressure perturbations on this time scale. They conclude that pressure-surge activity on this time scale interacts with convective activity on MJO time scales, albeit through a rather complex, dynamical interaction. A similarly complex picture of pressure surge activity and MJO interaction is discussed in Compo et al. (1999), but nevertheless provides a link between quasi-stochastic high-frequency pressure surge activity and MJO-related convection. The existence of such a link seems plausible given the seasonal variations in MJO activity.

In this chapter, we explored the nature of the wave motions that fill the atmosphere,
from both theoretical and observational perspectives. The basic theory of planetary waves, governing even the tropical waves that were rediscovered by Matsuno in 1966, was worked out by Laplace.

The waves of largest scale are by far the most energetic, and produce the lion’s share of the fluxes of heat, moisture, and momentum that affect the zonally averaged flow.

Monsoons and the Walker Circulation can be regarded as thermally forced eddies.

Transient and stationary eddies are both important components of the general circulation of the atmosphere. Their effects on the mean flow are the subject of the next Chapter.

Problems

1. Show that for an isothermal atmosphere the static stability, $S_p$, increases strongly upward.

2. Prove that an isothermal atmosphere has only one equivalent depth for free oscillations, given by

$$\hat{h} = \gamma H,$$

where $\gamma \equiv c_p/c_v$, and $H = RT/g$.

3. Show that for a resting isentropic basic state with no heating and no gravitational forcing, perturbations satisfy

$$\frac{\partial z'}{\partial t} + H_0 \nabla_p^2 \chi = 0,$$  \hspace{1cm} (8.101)

for all $p$, where $H_0 = RT_0/g$ and subscript zero denotes a surface value. Notice that (8.101) looks very much like the continuity equation for shallow water.

4. Show that the FOSC with the nondivergent approximation ($\chi = 0$) must satisfy

$$\nabla_p \cdot (f \nabla_p \psi) = g \nabla_p^2 z'. $$  \hspace{1cm} (8.102)

This implies that the oscillation is not in exact geostrophic balance.

5. Planetary waves are often observed to “bend westward” on their equatorward sides (see Fig. 8.52). Give an explanation for this based on Rossby wave
6. Show that when there is a basic zonal current from the west, with a constant angular velocity $\lambda$, the apparent “phase speed” relative to the Earth’s surface of the free oscillation of the second class is given by

$$\frac{\sigma}{s} = \dot{\lambda} - \frac{2(\Omega + \dot{\lambda})}{n(n + 1)}.$$  

(8.103)

7. Laplace’s tidal equation has a special solution when the period of oscillation is $1/2$ day and the longitudinal wave number is zero.

i) Show that the solution is given by

$$\Theta_n = A \sin(\sqrt{\varepsilon_n} \mu) + B \cos(\sqrt{\varepsilon_n} \mu)$$  

(8.104)

where

$$\varepsilon_n = \frac{4\Omega^2 a^2}{gh_n}.$$  

(8.105)

j) Show that
Planetary-scale waves and other eddies

8. Show that the gravitational tidal potential \( \Phi \) at a point \( P \), due to the moon, is approximately given by

\[
\Phi = -\frac{3}{2} \frac{\gamma g M a^2}{D^3} \left(1 - \cos^2 \theta \right),
\]

(8.107)

where \( \gamma_g \) is the universal constant of gravitation, \( M \) is the mass of the moon, and \( D \) and \( \theta \) are as shown in Fig. 8.53.

9.

a) Work out an algebraic expression for the temperature profile, as a function of pressure, in an atmosphere that has a constant (vertically uniform) static stability.

b) Choose a “typical” positive constant for value for the static stability, and explain how you arrive at this choice. It should be “typical” for either the troposphere or the stratosphere, and you should state which. Plot the
temperature as a function of pressure using this value, for pressures ranging from 1000 mb to 1 mb, assuming a surface temperature of 288 K. Repeat, assuming a temperature at 1 mb of 270 K. Comment on the two plots -- are they at all realistic?

c) Find all of the equivalent depths for free oscillations of an atmosphere with constant static stability.

10. Derive (8.79) from (8.72). Verify by substitution that (8.83) is a solution of (8.79).

11. For the model of Charney and Eliassen, derive an expression for the "mountain torque" associated with stationary, frictionally damped, resonant waves (i.e. $K^2 = K_s^2$).
CHAPTER 9 | Wave–Mean Flow Interactions

9.1 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 are actually powerless to affect the mean flow. There are several related theorems that demonstrate this “non-interaction” of the eddies with the mean flow. They are called, reasonably enough, “non-interaction theorems.” The earliest such ideas were published by Eliassen and Palm (1961), and the following discussion is based on their paper. This same material is also discussed at length in Lindzen’s (1990) book.

Consider the equation of zonal motion in the simplified form

\[ \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + w \frac{\partial u}{\partial z} = -\frac{1}{\rho} \frac{\partial p}{\partial x}. \]  

(9.1)

We have omitted rotation, sphericity, friction, and \( y \)-variations for simplicity. Eq. (9.1) can apply, for example, to small-scale gravity waves forced by flow over topography. Let

\[ u = U + u', \quad U = U(z), \]

\[ w = w', \]

\[ p = \tilde{p} + p', \quad \tilde{p} = \tilde{p}(z), \]

\[ \rho = \bar{\rho} + \rho', \quad \bar{\rho} = \bar{\rho}(z). \]  

(9.2)

We interpret the primed quantities as small-amplitude wave-like perturbations with zero means. Recall that

\[ \rho \frac{\partial U}{\partial t} \sim -\frac{\partial}{\partial z} \rho w' u'. \]  

(9.3)
We are interested in what determines the wave momentum flux divergence, \( \frac{\partial}{\partial z} \rho w'u' \).

Substitute (9.2) into (9.1) and linearize, to obtain
\[
\bar{\rho} \frac{\partial u'}{\partial t} + \bar{\rho} U \frac{\partial u'}{\partial x} + \bar{\rho} w' \frac{\partial U}{\partial z} + \frac{\partial p'}{\partial x} = 0 .
\]
(9.4)

Assume that the perturbations are steady, so that
\[
\frac{\partial u'}{\partial t} = 0 .
\]
(9.5)

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
\[
\bar{\rho} U \frac{\partial u'}{\partial x} + \bar{\rho} w' \frac{\partial U}{\partial z} + \frac{\partial p'}{\partial x} = 0 .
\]
(9.6)

This is the form of the steady equation of motion that we will use.

Next, multiply (9.6) by (\( \rho Uu' + p' \)), to obtain
\[
\bar{\rho} U^2 \frac{\partial}{\partial x} \left( \frac{u'^2}{2} \right) + \bar{\rho} w'u' \frac{\partial}{\partial z} \left( \frac{U^2}{2} \right) + \bar{\rho} U \frac{\partial}{\partial x} (p'u') + \bar{\rho} w' \rho' \frac{\partial U}{\partial z} + \frac{\partial}{\partial x} \left( \frac{1}{2} p'^2 \right) = 0 .
\]
(9.7)

The terms involving \( \frac{\partial}{\partial x} \) vanish when integrated over the whole domain, leaving
\[
\left. \frac{\partial}{\partial z} \left( \frac{U^2}{2} \right) \right|_{-\infty}^{\infty} \bar{\rho} w'u'dx + \frac{\partial U}{\partial z} \int_{-\infty}^{\infty} w'p'dx = 0 ,
\]
(9.8)

which can be simplified to
\[
U \left. \int_{-\infty}^{\infty} \bar{\rho} w'u'dx + \int_{-\infty}^{\infty} w'p'dx = 0 ,
\]
(9.9)

provided that \( \frac{\partial U}{\partial z} \neq 0 \).
Eq. (9.9) is an important result. It means that the wave momentum flux, \( \int_{-\infty}^{\infty} \bar{\rho} w'u' dx \), and the wave energy flux, \( \int_{-\infty}^{\infty} w'p' dx \), are closely related. At a “critical” level, where \( U = 0 \), the wave energy flux must vanish; 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 (9.9) shows that it goes to zero at a critical level. This means that the wave does not exist above the critical level. The upward propagation of the wave is blocked at the critical level.

Let \( q \) be the total eddy energy associated with the wave (the sum of the eddy kinetic, eddy internal, and eddy potential energies). We can show that \( q \) satisfies

\[
\frac{\partial}{\partial x} (qU + p'u') + \frac{\partial}{\partial z} (p'w') = -\bar{\rho} u'w'\frac{\partial U}{\partial z} .
\] (9.10)

The right-hand side of (9.10) is a “gradient production” term that represents conversion of the kinetic energy of the mean state into the total eddy energy, \( q \). Eq. (9.10) simply says that the production term on the right-hand side is balanced by the transport terms on the left-hand side. Integration over the domain gives

\[
\frac{\partial}{\partial z} \int_{-\infty}^{\infty} p'w' dx = -\frac{\partial U}{\partial z} \int_{-\infty}^{\infty} \bar{\rho} u'w' dx .
\] (9.11)

This means that the wave energy flux divergence balances conversion to or from the kinetic energy of the mean flow.

By combining (9.9) and (9.11) we can show that

\[
U \frac{\partial}{\partial z} \left( \int_{-\infty}^{\infty} \bar{\rho} u'w' dx \right) = 0 .
\] (9.12)

This means that, when \( U \neq 0 \), the wave momentum flux \( \int_{-\infty}^{\infty} \bar{\rho} u'w' dx \) is independent of height. This is very important because, as shown by (9.3), it means that the wave momentum flux has no effect on \( U(z) \), except at the critical level where \( U = 0 \). The wave momentum flux is absorbed at the critical level. From (9.3), it follows that \( U \) will tend to change with time at the critical level, so \( U \) will become different from zero. This means that the critical level will move.
If we allowed the phase speed $c$ to be non-zero, we would find $U - c$ everywhere in place of $U$. The momentum would be absorbed at the critical level where $U = c$.

Since (9.12) tells us that $\int_{-\infty}^{\infty} \tilde{u}u'w'dx$ is independent of height (where $U \neq 0$), we see from (9.9) that the wave energy flux is just proportional to $U$. Alternatively, we can combine (9.9) and (9.12) to write

$$\frac{1}{U} \int_{-\infty}^{\infty} w'p'dx = \text{constant}.$$  (9.13)

The conserved quantity $\frac{1}{U} \int_{-\infty}^{\infty} w'p'dx$ is called the “wave action.” Note that (9.9) can be written as “wave action plus wave momentum flux = zero.”

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” (McFarlane, 1987). Most of the discussion so far has been on gravity waves forced by flow over topography, although recently gravity waves forced by convective storms are receiving a lot of attention (e.g. Fovell et al., 1992).

Fig. 9.1 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 using methods that we will not discuss, which are 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 actual response of the zonally averaged zonal wind is shown in Fig. 9.2. The changes are very large. In order for thermal wind balance to be maintained, there must be corresponding changes in the zonally averaged temperature; these are shown in Fig. 9.3. The polar troposphere has warmed dramatically, to be consistent with the weaker westerly jet. The changes shown in Fig. 9.2 and Fig. 9.3 make the model results more realistic than before, suggesting that gravity-wave drag is an important process in nature.

9.2 Vertical propagation of planetary waves

The following discussion is based on the famous paper by Charney and Drazin (1961), which deals with the vertically propagating planetary waves.

Let $T_S(p)$ be a standard temperature profile, and define $\alpha_S$, $\theta_S$, and $\rho_S$ accordingly. The quasigeostrophic form of the potential vorticity equation is

$$\left( \frac{\partial}{\partial t} + \mathbf{v}_g \cdot \nabla_p \right) q = 0 ,$$  (9.14)

where
Vertical propagation of planetary waves

9.2 Vertical propagation of planetary waves

is the quasigeostrophic pseudo-potential vorticity, $S_p = -\frac{\alpha S}{\theta S} \frac{\partial S}{\partial p}$ is the static stability, and in the last term of (9.15) $f$ has been replaced by $f_0$. [See Chapter 8 of Holton (1992).] We are working on a $\beta$-plane, such that $f = f_0 + \beta y$. Note that $q$ is essentially determined by the absolute vorticity and the change of temperature with height, and that (9.14) does not contain a vertical advection term.

From (9.14) we can derive

$$\frac{\partial}{\partial t} [q] = \frac{\partial}{\partial y} [v^* q^*].$$

**EXERCISE:** Show that
Eq. (9.17) is a very important relationship. It shows that 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., \( \frac{\partial}{\partial y} [v_g^* q_g^*] \), (9.17) will give us

\[
\frac{\partial}{\partial y} \left\{ \frac{\partial}{\partial y} [u_g^* v_g^*] \right\}.
\]

This affects the meridional shear of \([u]\). We will also get a term proportional to \( \frac{\partial}{\partial p} \left\{ \frac{\partial}{\partial y} [v_g^* T_g^*] \right\} \). This affects the static stability.

Figure 9.2: The actual change in the zonally averaged zonal 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\(^{-1}\). From McFarlane (1987).
9.2 Vertical propagation of planetary waves

We adopt the “log pressure” coordinate

\[ z(p) = -(\frac{RT_0}{g})\ln\left(\frac{p}{p_0}\right), \tag{9.18} \]

where \( T_0 \) is a constant reference temperature. With the use of (9.18), (9.15) can be rewritten as

\[ q = f + \nabla^2 \psi + \frac{1}{\rho_S} \frac{\partial}{\partial z} \left( \rho_S f_0^2 \frac{\partial \psi}{\partial z} \right), \tag{9.19} \]

where

\[ \psi \equiv \frac{\phi}{f_0} \tag{9.20} \]

is called the “geostrophic stream function,” and the Brunt-Vaisala frequency \( N \), satisfies

---

Figure 9.3: 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).
An Introduction to the General Circulation of the Atmosphere

Wave-Mean Flow Interactions

(9.21)

\[ N^2 \equiv \frac{g}{\theta_S} \frac{\partial \theta_S}{\partial z}. \]

Note that

(9.22)

\[ v_g = \frac{\partial \psi}{\partial x} \quad \text{and} \quad u_g = -\frac{\partial \psi}{\partial y}. \]

Linearizing (9.14) about the zonal-mean state gives

(9.23)

\[ \left( \frac{\partial}{\partial t} + [u] \frac{\partial}{\partial x} \right) q^* + v_g^* \frac{\partial}{\partial y} [q] = 0. \]

We look for solutions of the form

(9.24)

\[ \psi^* = \text{Re}\{\tilde{\psi}(y, z)e^{ik(x-ct)}\}, \]

(9.25)

\[ q^* = \text{Re}\{\tilde{q}(y, z)e^{ik(x-ct)}\}. \]

Substitution of (9.19), (9.24), and (9.25) into (9.23) gives

(9.26)

\[ ([u] - c) \tilde{q} + \tilde{\psi} \frac{\partial}{\partial y} [q] = 0 , \]

where

(9.27)

\[ \tilde{q} = -k^2 \tilde{\psi} + \frac{\partial^2 \tilde{\psi}}{\partial y^2} + \frac{1}{\rho_S} \frac{\partial}{\partial z} \left( \rho_S \frac{f_0^2}{N^2} \frac{\partial \tilde{\psi}}{\partial z} \right). \]

Using (9.27), we can rewrite (9.26) as

(9.28)

\[ \frac{\partial^2 \tilde{\psi}}{\partial y^2} + \frac{1}{\rho_S} \frac{\partial}{\partial z} \left( \rho_S \frac{f_0^2}{N^2} \frac{\partial \tilde{\psi}}{\partial z} \right) = -\left( \frac{1}{[u] - c} \right) \frac{\partial [q]}{\partial y} - k^2 \tilde{\psi}. \]

This is a fairly general form of the quasi-geostrophic wave equation that we want to analyze, but we will simplify it considerably before doing so.

As wave energy propagates up to higher levels, it encounters decreasing values of \( \rho_S \). The energy-density (energy per unit volume) scales like \( \rho_S(k\psi)^2 \), so if the energy density is constant with height, \( \psi \) must increase like \( \frac{1}{\sqrt{\rho_S}} \). Because of this effect, the
equations become simpler if we introduce a scaled value of $\hat{\psi}$:

$$\psi \equiv \sqrt{\frac{\rho_S}{N}} \hat{\psi}. \quad (9.29)$$

Note that here $\psi$ (no hat) is the scaled value; the meaning of $\psi$ now departs from that used in (9.20). We also note that

$$\frac{1}{\rho_S} \frac{\partial}{\partial z} \left( \rho_S^2 \frac{f_0^2}{N^2} \frac{\partial \psi}{\partial z} \right) = \frac{f_0^2}{\rho_S} \frac{\partial}{\partial z} \left\{ \sqrt{\frac{\rho_S}{N}} \frac{\partial}{\partial z} \left( \sqrt{\frac{\rho_S}{N}} \hat{\psi} \right) - \sqrt{\frac{\rho_S}{N}} \frac{\partial}{\partial z} \left( \sqrt{\frac{\rho_S}{N}} \right) \right\},$$

$$= \frac{f_0^2}{\rho_S} \frac{\partial}{\partial z} \left( \sqrt{\frac{\rho_S}{N}} \right) \frac{\partial}{\partial z} \left( \sqrt{\frac{\rho_S}{N}} \hat{\psi} \right) + \sqrt{\frac{\rho_S}{N}} \frac{\partial^2}{\partial z^2} \left( \sqrt{\frac{\rho_S}{N}} \right),$$

$$- \frac{\partial}{\partial z} \left( \sqrt{\frac{\rho_S}{N}} \right) \frac{\partial}{\partial z} \left( \sqrt{\frac{\rho_S}{N}} \hat{\psi} \right) + \sqrt{\frac{\rho_S}{N}} \frac{\partial^2}{\partial z^2} \left( \sqrt{\frac{\rho_S}{N}} \right).$$

Using (9.29) and (9.30), we can rewrite (9.28) as

$$\frac{\partial^2 \psi}{\partial y^2} + \frac{f_0^2}{N^2} \frac{\partial^2 \psi}{\partial z^2} = -\frac{f_0^2}{4H_0^2} \left( \frac{1}{[u]-c} \right) \frac{\partial [q]}{\partial y} - \frac{f_0^2}{\sqrt{\rho_S} \frac{\partial^2}{\partial z^2} \left( \sqrt{\frac{\rho_S}{N}} \right)},$$

where

$$n^2 \equiv \frac{4N^2 H_0^2}{f_0^2} \left\{ \frac{1}{[u]-c} \frac{\partial [q]}{\partial y} - k^2 - \frac{f_0^2}{\sqrt{\rho_S} \frac{\partial^2}{\partial z^2} \left( \sqrt{\frac{\rho_S}{N}} \right)} \right\}. \quad (9.32)$$

is called the “index of refraction.” Here $H_0 \equiv \frac{R T_0}{g}$, where $T_0$ is a reference temperature. Comparing (9.31) - (9.32) with (9.28), it seems that the left-hand side has been simplified but the right-hand side has become more complicated., Eq. (9.32) is a form of the
quasigeostrophic wave equation. When \( n^2 > 0 \), \( \psi \) is oscillatory (propagating), and when \( n^2 < 0 \), \( \psi \) is “evanescent” (exponentially decaying away from the source of excitation).

The index of refraction as given by (9.32) is pretty complicated. Consider a simplified special case: an isothermal atmosphere with \( T_S(p) \equiv T_0 = \text{constant} \). This is not unrealistic for the lower stratosphere. For this case, we can show that \( N \equiv \text{constant} \) and \( \rho_s \sim e^{-z/H_0} \), so that (9.32) can be simplified to

\[
n^2 = \frac{4N^2H_0^2}{f_0^2} \left( \frac{1}{[u]} - c \frac{\partial[\theta]}{\partial y} - k^2 \right) - 1 . \tag{9.33}
\]

Now we concentrate on stationary waves, for which the phase speed, \( c \), is zero. This type of wave can be forced by orography, for example, as discussed in Chapter 7. Then (9.31) and (9.33) become

\[
\frac{\partial^2 \psi}{\partial y^2} + \frac{f_0^2}{N^2} \frac{\partial^2 \psi}{\partial z^2} = -\frac{n^2}{4H_0^2} \psi . \tag{9.34}
\]

\[
n^2 = \frac{4N^2H_0^2}{f_0^2} \left( \frac{1}{[u]} \frac{\partial[\theta]}{\partial y} - k^2 \right) - 1 . \tag{9.35}
\]

To simplify \( n^2 \) even further, note from (9.19) that

\[
\frac{\partial}{\partial y} [\theta] = \beta - \frac{\partial^2 [u]}{\partial y^2} - \frac{1}{\rho_s} \frac{\partial}{\partial z} \left( \rho_s \frac{f_0^2}{N^2} \frac{\partial [u]}{\partial z} \right) , \tag{9.36}
\]

where \( \beta \equiv \frac{df}{dy} \). When the variations of \([u]\) are not too strong,

\[
\frac{\partial[\theta]}{\partial y} \equiv \beta \geq 0 . \tag{9.37}
\]

Then

\[
n^2 = \frac{4N^2H_0^2}{f_0^2} \left( \frac{\beta}{[u]} - k^2 \right) - 1 . \tag{9.38}
\]

From (9.38), we see the following:
To have vertical propagation \( (n^2 > 0) \), we need \( \beta / [u] > 0 \). Because \( \beta > 0 \) and \( [u] \) must be positive (westerly), \textit{Stationary Rossby waves cannot propagate through easterlies}. Recall that the summer hemisphere stratosphere is dominated by easterlies, while the winter hemisphere stratosphere is dominated by westerlies. Note, however, that large positive \( [u] \) also makes \( n^2 < 0 \). Waves cannot propagate through very strong westerlies. Fig. 9.4, from Charney and Drazin (1961), shows the vertical distribution of \( n^2 \) for summer and winter, averaged over the Northern Hemisphere middle latitudes, for three different wavelengths.

![Figure 9.4: 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).](image)

Even when \( \beta / [u] > 0 \), for a given \([u]\) waves with sufficiently large \( k \) (short wavelength) cannot propagate. Short waves are, therefore, “trapped” near their excitation levels. Since \([u]\) has a \textit{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 the stratosphere and mesosphere even more than they do in the troposphere.

- A level where \([u] = 0\) is called a “critical level” for stationary waves. Suppose that \([u] > 0\) below a critical level, and \([u] < 0\) above. Then, for waves excited at the lower boundary (e.g., by topography), upward propagation is completely blocked at the critical level.

Fig. 9.5 provides evidence that the theory is correct. It shows the geopotential height fields at 500 mb, 100 mb, and 10 mb, for Northern Hemisphere summer and winter. In winter, planetary waves clearly propagate upward to the 10 mb 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 shorter modes are trapped at lower levels while longer modes can continue to propagate upward to great heights.

Waves can also be trapped at critical latitudes where \([u] = 0\). We could therefore define critical surfaces in the y-z plane.

If we allowed \(c \neq 0\), we would find that the critical surfaces are those for which \([u] - c = 0\).

Matsuno (1970) used \([u]\) for the Northern Hemisphere winter to compute \(\frac{\partial q}{\partial \phi}\), the index of refraction, and the energy flow in the latitude-height plane for zonal wave number 1. His results are shown in Fig. 9.6.

Fig. 9.7 shows that planetary wave energy does in fact propagate from the troposphere into the stratosphere during northern winter.

Fig. 9.8 shows that the eddy kinetic energy in the stratosphere is supplied by the troposphere. KZ is converted into AZ, i.e. the meridional temperature gradient is increased by an indirect circulation. The general circulation of the lower stratosphere thus acts like a refrigerator.

### 9.3 Vertical and meridional fluxes due to planetary waves

Now we investigate under what conditions planetary waves can transport energy and momentum. The quasigeostrophic form of the thermodynamic energy equation is (e.g. Holton, 1992)

\[
\left(\frac{\partial}{\partial t} + V \cdot \nabla\right)\frac{\partial \phi}{\partial p} + \frac{\partial S}{\partial p} \omega = 0.
\]

Here \(V\) is the geostrophic wind, which is important for the following discussion, and \(\nabla\) is \(\nabla_p\). Eq. (9.39) can be written as
Figure 9.5: 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 mb, 100 mb, and 10 mb. From Charney (1973).
Figure 9.6: a) The latitudinal gradient of the potential vorticity, $\partial q / \partial \phi$, expressed as a multiple of the Earth’s rotation rate. b) An idealized but somewhat realistic basic state zonal wind distribution (m s$^{-1}$) in the Northern Hemisphere winter. c) The refractive index square $n^2$, for the $k = 0$ wave. d) Computed distribution of energy flow in the meridional plane associated with zonal wave number 1. From Matsuno (1970).
9.3 Vertical and meridional fluxes due to planetary waves

An Introduction to the General Circulation of the Atmosphere

331

\[ \frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla \psi_z + \frac{N^2}{f_0} w = 0 , \] (9.40)

where \( w \) is defined by \(-\omega/(\rho_g g)\). Here \( \psi_z \equiv \partial \psi / \partial z \), and \( z \) is the “log-p” coordinate defined by (9.18). Linearizing (9.40) gives

Figure 9.7: a) Daily vertical eddy flux of geopotential through 100 and 10 mb. Units: erg cm\(^{-2}\) s\(^{-1}\). b) Daily divergence of the vertical eddy flux of geopotential for 100-10 mb, and for the region 90° N to 20° N. Units: erg cm\(^{-1}\) s\(^{-1}\). Data for 1964. From Dopplick (1971).
Here we have used the thermal wind equation. Multiplying (9.41) by $\psi_z^*$, we obtain a form of the “temperature variance equation:”

$$\left(\frac{\partial}{\partial t} + [u] \frac{\partial}{\partial x}\right) \psi_z^* - v_+ \frac{\partial}{\partial z} \left[\psi_z^* \right] + \frac{N^2}{f_0} w^* = 0.$$  \hspace{1cm} (9.42)

Note the two gradient production terms.

