DISSERTATION

A COMPARISON OF VERTICAL COORDINATE SYSTEMS FOR NUMERICAL MODELING OF THE GENERAL CIRCULATION OF THE ATMOSPHERE

Submitted by

Ross Parker Heikes

Department of Atmospheric Science

In partial fulfillment of the requirement

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer, 2002

UMI Number: 3063992

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

JUNE 6, 2002

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY ROSS PARKER HEIKES ENTITLED A COMPARISON OF VERTICAL COORDINATE SYSTEMS FOR NUMERICAL MODELING OF THE GENERAL CIRCULATION OF THE ATMOSPHERE BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Wayne Schubert

Adviser

Department Head (acting)

ABSTRACT OF DISSERTATION

A COMPARISON OF VERTICAL COORDINATE SYSTEMS
FOR NUMERICAL MODELING
OF THE GENERAL CIRCULATION OF THE ATMOSPHERE

The Colorado State University general circulation model has been dramatically changed in the last few years. The horizontal grid has been changed from a latitude/longitude grid with an Arakawa C-grid staggering of mass, zonal and meridional wind to a geodesic grid in which mass, vorticity and divergence are predicted at cell centers. Throughout this change the vertical structure has remained unchanged. Now, we are poised to revamp the vertical structure of the model.

We have studied several possible alternative vertical coordinate systems in order to pick the one best suited for long-term climate prediction. After a brief historical review of various vertical coordinates used in numerical modeling, we will focus on three models -- one based on the co-called sigma-coordinate, on based on a pure isentropic coordinate and one based a hybrid σ-θ vertical coordinate. We will show how the continuous equations are transformed to finite-difference analogs, and how the transformation preserves many of the conservation properties of the continuous equations.

We will show how different vertical discretization of the prognostic variables -- Lorenz or Charney-Phillips vertical staggering -- can influence the quality of the numerical simulation. Many fundamental questions about the isentropic- and hybrid-coordinate need to be answered. Since the isentropic-coordinate surfaces behave as Lagrangian surfaces, the ideas of grid resolution from the Eulerian framework are not applicable. Also, in the isentropic- and hybrid- coordinates, there is a much greater

variation in layer thickness. Excessive vertical resolution in the extra-tropics where coordinate surfaces are close, may become inadequately coarse in the tropics as the coordinate surfaces spread apart. We will show how the quality of the Held and Suarez test case changes as a function of resolution.

We have performed a simulation of stratospheric sudden warming with the isentropic-coordinate model. We will compare our results with observations. We have also performed a simulation of baroclinic instability with the hybrid-coordinate model.

Ross P. Heikes
Atmospheric Science Department
Colorado State University
Fort Collins, CO 50523
Summer 2002

ACKNOWLEDGMENTS

First and foremost I would like express my very sincere gratitude to my advisor, David Randall. I am very fortunate to have been his advisee. He has been constant source of support and encouragement, and has demonstrated tremendous patience over the last few years. I would also like to thank my committee member Wayne Schubert, Graeme Stephens, Jim Thomas and surrogate committee member Chris Kummerow for their time and interest. I would very much like to thank Celal Konor. I have greatly enjoyed our conversations about numerical modeling which were invaluable in the development of the model, and I look forward to future conversations. Discussions with my officemate Todd Ringler were also very useful.

I would like to thank my family and my parents. I would like to thank my lovely wife, Jenny. Without her help, support and gentle prodding I would still be deciding which font to use. She is a profound blessing in my life.

This work was funded by a cooperative agreement with the department of energy.

TABLE OF CONTENTS

1. Introduction

In the design of numerical models there are many vertical coordinate systems from which to choose. Any property of air which varies monotonically between the earth's surface and the top of the atmosphere can be used to make a one-to-one mapping with physical height. We will examine the advantages and disadvantages of three of these systems. Building on the work of Heikes and Randall (1995 a, b), we have developed three global dry dynamical cores whose horizontal discretization employs spherical geodesic grids. The three vertical coordinates which we will consider in some detail are a sigma-coordinate based on the work of Arakawa and Konor (1996); an isentropic vertical coordinate based on the work of Hsu and Arakawa (1990); and a generalized vertical coordinate based on the work of Konor and Arakawa (1997).

The Colorado State University general circulation model has been dramatically changed in the last few years. The horizontal grid has been changed from a latitude/longitude grid with an Arakawa C-grid staggering of mass, zonal and meridional wind to a geodesic grid in which mass, vorticity and divergence are predicted at cell centers. Throughout this evolution the vertical structure has remained unchanged. Now, we are poised to revamp the vertical structure of the model. The purpose of the study is to investigate several possible alternative vertical coordinate systems, and identify the one best suited for long-term climate prediction.

There are several new contributions in this work. We have constructed -- more-or-less from scratch -- three dynamical cores combining the geodesic grid with the vertical coordinates mentioned above. Through extensive numerical testing we have shown that the geodesic grid and the isentropic/hybrid-coordinate can be combined and work well together. This testing includes a numerical experiment designed to explore the amount of vertical noise generated with the Lorenz staggering relative to the Charney-Phillips staggering. We also compare simulations using sigma-coordinates and hybrid-coordinates using various horizontal and vertical resolutions. The purpose of these experiments is to determine the resolution necessary in order for with hybrid model to provide a simulation comparable to that obtained with the sigma model. We show that the poleward transport of tracers is simpler and more straightforward when viewed in isentropic/hybrid coordinates, relative to other coordinate systems. We will demonstrate this result using model data. And finally, we show numerical simulations of stratospheric sudden warming and baroclinic instability performed with the isentropic- and hybrid-coordinate model. These results demonstrate the geodesic-isentropic/hybrid modeling system is robust and mature.

In section 2 we will briefly review the historical development of vertical coordinate systems used in numerical models of the atmosphere. This section will also serve to define many of the ideas and equations used in later sections. The equations will, for the most part, be presented in a continuous setting. Their finite-difference analogs will be shown section 3, and in more detail in the appendices. In this section we will focus on the derivation of the three models mentioned above. Useful formulas for changing coordinate system are derived in Appendix B. In particular, we will make use of equations (136), (145) and (147) in Appendix B. For completeness, section 4 will briefly review the ideas of the spherical geodesic grid. Numerical results will be shown in section 5. Conclusions are given section 6.

2. Vertical coordinates

2.1 Geometric height

Since we use geometric length to measure horizontal distances between grid points, it would seem that the most natural vertical coordinate is geometric height. Although geometric height has rarely been used for numerical models of the atmosphere, it is a starting point for the derivation of the equations in other coordinate systems.

The continuity equation is given by

$$\frac{\partial \rho}{\partial t} + \nabla \bullet (\rho \mathbf{V}) + \frac{\partial}{\partial z}(\rho w) = 0, \tag{1}$$

where ρ is density, $\mathbf{V} \equiv (u, v)$ is the horizontal velocity and w is vertical velocity.

The horizontal momentum equation is given by

$$\frac{D\mathbf{V}}{Dt} + f\hat{\mathbf{k}} \times \mathbf{V} = -\frac{1}{\rho}\nabla_z p, \tag{2}$$

where $f \equiv 2\Omega \sin\varphi$ is the Coriolis parameter (Ω is the angular velocity of the earth and φ is latitude), and p is pressure. The total derivative D/Dt can be written

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{V} \bullet \nabla_z + w\frac{\partial}{\partial z}. \tag{3}$$

The first law of thermodynamics is given by

$$\frac{D}{Dt}\ln\theta = \frac{Q}{c_p T}, \tag{4}$$

where $\theta \equiv T(p_0/p)^{\kappa}$ is potential temperature. With this definition and the equation of state,

$$p = \rho R T, \tag{5}$$

we can write an alternative form of the first law of thermodynamics as

$$\frac{D\rho}{Dt} - \frac{1}{\gamma R T}\frac{Dp}{Dt} = -\frac{\rho Q}{c_p T}. \tag{6}$$

This is shown in Appendix A.

With height coordinates, variables we predict are, for example, ρ, $\mathbf{V}$ and p (or T or θ) from (1), (2) and (6), respectively. For large-scale motions of the atmosphere, we can use the hydrostatic equation

$$\frac{\partial p}{\partial z} = -\rho g, \tag{7}$$

where g is the gravitational acceleration of the earth. On each surface of constant z, ρ and p must evolve so that (7) is satisfied. Since both (1) and (6) implicitly contain w, the value of w is formulated to satisfy (7). This idea was first proposed by Richardson (1922).

One major disadvantage of geometric height as a vertical coordinate is that the earth's surface is not a coordinate surface. Because of topography, the height of the earth's surface, measured from some fixed position, changes depending on location. So, when calculating horizontal finite-differences in regions where a coordinate surface intersects a mountain, a grid-point's neighbors might not exist. Kasahara and Washington (1971) discuss their attempt to incorporate topography into the NCAR general circulation model. Their model consisted of six layers, the bottom two of which could contain a mountain. The pressure-gradient force in their model was applied at the centers of the model layers. Since pressure is defined at layer edges, the gradients were first computed along edges and then averaged to the centers. In the event that a grid-point was coincident with the earth's surface, it was necessary to extrapolate pressure from above to determine the pressure on the coordinate surface at a neighboring grid-point within a mountain.