Figure 9.8: Annual energy cycle for the lower stratosphere between 100–10 mb and for the region 20°N to 90°N. Units: contents 10^7 erg cm^-2; conversions: erg cm^-2 s^-1. “B” indicates boundary effects (at the lower or equatorward boundaries), and “G” indicates generation. Overall, the stratosphere gains kinetic energy through interactions with the troposphere. The kinetic energy is converted to potential energy. This balances a loss of potential energy through radiative effects. From Dopplrick (1971).
Take the zonal mean of (9.42), so that the \( [u] \frac{\partial}{\partial x} \) term drops out. Rearrange to isolate the meridional energy flux by itself on the left-hand side:

\[
[v^* \psi_z^*] \frac{\partial [u]}{\partial z} = \frac{\partial}{\partial t} \left[ \frac{1}{2} (\psi_z^*)^2 \right] + N^2 \left[ \frac{w^* \psi_z^*}{f_0} \right].
\] (9.43)

Note that \( [w^* \psi_z^*]/f_0 > 0 \) implies an upward temperature flux, in either hemisphere. Also \( [v^* \psi_z^*] > 0 \) implies a poleward temperature flux, in either hemisphere.

First, consider a baroclinically amplifying wave, for which \( \frac{\partial}{\partial t} \left[ (\psi_z^*)^2 \right] > 0 \) and the wave temperature flux is upward. From (9.43), we see that such a wave produces a poleward temperature flux (in either hemisphere) when \( \frac{\partial [u]}{\partial z} > 0 \), i.e. when the temperature is decreasing towards the pole. The heat flux is downgradient, so the gradient-production term is positive.

Next, consider a neutral wave of the form \( e^{i(kx - ct)} \), for which \( \frac{\partial}{\partial t} = -c \frac{\partial}{\partial x} \), where \( c \) is real. Multiply (9.41) by \( \psi^* \) and take the zonal mean, to obtain

\[
([u] - c) [v^* \psi_z^*] = N^2 \left[ \frac{w^* \psi^*}{f_0} \right].
\] (9.44)

Note that \( [w^* \psi^*]/f_0 > 0 \) means an upward propagation of wave energy in either hemisphere. Recall also that \( [u] - c > 0 \) is needed in order for the wave to propagate. It follows that an upward-propagating neutral wave transports energy poleward. Such a wave might be forced, for example, by flow over mountains.

In summary, poleward energy transport is produced by either a baroclinically amplifying wave with \( \frac{\partial [u]}{\partial z} > 0 \) or a neutral wave that propagates upward.

Applying the eddy PV equation (9.23) to a neutral wave gives

\[
([u] - c) \frac{\partial q^*}{\partial x} + v^* \frac{\partial [q]}{\partial y} = 0.
\] (9.45)

Multiply (9.45) by \( \psi^* \) and take the zonal mean to show that

\[
[v^* q^*] = 0 \text{ except where } [u] = c
\] (9.46)
(the critical level). This very important result shows that neutral waves produce no potential vorticity flux except at a critical level. It follows from (9.16) that \textit{neutral waves do not affect $[q]$ except at a critical level}. This is a non-interaction theorem for planetary waves, analogous to the result obtained for gravity waves by Eliassen and Palm (1961).

From (9.17), \([v_*q_*] = 0\) means

$$- \frac{\partial}{\partial y} [u_*v_*] + \frac{f_0^2}{\rho_S} \frac{\partial}{\partial z} \left( \frac{\rho_S}{N^2} [v_*\psi^*_z] \right) = 0 . \quad (9.47)$$

As we will see later, the expression equal to zero in (9.47) is the divergence of the Eliassen-Palm flux.

Vertically integrate (9.47) through the depth of the atmosphere to obtain

$$\int_0^{p_S} \frac{\partial}{\partial y} [u_*v_*] dp = \int_0^{p_S} \frac{\rho_S}{N^2} [v_*\psi^*_z] d \psi . \quad (9.48)$$

for the neutral waves. The left-hand side 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 from (9.44) that an upward propagating neutral wave produces a poleward energy flux, i.e. \([v_*\psi^*_z] > 0\). It follows from (9.48) that

$$\int_0^{p_S} \frac{\partial}{\partial y} [u_*v_*] dp > 0 . \quad (9.49)$$

This means that the vertically integrated meridional momentum flux divergence tends to accelerate the vertically integrated \([u]\). In other words, the eddies feed the jet! This explains why we observe $KE \rightarrow KZ$. If the waves are also transporting temperature poleward, they will tend to reduce the meridional temperature gradient and so tend to reduce the strength of the westerlies. The momentum flux and heat flux thus have opposing effects 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 this, consider the zonal momentum equation, in Cartesian coordinates for simplicity:

$$\begin{align*}
\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} (uu) + \frac{\partial}{\partial y} (uv) + \frac{1}{\rho_S} \frac{\partial}{\partial z} (\rho_S u w) &= - \frac{\partial \phi}{\partial x} + f v .
\end{align*} \quad (9.50)$$

We have neglected the metric term and assumed no friction above the boundary layer. Taking the zonal mean of (9.50) gives
Here advection of $[u]$ by $[v]$ and $[w]$ is neglected; this is justified in the midlatitude winter. To the extent that $[v]$ is geostrophic, it vanishes anyway. Now assume $\frac{\partial[u]}{\partial t} = 0$, consistent with $\frac{\partial[q]}{\partial t} = 0$. This leads to

$$\frac{\partial}{\partial y} [u^* v^*] = -\frac{1}{\rho_S} \frac{\partial}{\partial z} (\rho_S [u^* w^*]) + f[v].$$  \hspace{1cm} (9.52)

Integrating (9.52) vertically with respect to mass, using $\int_0^\infty \rho_S [v] dz \equiv 0$, and employing (9.49), we find that

$$\int_0^\infty \frac{\partial}{\partial z} (\rho_S [u^* w^*]) dz > 0. \hspace{1cm} (9.53)$$

We know that $\rho_S [u^* w^*]$ must vanish at great height, so we conclude that

$$\rho_S [u^* w^*]_{S < 0}. \hspace{1cm} (9.54)$$

This shows that, near the lower boundary, friction and/or mountain torque must carry westerly momentum into the Earth's surface, in the presence of an upward propagating planetary wave. An alternative interpretation is that frictional and/or mountain torque, in a belt of westerlies where (9.54) is satisfied, will produce an upward-propagating planetary wave that transports energy poleward.

Compare (9.49) and (9.54). The meridional momentum flux accelerates the westerlies, while the vertical momentum flux decelerates them.

### 9.4 Sudden warmings

The preceding discussion is highly relevant to stratospheric sudden warmings. This is how stratospheric sudden warmings work: The polar stratosphere is observed to warm up, while the lower latitudes cool off, as shown in the upper panel of Fig. 9.9. There is no significant change of the *area-averaged* temperature, however. This suggests that the sudden warming is due to a poleward redistribution of sensible heat. The polar westerlies are observed to weaken, and in some cases they give way to easterlies. There are two mechanisms that can change the zonally averaged temperature by meridional redistribution of sensible heat. The possibilities are:

- A *mean meridional circulation*. Sinking near the poles can produce adiabatic
warming, while the compensating rising motion in lower latitudes gives adiabatic cooling. This MMC would be *direct*, at least to start with.

- **Poleward eddy energy transport.** In this case, eddies produce warming near the pole and cooling in lower latitudes. An MMC would be produced by the “apparent” heating and cooling due to the eddies. This MMC would feature rising at the pole to counteract the eddy warming there, and sinking in lower latitudes to counteract the eddy cooling there. This means that the MMC would be *indirect*.

In the first scenario, the westerlies will increase aloft. (Why?) Since the westerly shear must decrease during a sudden warming (Why?), there must be an even stronger intensification of the westerlies below. In the second scenario, on the other hand, the westerlies will weaken above. *The observed transition to easterlies in the stratosphere supports the second hypothesis.* We conclude that stratospheric sudden warmings are produced by poleward eddy heat fluxes.

Sudden warmings are characterized by intense wave activity, of planetary scale. The longitudinal phase is stationary, suggesting orographic forcing. Sudden warmings are relatively infrequent and weak in the Southern Hemisphere, where there is little orography.

There is a strong flux of wave energy from the troposphere to the stratosphere during winter. From our previous analysis, we know that such a wave will transport energy poleward. A sudden warming is triggered by the rapid growth of a quasistationary planetary wave in the troposphere, and its subsequent upward propagation into the stratosphere. As shown in Fig. 9.9 and Fig. 9.10, there is observational evidence for such wave growth, often in association with the formation of a blocking high (discussed later).

As it propagates upward, the planetary wave transports energy poleward, warming the pole and consequently weakening the westerlies aloft. In extreme cases, the westerlies are actually changed to easterlies. When easterlies form above, the wave propagation is blocked, so the wave energy is concentrated near the critical level. This leads to *really* sudden warming.

Matsuno (1971) was the first to successfully simulate sudden warmings with a numerical model.

### 9.5 Eliassen–Palm Theorem–Reprise

Previously we discussed non-interaction theorems for pure gravity waves and for quasi-geostrophic waves on a $\beta$-plane. It was discovered during the 1970’s that non-interaction theorems can be derived for very general balanced flows. The following discussion provides an example. The discussion is based on Andrews et al. (1987).

The zonally averaged momentum equations in spherical coordinates can be written as
Figure 9.9: a) Latitude–time section of zonal mean temperature (K) measured by channel A for 31 December 1970 to 16 January 1971. Regions of temperature lower than 258 K are shaded. A major warming had a peak at this level on 9 January at 80°N. b) Latitude–time section of amplitude of zonal wavenumber one of channel A temperature (K) for 31 December 1970 to 16 January 1971. Maximum amplitude occurred on 4 January at 65°N. From Barnett (1974).
Figure 9.10: 10 mb charts during a sudden warming in January 1963. Height contours (solid lines) at 32 dm intervals. Isotherms (dashed lines) at 10 K intervals. Note that the process spans 9 days. From Sawyer (1965).
Here \( z = -H \log(p/p_S) \) is the vertical coordinate, and \( w \equiv Dz/Dt \). The scale height \( H \) is \( \frac{RT_0}{g} \), where \( T_0 \) is a constant. The zonally averaged thermodynamic energy equation is

\[
\frac{\partial [\theta]}{\partial t} + \frac{1}{a \cos \phi} \left[ \frac{\partial [\theta]}{\partial \phi} + [w] \frac{\partial [\theta]}{\partial z} \right] = \frac{1}{\rho_s} \frac{\partial}{\partial z} \left( \rho_s [w \theta^*] \right) - [Q]
\]

\[
= \frac{-1}{a \cos \phi} \left( [v^*] \cos \phi \right) - \frac{1}{\rho_s} \frac{\partial}{\partial z} \left( \rho_s [v^* w^*] \right) \quad (9.57)
\]

Here \( Q \) represents a heating process. Finally, we will need the zonally averaged continuity equation,

\[
\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \left( \rho_s [v \cos \phi] \right) + \frac{\partial}{\partial z} \left( \rho_s [w] \right) = 0 \quad (9.58)
\]

and hydrostatics:

\[
\frac{\partial [\phi]}{\partial z} - \frac{R[\theta]}{H} \frac{\kappa z}{H} = 0 \quad (9.59)
\]

In the above equations, \( \rho_s(z) \equiv \rho_0 e^{-z/H} \), where \( \rho_0 \) is a constant. We assume that (9.56) can be approximated by gradient wind balance, i.e.
This assumed balance is essential to the following argument.

We define a “residual mean meridional circulation” \((0, V, W)\) by

\[
V \equiv [v] - \frac{1}{\rho_s} \frac{\partial}{\partial z} \left( \frac{\rho_s [v^* \theta^*]}{\partial (\theta)} \right),
\]

\[
W \equiv [w] + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \left( \cos \phi \frac{\rho_s [v^* \theta^*]}{\partial (\theta)} \right).
\]

In the absence of eddies, \(V = [v]\) and \(W = [w]\). Substitution shows that \(V\) and \(W\) satisfy a continuity equation analogous to (9.58). Use of (9.61) and (9.62) to eliminate \([v]\) and \([w]\) in favor of \(V\) and \(W\) allows us to rewrite (9.55) and (9.57) as:

\[
\frac{\partial [u]}{\partial t} + V \left\{ \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \left( \left[ u \right] \cos \phi \right) - f \right\} + W \frac{\partial [u]}{\partial z} - [F_x] = \frac{1}{\rho_s a \cos \phi} (\nabla \cdot EPF),
\]

and

\[
\frac{\partial [\theta]}{\partial t} + V \frac{\partial [\theta]}{\partial \phi} + W \frac{\partial [\theta]}{\partial z} - [Q] = \frac{-1}{\rho_s} \frac{\partial}{\partial z} \left\{ \left( \frac{\partial [\theta]}{\partial z} \right)^{-1} \rho_s [v^* \theta^*] \frac{a}{\partial \phi} \frac{\partial [\theta]}{\partial \phi} + \rho_s [w^* \theta^*] \right\},
\]

respectively where

\[
EPF \equiv [0, (EPF)_\phi, (EPF)_z]
\]

is the “Eliassen-Palm flux,” whose components are
In (9.66), the term is dominant, and in (9.67) the term is dominant. Compare (9.66) and (9.67) with (9.17) and (9.47). When the EPF points upward, the meridional energy flux is in control. When it points in the meridional direction, the meridional flux of zonal momentum is in control. From (9.63) we see that a positive Eliassen-Palm flux divergence tends to increase .

The preceding derivation appears to be nothing more than an algebraic shuffle. We wrote down (9.61) and (9.62) without any explanation or motivation. What is the point of all this? The point is that for steady linear waves with and , it can be shown that . (9.68)

Recall that this follows essentially from . It turns out that the eddy forcing term of (9.64) is zero under the same conditions, i.e.

\[ \frac{1}{\rho S} \frac{\partial}{\partial z} \left( \rho S \left[ v^* \theta^* \right] - \left[ u^* v^* \right] \right) = 0 . \]  

(9.69)

This follows essentially from the facts that: 1) \([u]\) does not change, and 2) thermal wind balance is maintained.

For the case of steady, linear waves, in the absence of friction and heating, our system of equations reduces to

\[ \frac{\partial [u]}{\partial t} + \nu \left\{ \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} ([u] \cos \phi) - f \right\} + W \frac{\partial [u]}{\partial z} = 0 , \]

\[ [u] \left( f + [u] \tan \frac{\phi}{a} \right) + \frac{1}{a} \frac{\partial [\phi]}{\partial \phi} = 0 , \]
Wave–Mean Flow Interactions

An Introduction to the General Circulation of the Atmosphere

This system has the following steady solution:

\[ \frac{\partial [\theta]}{\partial t} + \frac{V}{a} \frac{\partial [\theta]}{\partial \phi} + W \frac{\partial [\theta]}{\partial z} = 0, \]

\[ \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} (\rho_s V \cos \phi) + \frac{\partial}{\partial z} \rho_s W = 0, \]

\[ \frac{\partial [\phi]}{\partial z} - \frac{\kappa_z}{H} = 0. \] (9.70)

This system has the following steady solution:

\[ \frac{\partial [u]}{\partial t} = 0, \quad u \text{ in gradient-wind balance,} \]

\[ V = 0, \quad W = 0, \]

\[ \frac{\partial [\theta]}{\partial t} = 0, \quad [\theta] \text{ specified from the past history or radiative-convective equilibrium.} \] (9.71)

This is essentially the same solution that we discussed in Chapter 4. (There \([\theta]\) was determined by the specified “equilibrium” distribution, \(\theta_E\).) From the definitions of \(V\) and \(W\), we can find the mean meridional circulation implied by \(V = 0\) and \(W = 0\):

\[ \rho_s [v] = \frac{\partial}{\partial z} \left( \rho_s \left[ \frac{[v]}{[\theta]} \right] \right), \] (9.72)

\[ [w] = -\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \left( \cos \phi \left[ \frac{[v]}{[\theta]} \right] \right). \] (9.73)

Suppose we have a solution with no eddies at all. “No eddies” certainly qualify as “steady linear waves.” The above argument therefore applies, so we can get \([u]\), \([\theta]\), and from (9.72) and (9.73) we conclude that the MMC will vanish. This is essentially the same solution that we discussed in Chapter 4, for the case of no friction.

Now add steady linear eddies, so that \(\nabla \cdot (EPF)\) continues to be zero. Then exactly the same \([u]\) and \([\theta]\) will satisfy the equations! Of course, \([v]\) and \([w]\) will be different, i.e., the MMC will be different. The MMC will in fact have to be just what it takes to ensure that \(V = W = 0\), i.e. to satisfy (9.72) and (9.73). We can say that this MMC is “induced by”
An Introduction to the General Circulation of the Atmosphere

9.5 Eliassen-Palm Theorem—Reprise

the eddies. The system produces this MMC in order to prevent the eddies from disrupting the thermal 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 \([u]\) and \([\theta]\) by applying eddy forcing such that \(\nabla \cdot (\mathbf{EPF}) = 0\) (no potential vorticity flux), 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 will induce an MMC that exactly cancels the direct effects of the eddies on \([u]\) and \([\theta]\).

When the eddies are unsteady, the residual circulation is different from zero, and \([u]\) and \([\theta]\) are modified by the combined effects of the eddies and/or the eddy-induced MMC. Cancellation of the effects of the eddies and the MMC still tends to occur, but the cancellation is incomplete.

Edmon et al. (1980) discussed the quasi-geostrophic form of the non-interaction theorem, and used it to analyze the data of Oort and Rasmussen (1971). As a reminder [see (9.47)], the meridional component of the quasigeostrophic \(\mathbf{EPF}\) is

\[
(EPF)_\phi = -a \cos(\phi)[u^*v^*],
\]  

(9.74)

and the vertical component is

\[
(EPF)_p = f a \cos(\phi) \frac{\partial (\frac{1}{\rho} \frac{\partial \theta}{\partial p})}{\partial \theta}.
\]  

(9.75)

[Note: Compare (9.74) and (9.75) with (9.66) and (9.67), respectively.] Fig. 9.11 shows the contribution of the transient eddies to the Eliassen-Palm fluxes. First consider the winter results, shown in the upper panel. Near the surface in middle latitudes, we see arrows pointing strongly upward, indicating an intense poleward potential temperature flux. Near the tropopause, the arrows curve over and become horizontal, pointing towards the tropics. This indicates a strong poleward eddy momentum flux. The contours in the figure show the divergence of the Eliassen-Palm flux. Keep in mind that \(\nabla \cdot \mathbf{EPF} > 0\) means \(\frac{\partial [u]}{\partial t} > 0\), i.e. a positive \(\mathbf{EPF}\) 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 \(\mathbf{EPF}\) convergence results mainly from the upward decrease of the upward component of the flux, i.e. it is mainly due to the energy flux.

The results for summer are quite similar, except that everything is weaker, and shifted poleward.

Fig. 9.12 shows the corresponding results for the stationary waves. In winter, the “strong” arrows are pointing nearly straight up everywhere, indicating that the poleward eddy potential temperature flux is playing a much more important role than the eddy momentum
flux. The westerlies are decelerated aloft, near 50° N, but they are accelerated near the surface. In summer the arrows point downward. The eddy momentum flux is important near the summer tropopause, but again the eddy potential temperature flux is more important overall. The westerlies are strongly decelerated near the surface in the subtropics, and they are actually accelerated at 200 mb near 35° N.

An Introduction to the General Circulation of the Atmosphere
Fig. 9.13 shows the combined effects of the transient and stationary eddies. Note that the transient eddies dominate, in both seasons. Finally, Fig. 9.14 shows the residual circulation, $VW$, for summer and winter. In winter, the residual circulation looks suspiciously like a giant Hadley Cell, extending from the tropics to the poles. This is reminiscent of the mean meridional circulation as seen in isentropic coordinates. In summer, we seem to see the northern edge of a Hadley Cell extending into the Southern Hemisphere.

Figure 9.12: Contribution of stationary eddies to the seasonally averaged Eliassen-Palm cross sections for the troposphere: (a) 5-year average from Oort & Rasmusson (1971) for winter; (b) the same, respectively, for summer. The contour interval is $1 \times 10^{15} \text{m}^3$ for both panels. The horizontal arrow is indicated at bottom right. From Edmon et al. (1980).
Clearly, we can regard the residual circulation as a response to heating.

9.6 The Eliassen–Palm theorem in isentropic coordinates

The Eliassen-Palm theorem is somewhat simpler and easier to interpret when we use isentropic coordinates. Following Andrews (1983), we begin with the flux form of the
The Eliassen–Palm theorem in isentropic coordinates

(9.76) Angular momentum equation in isentropic coordinates, neglecting friction:

\[
\frac{\partial}{\partial t}(Mu) + \frac{1}{a \cos \phi} \left[ \frac{\partial}{\partial \lambda} (muM) + \frac{\partial}{\partial \phi} (m\cos \phi) \right] = -m \frac{\partial s}{\partial \lambda} - \frac{\partial}{\partial \theta} (mM) .
\]
Here $m$ is the pseudo-density:

$$m = -\frac{1}{g\partial\theta} \frac{\partial p}{\partial \theta},$$

(9.77)

and

$$s = c_p T + g z$$

$$= \Pi \theta + g z,$$

(9.78)

where

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

(9.79)

is the Exner function. Using the hydrostatic equation in isentropic coordinates, i.e.,

$$\frac{\partial s}{\partial \theta} = \Pi,$$

(9.80)

we obtain

$$m \frac{\partial s}{\partial \lambda} = -\frac{1}{g\partial\theta} \frac{\partial p}{\partial \theta} \frac{\partial s}{\partial \lambda}$$

$$= -\frac{1}{g\partial\theta} \left( \frac{p}{\partial \lambda} \right) + \frac{p}{g\partial\lambda} \left( \frac{\partial s}{\partial \theta} \right)$$

$$= -\frac{1}{g\partial\theta} \left( \frac{p}{\partial \lambda} \right) \left( \Pi \theta + g z \right) + \frac{p}{g\partial\lambda} \frac{\partial \Pi}{\partial \theta}$$

(9.81)

$$= -\frac{1}{g\partial\theta} \left( \frac{p}{\partial \lambda} \theta \frac{\partial \Pi}{\partial \lambda} + p \frac{\partial z}{\partial \lambda} \right) + \frac{p}{g\partial\lambda} \frac{\partial \Pi}{\partial \theta}$$

$$= -\frac{1}{g\partial\theta} \left( \theta \frac{p}{\partial \lambda} \frac{\partial \Pi}{\partial \lambda} + p \frac{\partial \Pi}{\partial \theta} \frac{\partial z}{\partial \lambda} \right)$$

$$= -\frac{\theta}{g\partial\theta} \left( p \frac{\partial \Pi}{\partial \lambda} \right) - \frac{\partial}{\partial \theta} \left( p \frac{\partial z}{\partial \lambda} \right).$$

Substituting back, the momentum equation becomes
When we take the zonal mean of (9.82), the first term on the right-hand side vanishes (Why?) and we are left with

\[
\frac{\partial}{\partial t}(mM) + \frac{1}{a \cos \phi} \left\{ \frac{\partial}{\partial \lambda} (muM) + \frac{\partial}{\partial \phi}(mvM \cos \phi) \right\} = \left\{ \theta \frac{\partial}{\partial \theta} \left[ p \frac{\partial \Pi}{\partial \phi} \right] + \frac{\partial}{\partial \theta} \left( p \frac{\partial z}{\partial \lambda} \right) \right\} - \frac{\partial}{\partial \theta}(m\theta M).
\]

(9.82)

The pressure-gradient term of (9.83) has a very simple and interesting form: It is proportional to the change with \( \theta \) of the zonal mean of the product of the pressure and the slope of the height of the isentropic surface. The expression \( \left[ p \frac{\partial z}{\partial \lambda} \right] \) can be interpreted as “form drag” on the isentropic surface, analogous to the drag on mountains discussed earlier in the course. Here the “mountains” are upward bulges of the isentropic surfaces, associated with blobs of cold air at a given pressure level; the “valleys” are downward bulges of the isentropic surfaces, associated with blobs of warm air at a given pressure level. We can say that

\[
\text{upward flux of zonal momentum due to the wave} = - \left[ p \frac{\partial z}{\partial \lambda} \right].
\]

(9.84)

This shows that, 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 vanishes in isentropic coordinates in the absence of 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 upper side. It is the difference between these two forces that tends to produce a net acceleration of the layer. That is why we see \( \frac{\partial}{\partial \theta} \left[ p \frac{\partial z}{\partial \lambda} \right] \) in (9.83).