This problem with the lower boundary condition can be addressed by using a normalized height or transformed height coordinates. Kasahara (1974) defines such a coordinate with a continuous, monotonic function $\tilde{z} \equiv F(z, z_S, z_T)$ where z_T is a constant height at the top of the model and z_S is the height of the earth's topography. For example,

$$\tilde{z} = \frac{z_T - z}{z_T - z_S} \tag{8}$$

has the property $0 \le \tilde{z} \le 1$, and the coordinate surface $z = 1$ is coincident with the earth's surface. It is assumed that $\tilde{z}$ is independent of time. This is in contrast to the σ-coordinate discussed in a bit. Using (145), the continuity equation can be written

$$\left(\frac{\partial}{\partial t}\right)_{\tilde{z}}\frac{\partial p}{\partial \tilde{z}} + \nabla_{\tilde{z}} \bullet \left(\frac{\partial p}{\partial \tilde{z}}\mathbf{V}\right) + \frac{\partial}{\partial \tilde{z}}\left(\frac{\partial p}{\partial \tilde{z}}\dot{\tilde{z}}\right) = 0. \tag{9}$$

2.2 Pressure

2.2.1 *Standard pressure*

Pressure monotonically decreases with increasing height since the amount of mass in the column above decreases with increasing height, so pressure is a suitable candidate for a vertical coordinate. The use of pressure was popular in the 1940s and 1950s, and it is still widely used in analytical studies. It was shown by Sutcliffe (1947) and others that the continuity equation reduces to a diagnostic equation, and that there is a simple expression for vertical velocity in pressure coordinates. Pressure coordinates have been used in numerical models by Charney and Phillips (1953) in the quasi-geostrophic setting, and by Leith (1965) with the primitive equations.

Using (145) we can write the continuity equation in p-coordinates as

$$\nabla_p \bullet \mathbf{V} + \frac{\partial \omega}{\partial p} = 0, \tag{10}$$

where $\omega \equiv \dot{p}$. Since $\partial p / \partial p \equiv 1$, the time tendency term goes to zero, and the continuity equation becomes a diagnostic equation. Suppose p_T is the pressure at the top of the model. Integrating (10) from p_T to an arbitrary point within the atmosphere, we get

$$\omega = -\int_{p_T}^{p} \nabla_p \bullet \mathbf{V} dp, \tag{11}$$

which is an expression for the vertical velocity.

With (147), the pressure-gradient term transforms very cleanly, and the momentum equation becomes

$$\frac{D\mathbf{V}}{Dt} + f\hat{\mathbf{k}} \times \mathbf{V} = -\nabla_p \Phi. \tag{12}$$

2.2.2 Normalized pressure

Pressure suffers the same drawbacks as geometric height, since the earth's surface is not a constant pressure surface. However, in this case, there is an additional complication in that pressure is also a function of time, so lines along which pressure surfaces intersect the ground are constantly moving. Analogous to the transformed height coordinate, Phillips (1957) first introduced the so-called σ-coordinate

$$\sigma \equiv p / p_S, \tag{13}$$

where p_S denotes pressure at the earth's surface. More generally, if p_T denotes a globally constant pressure at the top of the model, and p_S denotes the pressure at the earth's surface, then we can also define

$$\sigma = \frac{p - p_T}{p_S - p_T}. \tag{14}$$

So if $p_T = 0$, then (14) reduces to the original formulation by Phillips. It is clear from both (13) and (14) that $\sigma = 1$ is a surface which is coincident with the earth's surface. This is now the most common coordinate used in general circulation modeling.

A family of sigma coordinates can be defined with

$$\sigma \equiv a(p, p_S)\frac{p - p_T}{p_S - p_T} + [1 - a(p, p_S)]f(p), \tag{15}$$

where $f(p)$ is some function of pressure. Suppose the weighting function $a(p, p_S)$, where $0 \le a(p, p_S) \le 1$, transitions from $a = 1$ at the earth's surface to $a = 0$ away from the surface. Then σ is terrain-following for values of p which are sufficiently large to be surface values and transitions to $f(p)$ for smaller values of pressure. This continuous transformation from sigma to pressure is called a hybrid coordinate. Later, we will see hybrid coordinates which transition from sigma to potential temperature. It can be shown that the sigma coordinate presented by Simmons and Burridge (1981) is a member of this general family.

Next we look at the continuity equation and momentum equation in sigma coordinates. Define $p_* \equiv p_S - p_T$ to be the pressure thickness of the entire model column. With (145) we can write the continuity equation as

$$\frac{\partial p_*}{\partial t} + \nabla_\sigma \bullet (p_* \mathbf{V}) + \frac{\partial}{\partial \sigma}(p_* \dot{\sigma}) = 0. \tag{16}$$

With (16) we can predict p_* and then use (14) or (15) to diagnose pressure at each level within a model column.

Applying (136) to σ defined by (14) gives

$$\nabla_\sigma \sigma = \nabla_p \sigma + (\nabla_\sigma p)\frac{\partial \sigma}{\partial p} \tag{17}$$
$$= \nabla_p \sigma + (\nabla_\sigma p)\frac{1}{p_*} = \frac{1}{p_*}(p_* \nabla_p \sigma + \nabla_\sigma p) = \frac{1}{p_*}(\nabla_\sigma p - \sigma \nabla_p p_*) = 0.$$

Since p_* is constant throughout a column it is not a function of any vertical coordinate, hence the gradient of p_* is the same evaluated along any surface, and the subscript is dropped from the gradient operator. We will assume it is evaluated along the earth's surface. With the equation of state, we can write (17) as

$$-\frac{1}{\rho}\nabla_\sigma p = -\frac{\sigma RT}{p}\nabla p_* = -\frac{\sigma RT}{(p - p_T) + p_T}\nabla p_* = -\frac{\sigma RT}{\sigma p_* + p_T}\nabla p_*. \tag{18}$$

Again, using (134) acting on p and combining the result with the hydrostatic equation, we can write the momentum equations

$$\frac{D\mathbf{V}}{Dt} + f\hat{\mathbf{k}} \times \mathbf{V} = -g\nabla_\sigma z - \frac{\sigma RT}{\sigma p_* + p_T}\nabla p_*. \tag{19}$$

Using hydrostatic equation in the form (143) we can write

$$g\frac{\partial z}{\partial \sigma} = -\frac{1}{\rho}\frac{\partial p}{\partial \sigma} = -\frac{RTp_*}{\sigma p_* + p_T}, \tag{20}$$

which can be integrated from the earth's surface $(\sigma = 1)$ to some point in a model column,

$$gz = gz_S + Rp_* \int_\sigma^1 \frac{T}{\sigma p_* + p_T} d\sigma, \tag{21}$$

to get the geopotential used in (19).

There are numerical difficulties associated with the use of σ-coordinates in the vicinity of steeply sloping terrain. We can see from (19) that the pressure-gradient force is calculated from the sum of two terms. The first term is the gradient of the geopotential taken along a σ-surface. The second term is known as the hydrostatic correction term. Near steep topography, both terms are large in magnitude, and their difference, which is comparatively small, may contain large errors. Gary (1973) performed several numerical experiments in which he specified three different analytic functions for pressure as a function of temperature. In the presence of an artificial mountain, Gary found large errors in both σ-coordinates and the transformed height coordinates not only near the mountain, but throughout the entire model column above the mountain.

Simmons and Burridge (1981) present a model based on a form for sigma similar to that shown in (15) which conserves energy and angular momentum. If temperature is a function of only pressure, then, in continuous form, the two parts of the pressure-gradient cancel. In the finite-difference realm, however, they do not cancel. The authors specify an analytic profile of temperature as a function of pressure where temperature linearly decreases from the ground to the tropopause, and then increases again in the stratosphere. They show that the *kink* in temperature at the tropopause can produce an error similar to the error over steep topography. They show that their modified sigma coordinate, since it is more like pressure near the tropopause, behaves better than the Phillips sigma coordinate. The modified sigma coordinate became the vertical coordinate of the forecast model of the European Centre for Medium-Range Weather Forecasts.

Arakawa and Konor (1996) constructed a model based on a generalized σ-coordinate. It is called *generalized* because σ is not limited to the form given in (13) or (14). Instead, σ can be any monotone function of p and p_S, $\sigma \equiv \sigma(p, p_S)$. This is the first of three models that will be discussed in greater detail in the next section.

2.3 Potential temperature

2.3.1 Advantages and disadvantages

The potential temperature, $\theta \equiv T(p_0/p)^{\kappa}$, is strictly monotonically increasing upward in a stably stratified atmosphere. Here, p_0 is a reference pressure, typically $p_0 = 1000$ hPa, and $\kappa \equiv R/c_p$ where R is the gas constant for dry air and c_p is the specific heat at constant pressure. The conservation equation for θ is derived and discussed in Appendix A. The observed structure of potential temperature in a latitude-pressure cross-section is shown in Figure 1. The motions of the free atmosphere above the planetary boundary layer are to a good approximation adiabatic. In this case, the potential temperature of a parcel of air is conserved following its motion. Coordinate surfaces behave as Lagrangian surfaces since parcels under adiabatic advection cannot not cross constant potential temperature surfaces. There is no vertical advection in response to physical processes such as gravity waves. The surfaces simply oscillate upward and downward following the wave. When vertical advection does occur, it is due exclusively to heating. Understanding physical processes is simplified because the dynamics is isolated within the horizontal advection and pressure-gradient terms, and heating within the vertical advection terms.

The idea of a model with an Eulerian horizontal structure and a Lagrangian vertical structure was first proposed by Starr (1944). The coordinate surfaces are semi-material or quasi-Lagrangian surfaces since, unlike pure material surfaces, they cannot fold. If this were to happen there would be regions in which $\partial\theta/\partial z < 0$, violating our assumptions about monotonicity. This is not a serious limitation since unstable layers exist in the real atmosphere only on relatively small time and length scales, smaller than the phenomena we are interested in modeling. Numerical models may develop unstable

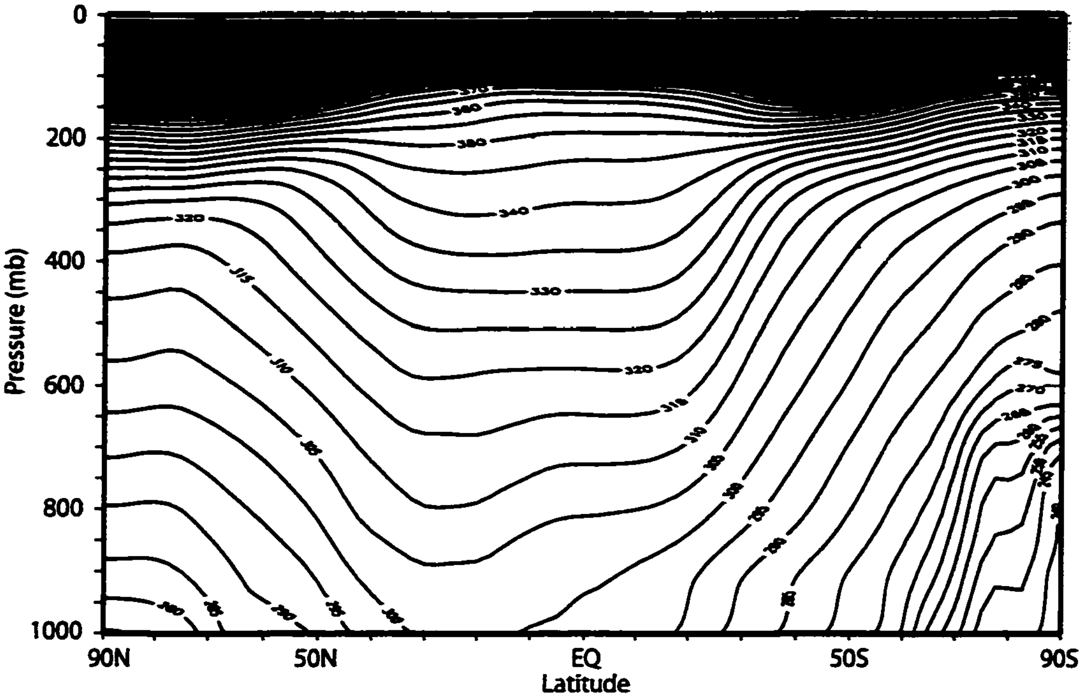


Figure 1. The observed, zonally-averaged distribution of potential temperature for July 1999. Figure provided by Dr. Konor.

columns, but these columns are brought back to neutral stability with a convective adjustment process.

The use of potential temperature provides superior resolution of fronts. As a front forms, θ-surfaces pack tightly together parallel to the front and perpendicular to the strong horizontal temperature gradient. So, a θ-coordinate model with a relatively coarse horizontal resolution can still accurately simulate the shape and intensity of a front. The slope of the coordinate surfaces can become quite steep, so increasing the vertical resolution of a model has the added benefit of increasing the horizontal resolution. Arakawa *et al.* (1992) demonstrates the utility of isentropic coordinates in simulating the evolution of mid-troposphere mid-latitude baroclinic events including tropopause folding. The structure of the atmosphere and mechanisms which contribute to these events are considered.

The pressure-gradient force in isentropic coordinates is given by $-\nabla_p\Phi = -\nabla_\theta M$, which is irrotational along a surface of constant θ. Hence, the pressure-gradient force does not produce a horizontal circulation along a coördinate surface, and there are no pressure-gradient terms in the vorticity equation. A vorticity-divergence representation of momentum coupled with the θ-coordinate is a particularly good combination. Since the mass or pseudo-density is defined $m = -\partial p/\partial\theta$, and the Ertel potential vorticity is defined as

$$PV \equiv \frac{f + \hat{\mathbf{k}} \bullet \nabla_\theta \times \mathbf{V}}{\partial p/\partial\theta}, \tag{22}$$

it is easy to see that a model that conserves absolute vorticity under horizontal advection will also conserve the mass-weighted potential vorticity under horizontal advection.

The disadvantages of the θ-coordinate are primarily realized at the lower boundary. In the real atmosphere, the planetary boundary layer or PBL is well-mixed due to turbulent mixing caused by heating at the earth's surface. Properties of the air are constant throughout the depth of the PBL. In the case $\partial\theta/\partial z = 0$, the isentropic surfaces become vertical, which means that we have no vertical resolution with θ-coordinates. Several authors have suggested that the PBL be separated from the free atmosphere. This approach is discussed in greater detail in the next section.

With the isentropic coordinate the earth's surface is not a constant θ surface as shown in Figure 1. It is possible to normalize θ by the surface value of θ, but this eliminates the Lagrangian properties of the pure θ-coordinate. There are several methods to handle this problem, including subterranean extrapolation and massless layers, which will be discussed later.

2.3.2 Historical review of isentropic coordinates

This review will be organized around the various strategies which have been devised to handle the problem at the lower boundary. The first models based on this coordinate were constructed by Eliassen and Raustein (1968). Their primitive equation model consisted of two layers. The coordinate surface separating them was allowed to intersect the ground. The potential temperature at the earth's surface, θ_S, is governed by

$$\left(\frac{\partial\theta}{\partial t}\right)_S + \mathbf{V}_S \bullet \nabla_S\theta_S = \dot{\theta}_S. \tag{23}$$

In a given model column, coordinate surfaces with values of θ greater than θ_S will be up in the air. The questions then are how to treat those surfaces with values of θ less than θ_S, and how to compute differences along a surface where it is intersecting the ground. Eliassen and Raustein employ a method called subterranean extrapolation. This is shown in Figure 2. Here, three coordinate surfaces are shown (solid lines) as well as the earth's

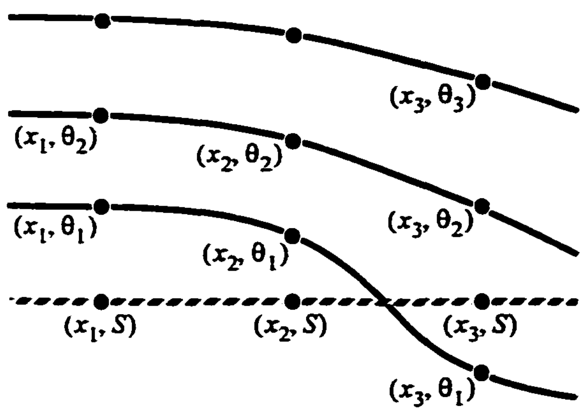


Figure 2. Subterranean extrapolation. To compute centered numerical derivatives at the point (x_2,θ_1), we need values at its neighbors, (x_1,θ_1) and (x_3,θ_1). By predicting the surface potential temperature at the point (x_3,S) we can see that the point (x_3,θ_1) is *underground* since the surface θ is greater than θ_1. Values at (x_3,θ_1) are approximated by linear extrapolation from points (x_3,S) and (x_3,θ_2) (Eliassen and Raustein (1968)) or from points (x_3,θ_2) and (x_3,θ_3) (Shapiro (1975)).

surface (dashed line). In the event that the earth's surface warms, the difference between the surface θ_S and θ_2 will become small, and the extrapolation will be poorly behaved. To counteract this problem, Bleck (1974) interpolates from surrounding surfaces to an auxiliary point a fixed $\Delta\theta$ above the ground and then uses this value and the surface value to compute the subterranean value.

The two-layer model was extended to 20 layers by Shapiro (1975). Starting from an analytic initial condition, the model simulates upper-level frontogenesis. Shapiro notes the model's ability to compress isentropic surfaces in the region of frontogenesis, and to stretch them apart above and below the front. In this way, the model has a variable-resolution coordinate system. Because potential temperature increases rapidly with increasing height in the stratosphere, the denominator of (22) will be small there, and so stratospheric values of PV are large compared to tropospheric values. The model simulates the process of tropopause folding in which a tongue high potential vorticity air is extruded from the stratosphere into the upper troposphere.

Bleck (1974) discusses two models, one based on the geostrophic balance equations which conserve the geostrophic potential vorticity, and the other based on the primitive equations. Bleck was interested in demonstrating one model's superiority to the other. Since the geostrophic model does not allow gravity waves, it was thought that noise generated by the lower boundary would contaminate the results less. The subterranean extrapolation approach was only marginally successful. Models based on the primitive equations had difficulty suppressing gravitational noise near the earth's surface, and he was unable to construct discrete models which conserve integral quantities such as total energy and angular momentum.