Before completing our discussion of the Eliassen-Palm theorem in isentropic coordinates, it is useful to recall the form of the mechanical energy equation in isentropic coordinates, which was given earlier in the course and is repeated here for your convenience:

\[
\left\{ \frac{\partial}{\partial t}(mK) \right\}_\theta + \nabla_\theta \cdot \left\{ m \nabla(K + \phi) \right\} + \frac{\partial}{\partial \theta} \left\{ m\dot{\theta}(K + \phi) \right\} + m\alpha \nabla \cdot (\mathbf{F} \cdot \nabla)
\]

\[
= -\frac{\partial}{\partial \theta} \left\{ -z \left( \frac{\partial p}{\partial \theta} \right)_\theta \right\} - m\omega \alpha - m\delta.
\]

(9.85)
The first term on the right-hand side of (9.85) represents the vertical transport of energy via "pressure-work." The upward flux of wave energy is, therefore, given by \([-z^* \left( \frac{\partial \rho^*}{\partial t} \right) \theta] \). Recall that for a neutral wave propagating zonally and vertically, with zonal phase velocity \( c \), \( \frac{\partial}{\partial t} = \left( \frac{[u] - c}{a \cos \phi} \right) \frac{\partial}{\partial \lambda} \). It follows that

\[
\text{upward wave energy flux} = \left( \frac{[u] - c}{a \cos \phi} \right) \left( p^* \frac{\partial z^*}{\partial \lambda} \right) \theta. \quad (9.86)
\]

Comparing (9.85) with (9.86), we conclude that

\[
\text{upward wave energy flux} = \left( \frac{[u] - c}{a \cos \phi} \right) \text{times the upward wave momentum flux.} \quad (9.87)
\]

This shows that for neutral waves propagating towards the east relative to the air \((-([u] - c) > 0)\), the momentum flux and the energy flux have the same sign, while for neutral waves propagating towards the west relative to the air \((-([u] - c) < 0)\) they have opposite signs. As we know, Rossby waves always propagate west relative to the air, so for Rossby waves the momentum flux is always opposite in direction to the energy flux.

Now we relate the preceding analysis to the Eliassen-Palm theorem, following Andrews (1983) and Andrews et al. (1987). Recall that

\[
[mA] = [m][A] + [m^* A^*], \quad (9.88)
\]

for an arbitrary variable \( A \). Using (9.88) in the time-rate-of-change term of (9.83), we obtain

\[
\frac{\partial}{\partial t}([m][M]) + \frac{1}{a \cos \phi} \left\{ \frac{\partial}{\partial \phi}([m v M] \cos \phi) \right\} = -\frac{\partial}{\partial t}[m^* M^*] + \frac{\partial}{\partial \theta} \left[p^* \frac{\partial z^*}{\partial \lambda} \right] - \frac{\partial}{\partial \theta}[m M \theta]. \quad (9.89)
\]

Here the "eddy part" of the time-rate-of-change term has been moved to the right-hand-side of the equals sign; this will be discussed later. We want to derive an "advective form" of (9.89), so we bring in the zonally averaged continuity equation in isentropic coordinates, which can be written as

\[
\frac{\partial}{\partial t}[m] + \frac{1}{a \cos \phi \partial \phi} ([m v] \cos \phi) = -\frac{\partial}{\partial \theta}[m \theta]. \quad (9.90)
\]
Subtract \([M]\) times (9.90) from (9.89), to obtain

\[
[m] \frac{d[M]}{dt} + \frac{1}{a \cos \phi} \left\{ \frac{\partial}{\partial \phi} ([m v M](m v M) \cos \phi) \right\} - \frac{[M]}{a \cos \phi \partial \phi} ([m v M] \cos \phi)
\]

\[
= - \frac{\partial}{\partial t} [m^* M^*] + \frac{\partial}{\partial \theta} \left[ p^* \frac{\partial z^*}{\partial \lambda} \right] + [M] \frac{\partial}{\partial \theta} [m \hat{\theta}] - \frac{\partial}{\partial \theta} [m M \hat{\theta}] .
\] (9.91)

We cannot yet combine the meridional and vertical derivative terms to obtain an advective form.

At this point we need to introduce a mass-weighted zonal mean, defined by

\[
\hat{A} = \frac{[m A]}{[m]} .
\] (9.92)

Using the definition (9.92), we can write

\[
m A = [m A] + (m A)^* = [m] \hat{A} + (m A)^* ,
\] (9.93)

and

\[
[m A B] = [m A] [B] + [(m A)^* B^* ]
\]

\[
= [m] \hat{A} [B] + [(m A)^* B^* ] ,
\] (9.94)

where \(B\) is a second arbitrary variable. Using (9.93) and (9.94), we can rewrite (9.91) as

\[
[m] \frac{d[M]}{dt} + \frac{1}{a \cos \phi \partial \phi} \left\{ ([m] \hat{v} [M] + [(m v M^*)]) \cos \phi \right\} - \frac{[M]}{a \cos \phi \partial \phi} ([m] \hat{v} \cos \phi)
\]

\[
= - \frac{\partial}{\partial t} [m^* M^*] + \frac{\partial}{\partial \theta} \left[ p^* \frac{\partial z^*}{\partial \lambda} \right] + [M] \frac{\partial}{\partial \theta} [(m \hat{\theta})] - \frac{\partial}{\partial \theta} \left\{ [m] \hat{\theta} [M] + [(m \hat{\theta})^* M^*] \right\}.
\] (9.95)

The meridional and vertical derivatives can now be combined. We also divide by \([m]\), simplify, and rearrange, obtain
Wave–Mean Flow Interactions

An Introduction to the General Circulation of the Atmosphere

Here all of the eddy terms have been collected on the right-hand side, and the non-eddy terms have been collected on the left-hand side.

Now define the isentropic Eliassen-Palm flux as

\[
\mathbf{F} \equiv (0, F_\varphi, F_\theta),
\]

\[
F_\varphi \equiv -[(mv)^* M^*],
\]

\[
F_\theta \equiv \left[p^* \frac{\partial z^*}{\partial \lambda}\right] - [(m\dot{\theta})^* M^*].
\]

Its divergence is given by

\[
\nabla \cdot \mathbf{F} = -\frac{1}{\cos \varphi \partial \varphi} \frac{\partial}{\partial \varphi} \{(mv)^* M^* \cos \varphi\} + \nabla \left[p^* \frac{\partial z^*}{\partial \lambda}\right],
\]

where it is understood that the meridional derivative is taken along an isentropic surface.

With these definitions, (9.96) can be written as

\[
\frac{\partial \hat{M}}{\partial t} + \frac{[v]}{a \cos \varphi \partial \varphi} ([M] \cos \varphi) + \hat{\dot{\theta}} \frac{\partial \hat{M}}{\partial \theta} = \frac{1}{[m]} \left( -\frac{\partial}{\partial t} [m^* M^*] + \nabla \cdot \mathbf{F} \right).
\]

The “eddy mass flux” \((mv)^*\), which was introduced in (9.95) and appears in (9.99), obviously has zonal mean equal to zero. This means that it does not transport any mass on the average. It is a “mixing” or “diffusive” or “sloshing” mass flux. A similar comment applies to the eddy vertical mass flux, \((m\dot{\theta})^*\).

Consider a steady state (or time average) with no heating. Then the continuity equation (9.90) reduces to

\[
\frac{\partial}{\partial \varphi} ([mv] \cos \varphi) = 0.
\]

Since \([mv] \cos \varphi = 0\) at both poles, we conclude that
from which it follows that

\[ \hat{\nu} = 0 \text{ for all } \varphi. \quad (9.102) \]

This means that for steady flow the meridional advection term of (9.99) vanishes, which is quite amazing. Naturally the tendency term of (9.99) is also zero in this case. It follows that for steady flow in the absence of heating, the Eliassen-Palm flux is non-divergent:

\[ \nabla_\theta \cdot \mathbf{F} = 0 \text{ for steady flow without heating.} \quad (9.103) \]

Another way of saying this is that, for steady flow in the absence of heating, the zonally averaged meridional angular momentum transport 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.

### 9.7 Potential vorticity fluxes

Consider the momentum equation in isentropic coordinates:

\[
\frac{\partial \mathbf{V}}{\partial t} + \mathbf{k} \times (\eta \mathbf{V}) + \nabla_\theta (K + s) + m\frac{\partial \mathbf{V}}{\partial \theta} + \mathbf{F} = 0.
\]

(9.104)

Here \( s \) is the Montgomery Stream Function,

\[ \eta \equiv \zeta + f \quad (9.105) \]

is the absolute vorticity, where

\[ \zeta \equiv \mathbf{k} \cdot (\nabla_\theta \times \mathbf{V}) \quad (9.106) \]

and \( \mathbf{F} \) is the friction vector. Note that \( \zeta \) is the vorticity along an isentropic surface.

To derive the vorticity equation, we apply \( \mathbf{k} \cdot \nabla_\theta \times \) to (9.104). Starting from standard vector identities, we can show that

\[ \mathbf{k} \cdot \nabla \times (\mathbf{k} \times \mathbf{A}) = \nabla \cdot \mathbf{A}, \quad (9.107) \]

and

\[ \mathbf{k} \cdot \nabla \times \mathbf{A} = -\nabla \cdot (\mathbf{k} \times \mathbf{A}), \quad (9.108) \]
where $\mathbf{A}$ is an arbitrary horizontal vector. With these relations, and using the fact that $f$ is independent of time, we can show that

$$\frac{\partial \eta}{\partial t} + \nabla_\theta \cdot (\nabla \eta) = \nabla \cdot \left\{ k \times \left( m \frac{\partial \mathbf{V}}{\partial \theta} + \mathbf{F} \right) \right\}. \quad \text{(9.109)}$$

We now define

$$q = \frac{\eta}{m} \quad \text{(9.110)}$$

as the Ertel potential vorticity. Here

$$m \equiv \frac{\partial p}{\partial \theta} > 0 \quad \text{(9.111)}$$

is the pseudo-density. Then (9.109) becomes

$$\frac{\partial (mq)}{\partial t} + \nabla_\theta \cdot (m \mathbf{V} q) = \nabla \cdot \left\{ k \times \left( m \frac{\partial \mathbf{V}}{\partial \theta} + \mathbf{F} \right) \right\}. \quad \text{(9.112)}$$

This equation was derived and discussed by Haynes and McIntyre (1987). According to (9.109), the average over the sphere of the mass-weighted potential vorticity on an isentropic surface cannot change, even in the presence of heating and friction. The amazing result that neither vertical advection of momentum nor friction can alter the mass-weighted average PV on an isentropic surface is called the impermeability theorem.

Now take the zonal average of (9.112), which yields

$$\frac{\partial [mq]}{\partial t} + \frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} \left( [mq] \cos \varphi \right) = \frac{-1}{a \cos \varphi} \frac{\partial}{\partial \varphi} \left( m \frac{\partial u}{\partial \theta} + F_\lambda \right) \cos \varphi. \quad \text{(9.113)}$$

Using (9.88), (9.93) and (9.94), we can rewrite (9.113) as

$$\frac{\partial [mq]}{\partial t} + \frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} \left( \left[ ([m][\hat{v}][q] + [(m \mathbf{v})^* \mathbf{q}^*]) \cos \varphi \right) \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \right. \r
which implies that

\[(mv)^{\ast}q^{\ast} = 0 \text{ for all } \varphi,\]

(9.116)
i.e., in the absence of heating and friction the meridional flux of potential vorticity due to the eddies vanishes at all latitudes. Compare with (9.16), which was derived for the quasi-geostrophic case.

9.8 The quasi-biennial oscillation

The zonal winds of the tropical stratosphere undergo an amazing oscillation with a period of about 26 months, called the “Quasi-Biennial Oscillation” or QBO. The QBO was discovered by Richard Reed in about 1960 (see the early review by Reed, 1965). The early 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 exists, as shown in Fig. 9.15. The observations show that a “ring” of air in the tropical stratosphere, extending over all longitudes, reverses the direction of its zonal motion roughly every two years. The oscillation is like a giant zonally oriented Ferris Wheel that periodically reverses the direction of its rotation. The winds shift from about 20 m s\(^{-1}\) westerly to 20 m s\(^{-1}\) easterly, and back again, so the changes are quite large. They are observed to propagate down from the middle stratosphere to near the tropopause. Corresponding oscillations are seen in other stratospheric fields, and to a much lesser extent also in the troposphere.

A theory of the QBO was proposed by Lindzen and Holton (1968) and Holton and Lindzen (1972). According to this theory, the oscillation is due to the interactions of upward-propagating Kelvin and Yanai waves with the mean flow. More recently, it has been proposed that eastward- and westward-propagating gravity waves play an important role, and the Kelvin and Yanai waves are now being de-emphasized.

Both Kelvin waves and Yanai waves, which were discussed in Chapter 7, are observed to produce upward energy propagation into the stratosphere. The source of wave energy must, therefore, be in the troposphere, and is believed to be associated with cumulus convection. Each type of wave can be blocked by a critical level where \([u] - c = 0\); here \(c\) is the phase speed of the wave. Kelvin waves, with \(c > 0\), can propagate through easterlies, for which \([u] - c < 0\), but not through westerlies strong enough to make \([u] - c > 0\). Because Kelvin waves propagate towards the east, relative to the mean flow, the “isentropic form drag” paradigm tells us immediately that upward propagating Kelvin waves transport westerly momentum upward, i.e., they deplete the westerly momentum at lower levels, and deposit this momentum aloft, thus tending to produce westerlies aloft and easterlies below. Therefore upward propagating Kelvin waves tend to produce a westerly acceleration at the base of a layer of westerlies, thus causing the westerlies to descend with time, as observed in the QBO. Recall that Kelvin waves do not involve fluctuations of the meridional wind; because of this they produce no meridional eddy transports of any variable. See Fig. 9.16.

Yanai waves, with \(c < 0\), can propagate through westerlies, for which \([u] - c > 0\), but not through easterlies strong enough to make \([u] - c < 0\). Yanai waves that transport energy upward transport westerly momentum downward. The westward-propagating Yanai

---

1. Here “upward propagating” means that the wave energy flux is upward.
waves are damped in easterlies at the level where $\frac{\partial u}{\partial z} - c = 0$, and so they produce an easterly acceleration near the base of a layer of descending easterlies. Westward-propagating
9.8 The quasi-biennial oscillation

Recent work suggests that they are in fact important for the QBO.

Plumb (1984) produced a remarkable laboratory simulation of the QBO in an annular tank full of stratified salt water. In his 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.

For many years, atmospheric general circulation models failed to simulate the QBO. Recently, however, M. Takahashi (1996) has produced reasonably successful simulations of the QBO using a full atmospheric general circulation model of the troposphere and stratosphere. Earlier Carriolle et al. (1993) had produced a much weaker and less realistic but still promising simulation, and Takahashi and Shiobara (1995) had produced a QBO in a simplified GCM. 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

Figure 9.16: Energy and momentum fluxes associated with a Kelvin wave. The wavy lines show the heights of isentropic surfaces. Energy is propagating both upward and downward, away from the energy source. Where the energy flux is upward, the momentum flux is also upward, and vice versa. Westerly momentum is transported away from the energy source region, which feels an easterly acceleration to compensate.

inertia-gravity waves will produce a similar effect, and recent work suggests that they are in fact important for the QBO.
There is some evidence that a phenomenon similar to the QBO, with a period of about four Earth years, is at work in the atmosphere of Jupiter (Orton et al., 1991; 1994; Leovy et al., 1991; Friedson, 1999; Flasar et al., 2004).

### 9.9 Blocking

Blocking (Rex, 1950 a, b) is a low-frequency, middle-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. (A high index regime is strongly zonal.) Blocking highs tend to fluctuate in intensity, weakening briefly and then re-intensifying. They tend to 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, and specifically in the eastern North Atlantic, the eastern North Pacific, and northern Asia. Southern Hemisphere blocking does occur, most commonly near New Zealand. Blocking can occur in both summer and winter, but is more common in winter. Wintertime blocking events sometimes appear to be associated with stratospheric sudden warmings (Quiroz, 1986). The formation of a block can be associated with upward propagation of a Rossby wave, which then interacts with the stratospheric zonal flow to produce a breakdown of the polar night vortex, i.e. a “Sudden Warming” event.

Blocks strongly influence weather patterns for one or more weeks at a time, sometimes in association with such things as persistent droughts (e.g. Green, 1977). If the formation and dissipation of blocks can be predicted, then weather patterns can be predicted with improved skill. 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 improved dramatically, apparently as a result of increased model resolution.

What causes the formation of a blocking anticyclone? The current view is that blocks originate when particularly intense cyclones advect air with low Ertel potential vorticity from the subtropics (where low potential vorticity is the norm) into middle latitudes (where low potential vorticity is an anomaly). An example of this is shown in Fig. 9.17 taken from Shutts (1986). At the same time, a region of high potential vorticity air is advected into position on the equatorward side of the low potential vorticity anomaly. The block thus has a dipole structure in terms of potential vorticity. Near the longitude of the block, the potential vorticity 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. 9.18 shows the corresponding sea level pressure and 500 mb height fields. The dipole structure 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.

In this example, a vigorous, short-lived, and small-scale cyclogenesis event (not shown in the figures) created a larger-scale anticyclonic perturbation that persisted for about 8 days, while the cut-off low survived for only about two days. 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
Figure 9.17: Contours of the fourth root of the Ertel potential vorticity on the 320 K isentropic surface, which slopes from 200 mb near the pole to 600 mb in the tropics. Panels a–e are for successive days. The contour interval is 0.001 in SI units. The bold contour is 0.005. From Shutts (1986).
lows are disrupted by convection, while highs can survive because convection is suppressed there. Observations suggest that the smaller-scale transient eddies that are steered around a blocking high actually help to maintain the high. Blocks may therefore be examples of upscale energy transport, which, as discussed in Chapter 10, is expected on theoretical grounds in two-dimensional turbulence. It has been suggested that blocking anticyclones in the Earth’s atmosphere are dynamical cousins of the “Great Red Spot” that has persisted in the atmosphere of Jupiter for at least several hundred Earth years.

There are two main theories of blocking dynamics, which look quite different but are not necessarily in conflict with each other. These are briefly summarized below.

Charney and DeVore (1979; hereafter CDV) suggested that in the presence of topography the large-scale circulation can adopt either of two equilibrium states, and that one of these corresponds to blocking, while the other represents a more typical and more zonal flow. The basic idea is that when the configuration of the mean flow is “right” there can be

---

Figure 9.18: (a) Mean sea level pressure field for 15 February 1983, 12Z. Contour interval: 5 mb. (b) Height of the 500-mb surface for 15 February, 12Z. Contour interval: 8 dam. From Shutts (1986).
stationary waves that are resonantly forced by the topography. The waves feed back on the mean flow, however. 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 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 configurations. These are referred to as “multiple equilibria.” Because the CDV 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.

\[ h(x) \equiv h_T \cos(K_m x) . \]  

(9.117)

The motion is described by a stream function, \( \psi \), which is assumed to be of the form

\[ \psi(x, y, t) = -U(t)y + A(t) \cos(K_m x) + B(t) \sin(K_m x) . \]

(9.118)

Think of this as a highly truncated functional expansion. The stream function is defined by the relations \( u = -\frac{\partial \psi}{\partial y} \) and \( v = \frac{\partial \psi}{\partial x} \); using these formulae, we see that the zonal component of the flow described by (9.118) 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. On the other hand, for \( B > 0 \) 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 \( t \):

\[ v(x, t) = K_m \{ -A(t) \sin(K_m x) + B(t) \cos(K_m x) \} . \]

(9.119)

By substituting (9.118) into a suitable nonlinear vorticity equation describing a Rossby wave flow with friction, CDV showed that the wave motion satisfies

\[ \frac{1}{K_m} \frac{dA}{dt} + \left( \frac{v}{K_m} \right) A + \left( U - \frac{\beta}{K_m^2} \right) B = 0 , \]

(9.120)

\[ \frac{1}{K_m} \frac{dB}{dt} - \left( U - \frac{\beta}{K_m^2} \right) A + \left( \frac{v}{K_m} \right) B + \left( \frac{f_0 h_T}{K_m H} \right) U = 0 , \]

(9.121)

while the zonal flow obeys

\[ \frac{dU}{dt} = \left( \frac{f_0 h_T K_m}{4H} \right) B - v(U - U^* ) . \]

(9.122)
In (9.121) - (9.122), the terms involving $v$ represent friction. In (9.122), the “$B$” term represents the orographic form drag (or “mountain torque”) that the mountain exerts on the mean flow when the wave is oriented with the trough over the mountain, and the $U^*$ term represents a “momentum forcing” that maintains the mean flow against friction. Note that (9.120) and (9.121) “blow up” if the wave number of the topography, $K_m$, is equal to zero. This simply means that the wave solution does not occur in the absence of topography, i.e. the wave is topographically forced. Also note that (9.120) and (9.121) 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 (9.120)-(9.122). These equilibria can be found by setting the time-rate of change terms to zero, solving the resulting linear system (9.120)-(9.121) for $A$ and $B$ as functions of $U$, and selecting the appropriate value of $U$ by requiring that (9.122) also be satisfied. Fig. 9.20 shows an example, in which the straight line represents the $(B, U)$ pairs that satisfy (9.122), while the peaked line represents the $(B, U)$ pairs that satisfy (9.120)-(9.121). 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 (9.122) represents a drag on $U$. 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

---

An Introduction to the General Circulation of the Atmosphere
An Introduction to the General Circulation of the Atmosphere

363

9.9 Blocking

Blocking 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 of them) is an artifact. Nevertheless the theory continues to be cited frequently as an important background concept for the interpretation of blocking. A number of authors have searched for evidence of multiple discrete equilibria in various datasets, with results that are suggestive but not entirely convincing.

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 property of modons is that they are resistant to disruption by perturbations. A modon “translates” relative to the mean

Figure 9.20: From Speranza (1986). Equilibrium solutions of (9.120)–(9.122), found by setting the time-rate of change terms to zero, solving the resulting linear system (9.120)–(9.121) for $A$ and $B$ as functions of $U$, and selecting the appropriate value of $U$ by requiring that (9.122) also be satisfied. The straight line represents the $B$, $U$ pairs that satisfy (9.122), while the peaked line represents the $B$, $U$ pairs that satisfy (9.120)–(9.121). The figure shows that there are three equilibrium solutions, but it can be shown that the middle one is unstable.

An Introduction to the General Circulation of the Atmosphere
flow. 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.

We can identify the following issues in connection with blocking, among others:

- 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?
- 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. We do not yet have a full understanding of blocking, however.

9.10 Summary

Waves and other eddies produce important effects on the large-scale circulation of the atmosphere. Important fluxes are associated with a wide variety of waves, including gravity waves, Rossby waves, and Kelvin waves. Momentum fluxes and temperature fluxes can tend to produce mutually counteracting effects, so that the mean zonal flow and temperature may not be altered. The appreciable effects of the eddies on the mean flow are typically associated with developing or decaying eddies, rather than steady, equilibrated eddies.

Two particularly important examples of wave mean-flow interactions in the general circulation are stratospheric sudden warmings and the quasi-biennial oscillation. Both involve wave propagation from the troposphere into the stratosphere.

Problems

1. Solve the one-dimensional Charney-Devore model, as given by the steady-state versions of Eqs. (9.120) - (9.122). Make a plot like that shown in Fig. 9.20. In order to do this you will have to choose appropriate values for the various parameters of the model. Explain your choices.

2. Show that $\psi_z^* = \frac{g T^*}{f_0 T_S}$.

3. Derive Eq. (9.17) of the notes.
CHAPTER 10

The general circulation as turbulence

10.1 Energy and enstrophy cascades

Consider small-scale three-dimensional motion. The relevant momentum equation is

$$\frac{\partial \mathbf{V}}{\partial t} = - (\mathbf{V} \cdot \nabla) \mathbf{V} - \nabla \left( \frac{\delta p}{\rho_0} \right) + g \frac{\delta \theta}{\theta_0} \mathbf{k} + \nu \nabla^2 \mathbf{V}. \tag{10.1}$$

Here \( \nu \) is the (molecular) kinematic viscosity, which we assume to be a constant; we have used the Boussinesq approximation, for simplicity; and we have neglected the effects of rotation. We also adopt the Boussinesq form of the continuity equation:

$$\nabla \cdot \mathbf{V} = 0. \tag{10.2}$$

Taking \( \nabla \cdot \) (10.1), and using (10.2), we obtain a diagnostic equation for \( \delta p \):

$$\nabla^2 \left( \frac{\delta p}{\rho_0} \right) = \nabla \cdot \left[ - (\mathbf{V} \cdot \nabla) \mathbf{V} + g \frac{\delta \theta}{\theta_0} \mathbf{k} + \nu \nabla^2 \mathbf{V} \right]. \tag{10.3}$$

This shows that \( \delta p \) is entirely determined by the motion. It plays only a passive role. We can eliminate it by taking \( \nabla \times \) (10.1):

$$\frac{\partial \omega}{\partial t} = - (\mathbf{V} \cdot \nabla) \omega + (\omega \cdot \nabla) \mathbf{V} + \nabla \times \left( g \frac{\delta \theta}{\theta_0} \mathbf{k} \right) + \nu \nabla^2 \omega. \tag{10.4}$$

Here \( \omega \) is the three-dimensional vorticity vector, and we have used (10.2).