In a sense the underground extrapolation method is thinking about the problem in z-space, in that the ground is thought of as flat, a constant-z surface, which the θ-surfaces intersect. With the massless layer approach, the θ-surfaces are more like a uniform grid, on which the distribution of mass determines not only the thermodynamic structure, but also the location of the ground. To motivate the idea of massless layers, consider the model layer sandwiched between the $\theta = 250$ K and $\theta = 260$ K coordinate surfaces. In the real world, one would expect to find a parcel of air with a value of $\theta = 255$ K near the ground at high latitudes. In a numerical model, this parcel would be between the two specified θ-surfaces. On the other hand, in a tropical column, there are no parcels of air with $\theta = 255$ K, and the 250 K and 260 K coordinate surfaces, which we assume cover the entire globe, have no mass between them. This tropical portion of the model layer is called massless. In a latitude-height cross section, the coordinate surfaces intersect the earth's surface somewhere in mid-latitudes, and we imagine them plastered against the ground throughout the tropics, massless, and emerging again somewhere in the opposite hemisphere. There is no equation for potential temperature at the earth's surface. Instead, the thermodynamic structure is determined solely by the distribution of mass in θ-space. If $\dot{\theta} = 0$, then the three-dimensional model is really just a stack of two-dimensional shallow-water models. The "topography" at the base of a given layer is the free surface of the layer below. The spirit of this idea was first proposed by Lorenz (1955), and was further refined by Bleck (1984). Since mass must be positive-definite, numerical methods such as flux-corrected transport (FCT; Boris and Book, 1973) are required to prevent the formation of negative mass in the massless regions. If we assume the continuity equation is written in flux form, the idea of FCT is to use a linear combination of a high-order flux and a low-order flux. In regions where the mass field is smooth and bounded away from zero, the linear combination of fluxes is exclusively the high-order flux. On the other hand, in regions of sharp gradients where dispersion errors may cause negative mass to form, the linear combination selects the first-order flux to prevent the formation of negative mass.

Hsu and Arakawa (1990) built an isentropic-coordinate model based on the massless layer approach. This is the second of three models that we will discuss in detail in the next section. Their model lives in a zonal channel on a β-plane. The initial condition consists of a zonally uniform jet and geostrophically balanced pressure field with a superimposed small-amplitude wavenumber one perturbation. The purpose of the simulation was primarily to demonstrate the feasibility and performance of the coordinate.

The model simulates well the baroclinic growth of wavenumber one. The authors comment on the surprisingly good quality of the dynamical features despite the relatively coarse horizontal resolution. They also draw attention to some difficulties with the model. It is difficult to determine the surface air temperature in the model, making it difficult to compute the surface heat sensible flux. It is also difficult to determine a smooth horizontal distribution of surface temperature. As a result gradients of the Montgomery potential obtained from the discrete hydrostatic equation are nearly discontinuous. This problem is discussed more in Randall *et al.* (2000).

2.4 Hybrid vertical coordinate

Several authors have proposed a marriage of sigma and potential temperature. The main idea is to combine the terrain-following properties of the σ-coordinate near the earth's surface, and the quasi-Lagrangian properties of the isentropic coordinate in the free atmosphere. The coordinate is constructed so that it will quickly transition from sigma to potential temperature with increasing height, so that the advantages of the isentropic coordinate are realized throughout the majority of the model atmosphere.

2.4.1 Composites of σ and θ

In this approach, the atmosphere is partitioned into two concentric shells. The σ-coordinate is used in the shell adjacent to the earth surface, and the outer shell, the "free atmosphere" shell, uses the θ-coordinate. The question is how to define the internal boundary conditions at the interface between the two domains.

One possibility is for the surface to be initially coincident with a constant pressure surface at about 700 hPa, and then evolve as a material surface. That is, no mass crosses the interface. This was the approach proposed by Deaven (1976). Suppose that p_I is the pressure at the interface between the two domains. The definition of sigma is then

$$\sigma \equiv \frac{p - p_I}{p_S - p_I}. \tag{24}$$

Integrating (145) with $z_2 \equiv \theta$ from the top of the isentropic layer, θ_T, down to the interface θ_I, we have

$$\left(\frac{\partial p_I}{\partial t}\right)_\theta = -\int_{\theta_T}^{\theta_I} \nabla_\theta \bullet \left(\frac{\partial p}{\partial \theta}\mathbf{V}\right) d\theta. \tag{25}$$

Similarly, integrating (145) over the sigma domain, we can write

$$\left(\frac{\partial p_S}{\partial t}\right)_\sigma = \left(\frac{\partial p_I}{\partial t}\right)_\sigma - \int_0^1 \nabla_\sigma \bullet \left(\frac{\partial p}{\partial \sigma}\mathbf{V}\right) d\sigma, \tag{26}$$

since $\dot{\sigma} = 0$ at $\sigma = 0$ and $\sigma = 1$. With (135), the local time rate of change at the interface on an isentropic surface and a sigma surface are related with

$$\left(\frac{\partial p_I}{\partial t}\right)_\sigma = \left(\frac{\partial p_I}{\partial t}\right)_\theta + \left(\frac{\partial p}{\partial \theta}\right)\left(\frac{\partial \theta}{\partial t}\right)_\sigma. \tag{27}$$

Of course the isentropes still plunge down through the interface and into the sigma domain. Deaven extrapolates from the isentropic domain, in a fashion exactly analogous to subterranean extrapolation, to approximate values in the sigma domain. In hybrid coordinate models this problem is less severe than in the pure isentropic coordinate, since in the hybrid models the angle of the intersection with the interface is more shallow than the angle of the intersection with the earth's surface.

Since the serious problems associated with the pure isentropic coordinate are related to the lower boundary, Deaven devised a set of numerical experiments to emphasize the planetary boundary layer. In the first experiment, wind is directed over a bell shaped mountain. The potential temperature field evolves to form standing waves which tilt upstream with increasing height. This numerical experiment agrees well with observations. In a second experiment, Deaven applies a heating in a limited region at the model surface. Two columns of air with large vertical velocities form at the lateral edges of the heated region. These rising columns act as a barrier to the horizontal flow in much the same way as a mountain. Standing waves similar to lee waves are formed in response to the thermodynamically induced obstacle.

Uccellini *et al.* point out some limitations of Deaven's material-surface strategy. Deep tropospheric circulations are characterized by inflow in the boundary layer, upward transport throughout the entire vertical column and outflow in the upper troposphere. With the interface as a material surface the sigma domain would eventually constitute the entire troposphere.

Friend *et al.* (1977) set the interface to be a constant p_I (=700 hPa) surface. Within a model column, consider the first isentrope above the 700 hPa surface. With (138), the vertical velocity on the isentrope is given by

$$\omega(\theta) \equiv \dot{p} = \left(\frac{\partial p}{\partial t}\right)_\theta + \mathbf{V} \bullet \nabla_\theta p, \tag{28}$$

where we assume the isentropic domain is adiabatic. Integrating (10) from the first isentropic above the interface down to the interface gives

$$\omega(700) = \omega(\theta) - \int_{p=700}^{p=p(\theta)} \nabla \bullet \mathbf{V} dp. \tag{29}$$

We can substitute for $\omega(\theta)$ from (28). Analogous to (16), we define $p_* \equiv p_S - p_I$ to be the pressure thickness of the sigma domain. Then, with (24) and the fact that $\sigma = 0$ at $p = p_I$,

$$\dot{\sigma}(700) = \frac{1}{p_*}[\dot{p} - \sigma\dot{p}_*] = \frac{1}{p_*}\omega(700) \tag{30}$$

gives an expression for the vertical velocity at the interface.

Friend *et al.* point out that thermal instability may occur at the interface. For example, suppose within a model column, the value of θ on an isentropic coordinate surface is less than the value of θ at the interface, indicating that the isentrope had plunged into the sigma domain. It is still possible that the pressure predicted on the isentrope is less than $p_I = 700$ hPa. From a pressure point of view, the isentrope is still above the interface. And, of course, the opposite situation can also occur. That is the predicted pressure on a theta coordinate surface may be greater than p_I, while θ at the interface is greater than θ of the coordinate surface. Friend *et al.* describe an adjustment process like dry convective adjustment to remove this instability.

Friend *et al.* test the model with a simulation whose domain covers the North American continent. They show one plot of sea level pressure from a 12 hour simulation which was initiated from observations at 00 GMT 20 January 1973. Their simulation captures the basic features of the corresponding observations. However, the purpose of their numerical results was more to demonstrate the feasibility of the model, rather than produce an accurate simulation.

Yet another candidate for the interface is $p_I = p_S - \Delta p$, where Δp is constant. This idea was proposed by Uccellini *et al.* (1979). Their model is based on the flux form of the primitive equations. By matching the boundary conditions at the interface between the sigma and isentropic models, they were able to design a model which, in the continuous case, conserves mass, momentum and energy with respect to transport across the interface. These conservation properties are lacking in the previous two models.

The authors write expressions for the mass-weighted integral of an arbitrary quantity, f, in the isentropic domain, F_2, and the sigma domain, F_1,

$$F_2 = \int_A \int_{\theta_I}^{\theta_T} \left(-\frac{1}{g}\frac{\partial p}{\partial \theta}\right) f d\theta dA \quad \text{and} \quad F_1 = -\int_A \int_1^{\sigma_I} \left(\frac{1}{g}\frac{\partial p}{\partial \sigma}\right) f d\sigma dA . \tag{31}$$

The limits of integration are invariant except for the lower edge of the isentropic domain, θ_I. Leibniz's rule is used to differentiate F_1 and F_2 with respect to time. The authors then substitute the appropriate form of the continuity equation, and compare the terms responsible for vertical mass transport to show that the transport out of the sigma domain is equal and opposite that into the isentropic domain. They then use these integral equations to construct the flux form finite-difference equations for mass and momentum in their model.

In the event that an isentropic surface intersects the interface, Uccellini *et al.* use values from within the sigma domain to interpolate to the submerged isentropic surface in order to compute gradients, rather than extrapolating from the isentropic domain.

The model was tested with a simulation of a jet streak, that is, a jet with zonally varying intensity. While their results compares well with results from similar simulations, their primary interest is in the model's conservation properties. In the isentropic domain, mass conservation can be checked by calculating the area-weighted average pressure on each isentrope. They found that the area-averaged pressure on isentropes which never intersected the domain interface remains constant. However, the area-averaged pressure on those isentropes which did intersect the interface gradually increased. In their model, no quantities are predicted at the interface. They attribute the increasing mass to the interpolation of θ to the interface, and interpolation of p to isentropes in the sigma domain.

This vertical coordinate system has enjoyed some success over the years, and was used in the University of Wisconsin global climate model. This model is discussed in Zapotocny *et al.* (1994).

Yet a fourth definition for the interface was proposed by Bleck (1978). Bleck sets the interface to be a constant potential temperature surface. Since the surface should never touch the Earth's surface, its value of theta needs to be larger than any value found at the ground. Bleck chooses 310 K. In an operational, global model this value would have to be much larger in order to exceed the very high values of surface potential temperature over regions such as the Tibetan plateau. The main advantage of this scheme, of course, is that it avoids the problem of isentropes intersecting the interface. A big disadvantage is that a large portion of the troposphere will be in the sigma domain even with a relatively small value of θ at the interface.

Bleck (1978) compares the four methods outlined above. The initial conditions represent a growing baroclinic wave. To make the simulation more challenging, he adds a large Gaussian-shaped mountain, and checkerboard noise to the initial wind fields. Bleck suggests a diagnostic measure of the $2\Delta t$ noise in a model, and uses this diagnostic tool to conclude the hybrid models are 30-40% *quieter* than a model based on a pure theta coordinate.

2.4.2 ALE

A second type of hybrid coordinate has been adopted from the ocean modeling community. This coordinate is know as Arbitrary Lagrangian-Eulerian or *ALE*. In the oceans, the quantity analogous to potential temperature is density. Surfaces of constant density are called isopycnals. Isopycnic coordinates in the ocean enjoy many of the same benefits as isentropic coordinates do in the atmosphere, such as improved resolution near fronts. They also are burdened by similar drawbacks in that isopycnal surfaces intersect or outcrop from the sea surface and also intersect the ocean bottom. Bleck and Boudra (1981) propose a coordinate in which isopycnals are used in the deep oceans and physical depth is used near the sea surface. In their model a single coordinate surface can have both isopycnal and depth properties.

The idea is this. Suppose we are considering a coordinate system s. Then the continuity equation (145) becomes

$$\frac{\partial}{\partial t}\left(\frac{\partial p}{\partial s}\right) + \nabla_s \bullet \left(\mathbf{v}\frac{\partial p}{\partial s}\right) + \frac{\partial}{\partial s}\left(\dot{s}\frac{\partial p}{\partial s}\right) = 0. \tag{32}$$

Integrating (32) from $s = s_T$ at the top of the model to an arbitrary point in a column, $s = s_1$, gives

$$\left.\frac{\partial p}{\partial t}\right|_{s_1} + \left.\dot{s}\frac{\partial p}{\partial s}\right|_{s_1} = \int_{s_1}^{s_T} \nabla_s \bullet \left(\mathbf{v}\frac{\partial p}{\partial s}\right) ds. \tag{33}$$

So, the horizontal convergence or divergence of mass in the column above the point s_1 is balanced by the vertical displacement of the coordinate surface, $\partial p / \partial t$, and the vertical motion relative to the coordinate surface, $\dot{s}\partial p / \partial s$. The important point is that the above analysis was made without the need to provide a definition for s. Hence, not only can the definition of s change depending on the position of a coordinate surface within a model column, as was the case in the composite models, but the definition can also change along an individual coordinate surface.

Bleck describes the algorithm for positioning the coordinate surfaces as follows. Imagine an ocean basin which is initially at rest with no horizontal gradients of salinity. In this case, the initial the s surfaces will be coincident with isopycnals. If a wind stress is applied to the surface, as the model runs, upwelling with eventually try to draw an s surface from depth to the sea surface. However, the algorithm forces the surface to stop some distance below the sea surface, and allows the dense water to pass through the surface. This will create a positive density anomaly in the upper layer. If the upwelling continues, as more s surfaces are drawn from depth, they are stopped before touching the surface above them, so as to maintain a prescribed minimum layer thickness. So the coordinate surfaces are constant density surfaces in the deep ocean, and constant geometric depth surfaces near the sea surface. A similar coordinate is currently employed in the mesoscale analysis and prediction system (MAPS) at NOAA/FSL. This model is used to make short-term regional weather predictions.

2.4.3 Continuous transformation

In the 1990's the discontinuous interface between the sigma domain and theta domain, characteristic of the composite models, was replaced with a continuous transition. Zhu *et al.* (1992) extend the ideas of Simmons and Burridge from a σ-p coordinate to a σ-θ-p coordinate. In a model column, they define the pressure on each coordinate surface so that it is a function of temperature and surface pressure. That is,

$$c_4 p = c_1 + c_2(p_S - p_T) + c_3 p T^{-1/\kappa}. \tag{34}$$

The coefficients c_1, c_2, c_3 and c_4 define the coordinate system. These coefficients change depending on the position of a coordinate surface within a model column. In this way, the coordinate can smoothly transition from σ at the earth's surface to θ and to p at the top of the model. The vertically integrated continuity equation gives a prognostic equation for the surface pressure. This coupled with the thermodynamic equation allows pressure at layer edges to be diagnosed with (34).

Konor and Arakawa (1997) propose a more elegant hybrid vertical coordinate called the generalized vertical coordinate. This is the third of three models that we will discuss in detail in the next section. In the so-called generalized vertical coordinate, denoted ζ, the vertical coordinate can be any monotonic function of pressure, surface pressure and potential temperature, i.e. $\zeta \equiv F(\theta, p, p_S)$.

3. A detailed look at three models

In this section we will discuss three of the models surveyed in the previous section. Each vertical coordinate implies a different set of discrete prognostic equations. We predict mass or a thermodynamic quantity (or both) and momentum. For each coordinate we will briefly outline the resulting discrete equations and leave many of the details of the derivations for the appendices.

3.1 Vertical discretization

In a numerical model the atmosphere is divided into discrete layers. Each layer is separated from its neighbors by an edge. These edges are coordinate surfaces. We can imagine another surface in the middle of each layer sandwiched between each pair of successive edges. Variables defined at the layer centers describe the properties of the layer as a whole. We denote layer centers with a dashed line, and layer edges with a solid line. The top and bottom of a column are layer edges. Suppose a column consists of L layers. Our convention will be to label the layer centers with whole-integers $\{1, 2, \ldots, l-1, l, l+1, \ldots, L\}$ and layer edges will be labeled with half-integer values $\{1/2, 3/2, \ldots, l-1/2, l+1/2, \ldots, L+1/2\}$. Note that layer index increases down.

A finite-difference discretization of the continuous equations requires specification of the vertical distribution of prognostic variables. Pressure and vertical velocity are placed at the layer edges, and momentum is placed at the layer centers. Note that the vertical velocity depicted in Figure 3 is $\dot{\sigma}$, which implies sigma coordinates. Alternatively, it could also have been $\dot{\theta}$ or $\dot{\zeta}$. Historically there have been two distributions for the thermodynamic variable. The Lorenz staggering (L-grid) places temperature at the layer centers with momentum. Lorenz (1960) constructed a model based on the primitive equations. His model conserves total energy, mean potential temperature and potential temperature variance under reversible adiabatic processes. The L-grid is currently used in many general circulation models including the CSU general circulation model. On the other hand, the Charney-Phillips (CP-grid) staggering (Charney and Phillips, 1953) places temperature at the layer edges with pressure. All three models featured in this section are based on the CP-grid.