When \( \nu \) is sufficiently small (i.e. when the Reynolds number, \( \frac{VL}{\nu} \), is sufficiently large), the flow described by (10.4) becomes turbulent, due to shearing instability. The mechanism of shearing instability is illustrated in Fig. 10.1. The upper panel of the figure shows a balanced vortex sheet which is considered to extend to infinity in both directions. The sheet is in balance 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. If one vortex is perturbed
The general circulation as turbulence

upward, however, as in the middle panel, it is carried to the left by the combined effects of the vortices left behind. After being displaced to the left, it experiences a net upward advection, away from the sheet. This means that the initial upward perturbation is amplified, and so 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.

\[ \text{Infinite vortex sheet (balanced)} \]

\[ \cdots \circ \circ \circ \circ \circ \circ \cdots \]

Disturb one vortex...

\[ \cdots \circ \circ \circ \circ \circ \circ \cdots \]

This leads to:

\[ \cdots \circ \circ \circ \circ \circ \circ \cdots \]

Figure 10.1: The basic mechanism of shearing instability.

The simplified discussion of shearing instability given above refers only to vorticity advection, i.e. to the \(- (\nabla \cdot \nabla) \omega\) term of (10.4). The buoyancy term of (10.4) 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;” buoyancy acts as an indirect source of turbulence, but not as a direct source. The buoyancy term of (10.4) is linear; as discussed later, turbulence is intrinsically nonlinear. 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 shearing instability. In short, buoyancy creates conditions in which turbulence can

An Introduction to the General Circulation of the Atmosphere
develop through shearing instability, but buoyancy does not actually produce the turbulence itself. Further discussion of the buoyancy term of (10.4) is outside the scope of this course.

The \((\omega \mathbf{\bullet} \nabla)\mathbf{V}\) term of (10.4) represents both stretching and twisting; it vanishes in two-dimensional flows, and is not essential for the generation of turbulence. It can be written as

\[
(\omega \mathbf{\bullet} \nabla)\mathbf{V} = |\omega| \partial \mathbf{V} / \partial s ,
\]

where \(s\) is a curvilinear coordinate in the direction of the vorticity vector. Note that the vector velocity appears. We can subdivide \(|\omega| \partial \mathbf{V} / \partial s\) into two effects:

i) Stretching, \(e_s |\omega| \partial v_s / \partial s\), where \(v_s\) is the component of \(\mathbf{V}\) in the direction of \(\omega\) and \(e_s\) is a unit vector parallel to \(\omega\); see Fig. 10.2.

\[\begin{align*}
&v_n \quad v_s \\
&v_s \quad v_n \quad \omega
\end{align*}\]

\[\text{Figure 10.2: The components of the velocity normal and tangent to the vorticity vector, and the stretching and twisting processes associated with these velocity components.}\]

Positive stretching \((\partial v_s / \partial s > 0)\) causes \(\partial \omega / \partial t > 0\). There is a tendency for the vorticity field and \(\partial v_s / \partial s\) to be positively correlated, because viscosity causes convergence into regions of positive vorticity, and divergence from regions of negative vorticity. This will be proven below. A consequence is that stretching causes \(\omega\) to increase in an average sense. As demonstrated below, we can also say that the stretching term causes the mean of the squared vorticity to increase; the squared vorticity is called the “enstrophy.”

ii) Twisting, \(e_n |\omega| \partial v_n / \partial s\), where \(v_n\) is the component of \(\mathbf{V}\) normal to the vector \(\omega\) and \(e_n\) is a unit vector perpendicular to \(\omega\). Twisting changes the direction of \(\omega\), but not its magnitude. You should prove this as an exercise.
To derive an equation for the time change of the enstrophy, we dot (10.4) with the vorticity vector, \( \omega \):

\[
\frac{\partial}{\partial t}\left( \frac{1}{2}|\omega|^2 \right) = - (V \cdot \nabla) \left( \frac{1}{2} |\omega|^2 \right) + |\omega|^2 \frac{\partial v_s}{\partial s} + \omega \cdot \left[ \nabla \times \left( \frac{g \delta \theta}{\theta_0} \right) \right].
\]

(10.6)

Here we have exposed the effects of vortex stretching on the enstrophy, using

\[
\omega \cdot [(\omega \cdot \nabla) V] = |\omega|^2 \frac{\partial v_s}{\partial s},
\]

(10.7)

which you should prove as an exercise, and also we have used

\[
\omega \cdot [v \nabla^2 \omega] = v \omega \cdot [\nabla \cdot (\nabla \omega)]
\]

(10.8)

\[
= v \{ \nabla \cdot [\omega \cdot (\nabla \omega)] - ((\nabla \omega) \cdot \nabla) \cdot \omega \}. \]

The term \(-v[(\nabla \omega) \cdot \nabla] \cdot \omega \) represents enstrophy dissipation, and is always an enstrophy sink.

Now use the continuity equation to rewrite (10.6) in flux form, drop the buoyancy term, and combine the two flux-divergence terms:

\[
\frac{\partial}{\partial t}\left( \frac{1}{2}|\omega|^2 \right) = \nabla \cdot \left\{ -v \frac{1}{2}|\omega|^2 + v \omega \cdot (\nabla \omega) \right\} + |\omega|^2 \frac{\partial v_s}{\partial s}
\]

(10.9)

\[
- v[(\nabla \omega) \cdot \nabla] \cdot \omega + \omega \cdot \left[ \nabla \times \left( \frac{g \delta \theta}{\theta_0} \right) \right].
\]

In a steady state, the left-hand side of (10.9) vanishes, and when we average over the whole domain the flux divergence terms on the right-hand side vanish as well, so that we are left with

\[
\int_M \left[ |\omega|^2 \frac{\partial v_s}{\partial s} - v[(\nabla \omega) \cdot \nabla] \cdot \omega \right] dM = 0.
\]

(10.10)

Because the viscous term of (10.10) is a sink of enstrophy, we can conclude that, on the average, the stretching term has to be a source of enstrophy, i.e.

\[
\int_M |\omega|^2 \frac{\partial v_s}{\partial s} dM > 0.
\]

(10.11)
10.2 Nonlinearity and scale interactions

Note that this follows from our assumption of a steady state. Recall that the rate of production of vorticity by vortex stretching is \( |\omega| \frac{\partial \nu_s}{\partial s} \). Eq. (10.11) implies that in a statistically steady turbulence there is a tendency for \( \frac{\partial \nu_s}{\partial s} \) to be positive where \( |\omega| \) is larger than average, and for \( \frac{\partial \nu_s}{\partial s} \) to be negative where \( |\omega| \) is smaller than average. This means that the rich get richer -- strong vortices are stretched and made even stronger.

We conclude that the net effect of \( (\omega \cdot \nabla)\mathbf{V} \) is to increase \( |\omega| \) 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 affecting the kinetic energy tends to shift the spectrum of kinetic energy towards shorter scales. To see this, note that the ratio of the kinetic energy to the enstrophy has the units of a length squared:

\[
\frac{1}{2} \frac{|V|^2}{|\omega|^2} \sim L^2.
\]  

(10.12)

We can interpret \( L \) as the width of the most energetic vortices. Circulations with lots of vorticity per unit kinetic energy have their energy concentrated in relatively small vortices. When \( |\omega|^2 \) increases while \( \frac{1}{2} |V|^2 \) remains constant, \( L \) must decrease. This systematic migration of the kinetic energy from larger to smaller scales, due to the action of the stretching term, is called a kinetic energy “cascade”. This term evokes a waterfall in which a single stream falls off a cliff and repeatedly splits on the way down, as the water strikes rocks or other obstructions. 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 leads to kinetic energy dissipation.

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 modes given by

\[
A(x) = \hat{A} e^{ikx} \quad \text{and} \quad B(x) = \hat{B} e^{ilx},
\]  

(10.13)

respectively. Here the wave numbers of modes \( A \) and \( B \) are denoted by \( k \) and \( l \), respectively. If we combine \( A \) and \( B \) linearly, e.g., form

\[
\alpha A + \beta B,
\]  

(10.14)

where \( \alpha \) and \( \beta \) are spatially constant coefficients, then no “new” waves 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, \( k + l \):

\[
\alpha A \beta B,
\]
Other nonlinear operations such as division, exponentiation, etc., will also generate new wave
numbers.

### 10.3 Two-dimensional turbulence

In two dimensions, the stretching/twisting term of (10.4) is zero, because $\frac{\partial}{\partial s}$ is zero. It follows that *vorticity and enstrophy are both conserved under inertial processes in two-dimensional turbulence*. Here the term “inertial processes” refers to advection and rotation, only. Of course, kinetic energy is also conserved under inertial processes. Since both kinetic energy and enstrophy are conserved under inertial processes in two-dimensional flows, $L$ is also conserved. The implication is that *kinetic energy does not cascade in frictionless two-dimensional flows*; it “stays where it is” in wave number space.

When the effects of viscosity are included in two-dimensional turbulence, they act most effectively on the smallest scales, which are the scales on which 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 $L$ will tend to increase as a result of enstrophy dissipation. This means that the scale of the most energetic eddies 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.

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, as shown in Fig. 10.3. By “equally spaced” we mean that

$$\lambda_2 - \lambda_1 = \lambda_3 - \lambda_2 \equiv \Delta \lambda.$$ 

(10.16)

The enstrophy, $E$, is

\[ \hat{A} \hat{B} = \hat{A} \hat{B} e^{i(k+l)x}. \] 

(10.15)
Two-dimensional turbulence

\[ E = E_1 + E_2 + E_3 , \quad (10.17) \]

and the kinetic energy is

\[ K = K_1 + K_2 + K_3 . \quad (10.18) \]

It can be shown that

\[ E_n = \lambda_n^2 K_n , \quad (10.19) \]

where \( \lambda_n \) is a wave number, and the subscript \( n \) denotes a particular Fourier component.

Consider an inertial process such that kinetic energy and enstrophy are redistributed in a two-dimensional frictionless flow, i.e.

\[ K_n \rightarrow K_n + \delta K_n , \quad (10.20) \]

\[ E_n \rightarrow E_n + \delta E_n . \quad (10.21) \]

Because kinetic energy and enstrophy are both conserved under two-dimensional inertial processes, we have

\[ \sum \delta K_n = 0 , \quad (10.22) \]

\[ \sum \delta E_n = 0 . \quad (10.23) \]

From (10.22) we see

\[ \delta K_1 + \delta K_3 = -\delta K_2 , \quad (10.24) \]

Note from (10.19) that

\[ \delta E_n = \lambda_n^2 \delta K_n . \quad (10.25) \]

From (10.23) and (10.25), we obtain

\[ \lambda_1^2 \delta K_1 + \lambda_3^2 \delta K_3 = -\lambda_2^2 \delta K_2 \]

\[ = \lambda_2^2 (\delta K_1 + \delta K_3) . \quad (10.26) \]

Collecting terms in (10.26), we find that
The general circulation as turbulence

Using (10.16), we can simplify (10.27) to

\[
\frac{\delta K_3}{\delta K_1} = \frac{\lambda_2^2 - \lambda_1^2}{\lambda_3^2 - \lambda_2^2}.
\]  

(10.27)

This shows that the energy transferred to higher wave numbers (\(\delta K_3\)) is less than the energy transferred to lower wave numbers (\(\delta K_1\)). This conclusion rests on both (10.22) and (10.23), 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. This is sometimes called an “anti-cascade.”

We now perform a similar analysis for the enstrophy. As a first step, we note from (10.25) and (10.28) that

\[
\frac{\delta E_3}{\delta E_1} = \frac{\lambda_2^2 (\lambda_2 + \lambda_1)}{\lambda_1^2 (\lambda_3 + \lambda_2)}
\]

\[
= \frac{(\lambda_2 + \Delta \lambda)^2 (\lambda_2 - \frac{1}{2} \Delta \lambda)}{(\lambda_2 - \Delta \lambda)^2 (\lambda_2 + \frac{1}{2} \Delta \lambda)}.
\]

(10.29)

To show that this ratio is greater than one, we demonstrate that \(\frac{\delta E_3}{\delta E_1}\) can be written as \(a \cdot b \cdot c\), where \(a, b,\) and \(c\) are each greater than one. We can choose:

\[
a = \frac{\lambda_2 + \Delta \lambda}{\lambda_2 + \frac{1}{2} \Delta \lambda} > 1, \quad b = \frac{\lambda_2 - \frac{1}{2} \Delta \lambda}{\lambda_2 - \Delta \lambda} > 1, \quad c = \frac{\lambda_2 + \Delta \lambda}{\lambda_2 - \Delta \lambda} > 1.
\]

(10.30)

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.

10.4 Quasi–two–dimensional turbulence

Large-scale motions are quasi–two–dimensional, so it is reasonable to suppose that an enstrophy cascade occurs for them just as it does for purely two-dimensional motion. Consider the potential vorticity equation derived earlier:
An Introduction to the General Circulation of the Atmosphere

10.4 Quasi-two-dimensional turbulence

\[ \frac{\partial (mq)}{\partial t} + \nabla_\theta \cdot (mVq) = \nabla \cdot \left( m\dot{\theta} \frac{\partial V}{\partial \theta} + F \right). \]  

(10.31)

and the continuity equation in isentropic coordinates

\[ \frac{\partial m}{\partial t} + \nabla_\theta \cdot (mV) + \frac{\partial (m\dot{\theta})}{\partial \theta} = 0. \]  

(10.32)

By combining (10.31) and (10.32), we obtain an advective form of the potential vorticity equation:

\[ m \frac{\partial q}{\partial t} + mV \cdot \nabla_\theta q = q \frac{\partial (m\dot{\theta})}{\partial \theta} + \nabla \cdot \left( m\dot{\theta} \frac{\partial V}{\partial \theta} + F \right). \]  

(10.33)

According to (10.33), in the absence of heating and friction the potential vorticity is conserved following a particle; keep in mind that in the absence of heating the particle moves along an isentropic surface. 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 a negative anomaly is generated in the upper troposphere.

Multiplying (10.33) by \( q \) gives

\[ m \frac{\partial (q^2/2)}{\partial t} + mV \cdot \nabla_\theta (q^2/2) = q^2 \frac{\partial (m\dot{\theta})}{\partial \theta} + q \nabla \cdot \left( m\dot{\theta} \frac{\partial V}{\partial \theta} + F \right). \]  

(10.34)

The quantity \( q^2 \) is called the “potential enstrophy.” Eq. (10.34) shows that potential enstrophy tends to be generated locally where the heating rate increases upward, and destroyed where the heating rate decreases upward. 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, potential enstrophy is generated in the lower troposphere and destroyed in the upper troposphere.

Using (10.32) in (10.34), we can convert back to flux form:

\[ \frac{\partial (mq^2/2)}{\partial t} + \nabla_\theta \cdot \left( mVq^2/2 \right) = \frac{q^2}{2} \frac{\partial (m\dot{\theta})}{\partial \theta} + q \nabla \cdot \left( m\dot{\theta} \frac{\partial V}{\partial \theta} + F \right). \]  

(10.35)

In the absence of heating and friction, (10.35) implies that

\[ \frac{\dot{\theta}}{mq^2} = \text{constant}. \]  

(10.36)

Here \( \frac{\dot{\theta}}{mq^2} \) denotes an average over an isentropic surface. For a nondivergent flow (i.e., nondivergent on an isentropic surface) with no heating, (10.32) can be written as
\[ \frac{Dm}{Dt} = 0, \]  
\[ (10.37) \]
i.e., \( m \) is constant following the motion. Since  
\[ mq^2 = \frac{\eta^2}{m}, \]  
\[ (10.38) \]
(10.32) reduces to  
\[ \eta^2 = \text{constant for nondivergent flow with no heating.} \]  
\[ (10.39) \]
It follows that, in this limiting case,  
\[ L^2 \equiv \frac{K^\theta}{\eta^2} = \text{constant}. \]  
\[ (10.40) \]
This is the same conclusion that we reached earlier, for the case of two-dimensional turbulence.

For a more general divergent flow, we can write  
\[ \eta^2 < \left( \frac{m_{\text{max}}}{m} \right) \eta_{\text{max}}^2, \]  
\[ (10.41) \]
where \( m_{\text{max}} \) is an appropriately chosen constant upper bound on \( m \). Since \( m_{\text{max}} \) is a constant, (10.41) can be written as  
\[ \eta^2 < m_{\text{max}} \left( \frac{\eta^2}{m} \right)^\theta = m_{\text{max}} (mq^2)^\theta = \text{constant}. \]  
\[ (10.42) \]
According to (10.41) \( \eta^2 \) has an upper bound. Then from (10.40) we see that \( L^2 \) has a lower bound. This shows that even for divergent motion the kinetic energy cannot cascade into arbitrarily small scales; it tends to stay in the large scales. This is a partial explanation for the “smoothness” of the general 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
effect is that $L$ must increase. (Our conclusion above 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, “eddy mergers”
tend to occur in two-dimensional or geostrophic turbulence.

In an idealized numerical simulation, McWilliams (1984) showed that through vortex
mergers pure two-dimensional turbulence gradually organizes itself into a single large
cyclone-anticyclone pair. His results are shown in Fig. 10.4.

As mentioned earlier, it has been hypothesized that in a real atmosphere, the cyclonic
member of the pair would be killed off by convective mixing, so that only the anticyclonic
member would survive. 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.

### 10.5 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 $\varepsilon$ be the rate of kinetic
energy dissipation per unit mass, and $\nu$ the kinematic viscosity:

$$\varepsilon \sim L^2 T^{-3},$$

$$\nu \sim L^2 T^{-1}.\quad (10.43)$$

Kolmogorov (1941) hypothesized that for locally homogeneous and isotropic three-
dimensional turbulence, the turbulence statistics are determined by $\varepsilon$ and $\nu$; according to his
hypothesis, if $\varepsilon$ and $\nu$ are known, no other information is needed. Note that $\varepsilon$ is a property of
the flow, while $\nu$ is a property of the fluid.

The **viscous subrange** consists of scales for which both $\varepsilon$ and $\nu$ 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 $\nu$ and $\varepsilon$:

$$\lambda_k = \left(\frac{\nu^3}{\varepsilon}\right)^{\frac{1}{4}}.\quad (10.44)$$

This is called the **Kolmogorov microscale**.

If $\varepsilon$ increases for a given $\nu$, then $\lambda_k$ has to become smaller. For the Earth’s
atmosphere, a typical value of $\lambda_k$ is in the range $10^{-2}$ m to $10^{-3}$ m.

Kolmogorov further hypothesized that there exists an **inertial subrange** -- a range of
scales -- within which 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. The inertial subrange consists of homogeneous isotropic eddies that are larger than $\lambda_k$. 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, $\varepsilon$. Note that $\varepsilon$ represents dissipation that occurs on scales outside of (shorter than those of) the inertial subrange.

In the Earth’s atmosphere, kinetic energy is typically generated at lower wave numbers, and migrates to higher wave numbers by way of the inertial subrange, and is finally dissipated at the smallest scales. This is illustrated in Fig. 10.5. We want to find $K(k)$, the spectrum of energy for an inertial subrange. This is essentially the Fourier transform of the spatial distribution of energy. It has units of energy per unit mass per unit wave number. Dimensionally,

$$K(k) \sim L^3 T^{-2}.$$  \hspace{1cm} (10.45)

Since the turbulence statistics in the inertial subrange depend only on $\varepsilon$, $K(k)$ can depend only on $\varepsilon$ and $k$. Assuming that

$$K(k) \sim \alpha \varepsilon^m k^n,$$  \hspace{1cm} (10.46)

where $\alpha$ is a nondimensional constant, we find that $m = \frac{2}{3}$, and $n = -\frac{5}{3}$. In other words,

$$K(k) \sim \alpha \varepsilon^{\frac{2}{3}} k^{-\frac{5}{3}}.$$  \hspace{1cm} (10.47)

This conclusion is well supported by measurements.

We know that kinetic energy does not cascade in two-dimensional turbulence, but we have seen that 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 in this inertial subrange the turbulence statistics are determined by the rate of *enstrophy* dissipation. Using methods similar to those used to derive (10.47), we can show that the kinetic energy spectrum follows $k^{-3}$. This conclusion is supported by numerical simulations of two-dimensional turbulence. Note that $k^{-3}$ falls off 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. This is 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 (10.47) 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.

Now look at Fig. 10.6, which is a more refined version of Fig. 10.5. In Fig. 10.6, 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. For $k$ less than $k_B$, energy flows upscale, away from the baroclinic energy source. 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 $k$ slightly less than $k_C$, we have a three-dimensional inertial range with $K(k) \sim k^{-5/3}$ and energy flowing upscale. Clearly, the shape of the spectrum has to change somewhere between $k_B$ and $k_C$.

Lilly (1998) discussed observations that appear to show such a change in the spectral

Figure 10.6: Schematic indicating the kinetic energy spectrum in the Earth’s atmosphere, as implied by dimensional analysis. Baroclinic instability adds energy at wave number $k_B$, and convection adds energy at wave number $k_C$. See text for details.
slope, as seen in Fig. 10.7. The spectrum follows $k^{-5/3}$ for scales less than about 100 km, and $k^{-3}$ 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. 10.6 and Fig. 10.7 should be clear.

Consider a planet on which baroclinic instability is inactive, perhaps because solar heating is weak and internal heat sources are dominant, primarily driving small-scale convection. In such a case, $k_B$ drops out of the problem, and we might expect to see $K(k) \sim k^{-5/3}$ all the way from $k_C$ up to near planetary scales. Jupiter’s atmosphere may
work something like this.

There is another complication. Rhines (1975) pointed out that sufficiently large vortices will feel the $\beta$ effect, which exerts a “restoring force” opposing large meridional excursions by the particles making up the vortex. This means that the $\beta$ effect will tend to resist meridional widening of vortices beyond some limit. 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

$$k \sim k_\beta \equiv \frac{\beta}{\sqrt{U}}.$$  \hspace{1cm} (10.48)

This scale is sometimes called the “Rhines barrier.” For $k < k_\beta$, further meridional broadening is resisted by $\beta$, while longitudinal broadening can continue. The eddies therefore become elongated in the zonal direction and ultimately give rise to alternating zonal jets of width $k_\beta^{-1}$, like those seen on Jupiter. In recent years 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.

### 10.6 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. 10.8, for kinetic energy and enstrophy only. The slope of the kinetic energy spectrum is plotted as a function of height in Fig. 10.9. 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. 10.10; 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.

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 8, 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).
10.7 The general circulation as a blender

10.7.1 What does the blender blend?

To the extent that the eddies of the general 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 conservative variable. 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 $M \equiv a \cos \phi (u + \Omega a \cos \phi)$, is somewhat more conservative, because it is not affected by the Coriolis acceleration; a little thought shows 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 very conservative dynamical variable. In the absence of heating, particles stay on
Figure 10.9: 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 10.10: Observed nonlinear kinetic energy exchanges as a function of height. From Boer and Shepherd (1983).
their isentropic surfaces, so we would expect the large-scale turbulence to homogenize potential vorticity (and other conservative variables) along isentropic surfaces. Fig. 10.11, from Sun and Lindzen (1994), provides evidence that this does happen.

10.7.2 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:

$$\frac{d\mathbf{V}}{dt} + q\mathbf{k} \times (m\mathbf{V}) + \nabla \left( \frac{1}{2} \mathbf{V} \cdot \mathbf{V} + s \right) = 0. \quad (10.49)$$

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:

$$\frac{d\mathbf{V}}{dt} + (q - D)\mathbf{k} \times (m\mathbf{V}) + \nabla \left( \frac{1}{2} \mathbf{V} \cdot \mathbf{V} + s \right) = 0 \quad (10.50)$$

We can interpret $(q - D)$ as a modified potential vorticity. When we dot (10.50) with $\mathbf{V}$ to form the kinetic energy equation, the term involving $D$ drops out, regardless of the 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 (10.50) and using continuity:
\[
\frac{Dq}{Dt} = \frac{1}{m} \nabla \cdot (Dm \nabla V) .
\]

(10.51)

The right-hand side of (10.51) vanishes, as it should, for \( D = 0 \). Let the potential enstrophy averaged over an entire isentropic surface (globally) be given by

\[
Z(\theta) = \frac{1}{S} \int_{S} \frac{q^2}{2} m \, dS .
\]

(10.52)

Then (10.51) implies that

\[
\frac{dZ(\theta)}{dt} = -\int_{S} (Dm \nabla \cdot \nabla q) \, dS .
\]

(10.53)

To guarantee dissipation of \( Z(\theta) \), we choose

\[
D = \tau \nabla \cdot \nabla q ,
\]

(10.54)

with \( \tau \geq 0 \). By substituting (10.54) into (10.53), we find that

\[
\frac{dZ(\theta)}{dt} = -\int_{S} (m \nabla \cdot \nabla q)^2 \, dS .
\]

(10.55)

This guarantees that \( Z(\theta) \) decreases with time for \( \tau > 0 \). We can recover potential vorticity conservation and potential enstrophy conservation by setting \( \tau = 0 \). From (10.54), we see that

\[
q - D = q - \tau (\nabla \cdot \nabla q) \equiv q_{\text{anticipated}} .
\]

(10.56)

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. (10.56) is equivalent to

\[
\frac{q_{\text{anticipated}} - q}{\tau} = -\nabla \cdot \nabla q .
\]

(10.57)

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 this approach gives realistic kinetic energy and enstrophy spectra in a numerical model.