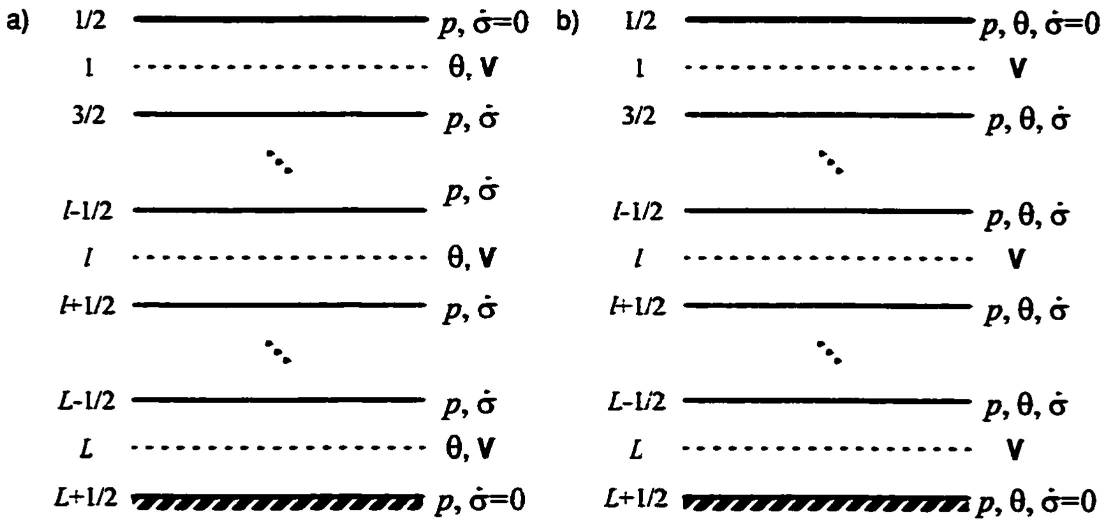


Figure 3. Vertical staggering of variables in σ-coordinates. a) Lorenz staggering. b) Charney-Phillips staggering.

3.2 Model design

The discrete finite-difference equations should be designed to mimic the continuous equations. In a landmark work, Arakawa and Lamb (1977) present a set of integral constraints which guide the design of finite-difference schemes through, for example, conservation to total energy and mass weighted potential temperature. This work and that of Arakawa and Suarez (1983) form the foundation of the current CSU general circulation model, which uses the sigma-coordinate. Arakawa and Lamb impose four constraints in their design of the UCLA general circulation model. These include:

1) The pressure gradient force generates no circulation of vertically integrated momentum along a contour of the surface topography.

2) The energy conversion term has the same form in the kinetic energy and thermodynamic energy equations with opposite signs.

3) the global mass integral of the potential temperature is conserved under adiabatic processes.

4) The global mass integral of a function of potential temperature, such as θ^2 or $\ln\theta$, is conserved under adiabatic processes.

We will see that the three models are designed with these principles in mind. As a result of enforcing these constrains, the hydrostatic equation is non-local, that is, the thickness of the bottom layer depends on the thicknesses within the rest of the model column. Arakawa and Suarez (1983) show that non-local dependence can affect the local accuracy of the hydrostatic equation. With a minor relaxation on the constraints for conservation of potential temperature as presented in Arakawa and Lamb, Arakawa and Suarez obtain a family of schemes for which the hydrostatic equation is local. Each model presented here has a local hydrostatic equation.

3.3 Charney-Phillips generalized sigma-coordinate

We have already introduced the sigma coordinate in which pressure is normalized with surface pressure. Arakawa and Konor (1996) generalize this idea and define σ to be any monotonic function of pressure and surface pressure. This function is denoted by

$$\sigma \equiv F(p, p_S). \tag{35}$$

The form of (35) allows p to be expressed as a function of σ and p_S, then we need only predict p_S in order to have p at each layer edge. The other two coordinates systems we will consider require prediction of mass throughout the entire model.

3.3.1 *Continuity*

Define the pseudo-density

$$m \equiv \frac{\partial p}{\partial \sigma}. \tag{36}$$

So, m tells how much mass is sandwiched between two constant σ surfaces. We have $m > 0$ since we assume that σ decreases with increasing height. With (145), the time rate of change of m is given by

$$\left(\frac{\partial m}{\partial t}\right)_\sigma + \nabla_\sigma \bullet (m\mathbf{V}) + \frac{\partial}{\partial \sigma}(m\dot{\sigma}) = 0, \tag{37}$$

where the subscript σ indicates the differentiation takes place on a surface of constant σ. In discrete form (37) is given by

$$\frac{\partial m_l}{\partial t} + \nabla \bullet (m\mathbf{V})_l + \frac{1}{(\delta\sigma)_l}[(m\dot{\sigma})_{l+1/2} - (m\dot{\sigma})_{l-1/2}] = 0 \text{ for } l = 1, 2, \ldots, L, \tag{38}$$

where

$$m_l \equiv \frac{p_{l+1/2} - p_{l-1/2}}{(\delta\sigma)_l} \text{ for } l = 1, 2, \ldots, L \tag{39}$$

is the discrete form of (36) and $(\delta\sigma)_l = \sigma_{l+1/2} - \sigma_{l-1/2}$ for $l = 1, 2, \ldots, L$. The vertical mass flux terms in (38) are derived in Appendix D. With the boundary conditions $(m\dot{\sigma})_{1/2} = (m\dot{\sigma})_{L+1/2} = 0$, (38) can be vertically summed as follows

$$\frac{\partial p_S}{\partial t} = -\sum_{l=1}^{L} \nabla \bullet (m\mathbf{V})_k (\delta\sigma)_k \tag{40}$$

to give an equation for p_S.

With Helmholtz's equation, the horizontal velocity, $\mathbf{V}$, can be written in terms of streamfunction, ψ, and velocity potential, χ, as follows: $\mathbf{V} = \hat{\mathbf{k}} \times \nabla\psi + \nabla\chi$. In this case the horizontal advection terms in (38) becomes

$$\frac{\partial m_l}{\partial t} - J(m_l, \psi_l) + \nabla \bullet (m_l \nabla \chi_l) + \frac{1}{(\delta\sigma)_l}[(m\dot{\sigma})_{l-1/2} - (m\dot{\sigma})_{l+1/2}] = 0 \qquad (41)$$

$$\text{for } l = 1, 2, \ldots, L.$$

The details are shown in (284) of Appendix O.

3.3.2 Momentum

The momentum equation can be written

$$\frac{D\mathbf{V}}{Dt} = \left(\frac{\partial \mathbf{V}}{\partial t}\right)_\sigma + \mathbf{V} \bullet \nabla_\sigma \mathbf{V} + \dot{\sigma}\frac{\partial \mathbf{V}}{\partial \sigma} = -(\nabla_p \Phi) - f\hat{\mathbf{k}} \times \mathbf{V} + \mathbf{F}, \qquad (42)$$

where $\mathbf{V}$ is the horizontal velocity. The Coriolis parameter is defined $f \equiv 2\Omega\sin\varphi$ where Ω is the angular velocity of the earth, and φ is latitude. The unit vector $\hat{k}$ is in the vertical direction, and $\mathbf{F}$ is the force due to friction. Note that the first equality in (42) defines the total derivative D/Dt.

The horizontal advection term in (42) can be transformed with the vector identity

$$\mathbf{V} \bullet \nabla\mathbf{V} = (\nabla \times \mathbf{V}) \times \mathbf{V} + \nabla[(\mathbf{V} \bullet \mathbf{V})/2] \qquad (43)$$

to give

$$\frac{\partial \mathbf{V}}{\partial t} + (f + \zeta)\hat{\mathbf{k}} \times \mathbf{V} + \nabla[(\mathbf{V} \bullet \mathbf{V})/2] + \dot{\sigma}\frac{\partial \mathbf{V}}{\partial \sigma} = -(\nabla_p \Phi) + \mathbf{F}, \qquad (44)$$

where ζ is relative vorticity. Let $\eta = f + \zeta$ be the absolute vorticity, and $K = (\mathbf{V} \bullet \mathbf{V})/2$ be the kinetic energy. Again, we write the horizontal velocity in terms of a streamfunction and velocity potential, $\mathbf{V} = \hat{\mathbf{k}} \times \nabla\psi + \nabla\chi$. Taking the curl of (44) gives an equation for the tendency of vorticity,

$$\hat{\mathbf{k}} \bullet \left[\nabla \times \frac{\partial \mathbf{V}}{\partial t}\right] = \frac{\partial \eta}{\partial t} = \ldots - \hat{\mathbf{k}} \bullet [\nabla \times (\eta \hat{\mathbf{k}} \times \mathbf{V} + \nabla K)] = \ldots + J(\eta, \psi) - \nabla \bullet (\eta \nabla \chi). \quad (45)$$

Note that the vertical advection, pressure-gradient and friction have been neglected. The details are shown in (286) of Appendix O. Taking the divergence of (44) gives an equation for the tendency of divergence.

$$\nabla \bullet \left[\frac{\partial \mathbf{V}}{\partial t}\right] = \frac{\partial \delta}{\partial t} = \ldots - \nabla \bullet (\eta \hat{\mathbf{k}} \times \mathbf{V} + \nabla K) = \ldots + J(\eta, \chi) + \nabla \bullet (\eta \nabla \psi) - \nabla^2 K. \quad (46)$$

The details are shown in (289) of Appendix O.