**10.7.3 The Gent–McWilliams theory of tracer transports along isentropic surfaces**

In isentropic coordinates, the conservation of an arbitrary tracer of mixing ratio \( \tau \) is
expressed by

\[ \left( \frac{\partial}{\partial t} \right)_\theta (m\tau) + \nabla_\theta \cdot (mV\tau) + \frac{\partial}{\partial \theta} (m\theta\tau) = mS(\tau), \quad (10.58) \]

\( S(\tau) \) is the source or sink of tracer per unit mass and per unit time. Putting \( \tau = 1 \) and \( S(1) = 0 \), we recover the continuity equation:

\[ \left( \frac{\partial}{\partial t} \right)_\theta m + \nabla_\theta \cdot (mV) + \frac{\partial}{\partial \theta} (m\theta) = 0. \quad (10.59) \]

In the absence of heating, (10.59) reduces to

\[ \left( \frac{\partial}{\partial t} \right)_\theta m + \nabla_\theta \cdot (mV) = 0. \quad (10.60) \]

We can write the mass flux \( mV \) as the sum of its zonal mean and the departure from the zonal mean:

\[ mV = [mV] + (mV)^* \]
\[ = [m][V] + [m^*V^*] + (mV)^*. \quad (10.61) \]

Here we have used

\[ [mV] = [m][V] + [m^*V^*], \quad (10.62) \]

which is similar to a relation discussed way back in Chapter 2. Recall that \([m^*V^*]\) is called the “bolus mass flux,” and \([m] \) is called the “bolus velocity.” For a steady state in the absence of heating, we can show (from the continuity equation) that

\[ [m^*V^*] = -[m][V]. \quad (10.63) \]

The flux of an advected quantity \( \tau \) can be written as

\[ mV\tau = ([mV] + (mV)^*)([\tau] + \tau^*) \]
\[ = [mV][\tau] + [mV]\tau^* + (mV)^* [\tau] + (mV)^* \tau^* \quad (10.64) \]

Zonal averaging gives
Substituting from (10.62), we obtain

\[
[mV \tau] = (mV)[\tau] + [(mV)^* \tau^*].
\] (10.65)

This shows that the mean value of \([\tau]\) is advected by both the mean mass flux and the bolus mass flux. The third term can be interpreted as “diffusion,” because the mass flux involved, i.e. \((mV)^*\), has zero mean. Substituting (10.66) into the zonal mean of (10.59), we obtain

\[
\left(\frac{\partial}{\partial \theta}\right)[m\tau] + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} [([m][v] + [m\nu^*][\tau]) \cos \phi] + \frac{\partial}{\partial \theta}[m\theta \tau]
= [mS(\tau)] - \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{([mV]^* \tau^* \cos \phi) \cos \phi\}. \] (10.67)

Similarly, the zonally averaged continuity equation can be written as

\[
\left(\frac{\partial}{\partial t}\right)[m] + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{([m][v] + [m\nu^*]) \cos \phi\} + \frac{\partial}{\partial \theta}[m\theta] = 0. \] (10.68)

In order to make use of these ideas, we have to determine \([m\nu^*]\). Gent and McWilliams (1990) suggested

\[
[m\nu^*] = -\frac{\partial}{\partial \theta} \left( \frac{\kappa}{a} \frac{\partial}{\partial \phi} [p] \right), \] (10.69)

where \(\kappa\) is a non-negative parameter. Substitution of (10.69) into (10.68) gives

\[
\left(\frac{\partial}{\partial t}\right)[m] + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \{([m][v] \cos \phi) + \frac{\partial}{\partial \theta}[m\theta]\}
= \frac{1}{a^2 \cos \phi} \frac{\partial}{\partial \phi} \left\{ \frac{\partial}{\partial \theta} \left( \frac{\kappa}{\partial \phi} [p] \right) \cos \phi \right\}. \] (10.70)

In case \(\kappa\) is spatially constant, this reduces to
10.8 The limits of deterministic weather prediction

\[
\left( \frac{\partial}{\partial t} \right)_\theta [m] + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} ([m] [v] \cos \phi) + \frac{\partial}{\partial \theta} [m \dot{\theta}]
\]
\[
= \frac{\kappa}{a^2 \cos \phi} \frac{\partial}{\partial \phi} \left( \frac{\partial}{\partial \phi} \left( m [p] \right) \cos \phi \right).
\]

(10.71)

Since

\[
[m] = -\frac{1}{g \theta} [p],
\]

(10.72)

this is equivalent to

\[
\left( \frac{\partial}{\partial t} \right)_\theta [m] + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} ([m] [v] \cos \phi) + \frac{\partial}{\partial \theta} [m \dot{\theta}]
\]
\[
= -\frac{g \kappa}{a^2 \cos \phi} \frac{\partial}{\partial \phi} \left( \frac{\partial}{\partial \phi} [m] \right) \cos \phi.
\]

(10.73)

This equation describes “diffusion of \([m]\);” it shows that the effect of (10.69) is to drive \([m]\) towards uniformity on \(\theta\)-surfaces, which implies driving \([p]\) towards uniformity on \(\theta\)-surfaces. Recall that a state in which \([p]\) is uniform along \(\theta\)-surfaces is one in which the available potential energy is zero. This means that the effect of (10.69) is to remove available potential energy from the system. An interpretation is that the available potential energy has been converted into subgrid-scale kinetic energy, through subgrid-scale baroclinic instability.

The theory outlined above, due to Gent and McWilliams (1990), is being used to parameterize eddy-transports in ocean models, with good results.
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:

1) 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.

2) The observing system

   • Imperfect measurements of the initial conditions (e.g., imperfect thermometers)

   • Imperfect spatial coverage of the initial conditions

   • Mistakes

3) 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 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 nonperiodic solutions. The equations of the model are remarkably simple:
Here, \( \sigma \), \( \tau \), and (10.75) are parameters. The numerical values given in (10.75) are particular choices (not unique ones) that lead to nonperiodic behavior. A solution of (10.74) is shown in Fig. 10.12. 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 very 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. 10.13. 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. This is a second reason for calling Fig. 10.13 the

\[
\begin{align*}
\dot{X} &= -\sigma X + \sigma Y, \\
\dot{Y} &= -XZ + rX - Y, \\
\dot{Z} &= XY - bZ.
\end{align*}
\]

(10.74)

Here

\[
\sigma = 10, \ b = \frac{8}{3}, \text{ and } r = 24.74
\]

(10.75)
Butterfly Attractor.

There are many kinds of instability, acting on virtually all spatial scales. Small-scale shearing 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 as a result of instabilities, deterministic physical laws do not necessarily permit deterministic predictions (Harman, 1998, 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 in Chapter 7. 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. 10.14. 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” which is a limit beyond which the weather is unpredictable in principle. This limit is a property of the atmosphere itself, and so it cannot be circumvented no matter what method we use to make our forecast; it is not a limitation that is intrinsic to numerical weather prediction or any other specific forecast technique. It applies to all methods.

An Introduction to the General Circulation of the Atmosphere
Figure 10.12: The Butterfly Attractor of Lorenz (1963), obtained as the solution of (10.74). From Drazin (1992).
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. 10.15.

Figure 10.15: 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).

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 circulation is said to be turbulent if its predictability time is shorter than the time scale of interest. For example, baroclinic storms in midlatitude winter 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. With this definition, turbulence is in the eye of the beholder.

10.9 Quantifying the limits of predictability

Three approaches to determine the limits of predictability were discussed by Lorenz (1969):

10.9.1 The dynamical approach

In this 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 general circulation models to study the growth of small perturbations. Fig. 10.16 shows some of their results, obtained with an early version of the

Figure 10.16: 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).
UCLA general circulation model (GCM). The rms 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 more modern example of the dynamical approach is was discussed by Shukla (1981, 1985). Fig. 10.17 is taken from the paper of Shukla (1985). The results shown in the figure are based on computations with a GCM. The model is used to perform multiple simulations, differing only through very small perturbations of the initial conditions. The results given in 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, 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. Note that at all latitudes the saturation values are close to one. This means that the errors stop growing when they become as large as the standard deviation.

Fig. 10.18 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 is no better than a forecast based on persistence. After 30 days, the numerical model still has some skill at the lower wave numbers. For higher wave numbers, the model’s skill disappears more quickly.

10.9.2 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 mb 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 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.
The “dynamical-empirical” approach is the most difficult of the three to understand. The basic idea, as first conceived by Lorenz (1969), is to derive an equation for the time change of “error kinetic energy” (using a model). We then Fourier transform the error energy...
The general circulation as turbulence

An Introduction to the General Circulation of the Atmosphere

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 (~40 m). We then use the resulting semi-empirical equation to draw conclusions about error growth. The dynamical-empirical approach shows that errors in the smallest scales amplify most quickly and soon dominate. Again, the dynamical-empirical approach suggests that the limit of deterministic predictability is about two to three weeks.

A 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. 10.19. 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

Figure 10.18: 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-mb 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).]
10.9 Quantifying the limits of predictability

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:

$$E_{1,2}^2 = N^{-1} S^{-1} \sum_{i=1}^{N} \int (z_{i,1} - z_{i,2})^2 dS .$$

(10.76)

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 forecasts were perfect, $E_{1,2}$ would be zero. If the 1-day forecasts were perfect but the two-day forecasts were not, $E_{1,2}$ would be positive, etc.

Figure 10.19: 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

$$E_{j,k}^2 = N^{-1} S^{-1} \sum_{i=1}^{N} \int (z_{i,j} - z_{i,k})^2 dS .$$

(10.77)

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. If $E_{j,k} > 0$, the implication is that $k$-day forecasts are less skillful, on the average, than $j$-day forecasts. 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. 10.20. Even if the model’s simulated climate is perfect, $E_{0,k}$ will be different from zero because of the limit of deterministic predictability.

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 $E_{j,k}$ “should be” independent of $j$. It is not, at least not for
“small” $j \ (\leq 20)$. 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.

As a special case, illustrated in Fig. 10.21, $E_{k-1,k}$ compares $(k-1)$-day forecasts with $k$-day forecasts for the same day. For example, putting $k = 10$, $E_{9,10}$ 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, 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 $E_{j,k}$ to approach a constant, i.e., to become independent of the actual values of $j$ and $k$, so long as $j \neq k$. We can say that $\lim_{k \to \infty} E_{j,k}$ (for $j \neq k$ and $j \neq 0$) measures the variability of $z$ in the model’s climate.

For a perfect model, we would have

$$\lim_{k \to \infty} E_{0,k} = \lim_{k \to \infty} E_{k-1,k} \ ,$$

(10.78)

because the model’s climate would be identical to the true climate. For an imperfect model, on the other hand, we expect

$$\lim_{k \to \infty} E_{0,k} > \lim_{k \to \infty} E_{k-1,k} \ ,$$

(10.79)
because the model's climate differs from the true climate. This is a key point.

Figure 10.21: The growth of $E_{0,k}$ and $E_{k-1,k}$ with time.

To explore these ideas with real data, Lorenz chose 1 Dec 80 - 10 March 1981 (100 days) as the verification dates. Because he worked with 100 verification days, $N = 100$ in (10.77). 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. 10.22. Each point represents an average over 100 pairs of forecasts.

The thin curves in Fig. 10.22 appear to run nearly parallel to each other, in the sense that $\frac{dE}{dk} = f(E)$. In other words, for any given value of $E$ the thin curves all have about the same slope. This means that the rate of error growth is strongly influenced by the size of the error; of course we know that this 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. To do this, he essentially fit a curve to the error growth rates in the ECMWF forecasts. He hypothesized that, for a perfect model

$$\frac{dE}{dk} = aE - bE^2. \quad (10.80)$$

Here $E \equiv E_{j,k}$, where $k - j = \text{constant}$, and so $\frac{dE}{dk} \equiv (E_{j+1,k+1} - E_{j,k})$ per day. Notice that the second index minus the first index is the same for both $E$'s, so we are following the
The general circulation as turbulence

An Introduction to the General Circulation of the Atmosphere

“thin curves.” Eq. (10.80) says that the rate of error growth is determined by the site of the error. For small $E$, the exponential growth rate is $a$. For larger $E$, saturation occurs so that

$$E_{sat} = \frac{a}{b}. \quad (10.81)$$

This means that $a/b$ is the maximum error for the perfect model. Note from (10.80) 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 $f(E) = aE - bE^2$ against $\frac{dE}{dk}$ as evaluated from the data. See Fig. 10.23. In performing this curve fit, we find the “best” values of $a$ and $b$. This calculation can be interpreted as 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. 10.24 shows the same data as Fig. 10.22, plotted in a different way. The dots correspond to the thin curves in Fig. 10.22, and so represent the “perfect model.” The crosses correspond to the thick curve in Fig. 10.22, 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.

In a later publication, Lorenz discussed a later set of more skillful forecasts made with an improved model and started from better observed initial conditions. He found that the heavy curve Fig. 10.22 came down, 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 shift of the thin lines suggest that the newer model’s weather is less active than that of the old model.

Figure 10.23: Diagram illustrating the expected variation of the error growth rate \( \frac{dE}{dk} \) with the magnitude of the error, \( E \).

\[
\frac{dE}{dk} = aE - bE^2
\]

\( dE/dk > 0 \) even for \( E = 0 \).
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.
Is the climate state unique and deterministic, for a given set of “external parameters” such as the solar constant and the Earth’s present orbital configuration? A system possessing a single stable climate for a given set of external parameters is called “transitive.” Here the “external parameters” are any factors that influence the climate but are not normally considered part of the climate system. Examples are the rate of energy output of the sun, and the angular velocity of the Earth’s rotation. Some parameters, like the composition of “dry air,” may or may not be considered external.

A system possessing more than one stable climate for a given set of external parameters is called “intransitive.” For a given set of external parameters, an intransitive system can assume any of $N$ possible states, where $N > 1$, and will remain in the same state indefinitely unless forced out by some external cause.

A system that exhibits occasional unforced transitions between two or more distinct climates is called “almost intransitive.” It may exhibit multiple climates if finite time-averaging intervals are used, but, an almost-intransitive system has only one climate if an indefinitely long averaging interval is used.

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 have the potential to make seasonal forecasting and/or climate change prediction possible. First, the system has components with very long memories, including especially the ocean. At present, however, we lack the observations needed to fully initialize these components of the system.

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 in Fort Collins, is that prediction based on solution of an initial value problem? 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?

These examples illustrate that predictions can be based on knowledge of an external forcing (e.g., the seasonal cycle, or a change in solar output) that changes “slowly” and in a predictable way. Here we have in mind that the forcing changes slowly enough so that the climate system has time to adjust to the variations.

We can thus distinguish two types of climate prediction.

- **Prediction of the equilibrated state (after transient adjustment period) that results from a steady forcing anomaly.** This is not an initial value problem.

- **Prediction of the transient adjustment period itself, as a response to steady or time-varying forcing.** This may or may not be an initial value problem, depending on how rapidly the forcing varies, compared with the internal adjustment times of the system.

In “Dynamical Extended-Range Forecasting” (DERF), long-range forecasts are made for the statistics of the weather, by long-range forecasts are not made for individual synoptic events. The period of the forecast ranges upwards from 10 days to a year or so. The physical basis is the “long memory” of the external boundary conditions that provide persistently
anomalous forcing to the atmosphere, such as sea surface temperature, ground wetness, sea ice, etc. (e.g., Palmer, 1993; Palmer and Anderson, 1995; Barnett, et al. 1996; Stockdale, et al., 1998).

An example of the ideas underpinning DERF is shown in Fig. 10.25, which is taken from the work of Shukla (1985). The upper panel shows the observed sea-surface temperature anomaly for January 1983, relative to long-term climatology. The warm temperatures in the tropical eastern Pacific are characteristic of El Niño, as will be discussed in detail later. The middle panel shows the associated rainfall anomaly as simulated by a general circulation model, and the lower panel shows the observed rainfall anomaly as inferred from outgoing long-wave radiation during winter 1983. This study illustrates that anomalous precipitation in the tropics can be successfully predicted on the basis of anomalous sea surface temperature patterns of the type associated with El Niño. It does not necessarily follow that we can also predict the effects of El Niño on precipitation elsewhere in the world; nor does it follow that all types of sea surface temperature anomalies produce predictable effects on the climate; nor does it follow that all climate anomalies are forced by or even associated with sea surface temperature anomalies.

Fig. 10.26, which is taken from Charney and Shukla (1981), illustrates that DERF may be more feasible for the tropics than for middle latitudes. The two panels of the figure show the simulated and observed zonally-averaged standard deviations, $\sigma_m$ and $\sigma_o$, as functions of latitude, and their ratio, for the mean July sea-level pressure; and the rainfall. The observed values are for land stations. The simulation results are based on runs of an atmospheric general circulation model, and are for grid-points over land. The model runs used the same sea surface temperatures in all cases, but the initial conditions varied. For middle latitudes, the simulated and observed standard deviations are quite comparable, while in the tropics the simulated values are much smaller than observed. The implication is that something is causing the sea level pressure and rainfall to vary more in the real tropics than they do in the simulations. Charney and Shukla (1981) hypothesized that the explanation is that in the real world the sea surface temperatures vary noticeably from year to year, e.g. due to El Niño, whereas in the simulations no such variations occurred.

One important implication is that the observed variations of tropical sea level pressure and rainfall are largely due to changes in sea surface temperature; this suggests that DERF can work in the tropics. In the middle latitudes, however, it appears that most of the variability of sea level pressure and rainfall can be explained without resorting to sea surface temperature variations, since in fact the model was able to reproduce the observed variability of sea level pressure and rainfall even though no sea surface temperature perturbations were prescribed as forcing in the simulations. The implication is that in midlatitudes the “signal” of sea surface temperature forcing is swamped by other variability, which we might call “noise,” so that midlatitude DERF may be difficult.

Fig. 10.27, taken from Lau (1985), shows temporal variations of monthly indices of the observed sea surface temperature (top) and the simulated zonal wind at 200 and 950 mb, precipitation, east-west sea level pressure gradient across the South pacific, 200 mb 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 general circulation model. 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
controlled 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.

10.11 The World’s Simplest GCM

Further studies of atmospheric predictability have been performed by using highly simplified models. The “World’s Simplest GCM” (Lorenz, 1975, 1984, 1990) is described by the following equations:

Figure 10.26: Model and observed zonally-averaged standard deviations, $\sigma_m$ and $\sigma_o$, as functions of latitude, and their ratio, for: (a) mean July sea-level pressure; and (b) rainfall. Observed values are for land stations and model values are for grid-points over land. From Charney and Shukla (1981).
Figure 10.25: (a) Observed sea–surface temperature anomaly (K) for January 1983. (b) Simulated rainfall anomaly (mm day$^{-1}$). (c) Observed rainfall anomaly (mm day$^{-1}$) as inferred from outgoing long–wave radiation during winter 1983. From Shukla (1985).
Figure 10.27: Temporal variations of monthly indices of sea surface temperature, zonal wind at 200 and 950 mb, precipitation, east-west sea level pressure gradient across the South Pacific, 200 mb 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).
Here \( X \) 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”), and \( G \) represents land-sea contrast (another “forcing”). For \( F = G = 0 \), the model has the trivial steady solution \( X = Y = Z = 0 \). For \( G = 0 \), we can have a steady solution with \( Y = Z = 0 \) even though \( X \neq 0 \); in this case, \( X \) has the steady solution \( X = F \). This eddy-free solution can be unstable, however. Instability can lead to the growth of \( Y \) and \( Z \). We interpret “large” \( F \) (~8) as “winter,” and “small” \( F \) (~6) as “summer.” Compare (10.82) with (10.74).

For fixed \( F \), there can be one, two, or three steady solutions, depending on the value of \( G \), as shown in Fig. 10.28. Here \( F = 2 \), so this is “super summer.” The steady solutions shown in Fig. 10.28 are found most easily by fixing \( X \) and solving for \( G \), \( Y \), and \( Z \).

Fig. 10.29 demonstrates that the model exhibits sensitive dependence on 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. 10.30 shows two “summer” solutions started from 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. 10.31 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. 10.32 shows results from a six-year run, in which two different kinds of 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 $\sigma$ as the (nondimensional) standard deviation of $X$ within the period July through September. Active summers have large $\sigma$, and inactive summers have small $\sigma$. We can calculate one value of $\sigma$ for each summer, or in other words one for each year. Fig. 10.33 shows the variations of $\sigma$ (dimensionless) with time in years, in a 100-year numerical solution of (10.82), for the conditions of Fig. 10.32. Active and inactive summers alternate, at irregular intervals.

10.12 Pushing the attractors around

Palmer (1993, 1999) has argued that the climate system can be viewed as occupying a collection of attractors. 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 which 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 the model given by (10.74):
10.12 Pushing the attractors around

Figure 10.32: The variations of $X$ (dimensionless) with $t$ (months) in a 6-year numerical solution of (10.82), with $A = 0.25$, $b = 4.0$, $F = 7.0 + 2.0 \cos(2\pi t/\tau)$, 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).

![Figure 10.32](image)

Figure 10.33: The variations of $\sigma$ (dimensionless) with $t$ (years) in a 100-year numerical solution of (10.82), for the conditions of Fig. 6, where $\sigma$ is the standard deviation of $X$ within the period July through September. From Lorenz (1990).

![Figure 10.33](image)

Figure 10.32: The variations of $X$ (dimensionless) with $t$ (months) in a 6-year numerical solution of (10.82), with $A = 0.25$, $b = 4.0$, $F = 7.0 + 2.0 \cos(2\pi t/\tau)$, 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).

\[ X = -\sigma X + \sigma Y + f_0 \cos \theta , \]
\[ \dot{Y} = -XZ + rX - Y + f_0 \sin \theta , \]  
\[ \dot{Z} = XY - bZ . \]  

(10.83)
Here $f_0$ is a forcing which tries to push $X$ and $Y$ in the direction of the angle $\theta$ in the $(X, Y)$ plane. In (10.74), $f_0 = 0$. Fig. 10.34 shows how different choices of $\theta$ 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 $\theta$ changes, the locations of the maxima of the PDFs do not change much. This means that the attractors of the model are insensitive to $\theta$. Note, however, that the maxima become stronger or weaker as $\theta$ 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. Also notice that there is some tendency for the solution to be pushed in the direction in which the forcing acts.

10.13 Summary

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 shearing 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 (some nuts to the contrary notwithstanding). 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 predictably to a sufficiently strong external forcing, provided that the forcing itself is predictable.

**Problems**

1. Summarize the basic dynamical differences between two-dimensional and three-dimensional turbulence. In your answer, be sure to discuss:
   a) vortex stretching and twisting;
   b) vorticity conservation;
   c) the shape of the energy spectrum;
   d) the existence of an energy cascade.

2. Analyze the effects of buoyancy on a spherical bubble of warm air, with uniform potential temperature $\theta$, sitting in an infinite environment of uniform potential temperature $\theta - \Delta \theta$, where $\Delta \theta > 0$. Draw a sketch indicating:
   i) Where vorticity generation occurs.
   ii) The sign of the vorticity generation.
   iii) Where vorticity generation does not occur.

3. Show that in two-dimensional flow the variance of a passive scalar, per unit wave
4. Show that in two-dimensional flow the geopotential variance, per unit wave number, varies as $k^{-5}$.

5. Consider Lorenz’s butterfly model:

$$X = -\sigma (X - Y)$$

$$\dot{Y} = -XZ + rX - Y$$

$$\dot{Z} = XY - bZ$$

Here $\sigma$, $b$ and $r$ are constants. Find the stationary solutions for $X$, $Y$, and $Z$, and investigate the stability of these solutions as functions of $\sigma$, $b$, and $r$. Assume $r \geq 1$, $b \geq 0$.

6. A Fortran code for the “World’s Simplest GCM,” as discussed by Lorenz (1984), can be obtained from the instructor. The equations of the model are:

$$\dot{X} = -Y^2 - Z^2 - aX + aF,$$  \hspace{1cm} (10.84)

$$\dot{Y} = XY - bXZ - Y + G,$$  \hspace{1cm} (10.85)

$$\dot{Z} = bXY + XZ - Z.$$  \hspace{1cm} (10.86)

Time derivatives have been approximated using the fourth-order Runge-Kutta time-differencing scheme, which is discussed in many numerical analysis books, with 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 use 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 “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 $X = 2.5$, $Y = 1$, $Z = 0$.

a) Demonstrate that two identical runs of the “real world” give identical results.
b) Do a test to show that the “real world” exhibits sensitive dependence on initial conditions.

c) Make a 100-day run of the “real world” and save the results for $X$, $Y$, and $Z$ once per simulated day. (The code is already set up to do this.)

Starting from each of the first 90 days of this “real world” weather record, make a ten-day forecast by running the “model.” Save the results from each of these forecasts, once per simulated day.

Produce a plot similar to Fig. 10.22, with at least the curves $E_{0,k}$ and $E_{k-1,k}$. (You may want to do more than just these two.) Here use $X$ as the variable you study, just as Lorenz (1982) used the 500 mb height.

d) Still using $X$ as your variable, plot $\frac{dE}{dk}$ versus $E$, as in Fig. 10.24, for both the “perfect model” and the “imperfect model.” Estimate the value of $a$, the error e-folding rate.

7. Explain as clearly as you can why it is not possible to make skillful deterministic weather forecasts of indefinite range.

8. Describe the methods used by Lorenz to estimate the limits of predictability.

9. Find the equilibrium solutions of Lorenz’s “World’s Simplest GCM,” and analyze their stability.
CHAPTER 11

Tropical atmosphere–ocean interactions

11.1 Introduction

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 2 x 10$^{14}$ kg. The mass of the atmosphere is about 250 times less, roughly 0.8 x 10$^{12}$ kg. Not only is water dense, it has a very high heat capacity: about 4200 J kg$^{-1}$ K$^{-1}$. In contrast, the heat capacity of air (at constant pressure) is only about 1000 J kg$^{-1}$ K$^{-1}$. The total heat capacity of the oceans is thus about 1000 times larger than the total heat capacity of the atmosphere. When the oceans say “Jump,” the atmosphere says “How high?”