Kinetic energy in terms of stream function and velocity potential is given by

$$\begin{aligned} \frac{1}{2}(\mathbf{V} \bullet \mathbf{V}) &= \frac{1}{2}[(\hat{\mathbf{k}} \times \nabla\psi + \nabla\chi) \bullet (\hat{\mathbf{k}} \times \nabla\psi + \nabla\chi)] \\ &= \frac{1}{2}[\nabla\psi \bullet \nabla\psi + \nabla\chi \bullet \nabla\chi] + \hat{\mathbf{k}} \bullet (\nabla\psi \times \nabla\chi) \\ &= \frac{1}{2}[\nabla \bullet (\psi \nabla \psi) - \psi \nabla^2 \psi + \nabla \bullet (\chi \nabla \chi) - \chi \nabla^2 \chi] + J(\psi, \chi). \end{aligned} \quad (47)$$

With (136) acting upon Φ, the pressure-gradient force can be written $-\nabla_p \Phi = -\nabla_\sigma \Phi - (RT/p_S)\nabla p_S$. Arakawa and Konor write the discrete form of the pressure-gradient term as follows

$$-(\nabla_p \Phi)_l = -\frac{1}{m_l}\nabla_\sigma(m_l \Phi_l) + \frac{1}{m_l(\delta\sigma)_l}(\Phi_{l+1/2}\nabla_\sigma p_{l+1/2} - \Phi_{l-1/2}\nabla_\sigma p_{l-1/2}) \quad (48)$$

$$\text{for } l = 1, 2, \ldots, L.$$

Eq. (48) is derived in Appendix F. Unfortunately, the pressure-gradient term is not a pure gradient, so it contributes terms to the vorticity equation.

3.3.3 Thermodynamics

The thermodynamic energy equation is given by

$$c_p\left[\frac{\partial T}{\partial t}+\mathbf{V}\bullet\nabla T+\dot{\sigma}\frac{\partial T}{\partial\sigma}\right] = \alpha\omega+Q. \tag{49}$$

We can combine (37) and (49)

$$\frac{\partial}{\partial t}(mc_pT)+\nabla\bullet(mc_pT)+\frac{\partial}{\partial\sigma}(mc_pT\dot{\sigma}) = m\alpha\omega+mQ. \tag{50}$$

Eq. (50) can be vertically discretized as follows:

$$\frac{\partial}{\partial t}(m_{l+1/2}c_pT_{l+1/2})+\nabla\bullet[(m\mathbf{V})_{l+1/2}c_pT_{l+1/2}]+c_p\frac{(Tm\dot{\sigma})_{l+1}-(Tm\dot{\sigma})_l}{(\delta\sigma)_{l+1/2}} \tag{51}$$
$$= (m\alpha\omega)_{l+1/2}+(mQ)_{l+1/2} \text{ for } l = 1,2,\ldots,L-1,$$

where

$$(m\mathbf{V})_{l+1/2}\equiv\frac{1}{2(\delta\sigma)_{l+1/2}}[(\delta\sigma)_l(m\mathbf{V})_l+(\delta\sigma)_{l+1}(m\mathbf{V})_{l+1}] \text{ for } l = 1,2,\ldots,L-1 \tag{52}$$

is the horizontal mass flux defined at a layer edge. Note that according to (51) the thermodynamic T is predicted at layer edges as expected with the CP-grid. For the upper and lower boundary, (50) becomes

$$\frac{\partial}{\partial t}(m_{1/2}c_pT_{1/2})+\nabla\bullet[(m\mathbf{V})_{1/2}c_pT_{1/2}]+\frac{c_p}{(\delta\sigma)_{1/2}}(Tm\dot{\sigma})_1 = (m\alpha\omega)_{1/2}+(mQ)_{1/2} \tag{53}$$

and

$$\frac{\partial}{\partial t}(m_{L+1/2}c_pT_{L+1/2})+\nabla\bullet[(m\mathbf{V})_{L+1/2}c_pT_{L+1/2}]-\frac{c_p}{(\delta\sigma)_{L+1/2}}(Tm\dot{\sigma})_L \tag{54}$$
$$= (m\alpha\omega)_{L+1/2}+(mQ)_{L+1/2}$$

In eqs. (53) and (54), $(m\mathbf{V})_{1/2}\equiv m_1\mathbf{V}_1$ and $(m\mathbf{V})_{L+1/2} = m_L\mathbf{V}_L$. The terms of the form $(Tm\dot{\sigma})_l$ and the energy conversion terms $(m\alpha\omega)_{l+1/2}$ still need to be defined. By carefully constructing these term, Arakawa and Konor enforce conservation of total energy. This is discussed in Appendix E.

Arakawa and Konor (1996) discuss the growth and persistence of spurious waves due to a computational mode found in the L-grid. This computational mode is not present in the CP-grid as discussed in section 5.2. With a linearized version of each model (L and CP) Arakawa and Konor performed a simulation of horizontal standing waves. Starting with an initial isothermal state, they applied a small perturbation to the temperature field at the bottom of a model column. In the L-grid model, this perturbation is trapped near the surface, and the stationary wave does not propagate vertically, while in the CP-grid model the wave does propagate vertically and disperses.

In a second test they demonstrate another difficulty of the L-grid which is discussed in Arakawa and Moorthi (1988). This is the spurious growth of short waves due to false satisfaction of the criteria for baroclinic instability. To demonstrate this problem, they initialize both models with a baroclinic zonal jet in geostrophic balance. They add a random perturbation to θ with an amplitude of ± 1 K at all levels. The evolutions of the two simulations are dramatically different. The L-grid model develops small-scale noise that contaminates and almost destroys the solution, while the CP-grid simulation is dominated by the graceful, large-scale structures indicative of baroclinic instability.

3.4 Isentropic coordinate

This model is based on the work of Hsu and Arakawa (1990). The prognostic equations in isentropic coordinates are particularly simple and elegant. As mentioned above, there is no thermodynamic equation. In the continuity equation, the vertical mass flux is determined solely by the heating, and the pressure-gradient force is irrotational so it does not appear in the vorticity equation.

3.4.1 Continuity

Now we want the amount of mass between two isentropes. Hence, mass is defined as

$$m \equiv -\frac{\partial p}{\partial \theta}. \tag{55}$$

With (145), the change in m is given by

$$\left(\frac{\partial m}{\partial t}\right)_\theta + \nabla_\theta \bullet (m\mathbf{V}) + \frac{\partial}{\partial \theta}(m\dot{\theta}) = 0. \tag{56}$$

Here the subscript θ indicates that differentiation takes place on isentropes. Analogous to (41), the discrete form (56) is given by

$$\frac{\partial m_l}{\partial t} - J(m_l, \psi_l) + \nabla\bullet(m_l \nabla \chi_l) + \frac{(m\dot{\theta})_{l-1/2} - (m\dot{\theta})_{l+1/2}}{(\delta\theta)_l} = 0 \text{ for } l = 1, 2, \ldots, L, \tag{57}$$

where

$$m_l \equiv \frac{p_{l-1/2} - p_{l+1/2}}{(\delta\theta)_l} \text{ for } l = 1, 2, \ldots, L \tag{58}$$

is the discrete form of (55) and $(\delta\theta)_l = \sigma_{l-1/2} - \sigma_{l+1/2}$ for $l = 1, 2, \ldots, L$, The horizontal advection terms transform just as in the sigma coordinate. The *vertical velocity* $\dot{\theta}$ is given by

$$\dot{\theta} = \frac{Q}{\Pi}, \tag{59}$$

where Q is the heating rate per unit mass. The boundary conditions $(m\dot{\theta})_{1/2} = (m\dot{\theta})_{L+1/2} = 0$ apply at the top and bottom edges of the model.

We use a simple flux-corrected transport scheme to prevent the formation of negative mass. A FCT scheme to maintain a positive-definite tracer, τ, in the isentropic coordinate is discussed in Appendix L. If we set $\tau \equiv 1$ then we have a scheme to maintain positive-definite mass.

3.4.2 Momentum

With (136) the pressure-gradient force becomes $-\nabla_p \Phi = -\nabla_\theta M$ where $M \equiv c_p T + \Phi$ is the Montgomery potential. This is discussed more in Appendix G. Thus, the momentum equation becomes

$$\frac{D\mathbf{V}}{Dt} = \left(\frac{\partial \mathbf{V}}{\partial t}\right)_\theta + \mathbf{V} \bullet \nabla_\theta \mathbf{V} + \dot{\theta}\frac{\partial \mathbf{V}}{\partial \theta} = -\nabla_\theta M - f\hat{\mathbf{k}} \times \mathbf{V} + \mathbf{F}. \tag{60}$$

We can again substitute (43) for the horizontal advection term and take the curl and divergence giving

$$\frac{\partial \eta}{\partial t} = J(\eta, \psi) - \nabla \bullet (\eta \nabla \chi) + \hat{\mathbf{k}} \bullet \left[\nabla \times \left(\dot{\theta}\frac{\partial \mathbf{V}}{\partial \theta} + \mathbf{F}\right)\right], \tag{61}$$

and

$$\frac{\partial \delta}{\partial t} = J(\eta, \chi) + \nabla \bullet (\eta \nabla \psi) - \nabla^2 (M + K) + \nabla \bullet \left[\dot{\theta}\frac{\partial \mathbf{V}}{\partial \theta} + \mathbf{F}\right]. \tag{62}$$

These equations have the same form as the shallow-water vorticity and divergence equations. The Montgomery potential is obtained by integrating the hydrostatic equation in the form

$$\frac{\partial M}{\partial \theta} = \Pi, \tag{63}$$

where $\Pi \equiv c_p \left(\frac{p}{p_0}\right)^\kappa$ is the Exner function. The discrete form of the hydrostatic equation is also shown in Appendix G.