The density of the atmosphere is highly variable, especially with height. In contrast, the density of sea water varies by only a few percent throughout the entire ocean; it is a complex but fairly weak function of temperature, salinity, and pressure. Because of the near-incompressibility of water, pressure effects (called “thermobaric” effects) are relatively unimportant; variations of the density 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 the water are altered mainly near the surface; below the top hundred meters or so, the properties of water parcels remain nearly invariant, even over decades or centuries.

The ocean 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. The property of the oceans that most directly affects the atmosphere is the sea surface temperature, which affects the upward longwave radiation, the sensible heat flux, and the latent heat flux. The roughness of the sea surface can also affect the efficiency of near-surface turbulent exchange. The albedo of the ocean is determined in part by the turbidity of the near-surface water, and of course by sea ice, where present. Sea ice also affects the sensible and latent heat fluxes, by presenting to the air a surface which can be much colder than the sea water below; and it can affect the atmospheric turbulence through its roughness and the presence of leads and other breaks in the ice.

Many properties of the atmospheric column can affect the ocean. These include the near-surface temperature, which influences sensible heat exchange and downward long wave radiation; the near-surface humidity which influences latent heat exchange and downward longwave radiation,; and cloudiness, which can affect the surface solar and terrestrial radiation. Precipitation also affects the salinity of the surface waters.

An Introduction to the General Circulation of the Atmosphere
When the wind stress acts on the ocean, the Coriolis acceleration acts to turn the wind-induced surface current so that it does not have the same direction as the surface stress. In the Northern Hemisphere, the Coriolis acceleration causes the surface current to move to the right of the surface wind, while in the Southern Hemisphere it moves to the left.

The ocean circulation is primarily wind-driven, but it does have an important buoyancy-driven component, called the thermohaline 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 this equatorward drift out, 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 either 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.

As already mentioned, the tradewinds induce a westward current in the upper ocean, thus tending to pile up warm water on the western side of each basin. There is a highly concentrated west-to-east return flow very close to the Equator and slightly below the surface, called the Equatorial Undercurrent.

11.2 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 Walker Circulation is driven by the east-to-west sea surface temperature gradient, it also helps to maintain that gradient through mechanisms to be discussed later. For this reason, the Walker Circulation is best understood as a coupled ocean-atmosphere phenomenon. It undergoes strong interannual variability. Fig. 11.1 is a schematic illustration of the Walker Circulation and its relation to the surface wind field in the Southern Hemisphere. 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. Fig. 11.2 shows the observed longitude-height cross sections of the zonal wind and vertical velocity, for January.

The Hadley Circulation is defined in terms of zonal averages, and so a particle participating in the Hadley Circulation through motions in the latitude-height plane cannot "escape" by moving to a different longitude. In contrast, the Walker Circulation is restricted to a narrow band of tropical latitudes, so that a particle participating in the Walker Circulation can escape by moving off to a different latitude; in fact, such meridional escapes are to be expected in view of the strong meridional motions associated with the Hadley Circulation. For this reason, we should not think of the Walker Circulation as a closed "race track;" it is better to view the Hadley and Walker Circulations as closely linked. For instance, 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 journey in the boundary layer.

An early description of the Walker circulation is given by Troup (1965), who described a toroidal circulation spanning the equatorial Pacific ocean. Troup presented data in which the mean 500-300 mb geopotential thickness for 120-140° E was 20 m greater than
that for 80-100° W. By the thermal wind equation (e.g. Holton 1992), this east-west thickness difference implies that the meridional geostrophic wind must decrease with height. Because the vertical gradient of meridional geostrophic wind is related to the zonal temperature gradient by the thermal wind relation, it can be shown that the zonal temperature gradient drives an ascending ageostrophic flow from the warm region to cold region. As Troup noted, the resulting distribution of ageostrophic motions is complicated, because it depends on the varying temperature differences and vertical motions across the different regions of the circulation. Troup found that westerly flow between 500 and 200 mb originates over the Indonesian region and terminates over the central and eastern equatorial Pacific; this upper-level flow is balanced by an easterly return flow, which he described as frictional ageostrophic flow down the pressure gradient. He noted that descent occurs over the central and eastern Pacific, while ascent occurs over the western Pacific and Indonesia.

Bjerknes (1969) theorized that when the equatorial cold tongue is well developed, the cool, dry air just above the surface cannot ascend to join the Hadley circulation. Instead, it 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. Furthermore, the ascending motion in the west would adjust so as to cover a smaller surface area than the descending motion, which develops as a result of the balance between subsidence warming and radiative cooling. Mass continuity demands that the region of intense rising motion must occupy a smaller area than the region of weak sinking motion. 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 which depends on the strength of the angular momentum flux divergence. Apparently, Bjerknes was unaware of the paper by Troup (1965),

Figure 11.1: The Walker and Hadley Circulations. From Philander (1990).
Fig. 11.3 presents scaled $u$-$\omega$ wind vectors in the equatorial $x$-$z$ plane for January 1989 from the ECMWF reanalysis dataset. We define $u$ and $\omega$ as the zonal and vertical components of velocity in pressure coordinates. The vertical velocity was scaled by $-300 \text{ m Pa}^{-1}$ in order to account for the much smaller speed of the vertical motions as compared with the horizontal motions. The figure shows that the rising branch of the Walker circulation is centered on 125° E, while the sinking branch is spread across a wide region between the dateline and 80° W. Within this region, westerlies are present between 100 and 400 mb, and easterlies are confined below 700 mb. Upper-level easterly flow exists between 120°E and 160°E, and appears to be associated with the Australian winter monsoon. The flow is quite weak between 400 and 700 mb. NKVB also described additional cells in the equatorial plane which Bjerknes did not mention. Their data (not shown) indicated the presence of Walker-like circulations over the tropical Atlantic ocean, near the African sector, and the western branch of the Pacific Walker cell.

Fig. 11.4 shows that the 1000-mb winds above the tropical Pacific (between 10° N and 10° S) have an easterly component in both solstitial seasons. For both seasons, easterly flow near the equator occurs 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, a notable characteristic is the strong cross-equatorial flow in the eastern Pacific. The zone of convergence at 1000 mb during the NH summer has moved north

*Figure 11.2: The longitude–height cross–sections of the zonal wind (left, in m s$^{-1}$) and vertical velocity (right, in Pa s$^{-1}$) along the Equator, for January, as analyzed by ECMWF.*

as no reference was made to the earlier study.
11.2 The Walker Circulation

Figure 11.3: January 1989 Equatorial $u-\omega$ vectors in units of m s$^{-1}$. The $\omega$ values, which were originally in units of Pa s$^{-1}$, were scaled by $-300$ m Pa$^{-1}$.
of 10° N over the eastern Pacific. At the latitude of the ITCZ, the easterly fetch originates to the east of Central America during both seasons. If we consider the Walker circulation to occur at near-equatorial latitudes, then easterly flow at 1000 mb cannot originate over the continents because the mountains of Peru act as a vertical barrier on the eastern boundary of the ocean. However, we note that a strong southerly component is evident during both seasons.

Figure 11.4: Streamlines and horizontal wind vectors for the tropical Pacific at 1000 mb for a) January 1989 and b) July 1989. The units are m s$^{-1}$.

Lindzen and Nigam (1987) used a simple model to show that SST gradients are capable of forcing low-level winds and convergence in the tropics. They assumed that near the surface the Coriolis acceleration is balanced by the sum of the horizontal pressure-gradient force and wind stress; this is called an Ekman balance. 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; hereafter N96) compared water-vapor data from the Upper Atmosphere Research 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. 11.6 shows the vertical velocity fields at 300-mb

![Fig. 11.5: Contour plot of mean vertical velocity at 300mb (units, 10−2 Pas−1) from the ECMWF reanalysis dataset for a) January 1989 and b) July 1989. Contour interval is 2 x 10−2 Pa s−1; negative contours are dashed. Data were obtained from NCAR.](image)

from the ECMWF reanalysis dataset for the solstitial months. During January 1989, centers of ascending motion were located near 145° E at latitudes 5° N and 5° S. The SPCZ is clearly evident in the January 1989 data, with a large region of ascending motion that extends southeastward from 145° E to 160° W. A region of strong sinking motion straddles the equator and extends eastward from 160° E. During July 1989, the ascending region remains fixed at 145° E, but the NH and SH centers of ascending motion have merged on the equator. During the NH summer, the ITCZ is well developed at 5° N, and so the zone of sinking motion has slipped southward from its January position, particularly the zone over the central Pacific. 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 and 160° E. West of 160° E, the low-level equatorial winds are very weak. However, easterlies span the equatorial Pacific at 5° S and 5° N.

Fig. 11.6 shows 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 (NH), weak westerly flow appears between 170° W and 140° W poleward of 15° N. In the southern hemisphere (SH), a westerly component of the wind exists south of 5° S to the east of the dateline. An interpretation is that
11.3 The relationship between the Walker Circulation and the sea surface temperature.

The Walker Circulation is an atmospheric phenomenon, so you may be wondering why we discuss it in detail in a chapter dealing with atmosphere-ocean interactions. The reason is that the Walker Circulation is closely tied to east-west sea surface temperature gradients which are produced by atmospheric phenomena including aspects of the Walker Circulation itself. The Walker Circulation can thus be viewed as a phenomenon of the coupled atmosphere-ocean system.

Fig. 11.7 shows that a sea-surface temperature (SST) maximum occurs over the tropical region centered on 120° E, and for this reason the region is known as the tropical Warm Pool. The “cold tongue” is a band of relatively cold waters along the equator that stretches from South America westward to near 160° E. Although a noticeable SST gradient exists along and across the cold tongue, the temperature variation is still much smaller than that which is generally observed in extratropical or polar regions of the globe. The tropical climate is characterized by sea surface and horizontal air temperature gradients which are weak compared to the corresponding mid-latitude gradients. As explained by Charney (1963), for cloud-free regions of the tropics, pressure and temperature gradients must be small compared to those of midlatitudes.
The distribution of tropical convection is strongly related to both the local SST and the SST gradient. The tropical-Pacific Warm Pool is a region of intense deep convection. In Fig. 11.8, 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 convection. From the figure, we see that convection occurs throughout the Warm Pool, and in the South 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.

Ramanathan and Collins (1991) 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.

In the eastern tropical Pacific, stratus clouds in the boundary layer 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 cold-water upwelling in the eastern Pacific, and along the equator in the eastern and central Pacific. The equatorial cold tongue is an example of an effect of cold-water upwelling on the SST distribution.

11.4 Theories of the Walker Circulation

The driving force behind the Walker circulation is the zonal structure of the precipitation rate, and that these variations are balanced by adiabatic heating/cooling due to sinking/rising motions. Over the warm waters of the western Pacific, latent heat release due to intense convection is balanced by adiabatic cooling and ascending motion (Webster 1987). The deep, bright clouds limit the radiative cooling of the atmosphere over the Warm Pool. It is the inability of radiative processes to locally balance the latent heating which gives rise to a tropical circulation. Over the eastern tropical Pacific, where the SST is relatively cold, convection is infrequent, and so a balance between radiation and subsidence exists.

Pierrehumbert (1995; hereafter P95) presented a two-box model of a Hadley/Walker circulation which has strongly influenced recent studies of the tropical climate. Fig. 11.10 presents a schematic of the furnace/radiator-fin model in P95. The model has separate energy budgets for its Cold-Pool and Warm-Pool regions. SSTs for the Cold Pool and Warm Pool were constrained to be those which 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, even though it was discussed in detail. 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 at \( z = z_f/2 \), where \( z_f \) is the assumed height of the tropopause. 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 net radiative flux at the Warm-Pool TOA was assumed to be balanced by a horizontal energy transport to the Cold Pool. The Cold-Pool SST and radiating temperature were then computed under the constraint that the net diabatic cooling must balance the energy imported laterally from the Warm Pool.

The mass flux 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. The mass flux is proportional to the sum of the net diabatic cooling of and specified mid-latitude atmospheric energy transport from the Cold Pool. The potential temperature difference between the inflow and outflow regions of the Cold-Pool atmosphere was specified. Therefore, the mass flux responds only to changes to the net diabatic cooling of the Cold Pool. Because precipitation over the Cold Pool is neglected, the diabatic cooling of the Cold-Pool atmosphere is purely radiative and depends on the Warm-Pool and Cold-Pool SSTs, and on the emissivity of the Cold-Pool atmosphere, which is a prescribed parameter. P95 assumed a uniform vertical temperature profile for the atmosphere which is controlled by
11.4 Theories of the Walker Circulation

Warm-Pool SST, and so the radiating temperature of the Cold-Pool atmosphere depends on

Figure 11.9: Observed annual-mean low-cloud amount (upper panel) and the net effects of clouds on the Earth’s radiation budget (lower panel). Negative values in the lower panel indicate a cooling, i.e. shortwave reflection dominates longwave trapping.

An Introduction to the General Circulation of the Atmosphere
Warm-Pool SST. It can be shown that the horizontal heat transport by the Warm-Pool atmosphere, $F_{ah1}$, is proportional to the diabatic cooling of the Cold-Pool atmosphere and to the ratio of Cold-Pool area, $A_2$, and Warm-Pool area, $A_1$. This ratio is also a prescribed parameter of the model.

The column energy budget for the model is given by

$$H = (Q_{v1} + F_{aexp}) \frac{A_1}{A_1 + A_2} + (Q_{v2} + F_{aexp}) \frac{A_2}{A_1 + A_2}, \quad (11.1)$$

where $Q_v$ is the energy added to the Cold-Pool atmospheric column due to vertical flux convergence and $F_{aexp}$ is the specified, horizontal-mean, net horizontal, atmospheric energy transport into the column. The radiative effects of clouds were ignored. In equilibrium, $H = 0$, and so

$$(Q_{v1} + F_{aexp})A_1 + (Q_{v2} + F_{aexp})A_2 = 0. \quad (11.2)$$

For the case in which $F_{aexp} = F_{oh1} = F_{oh2} = 0$, where $F_{oh}$ is the horizontal energy transport by the ocean, the net horizontal, atmospheric heat transport for the Warm Pool is given by $F_{ah1} = -Q_{v1}$. Obviously, $F_{ah1}$ increases as the ratio $A_2/A_1$ increases for a fixed Cold-Pool radiating temperature, and the climate cools for a fixed radiating temperature. P95 argued that $Q_{vij}$ is bounded due to limits on the OLR in moist atmospheres, i.e. the increase of OLR...
with SST levels off due to the longwave-trapping effect of water vapor, which also increases with SST. As $A_2$ becomes much larger than $A_1$, the horizontal energy transport between boxes must vanish, and radiative equilibrium results. For small $A_2/A_1$, $Q_{v2}$ is not bounded because the radiative temperature of the Cold Pool can be increased as much as needed in order to yield a finite energy transport between boxes, no matter how small $A_2$ becomes. Since the Warm-Pool SST controls the Cold-Pool radiating temperature, the Warm Pool becomes extremely hot in this limit.

P95 also showed that for very small values of the Cold-Pool emissivity, the Warm-Pool SST runs away, because the Cold Pool cannot radiate enough energy to balance the Warm Pool. This is called a runaway greenhouse (Ingersoll 1969). As the Cold Pool’s emissivity is increased, the SSTs of the Warm Pool and Cold Pool decrease and increase, respectively.

For fixed, lateral atmospheric energy transport, a larger $A_2/A_1$ implies a smaller $Q_{v2}$, which can be seen from (11.2). It can be shown that a smaller $Q_{v2}$ leads to a reduced Cold-Pool mass flux. Because the Cold-Pool mass flux was assumed to be proportional to the Cold-Pool surface evaporation rate, increasing $A_2/A_1$ warms the Cold-Pool SST. On the other hand, as the Cold-Pool emissivity increases, the radiating temperature of the Cold-Pool atmosphere must decrease in order to radiate the same amount of energy. 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 Cold-Pool mass flux is realistic. A serious weakness of the model is that it fails to account for cloud-radiative effects in the Warm-Pool region.

Next, we discuss an appealingly simple box model developed by Sun and Liu (1996; hereafter SL) to demonstrate the role of dynamic ocean-atmosphere coupling in SST regulation. SL constructed a three-box model (Fig. 11.11) of the tropical Pacific ocean, which is coupled to a very simple model atmosphere. Two adjacent equal-volume boxes represent the surface layer of the eastern and western Pacific ocean regions, respectively, and a subsurface box represents the equatorial undercurrent. Water is assumed to be advected into the western box from the eastern box, which is fed by upwelling of the equatorial undercurrent. Water returns to the equatorial undercurrent by subduction from the western box. The temperature of the equatorial undercurrent, $T_e$, was specified based on observations. The temperature tendencies for the two surface-layer boxes are assumed to be controlled by dynamical ocean-atmosphere processes and by thermal advection in the ocean. SL crudely parameterized the dynamical processes as a relaxation toward an equatorial equilibrium temperature $T_e$, with inverse time scale $c$. In order to determine $T_e$, the feedbacks due to surface emission, the clear-sky greenhouse effect, the greenhouse effect of clouds, and the cloud shortwave forcing with respect to an SST perturbation at $T_0 = 300$ K (i.e. the partial derivatives) were estimated. Except for the value of the surface emission feedback which is easily calculated, values for the other feedbacks were taken from published estimates. The difference between $T_e$ and $T_0$ was then computed by taking the quotient of the net heating of the ocean-atmosphere column which was evaluated at $T_0$ and the summed feedbacks for an SST perturbation. The advective temperature tendency can be shown to be proportional to the temperature difference between water entering and departing each box. The advection proportionality parameter, $q$, was assumed in turn to be proportional to the temperature difference between the two surface boxes, with a specified constant of proportionality, $\alpha$. Thus, the advective temperature tendencies for the Warm and Cold Pools are given by $\alpha(T_1-$
Tropical atmosphere-ocean interactions

An Introduction to the General Circulation of the Atmosphere

\[ T_2^2 \text{ and } \alpha(T_1-T_2)(T_c-T_2), \] respectively. The rationale for this form of \( q \) is that the strength of the ocean currents is proportional to the surface wind speed, which is assumed to be proportional to the east-west SST gradient.

SL found that the solution of their model is completely determined by a non-dimensional parameter \( \beta = (\alpha/c)(T_c-T_e) \), which gives the strength of the dynamic coupling relative to the thermodynamic forcing. As seen in Fig. 11.12, the east-west SST gradient and the ocean currents are zero for \( \beta < 1 \), and hence the tropical Pacific ocean-atmosphere column is in radiative-convective equilibrium, with equilibrium temperatures \( T_1 = T_2 = T_e \). This contrasts with the results of P95 which show that, due to the strong greenhouse effect, the Warm-Pool ocean-atmosphere column cannot establish radiative-convective equilibrium unless the SST is very warm and the atmosphere is very dry. For \( \beta > 1 \), two solutions exist. The radiative-convective solution still holds, but is unstable to perturbations. As a numerical integration of the model shows, a perturbation of the radiative-convective equilibrium causes the system to evolve to the second solution, which features a finite east-west temperature difference. The SSTs of the two surface boxes are colder than the radiative-convective equilibrium. As shown in Fig. 11.12, ocean currents develop for \( \beta > 1 \), and advection of undercurrent water to the Cold Pool and Cold-Pool water to the Warm Pool leads to colder SSTs in the Cold Pool and Warm Pool, respectively. The “Warm” Pool becomes warmer than the “Cold” Pool because water advected from the Cold Pool to the Warm Pool is warmer than water upwelled to the Cold Pool from the undercurrent. As described in Liu and Huang (1997), this destabilization of the radiative-convective equilibrium can be interpreted as a wind-cold water upwelling positive feedback. SL argue that \( T_c \) increases and \( c \) decreases due to positive radiative feedbacks in the atmosphere, such as those from water vapor and clouds. Since \( \beta \) increases with increasing \( T_c \) and decreasing \( c \), SL assert that strong positive feedbacks of the atmosphere play a role in the evolution of the

Figure 11.11: Schematic of the Sun–Liu coupled model. The boxes represent the atmosphere (light hatching), the Warm Pool (heavy hatching), the Cold Pool (stippling), and the undercurrent (clear box). Heavy arrows denote ocean currents, light arrows denote local heating and the dashed arrow represents the surface winds.
system from radiative-convective equilibrium. The ocean circulation transports heat from the subsurface to surface ocean, and leads to cooler SSTs. SL also showed that as $T_e$ increases, the difference between Warm-Pool SST and $T_e$ increases. This represents a regulation of the Warm-Pool SST, and does not explicitly depend on an atmospheric circulation. Results from a simplified coupled ocean-atmosphere general circulation model (GCM) support these conclusions.

Miller (1997; hereafter M97) extended Pierrehumbert’s model by studying the radiative effects of low clouds in the Cold-Pool region. Extending the basic concepts of P95, Miller constructed a three-box model, which includes energy- and moisture-balance equations for the boundary layer and free troposphere and a surface energy budget for each of three boxes: the updraft region, the Warm Pool, and the Cold Pool. Taking advantage of the small surface area covered by the updrafts, M97 simplified the model for the limit of vanishing updraft surface area. M97 demonstrated that in this limit, the boundary layer and tropopause of the Warm Pool region must be connected by a moist adiabat. Miller assumed that the lapse

![Figure 11.12: The Equilibrium solution from the Sun–Lu coupled model for (a) current strength measured by $q/c$, and for (b) non-dimensionalized SST as given by $T^*_{1} = (T_1 - T_e)/(T_c - T_e)$ and $T^*_{2} = (T_1 - T_e)/(T_c - T_e)$, as functions of $\beta$.](image-url)
rate of the Warm Pool region is moist adiabatic. Following P95, atmospheric dynamics were implicitly included by assuming a uniform free-tropospheric temperature sounding across both the Warm and Cold Pool regions.

The main finding of M97 is that low clouds act as a thermostat for tropical SST. Without a realistic distribution of stratus clouds, the SST was too warm beneath the subsiding branch of the tropical circulation. Although low clouds reduce the surface-absorbed solar radiation locally, M97 also found that the temperature drop in the Warm Pool region was nearly as large as that in the Cold Pool region. In order to obtain a realistic Warm-Pool SST, an additional -5 W m\(^{-2}\) forcing for the Cold Pool, and -12 W m\(^{-2}\) forcing for the Warm Pool were needed. This suggests that additional cloud types contribute to the surface forcing. In contrast, cloud radiative forcing was not required for the model of P95 to simulate realistic SST. This discrepancy suggests that the crude radiative transfer parameterization adopted in P95 likely canceled the impact of not including cloud radiative effects.

11.5 El Niño and the Southern Oscillation

Many studies have explored the processes that cause SST anomalies to develop in the tropical Pacific ocean (Cane and Zebiak 1985; Zebiak and Cane 1987; Battisti 1988; Battisti and Hirst 1989; Neelin and Jin 1993; Anderson and McCreary 1985; Schopf and Suarez 1988; Yamagata and Masumoto 1989; Graham and White 1990; and many others). The phenomena of El Niño and the closely related Southern Oscillation have been beautifully discussed in the book of Philander (1990).

The Southern Oscillation is a systematic shifting of atmospheric mass, back and forth across the Pacific basin. It is illustrated in Fig. 11.13, which shows the correlation of the surface pressure at various locations in the Pacific basin with the surface pressure in Darwin. A large-scale dipole structure is clearly evident. Fig. 11.14 shows year-to-year variations in the sea-level pressures measured at Darwin and Tahiti; they are generally anticorrelated.
11.5 El Niño and the Southern Oscillation

In parallel with these sea-level pressure changes, we see fluctuations of the sea-surface temperature, as shown in Fig. 11.15. In certain years, the SST in the eastern Pacific is anomalously warm; these are called El Niño events. They tend to begin late in the calendar year. We also see some years with anomalously cold temperatures in the eastern Pacific; these are called La Niña events. El Niños are associated with weak upwelling west of South America, and La Niñas are associated with strong upwelling. In some cases, the SST anomalies tend to move westward, as would be expected given the eastward currents driven by the trade winds.

As shown in Fig. 11.16, there are also very strong year-to-year variations of the precipitation in various locations. Heavy precipitation in the central and eastern Pacific is generally correlated with anomalously light precipitation in the western Pacific. In El Niño years, the east is wet and the west is dry.

Fluctuations of the SST in 1982 and 1983 are shown in Fig. 11.17. These years saw a very strong El Niño. See also Fig. 11.18. During the El Niño, the cold tongue of water along
the Equator virtually disappeared. Warm water appears to migrate from east to west. On the other hand, Fig. 11.19 appears to show deep clouds migrating from west to east.
The response of the upper-tropospheric circulation to El Niño is shown in Fig. 11.20.

Figure 11.18: Maps of sea surface temperature anomalies during 1982 and 1983. From Philander (1990).

Twin anti-cyclones straddle the region of enhanced convection. The anticyclonic vorticity can be interpreted as a response to the divergent outflow from the convective region.

11.5.1 Sea surface temperature and thermocline slope
Consider the upper-ocean structure shown in Fig. 11.21. We let $z_T$ denote the depth of the thermocline. The density jump across the thermocline is $\Delta \rho \equiv \rho_D - \rho_S > 0$. The hydrostatic equation is given by

$$dz = \frac{dp}{\rho g}.$$  \hfill (11.3)

There is no minus sign because both depth and pressure increase downwards. If $p = p_{\text{ref}}$ = constant at $z = z_{\text{ref}}$, and with

$$p_S \equiv 0,$$  \hfill (11.4)
which means neglecting the weight of the atmosphere compared with the weight of the water in the column below, we get

Figure 11.20: Streamfunction of the 200 mb winds. The bottom panel is a difference plot. From Philander (1990).
This shows that \( z_S \) becomes more negative, i.e. the sea surface becomes "higher," as the thermocline becomes deeper.

Consider a sloping thermocline below a sloping sea surface, as shown in Fig. 11.22. Define a transformed vertical coordinate which measures the depth below the thermocline:
11.5 El Niño and the Southern Oscillation

The horizontal pressure gradient force can be written as

\[
\frac{-1}{\rho} \left( \frac{dp}{dx} \right)_{z} = \frac{-1}{\rho} \left[ \frac{dp}{dx} \right]_{z^{'}} - \frac{dp}{dz} \frac{dz}{dx}.
\]

This can be applied below and above the thermocline, as follows:

\[
\frac{-1}{\rho_D} \left( \frac{dp}{dx} \right)_{z_{T^+}} = \frac{-1}{\rho_D} \left[ \frac{dp}{dx} \right]_{z_{T^+}^{'}} - \frac{dp}{dz} \frac{dz}{dx}.
\]

\[
\frac{-1}{\rho_S} \left( \frac{dp}{dx} \right)_{z_{T^-}} = \frac{-1}{\rho_S} \left[ \frac{dp}{dx} \right]_{z_{T^-}^{'}} - \frac{dp}{dz} \frac{dz}{dx}.
\]

But we assume that

\[
\left. \frac{dp}{dx} \right|_{z_{T^-}^{'}} = \left. \frac{dp}{dx} \right|_{z_{T^-}},
\]

which means that the pressure is continuous across the thermocline, and

\[
\left. \frac{dp}{dx} \right|_{z_{T^+}^{'}} = 0.
\]

which means that there is no pressure gradient below the thermocline. In addition, we use the hydrostatic equation, (11.3). Then we obtain

Figure 11.22: A sloping thermocline below a sloping sea surface.

\[z' \equiv z - z_T(x).\]
Tropical atmosphere–ocean interactions

An Introduction to the General Circulation of the Atmosphere

\[ 0 = -\frac{1}{\rho_D} \left[ \left( \frac{dp}{dx} \right)_{z_T^+} - \rho_D g \frac{dz_T}{dx} \right] \]

\[ -\frac{1}{\rho_S} \left( \frac{dp}{dx} \right)_{z_T^-} = -\frac{1}{\rho_S} \left[ \left( \frac{dp}{dx} \right)_{z_T^-} - \rho_S g \frac{dz_T}{dx} \right] \]

(11.11)

Substituting across gives

\[ -\frac{1}{\rho_S} \left( \frac{dp}{dx} \right)_{z_T^-} = -g \frac{\Delta p}{\rho_S} \frac{dz_T}{dx}. \]

(11.12)

We find that

\[ -\frac{1}{\rho_S} \left( \frac{dp}{dx} \right)_{z_T^-} = -g \frac{\Delta p}{\rho_S} \frac{dz_T}{dx} \]

(11.13)

For \( \frac{dz_T}{dx} < 0 \) (i.e., the thermocline slopes upward towards the east), we find that

\[ -\frac{1}{\rho_S} \left( \frac{dp}{dx} \right)_{z_T^-} \geq 0 \]

(the water above the thermocline is pushed towards the east).

Assuming that \( \rho \) is independent of depth in the mixed layer, we can write

\[ p(z) = \rho g (z - z_S) \text{ for } z < z_{T^-}. \]

(11.14)

Differentiation gives

\[ \frac{dp}{dx} \big|_z = g \left[ \frac{dp}{dx} \right]_z (z - z_S) - \rho_S \frac{dz_S}{dx} \text{ for } z < z_{T^-}. \]

(11.15)

If the pressure gradient is vertically uniform in the mixed layer (which will be if \( \frac{dp_S}{dx} \equiv 0 \)), then

\[ \frac{dp}{dx} \big|_z = -\rho_S g \frac{dz_S}{dx} = g \Delta \rho \frac{dz_T}{dx}, \]

(11.16)

and so by comparing (11.13) and (11.16) we find that

\[ \frac{dz_S}{dx} = -\frac{\Delta \rho}{\rho_S} \frac{dz_T}{dx}. \]

(11.17)
11.6 Summary

This means that the sea surface slopes up where the thermocline slopes down. We can therefore use the slope of the sea surface to infer the slope of the thermocline. In addition, we can use the slope of the sea surface to infer the near-surface pressure-gradient force.

11.6 Summary
Tropical atmosphere–ocean interactions

An Introduction to the General Circulation of the Atmosphere
Table 2.1: Summary of the annually averaged top-of-the-atmosphere radiation budget.  ............... 9

Table 2.2: Components of the globally and annually averaged surface energy budget. A positive sign means that the surface is warmed. .............................................................................. 21

Table 2.3: The globally and annually averaged energy budget of the atmosphere. A positive sign means that the atmosphere is warmed. .......................................................................... 22

Table 6.1: Lateral energy transports on the poleward side of the ITCZ. Adapted from Riehl and Malkus (1958). ........................................................................................................................ 172

Table 8.1: Predicted and observed phase speeds of Rossby-Haurwitz modes, in degrees of longitude per day. .......................................................................................................................... .... 259

Table 8.2: Periods of the free oscillations on the sphere, as computed from theory. Note that the periods of the gravity waves are given in hours, while those of the Rossby waves are given in days. From Phillips (1963).......................................................................................................................... 260

Table 8.3: Power as a function of wave number and frequency for the winter season. Units: m$^2$ rad$^{-1}$ (15 days)$^{-1}$. ............................................................................................................................ 265
FIG. 1.1: Full disk image, looking down on North America. Many elements of the
general circulation can be seen in this picture, including the rain band in the eastern
North Pacific, the midlatitude baroclinic waves, and the low clouds associated with
the subtropical highs.

FIG. 1.2: This figure shows lidar backscatter from aerosols and clouds. The figure has
been created using data from LITE, a lidar that flew on the space shuttle in 1994.
The lidar cannot penetrate through thick clouds, which explains the vertical black
stripes in the figure. The reddish layer just above the Earth’s surface is the PBL,
which is visible because the aerosol concentration decreases sharply upward at the
PBL top.

FIG. 1.3: A shuttle photograph of tropical thunderstorms. The storms are topped by
thick anvil clouds. Much shallower convective clouds can be seen in the
foreground.

FIG. 1.4: This shuttle photograph shows cirrus clouds associated with the
jet stream.

FIG. 1.5: A beautiful baroclinic eddy over the North Atlantic Ocean in winter.

FIG. 1.6: This ship was used to collect meteorological and oceanographic data in and
above the Arctic ocean. The ship is surrounded by sea ice, and is enveloped by
foggy low-level Arctic clouds.

FIG. 1.7: Cold-looking clouds over the mountains of Antarctica.

FIG. 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 Imbrie and
Imbrie, 1979)

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

FIG. 2.3: The zonally averaged incident solar radiation, albedo, outgoing longwave
radiation, and net radiation at the top of the atmosphere, as observed in the Earth
Radiation Budget Experiment (Barkstrom 1989).

FIG. 2.4: The poleward energy transport by the atmosphere and ocean combined, as
inferred from the observed annually averaged net radiation at the top of the
atmosphere. A Petawatt is 10$^{15}$ W.

FIG. 2.5: The Earth’s orography, averaged onto a mesh 1 degree of longitude “wide”
by 1 degree of latitude “high.”

FIG. 2.6: The distributions of sea ice (shown in grey) for March and September. The
data represent averages for grid cells 5° of longitude wide and 4° of latitude.
FIG. 2.7: a) Sea surface temperature distribution for January. The contour interval is 2 K, and values greater than 26 °C are shaded. b) Same for July. c) The sea surface temperature for March, minus the sea surface temperature for September. The contour interval is 2 K. The zero line is heavy, and negative values are indicated by dashed contours.

FIG. 2.8: A simplified depiction of the distribution of vegetation on the land surface. The resolution is one degree. “Permanent” land ice is shown in white.

FIG. 2.9: Observations of the upward and downward solar (SW) and terrestrial (LW) radiation at the Boulder Atmospheric Observatory (BAO) tower near Boulder Colorado, for the years 1986 - 1988. Figure provided by Ellsworth Dutton.

FIG. 2.10: The zonally averaged (land and ocean) net solar radiation absorbed by the Earth’s surface. These estimates are uncertain by 5% or so. Based on Darnell et al. (1988, 1992). Note: These data are not true observations, although they are based on observations.

FIG. 2.11: The zonally averaged (land and ocean) net infrared cooling of the Earth’s surface. These estimates have large but difficult to quantify uncertainties. Based on Gupta (1989) and Gupta et al. (1992). Note: These data are not true observations, although they are based on observations.

FIG. 2.12: The zonally averaged net surface radiation, obtained by combining the data from Fig. 2.10 and Fig. 2.11. Note: These data are not true observations, although they are based on observations.

FIG. 2.13: The zonally averaged surface latent heat flux, positive upward, based on ECMWF analyses for 1987. Note: These data are not true observations, although they are based on observations.

FIG. 2.14: The zonally averaged surface sensible heat flux, positive upward, based on ECMWF analyses for 1987. Note: These data are not true observations, although they are based on observations.

FIG. 3.1: Sketch illustrating conservation of mass. Here V is a “control volume,” V is the velocity, and is a small element of area on the boundary of the control volume. Control volumes are always shaped like potatoes; hence the expression “Great Geophysical Potato.”

FIG. 3.2: An example of surface pressure plotted against surface height, for a particular observation time. Note that the vertical scale decreases upward.

FIG. 3.3: Sea level pressure maps. The contour interval is 3 mb.
FIG. 3.4: The observed zonally averaged sea level pressure for January (solid line) and July (dashed line). - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 38

FIG. 3.5: 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. 39

FIG. 3.6: Prof. Jule G. Charney, whose accomplishments include scale analyses of both extratropical and tropical motions, development of the quasigeostrophic model, development (in his Ph.D. thesis at UCLA) 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. Because he did so much, Charney’s work is frequently cited in this course. - - - - - - - - - - - - - - - - - - 41

FIG. 3.7: The departure of the 500 mb 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). - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 43

FIG. 3.8: Latitude-height section of the zonal wind. The contour interval is 5 m s$^{-1}$. Easterlies are shaded. - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 44

FIG. 3.9: Maps of the 850 mb zonal wind, for January and July. The contour interval used is 3 m s$^{-1}$. Easterlies are shaded. - - - - - - - - - - - - - - - - - - - - - - - - - - - 45

FIG. 3.10: Image of Jupiter obtained from the Cassini probe in autumn 2000. Note the zonal bands, and the Great Red Spot in the southern hemisphere. - - - - - - - - - - - - - - - - - - - - - - - - - - 46

FIG. 3.11: Maps of the 200 mb zonal wind for January and July. The contour interval is 5 m s$^{-1}$. Easterlies are shaded. - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 47

FIG. 3.12: Latitude-height sections of the meridional wind, for January and July. The contour interval is 0.5 m s$^{-1}$. - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 48

FIG. 3.13: The variation with season of the vertically integrated meridional velocity. The top panel shows variations with latitude, and the bottom panel shows the seasonal cycle at the Equator. From Trenberth et al. (1987). - - - - - - - - - - - 51

FIG. 3.14: 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). - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 52

FIG. 3.15: Maps of the 850 mb meridional wind, for January and July. The contour interval is 2 m s$^{-1}$. Northerlies are shaded. - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 53
FIG. 3.16: Maps of the 200 mb meridional wind, for January and July. The contour interval is 2 m s\(^{-1}\). Northerlies are shaded. 54

FIG. 3.17: Streamlines of the 850 mb wind, for January and July. 55

FIG. 3.18: Streamlines of the 200 mb wind, for January and July. 56

FIG. 3.19: Maps of the geopotential height at 850 mb. The contour interval is 25 m. 57

FIG. 3.20: Maps of the geopotential height at 200 mb. The contour interval is 40 m. 58

FIG. 3.21: The stream function of the mean meridional circulation. Positive values represent counterclockwise circulations, while negative values represent clockwise circulations. 9

FIG. 3.22: Seasonal change of the mean meridional circulation. 61

FIG. 3.23: 500 mb w, the vertical “pressure velocity,” in nb s\(^{-1}\). Negative values (corresponding to rising motion) are shaded. 62

FIG. 3.24: The absolute (left panels) and relative (right panels) atmospheric angular momentum per unit mass, for January and July, as analyzed by ECMWF. The units are 107 m\(^2\) s\(^{-1}\). 64

FIG. 3.25: Sketch illustrating the concept of “mountain torque.” High occurs on the upstream side of the mountain, and low on the downstream side. 67

FIG. 3.26: Changes in the length of day as inferred from the angular momentum of the atmosphere (solid line) and as determined from astronomical data (dotted line) since July 1981. The mean difference between the two quantities has been subtracted from the angular momentum curve. From Rosen et al. (1984). 68

FIG. 3.27: The zonal mean of the annual zonal wind stresses for all oceans combined (Pa): Solid line: Han and Lee (1983); dashed line: Hellerman. From Han and Lee (1983). 69

FIG. 3.28: Meridional profiles of the total surface torque and the mountain torque for annual mean conditions. The horizontal axis is labeled in HUs, which are “Hadley units.” A Hadley unit is 10\(^{18}\) kg m\(^2\) s\(^{-2}\). Note that the scales in the two panels are different. Also note the difference in sign conventions relative to Fig. 3.27. (From Oort 1989, after Newton, 1971.) 70

FIG. 3.29: Cross sections of: a) the zonally averaged zonal flow (m s\(^{-1}\)), and b) the stream function of the zonally averaged relative angular momentum transport in units of 10\(^{18}\) kg m\(^2\) s\(^{-2}\). From Oort 1989. 71
FIG. 3.30: a) An idealized midlatitude temperature profile, based on the U.S. Standard Atmosphere (1976). b) The observed vertical distribution of ozone at $35^\circ$ N, for April, normalized by the total amount. From Manabe and Wetherald (1967). 72

FIG. 3.31: Latitude-height cross sections of temperature, for January and July. The contour interval is 5 K. 73

FIG. 3.32: Latitude-height cross sections of potential temperature, for January and July. The contour interval is 10 K. 74

FIG. 3.33: Maps of the 850 mb temperature for January and July. The contour interval is 5 K. 75

FIG. 3.34: Maps of the 200 mb temperature for January and July. The contour interval is 2 K. 76

FIG. 3.35: January zonal and time-averages as a function of latitude and potential temperature of: a) pressure (mb); b) pseudodensity (kg m$^{-2}$ K$^{-1}$); c) zonal wind (m s$^{-1}$); and d) meridional wind (m s$^{-1}$). The shaded area at the bottom of each panel represents the “ground,” and the thick line represents the average tropopause. From Edouard et al. (1997). 78

FIG. 3.36: The zonally averaged Ertel potential vorticity plotted as a function of latitude and potential temperature. The units are “potential vorticity units” (PVUs), which are defined as $10^{-6}$ m$^2$ K s$^{-1}$ kg$^{-1}$. The shaded area at the bottom represents the “ground,” and the thick line represents the average tropopause. From Edouard et al. (1997). 79

FIG. 3.37: Illustrative vertical profiles of a) the relative humidity (from Manabe and Wetherald, 1967), and b) the water vapor mixing ratio (from Dutton, 1976). 81

FIG. 3.38: Latitude-height section of water vapor mixing ratio. The contour interval used is 1 g kg$^{-1}$. 83

FIG. 3.39: Maps of the 850 mb water vapor mixing ratio. The contour interval is 1 g kg$^{-1}$ shaded values > 10 g kg$^{-1}$. 84

FIG. 3.40: The observed latitude-height distribution of the zonally averaged moist static energy, in kJ kg$^{-1}$, as analyzed by ECMWF. 85

FIG. 3.41: The zonally averaged precipitation rate for January and July. 86

FIG. 3.42: Maps of the January and July precipitation rate, from the Global Precipitation Climatology Center. The contour interval is 1 mm day$^{-1}$. Values greater than 4 mm day$^{-1}$ are shaded. 87

FIG. 3.43: Maps of the surface sensible heat flux, based on ECMWF reanalyses. These
are not real observations, although they are based on observations. The contour interval is 10 W m\(^{-2}\). Values higher than 50 W m\(^{-2}\) are shaded.

**FIG. 3.44:** Momentum balance without friction (left) and with friction (right). The thin arrows represent the wind vector, and the thick arrows represent various forces and apparent forces. The effect of friction is to make the wind turn towards low pressure.

**FIG. 3.45:** Maps of the surface latent heat flux, based on ECMWF reanalyses. These are not real observations, although they are based on observations. The contour interval is 20 W m\(^{-2}\). Values higher than 150 W m\(^{-2}\) are shaded.

**FIG. 3.46:** Maps of the magnitude of the surface wind stress based on satellite data. The contour interval is 0.02 Pa. Values higher than 0.2 Pa are shaded.

**FIG. 3.47:** 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.

**FIG. 3.48:** An example to illustrate 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 eddy parts of the January-mean meridional wind and temperature at 850 mb, respectively. The bottom panel shows the product, and on the bottom right the zonal mean of the product. See text for details.

**FIG. 3.49:** Meridional flux of zonal momentum in the annual mean. The top panel shows the total flux, the second panel shows the transient-eddy flux, the third panel shows the stationary eddy flux, and the bottom panel shows the flux due to the mean meridional circulation (c). From Peixoto and Oort (1992).

**FIG. 3.50:** The meridional flux of temperature. The top panel shows the transient-eddy flux, the second panel shows the stationary eddy flux, and the bottom panel shows the flux due to the mean meridional circulation. In the bottom panel, the globally averaged temperature on each pressure surface is removed before computing the flux. From Peixoto and Oort (1992).

**FIG. 3.51:** A crude sketch of the variation of the eddy heat flux with latitude. Where the flux increases poleward, the air is cooled, and where it decreases poleward, the air is warmed.

**FIG. 3.52:** A crude sketch of the mean meridional circulation. The subtropical sinking motion occurs where the eddies cool, and the midlatitude rising motion occurs where the eddies warm.

**FIG. 3.53:** The stream function of the seasonally varying mean meridional circulation.
as seen in p-coordinates (right panels) and θ-coordinates (left panels). Units are $10^6 \text{ kg s}^{-1}$. 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. From Townsend and Johnson (1985).

FIG. 4.1: Sketch defining notation used in the text.

FIG. 5.1: Definition of a spherical coordinate system whose pole is at an angle with respect to the axis of the Earth’s rotation. See Problem 11.

FIG. 5.2: A planet shaped like a doughnut. The inner radius is and the outer radius is . Latitude is measured from the outer Equator. See Problem 12.

FIG. 6.1: Joanne Malkus (now Joanne Simpson), and the late Herbert Riehl.

FIG. 6.2: Flow chart for the numerical time integration in the radiative-convective equilibrium study of Manabe and Wetherald (1967).

FIG. 6.3: 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).

FIG. 6.4: Temperature profiles for radiative-convective equilibrium, as simulated by the physical parameterizations of the CSU GCM. The surface temperature was prescribed as a more or less realistic function of latitude, and for each latitude the annual mean insolation was used.

FIG. 6.5: (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).

FIG. 6.6: Representative observed soundings for Darwin, Australia, 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 curves plotted show the dry static energy, the moist static energy, and the saturation moist static energy. The panels on the right cover both the troposphere and lower stratosphere, while those on the right zoom in on the lower troposphere to show more detail. Values are divided by $c_p$ to give units in K.

FIG. 6.7: 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.

FIG. 6.8: Schematic diagram summarizing the relationships among the cloud regimes.
depicted in Fig. 6.6, and how they fit into the mean meridional circulation. From Schubert et al. (1995).

**FIG. 6.9:** Estimates of the clear-sky radiative cooling rate in the eastern and central Pacific, based on ECMWF data

**FIG. 6.10:** Douglas Lilly, who has done leading work on a wide range of topics including cumulus convection, numerical methods, gravity waves, and stratocumulus clouds.

**FIG. 6.11:** Representative observed soundings for Denver, Colorado and Barrow, Alaska, for July and January. The curves plotted show the dry static energy, the moist static energy, and the saturation moist static energy. The panels on the right cover both the troposphere and lower stratosphere, while those on the right zoom in on the lower troposphere to show more detail. Values are divided by cp to give units in K.

**FIG. 6.12:** Prof. Akio Arakawa, cruising along in mid-lecture.

**FIG. 6.13:** Observed time sequence of the CAPE during TOGA COARE, over the Intensive Flux Array. K. Emanuel’s code was used to construct these curves. X. Lin of CSU performed the computations.

**FIG. 7.1:** Prof. Edward N. Lorenz, who proposed the concept of available potential energy, and has published a great deal of additional very fundamental research in the atmospheric sciences and the relatively new science of nonlinear dynamical systems.

**FIG. 7.2:** Sketch illustrating the concept of “massless layers.”

**FIG. 7.3:** Sketch illustrating the transition from a given state to the A-state. Here increases downward and we assume that , which means that the atmosphere is statically stable.

**FIG. 7.4:** Sketch illustrating the transition from a given state to the S state used to define the gross static stability.

**FIG. 7.5:** A simple example used to explain the idea of gradient production. State B is obtained by homogenizing State A. In both State A and State B, the average of is . Variance seems to “disappear” in passing from State A to State B. In reality it is converted from the variance of the mean to an eddy variance.

**FIG. 7.6:** Sketch illustrating the flow of energy through the atmospheric general circulation. Generation produces APE, which is converted to KE, which in turn is dissipated.

**FIG. 7.7:** Energy cycle during winter (left) and summer (right) for both hemispheres.
Integrals between 1000 and 50 mb 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 Arpé et al. (1986).

**FIG. 7.8:** 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 Arpé et al. (1986).

**FIG. 7.9:** 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 Bar)$^{-1}$. From Arpé et al. (1986). 242

**FIG. 7.10:** Mean annual cycles of energy amounts and conversions in both hemispheres and contributions from wavenumber groups calculated from 12 hour forecasts. From Arpé et al. (1986).

**FIG. 8.1:** 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.

**FIG. 8.2:** The solution procedure for free oscillations. The vertical structure equation is solved first, yielding the equivalent depths as eigenvalues. The frequencies are then obtained as eigenvalues of the LTE.

**FIG. 8.3:** Chain of vortices along a latitude circle, illustrating the westward propagation of Rossby waves.

**FIG. 8.4:** Successive daily values of the phase angle for the 24 hour tendency field, at the 500 mb level for the components during the 90 day period beginning 1 December 1956. The ordinate is time, in days. The abscissa represents the number of westward circulations round the Earth after the first passage of the Greenwich meridian, and so it is a measure of longitudinal phase propagation. From Eliassen and Machenhauer (1965).

**FIG. 8.5:** Maps of the geopotential height for nine winters: (a) rms unfiltered, contour interval 10 m; (b) average, contour interval 50 m. From Blackmon (1976).

**FIG. 8.6:** Maps of the low-pass filtered rms fields (winter): (a) all waves, contour interval 10 m; (b) waves in Regime I, contour interval 5 m; (c) waves in Regime II, contour interval 5 m; (d) waves in Regime III, contour interval 5 m. From Blackmon (1976).

**FIG. 8.7:** Maps of the medium-pass filtered rms fields (winter): (a) all waves, contour

An Introduction to the General Circulation of the Atmosphere
interval 5 m; (b) waves in Regime I, contour interval 2 m; (c) waves in Regime II, contour interval 2 m; (d) waves in Regime III, contour interval 2 m. From Blackmon (1976).

**FIG. 8.8:** Upper panel: The total eddy heat transport across a latitude circle as a function of wave number for selected latitudes, for January 1962. Lower panel: The eddy momentum transport averaged with respect to pressure, as a function of zonal wave number, for selected latitudes, based on observations for January 1962. In both panels the gray scale at the bottom denotes the zonal wave number. From Wiin-Nielsen et al. (1963).

**FIG. 8.9:** Schematic illustrating shallow water flowing over a “mountain.”

**FIG. 8.10:** The mean-square height response, $\sigma^2$, in the Charney-Eliassen model, as a function of $\sigma$, for different values of the Rayleigh friction coefficient. The units are 104 m². The integers written above the curve indicate the values of $\sigma$ for which particular zonal wave numbers resonate. From Held (1983).

**FIG. 8.11:** Upper panel: The height response as a function of longitude in the Charney-Eliassen model for $\sigma = 5$ days and $v = 17$ m s⁻¹ (solid line) and the observed climatological 500 mb eddy heights at $45^\circ$N in January (dashed line). The lower panel shows the topography used. From Held (1983).

**FIG. 8.12:** Time-latitude sections of Northern Hemisphere climatological mean stationary wave kinetic energy at the 200 mb level, based on data of Oort and Rasmusson (1971). a) Zonal wind component; b) meridional wind component, in units of m² s⁻². Arrows in a) denote the direction and relative magnitude of the meridional flux of zonal momentum by the stationary waves. From Wallace (1983).

**FIG. 8.13:** A photograph of Prof. T. Matsuno, taken as he gave a lecture at UCLA in January 1998.

**FIG. 8.14:** Model and coordinate system. From Matsuno (1966).

**FIG. 8.15:** Units of time (left scale) and length (right scale) as functions of the phase velocity of a pure gravity wave. From Matsuno (1966).

**FIG. 8.16:** Schematic used to explain the two-level model represented by Eqs. (8.73).

**FIG. 8.17:** Frequencies as functions of wave number. Positive frequencies correspond to westward propagation. Thin solid line: eastward propagating inertia-gravity waves. Thin dashed line: westward propagating inertia-gravity waves. Thick solid line: Rossby (quasi-geostrophic) waves. Thick dashed line: The Kelvin wave. The westward moving wave with $n=0$ is the mixed Rossby-gravity or Yanai wave. It is
denoted by a dashed line for \( k \), and by a solid line for \( k \). From Matsuno (1966) 280

**FIG. 8.18:** Pressure and velocity distributions of solutions for \( k \) and \( k \). a) Eastward moving inertia-gravity wave. b) Westward moving Yanai wave, which for this value of \( k \) behaves like an inertia-gravity wave. Panel c shows the structure of the Yanai wave for \( k \) and \( k \), in which case the Yanai wave acts like a Rossby wave. For each mode, \( u \) is a maximum on the Equator and does not pass through zero anywhere. This is characteristic of \( k \). From Matsuno (1966). 281

**FIG. 8.19:** Left side: Pressure and velocity distributions of solutions for \( k \). a) Eastward propagating inertia-gravity wave. For each mode, on the Equator, as we expect for \( k \). b) Westward propagating inertia-gravity wave. c) Rossby wave. Right side: Corresponding results for \( k \). For each mode \( u \) is symmetrical across the Equator, as we expect for \( k \). From Matsuno (1966). 282

**FIG. 8.20:** Pressure and velocity distributions for \( k \) and \( k \). This is the Kelvin wave. From Matsuno (1966). 283

**FIG. 8.21:** The variability of the outgoing longwave radiation (OLR) as a function of frequency and zonal wave number for modes that are symmetric across the Equator (right panel) and anti-symmetric (left panel). Eastward propagation is associated with positive wave numbers, and vice versa. The boxes select particular wave types. From Wheeler and Kiladis (1999). 283

**FIG. 8.22:** Longitudinal propagation of the Madden-Julian Oscillation (MJO; discussed later), 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 wave numbers and frequencies that fall in the corresponding boxes in Fig. 8.21. This is what is meant by “filtering.” From Wheeler and Kiladis (1999). 284

**FIG. 8.23:** Longitudinal propagation of eastward- and westward-propagating inertia gravity waves, as seen in the OLR. The zero contour has been omitted. As in Fig. 8.22, the various modes are selected by including only the contributions from wave numbers and frequencies that fall in the corresponding boxes in Fig. 8.21. From Wheeler and Kiladis (1999). 285

**FIG. 8.24:** Stationary circulation pattern (lower panel) forced by the mass source and sink shown in the upper panel. From Matsuno (1966). 286

**FIG. 8.25:** Solution of Gill’s model for the case of heating symmetric about the Equator. The upper panel shows the heating field and the low-level wind field. The center panel shows the perturbation pressure field, which features low pressure along the Equator generally, with twin cyclones slightly off the Equator. The bottom panel shows the implied vertical motion and the zonal variation of the pressure along the Equator. From Gill (1980). 288
FIG. 8.26: The response to antisymmetric heating. On the left side, the top panel shows contours of the mid-level vertical velocity superimposed on the horizontal wind vectors for the lower layer. The lower panel 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 panel shows the latitude-height distributions of the zonal velocity and the stream function of the mean meridional circulation, as well as the meridional profile of the surface pressure. From Gill (1980).

FIG. 8.27: The response of Gill’s model to a combination of symmetric and antisymmetric heating. From Gill (1980).

FIG. 8.28: Average elevation of the monsoon region. Data was averaged to $1^\circ \div 1^\circ$, and then 9-point smoothed. Terrain over 3000 m high is shaded.

FIG. 8.29: Observed 850 mb wind vectors for a) January, and b) July.

FIG. 8.30: Observed JJA climatological 500 mb temperatures. Contour interval is 2 K.

FIG. 8.31: Dates of onset of the Asian monsoon near India in 1988 (actual) and mean (normal). From Krishnamurti et al. (1990).

FIG. 8.32: a) Observed daily rainfall along the southwest coast of India for the summer monsoon seasons of 1963 and 1971. From Webster (1987). b) Progress of a monsoon depression across India. Many such depressions occur throughout the summer monsoon. From Webster (1981).

FIG. 8.33: The observed JJA climatological precipitation (Legates and Willmott, 1990). The contour interval is 2 mm day$^{-1}$.

FIG. 8.34: 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 mb 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).

FIG. 8.35: Observed JJA climatological 500 mb vertical velocity. Contour interval is 20 mb day$^{-1}$. Areas with vertical velocities more negative than -60 mb day$^{-1}$ have light shading.

FIG. 8.36: JJA climatological 200 mb winds. The scale vector is 50 m s$^{-1}$.

FIG. 8.37: Latitude-pressure plot of the JJA climatological zonal winds at 77.5 E. The contour interval is 5 m s$^{-1}$.
FIG. 8.38: Observed Fourier power spectra of surface pressure. a) at 10°N, 77.5°E, near the southern tip of India, for oscillations with a period of 0-5 days; b) at 18°N, 77.5°E, for oscillations with a period of 0-5 days; c) at 26°N, 77.5°E, for oscillations with a period of 0-5 days; d) at 10°N, 77.5°E, for oscillations with a period of 0-90 days; e) at 18°N, 77.5°E, for oscillations with a period of 0-90 days; and f) at 26°N, 77.5°E, for oscillations with a period of 0-90 days.


FIG. 8.40: The longitude-height cross-sections of the zonal wind (left, in m s\(^{-1}\)) and vertical velocity (right, in Pa s\(^{-1}\)) along the Equator, for January, as analyzed by ECMWF.

FIG. 8.41: Streamlines and horizontal wind vectors for the tropical Pacific at 1000 mb for a) January 1989 and b) July 1989. The units are m s\(^{-1}\).

FIG. 8.42: Contour plot of mean vertical velocity at 300mb (units, 10\(^{-2}\) Pas\(^{-1}\)) from the ECMWF reanalysis dataset for a) January 1989 and b) July 1989. Contour interval is 2 x 10\(^{-2}\) Pa s\(^{-1}\); negative contours are dashed. Data were obtained from NCAR.


FIG. 8.44: Tropical skin temperature for January 1989 from the ECMWF reanalysis dataset obtained from NCAR. Resolution for this dataset is 2.5° and the contour interval is 2 K.

FIG. 8.45: Tropical OLR for January averaged over 1985 to 1988. Daily means from the Earth Radiation Budget Experiment (ERBE) were averaged and interpolated onto a 5° x 4° (longitude-latitude) grid. The units are W m\(^{-2}\).

FIG. 8.46: Observed annual-mean low-cloud amount (upper panel) and the net effects of clouds on the Earth’s radiation budget (lower panel). Negative values in the lower panel indicate a cooling, i.e. shortwave reflection dominates longwave trapping.

FIG. 8.47: Reproduction of Pierrehumbert’s schematic representation of the ‘furnace/radiator-fin’ model of the tropical circulation. The symbols E and TS represent the evaporation rate and SST, respectively. The subscripts 1 and 2 denote the Warm Pool or furnace and the Cold Pool or radiator fin.

FIG. 8.48: Variance spectrum for station pressures at Nauru Island, 0.4° S, 161.0° E. Ordinate (variance/frequency) is logarithmic and abscissa (frequency) is linear. The 40-50-day period 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). Taken from Madden and Julian.
FIG. 8.49: Schematic depiction of the time and space (zonal plan) variations of the disturbance associated with the 40-50-day 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 is 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). Taken from Madden and Julian (1972 a).

FIG. 8.50: Mean phase angles (in degrees), coherence squares, and background coherence squares for approximately the 36-to-50-day period 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).

FIG. 8.51: 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 SSC with smaller westward-moving cloud clusters that develop, grow to maturity, and decay in a few days. From Nakazawa (1988).

FIG. 8.52: Schematic illustration of the structure of a Rossby wave.

FIG. 8.53: Earth-moon geometry.

FIG. 9.1: 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}. From McFarlane (1987).

FIG. 9.2: The actual change in the zonally averaged zonal 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}. From McFarlane (1987).

FIG. 9.3: 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 K1. From McFarlane (1987).

**FIG. 9.4:** The square of the index of refraction for summer and winter, averaged between $30^\circ$ and $60^\circ$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).

**FIG. 9.5:** 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 mb, 100 mb, and 10 mb. From Charney (1973).

**FIG. 9.6:** a) The latitudinal gradient of the potential vorticity, $\zeta$, expressed as a multiple of the Earth’s rotation rate. b) An idealized but somewhat realistic basic state zonal wind distribution (m s$^{-1}$) in the Northern Hemisphere winter. c) The refractive index square, $n^2$, for the $m$ wave. d) Computed distribution of energy flow in the meridional plane associated with zonal wave number 1. From Matsuno (1970).

**FIG. 9.7:** a) Daily vertical eddy flux of geopotential through 100 and 10 mb. Units: erg cm$^{-2}$ s$^{-1}$. b) Daily divergence of the vertical eddy flux of geopotential for 100-10 mb, and for the region $90^\circ$ N to $20^\circ$ N. Units: erg cm$^{-1}$ s$^{-1}$. Data for 1964. From Dopplick (1971).

**FIG. 9.8:** Annual energy cycle for the lower stratosphere between 100-10 mb and for the region $20^\circ$N to $90^\circ$N. Units: contents 107 erg cm$^{-2}$; conversions: erg cm$^{-2}$ s$^{-1}$. “B” indicates boundary effects (at the lower or equatorward boundaries), and “G” indicates generation. Overall, the stratosphere gains kinetic energy through interactions with the troposphere. The kinetic energy is converted to potential energy. This balances a loss of potential energy through radiative effects. From Dopplick (1971).

**FIG. 9.9:** a) Latitude-time section of zonal mean temperature (K) measured by channel A for 31 December 1970 to 16 January 1971. Regions of temperature lower than 258 K are shaded. A major warming had a peak at this level on 9 January at $80^\circ$N. b) Latitude-time section of amplitude of zonal wavenumber one of channel A temperature (K) for 31 December 1970 to 16 January 1971. Maximum amplitude occurred on 4 January at $65^\circ$N. From Barnett (1974).

**FIG. 9.10:** 10 mb charts during a sudden warming in January 1963. Height contours (solid lines) at 32 dm intervals. Isotherms (dashed lines) at 10 K intervals. Note that the process spans 9 days. From Sawyer (1965).

**FIG. 9.11:** Contribution of transient eddies to the seasonally averaged Eliassen-Palm cross sections for the troposphere: (a) 5-year average from Oort and Rasmusson.
(1971) for winter; (b) the same for summer. The contour interval is $20 \times 10^{15} \text{ m}^3$ for (a), and $1 \times 10^{15} \text{ m}^3$ for (b). The horizontal arrow scale for the horizontal component in units of m$^3$ is indicated at bottom right; note that it is different from diagram to diagram. A vertical arrow of the same length represents the vertical component, in m$^3$ kPa, equal to that for the horizontal arrow multiplied by 80.4 kpa. From Edmon et al. (1980).

FIG. 9.12: Contribution of stationary eddies to the seasonally averaged Eliassen-Palm cross sections for the troposphere: (a) 5-year average from Oort and Rasmusson (1971) for winter; (b) the same, respectively, for summer. The contour interval is $1 \times 10^{15} \text{ m}^3$ for both panels. The horizontal arrow scale in units of m$^3$ is indicated at bottom right. From Edmon et al. (1980).

FIG. 9.13: Total (transient plus stationary) Eliassen-Palm cross sections for the troposphere: (a) 5-year average from Oort and Rasmusson (1971) for winter; (b) the same, respectively, for summer. The contour interval is $2 \times 10^{15} \text{ m}^3$ for (a), and $1 \times 10^{15} \text{ m}^3$ for (b). The horizontal arrow scale in units of m$^3$ is indicated at bottom right. From Edmon et al. (1980).

FIG. 9.14: The stream function of the seasonally averaged residual meridional circulations. (a) 5-year average from Oort and Rasmusson (1971) for winter, and (b) the same for summer. The contour interval is $7.5 \times 10^{16} \text{ m}^2 \text{s Pa}$. From Edmon et al. (1980).

FIG. 9.15: Time-height plot of the monthly mean zonal wind based on observations from various equatorial stations, namely Canton Island (January 1953 to August 1967), Maldives Islands (September 1967 to December 1975) and Singapore (January 1976 to May 1992). Contour interval 10 m s$^{-1}$, westerlies shaded. From James (1994).

FIG. 9.16: Energy and momentum fluxes associated with a Kelvin wave. The wavy lines show the heights of isentropic surfaces. Energy is propagating both upward and downward, away from the energy source. Where the energy flux is upward, the momentum flux is also upward, and vice versa. Westerly momentum is transported away from the energy source region, which feels an easterly acceleration to compensate.

FIG. 9.17: Contours of the fourth root of the Ertel potential vorticity on the 320 K isentropic surface, which slopes from 200 mb near the pole to 600 mb in the tropics. Panels a - e are for successive days. The contour interval is 0.001 in SI units. The bold contour is 0.005. From Shutts (1986).

FIG. 9.18: (a) Mean sea level pressure field for 15 February 1983, 12Z. Contour interval: 5 mb. (b) Height of the 500-mb surface for 15 February, 12Z. Contour interval: 8 dam. From Shutts (1986).
FIG. 9.19: The relationships between the A and B 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 B-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).

FIG. 9.20: From Speranza (1986). Equilibrium solutions of (9.120)-(9.122), found by setting the time-rate of change terms to zero, solving the resulting linear system (9.120)-(9.121) for and as functions of, and selecting the appropriate value of by requiring that (9.122) also be satisfied. The straight line represents the pairs that satisfy (9.122), while the peaked line represents the pairs that satisfy (9.120)-(9.121). The figure shows that there are three equilibrium solutions, but it can be shown that the middle one is unstable.

FIG. 10.1: The basic mechanism of shearing instability.

FIG. 10.2: The components of the velocity normal and tangent to the vorticity vector, and the stretching and twisting processes associated with these velocity components.

FIG. 10.3: Diagram used in the explanation of Fjortoft’s (1953) analysis of the exchanges of energy and enstrophy among differing scales in two-dimensional motion.

FIG. 10.4: A time sequence of the vorticity distribution in an idealized numerical simulation of quasi-two-dimensional turbulence, by McWilliams (1984). In panel A the vorticity is highly disorganized. Through a process of enstrophy dissipation and vortex mergers, the vorticity gradually organizes itself into a single pair of vortices -- one cyclonic and one anti-cyclonic.

FIG. 10.5: Sketch illustrating the flow of kinetic energy through wave-number space, from a low-wave number region with an energy source, to a high-wave number region in which dissipation removes kinetic energy.

FIG. 10.6: Schematic indicating the kinetic energy spectrum in the Earth’s atmosphere, as implied by dimensional analysis. Baroclinic instability adds energy at wave number , and convection adds energy at wave number . See text for details.

FIG. 10.7: 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 (1998), after Nastrom and Gage (1985).
FIG. 10.8: 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). 381

FIG. 10.9: 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). 382

FIG. 10.10: Observed nonlinear kinetic energy exchanges as a function of height. From Boer and Shepherd (1983). 382

FIG. 10.11: Potential vorticity (solid lines) and potential temperature (dashed lines) plotted as functions of pressure and latitude. From Sun and Lindzen (1994). 383

FIG. 10.12: 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 (1987). 389

FIG. 10.13: Sketch illustrating the role of instability in leading to error growth, and of nonlinearity in leading to the movement of error from small scales to larger scales. 390


FIG. 10.15: 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). 392

FIG. 10.16: 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). 393

FIG. 10.17: Zonal 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). 395

FIG. 10.18: Root-mean-square error (solid line) averaged for six pairs of control and perturbation runs and averaged for latitude belt 40-60$^\circ$N for 500-mb 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).]

**FIG. 10.19:** A long sequence of one-day and two-day forecasts.

**FIG. 10.20:** The growth of $A$ with time. Rapid initial growth is followed by saturation.

**FIG. 10.21:** The growth of $A$ and $B$ with time.

**FIG. 10.22:** Global root-mean-square 500-mb height differences $\frac{m}{500}$, in meters, between -day and -day forecasts made by the ECMWF operational model for the same day, for $A$, plotted against $B$. Values of $A$ are shown beside some of the points. Heavy curve connects values of $A$. The thin curves connect values of $A$ for constant $B$. From Lorenz (1982).

**FIG. 10.23:** Diagram illustrating the expected variation of the error growth rate $\alpha$ with the magnitude of the error.  

**FIG. 10.24:** Increases in global root-mean-square 500-mb height differences $\sigma$, plotted against average height differences $\bar{m}$, in meters, for each one-day segment of each thin curve in Fig. 1 (large dots), and increases $\delta$ plotted against average differences $\delta$, for each one-day segment of heavy curve in Fig. 1 (crosses). Parabola of “best fit” to large dots is shown.

**FIG. 10.25:** Model and observed zonally-averaged standard deviations, $\sigma_m$ and $\sigma_o$, as functions of latitude, and their ratio, for: (a) mean July sea-level pressure; and (b) rainfall.Observed values are for land stations and model values are for grid-points over land. From Charney and Shukla (1981).

**FIG. 10.26:** (a) Observed sea-surface temperature anomaly (K) for January 1983. (b) Simulated rainfall anomaly (mm day-1). (c) Observed rainfall anomaly (mm day-1) as inferred from outgoing long-wave radiation during winter 1983. From Shukla (1985).

**FIG. 10.27:** Temporal variations of monthly indices of sea surface temperature, zonal wind at 200 and 950 mb, precipitation, east-west sea level pressure gradient across the South Pacific, 200 mb 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).

**FIG. 10.28:** Variations of $A$ with $B$, for steady solutions of (10.82) with $A$, $B$, and $H$ represents the Hadley solution. From Lorenz (1984).

**FIG. 10.29:** Solutions of (10.82) extending for 100 time steps, starting from initial
values of 0.0999 (upper), 0.1000 (middle), and 0.1001 (lower). The straight-line segments joining consecutive points are solely for the purpose of making the chronological order easier to see. This figure illustrates sensitive dependence on initial conditions. From Lorenz (1975).

**FIG. 10.30:** The variations of (dimensionless) with t (months) in a numerical solution of (10.82), with , , , and (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).

**FIG. 10.31:** (a) The same as Fig. 8.19a, except that (winter conditions). (b) The same as panel b, except that the initial state is (2.5, 1.0, 0). From Lorenz (1990).

**FIG. 10.32:** The variations of (dimensionless) with (months) in a 6-year numerical solution of (10.82), with , , and , where . Each row begins on 1 January, and, except for the first, each row is a continuation of the previous one. From Lorenz (1990).

**FIG. 10.33:** The variations of (dimensionless) with (years) in a 100-year numerical solution of (10.82), for the conditions of Fig. 6, where is the standard deviation of within the period July through September. From Lorenz (1990).

**FIG. 10.34:** The impacts of various imposed forcings on the PDF of the butterfly model. state vector with running time mean, in the plane. From Palmer (1999).

**FIG. 11.1:** The Walker and Hadley Circulations. From Philander (1990).

**FIG. 11.2:** The longitude-height cross-sections of the zonal wind (left, in m s$^{-1}$) and vertical velocity (right, in Pa s$^{-1}$) along the Equator, for January, as analyzed by ECMWF.

**FIG. 11.3:** January 1989 Equatorial u-$\omega$ vectors in units of m s$^{-1}$. The $\omega$ values, which were originally in units of Pa s$^{-1}$, were scaled by -300 m Pa$^{-1}$. 

**FIG. 11.4:** Streamlines and horizontal wind vectors for the tropical Pacific at 1000 mb for a) January 1989 and b) July 1989. The units are m s$^{-1}$.

**FIG. 11.5:** Contour plot of mean vertical velocity at 300mb (units, 10-2 Pas$^{-1}$) from the ECMWF reanalysis dataset for a) January 1989 and b) July 1989. Contour interval is 2 x 10 -2 Pa s$^{-1}$; negative contours are dashed. Data were obtained from NCAR.

**FIG. 11.6:** Streamlines and horizontal wind vectors for the tropical Pacific at 200 mb for a) January 1989 and b) July 1989. The units are m s$^{-1}$.

**FIG. 11.7:** Tropical skin temperature for January 1989 from the ECMWF reanalysis dataset obtained from NCAR. Resolution for this dataset is 2.5° and the contour
interval is 2 K.

**FIG. 11.8:** Tropical OLR for January averaged over 1985 to 1988. Daily means from the Earth Radiation Budget Experiment (ERBE) were averaged and interpolated onto a $5^\circ \times 4^\circ$ (longitude-latitude) grid. The units are W m$^{-2}$. 425

**FIG. 11.9:** Observed annual-mean low-cloud amount (upper panel) and the net effects of clouds on the Earth’s radiation budget (lower panel). Negative values in the lower panel indicate a cooling, i.e. shortwave reflection dominates longwave trapping. 427

**FIG. 11.10:** Reproduction of Pierrehumbert’s schematic representation of the ‘furnace/radiator-fin’ model of the tropical circulation. The symbols $E$ and $TS$ represent the evaporation rate and SST, respectively. The subscripts 1 and 2 denote the Warm Pool or furnace and the Cold Pool or radiator fin. 428

**FIG. 11.11:** Schematic of the Sun-Liu coupled model. The boxes represent the atmosphere (light hatching), the Warm Pool (heavy hatching), the Cold Pool (stippling), and the undercurrent (clear box). Heavy arrows denote ocean currents, light arrows denote local heating and the dashed arrow represents the surface winds. 430

**FIG. 11.12:** The Equilibrium solution from the Sun-Lu coupled model for (a) current strength measured by $q/c$, and for (b) non-dimensionalized SST as given by $\beta$. 431

**FIG. 11.13:** The Southern Oscillation. From Philander (1990). 432

**FIG. 11.14:** Time series of sea level pressure at Darwin and Tahiti. From Philander (1990). 433

**FIG. 11.15:** Longitude-time section of the sea surface temperature. From Philander (1990). 433

**FIG. 11.16:** Time series of precipitation at selected locations. The black bars are El Niño years. From Philander (1990). 434

**FIG. 11.17:** Observed SSTs for November 1982 and November 1983. From Philander (1990). 434


**FIG. 11.19:** Anomalies of outgoing longwave radiation and the winds, for 19982 and 1983. From Philander (1990). 436

**FIG. 11.20:** Streamfunction of the 200 mb winds. The bottom panel is a difference
plot. From Philander (1990). 437

FIG. 11.21: The assumed density distribution in the upper ocean. The depth, $z$, increases downward from zero at a level near the surface. The surface depth, $z_s$, is not necessarily zero; it can be slightly positive or slightly negative as sea level varies. 438

FIG. 11.22: A sloping thermocline below a sloping sea surface. 439
References and Bibliography

A


B


Cheng, M.-D., and A. Arakawa, 1997: Inclusion of rainwater budget and convective


**D**


**E**


Edmon, H. J., B. J. Hoskins, and M. E. McIntyre, 1980: Eliassen-Palm cross sections for the troposphere. *J. Atmos. Sci.*, 37, 2600-2616. (See also Corrigendum, *J. Atmos. Sci.*, 38, 1115.)


**F**


H


Laplace, P. S., 1832: *Méchanique Céleste*, 4 vols, (Translated by N. Bowditch), Boston. The relevant section is Part I, Book IV, Section 3, p. 543.

Larson, K., D. L. Hartmann, and S. A. Klein, 1999: Climate sensitivity in a two box model of the tropics. Accepted by the *J. Climate*, In press.


Lau, K.-M., and H. Lim, 1982: Thermally driven motions in an equatorial β-plane: Hadley...


Lord, S. J., 1982: Interaction of a cumulus cloud ensemble with the large-scale

*An Introduction to the General Circulation of the Atmosphere*


Manabe, S., and T. Terpstra, 1974: The effects of mountains on the general circulation of
the atmosphere as identified by numerical experiments. *J. Atmos. Sci.*, **31**, 3-42.


McWilliams, J. C., 1984: The emergence of isolated coherent vortices in turbulent flow. *J. Fluid Mech.*, **146**, 21-43.


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. *J. Atmos. Sci.*, **43**, 961-971.


**N**


O


An Introduction to the General Circulation of the Atmosphere


Q -R


Raymond, D. J., 2000: The Hadley circulation as a radiative-convective instability. *J.


An Introduction to the General Circulation of the Atmosphere


Climate, 11, 1685-1702.


Sui, C.-H. and K.-M. Lau, 1992: Multiscale phenomena in the tropical atmosphere over


**T**


**U -V**


**W**


Whitlock, C. H., T. P. Charlock, W. F. Staylor, R. T. Pinker, I. Laszlo, R. C. DiPasquale,


X


*An Introduction to the General Circulation of the Atmosphere*
Y


Z