3.5 Generalized vertical coordinate

Konor and Arakawa (1996) have designed a generalized vertical coordinate which combines the advantages of the sigma coordinate and the isentropic coordinate. It is defined with a function of the form

$$\zeta \equiv F(\theta, p, p_S). \tag{64}$$

The function F is constructed so that it will quickly transition from $\zeta \equiv \sigma$ to $\zeta \equiv \theta$ with increasing height. In this way, the coordinate is terrain-following near the surface, and isentropic throughout the majority of the model atmosphere.

The sigma coordinate has a simple equation for the surface pressure from which we can diagnose pressure at a layer edge. However, it has a fully three-dimensional thermodynamic equation. On the other hand, the isentropic coordinate has a fully three-dimensional equation for mass and no thermodynamic equation. With the generalized vertical coordinate, we predict a three-dimensional p and a three-dimensional θ. They must be predicted in a consistent fashion. That is, on each coordinate surface $\zeta_{l+1/2}$, the prognostic variables $p_{l+1/2}$ and $\theta_{l+1/2}$ and the surface pressure, p_S, must evolve so that they satisfy $\zeta_{l+1/2} = F(\theta_{l+1/2}, p_{l+1/2}, p_S)$. Konor and Arakawa accomplish this by cleverly constructing the vertical mass flux $(m\dot{\zeta})_{l+1/2}$. This is discussed in Appendix I.

Difficulties can arise if pressure and temperature are predicted completely independently from each other. Recall the Zhu *et al.* (1992) model where pressure is given by (34). In the pure isentropic region of the model, the pressure on a coordinate surface is related to the temperature by $p_{l+1/2} = p_0(T_{l+1/2}/\theta_{l+1/2})^{1/\kappa}$ where $\theta_{l+1/2}$ is a constant. Since the temperatures at two adjacent layer edges, $T_{l-1/2}$ and $T_{l+1/2}$, are independently predicted, it is possible for $p_{l+1/2} < p_{l-1/2}$, that is, the pressure surfaces can cross. This will occur, for example, if there is a strong radiative cooling which causes $T_{l+1/2}$ to significantly decrease. This model is based on the L-grid, temperature is defined at layer centers, and the quantity $T_{l+1/2}$ is approximated with $T_{l+1/2} = (T_l + T_{l+1})/2$. If the $l+1$ layer is very thick, T_{l+1} may be physically distant from T_l and have a very different diabatic heating rate. Nevertheless, it may have an excessive influence on $T_{l+1/2}$. So, an anomalous cooling at level $l+1$ can cause a strong inversion to form at level l, and increase the danger of coordinate surfaces crossing. Zhu *et al.* mitigate this problem by smoothing the vertical profile of temperature through artificial enhancement of vertical diffusion in regions where a shape gradient of temperature is detected. There are other unintended side effects of interpolating temperature to layer edges as we shall see later.

In the Konor and Arakawa model, the CP-grid is used for the vertical staggering of θ. The model is designed so that the thermodynamic equation is trivially satisfied when $\zeta = \theta$. That is, in isentropic coordinates, the thermodynamic equation reduces to a simple diagnostic equation for $\dot{\theta}$. Since $\dot{\zeta} = \dot{\theta}$ is the vertical velocity in the pure isentropic domain, then $\dot{\theta}$ is required to be defined a layer edges. Therefore, θ is also defined at layer edges.

3.5.1 *The coordinate definition*

Let $\zeta \equiv F(\theta, p, p_S)$ and $\sigma \equiv \sigma(p, p_S)$ be monotonically increasing with height. We require that

$$\frac{\partial \zeta}{\partial \sigma} > 0 \quad \text{and} \quad \zeta = \begin{cases} \text{const} & \sigma = \sigma_S \\ \theta & \sigma = \sigma_T \end{cases}, \tag{65}$$

where σ_S is the bottom model edge and σ_T is the top model edge. Suppose ζ has the form

$$\zeta \equiv f(\sigma) + g(\sigma)\theta, \tag{66}$$

where

$$\left.\begin{array}{ll} g(\sigma) \to 0; & \sigma \to \sigma_S \\ f(\sigma) \to 0, g(\sigma) \to 1; & \sigma \to \sigma_T \end{array}\right\}. \tag{67}$$

With (65) and (66) we have

$$\frac{\partial f}{\partial \sigma} + \frac{\partial g}{\partial \sigma}\theta + g\frac{\partial \theta}{\partial \sigma} > 0, \tag{68}$$

Suppose we choose an appropriate lower bound for potential temperature, $\theta_{\min}$, and a lower bound for static stability, $(\partial\theta/\partial\sigma)_{\min}$. Note that $(\partial\theta/\partial\sigma)_{\min} < 0$ is fine. Substitution into (68) yields

$$\frac{\partial f}{\partial \sigma} + \frac{\partial g}{\partial \sigma}\theta_{\min} + g\left(\frac{\partial \theta}{\partial \sigma}\right)_{\min} = 0 \tag{69}$$

If we choose $g(\sigma)$, then (69) can be solved for $f(\sigma)$.

For example, suppose $\sigma \equiv (p_S - p)/(p_S - p_T)$. So $\sigma_S = 0$ and $\sigma_T = 1$. And suppose we let $g(\sigma) = (1 - e^{-\alpha\sigma})/(1 - e^{-\alpha})$ which satisfies (67). Typically $\alpha \sim 10$. The larger α, the faster the transition to pure isentropic coordinates. We can integrate (69) from σ to σ_T, that is,

$$\int_\sigma^{\sigma_T}\frac{\partial f}{\partial\sigma}d\sigma+\theta_{\min}\int_\sigma^{\sigma_T}\frac{\partial g}{\partial\sigma}d\sigma+\left(\frac{\partial\theta}{\partial\sigma}\right)_{\min}\int_\sigma^{\sigma_T}g(\sigma)d\sigma = 0$$

$$\int_\sigma^{\sigma_T}\frac{\partial f}{\partial\sigma}d\sigma+\theta_{\min}\int_\sigma^{\sigma_T}\frac{\partial g}{\partial\sigma}d\sigma+\left(\frac{\partial\theta}{\partial\sigma}\right)_{\min}\int_\sigma^{\sigma_T}g(\sigma)d\sigma = 0$$

$f(\sigma_T$

$$f(\sigma_T)-f(\sigma)+\theta_{\min}[g(\sigma_T)-g(\sigma)]+\left(\frac{\partial\theta}{\partial\sigma}\right)_{\min}\int_\sigma^{\sigma_T}g(\sigma)d\sigma = 0. \tag{70}$$

With the bound[illegible]

$$\begin{aligned} f(\sigma) &= \theta_{\min}[1-g(\sigma)]+\left(\frac{\partial\theta}{\partial\sigma}\right)_{\min}\int_\sigma^{\sigma_T}\frac{1-e^{-\alpha\sigma}}{1-e^{-\alpha}}d\sigma \\ &= \theta_{\min}[1-g(\sigma)]+\left(\frac{\partial\theta}{\partial\sigma}\right)_{\min}\frac{1}{1-e^{-\alpha}}\left[\int_\sigma^{\sigma_T}d\sigma-\int_\sigma^{\sigma_T}e^{-\alpha\sigma}d\sigma\right] \\ &= \theta_{\min}[1-g(\sigma)]+\left(\frac{\partial\theta}{\partial\sigma}\right)_{\min}\frac{1}{1-e^{-\alpha}}\left[1-\sigma-\frac{e^{-\alpha\sigma}-e^{-\alpha}}{\alpha}\right]. \end{aligned} \tag{71}$$

If we set $(\partial\theta/\partial\sigma)_{\min} = 0$, then (71) simplifies considerably. In this case we see

$$F(\theta,\sigma) = f(\sigma)+g(\sigma)\theta = \theta_{\min}[1-g(\sigma)]+g(\sigma)\theta = \theta_{\min}+g(\sigma)(\theta-\theta_{\min}), \tag{72}$$

and

$$\frac{\partial F}{\partial\theta}=g(\sigma) \text{ and } \frac{\partial F}{\partial p} = (\theta-\theta_{\min})\frac{dg}{d\sigma}\frac{\partial\sigma}{\partial p} = (\theta-\theta_{\min})\frac{\alpha e^{-\alpha\sigma}}{1-e^{-\alpha}}\frac{-1}{p_S-p_T} \tag{73}$$

3.5.2 *The observed distribution of ζ*

The observed distribution of generalize vertical coordinate surfaces is shown in Figure 4. The plot looks very similar to Figure 1 in the upper part of the domain where the generalized vertical coordinate surfaces are more-or-less pure isentropic surfaces. Near the Earth's surface, however, they become terrain-following.

3.5.3 *Continuity*

In continuous form, mass is defined

$$m\equiv\frac{\partial p}{\partial\zeta}. \tag{74}$$

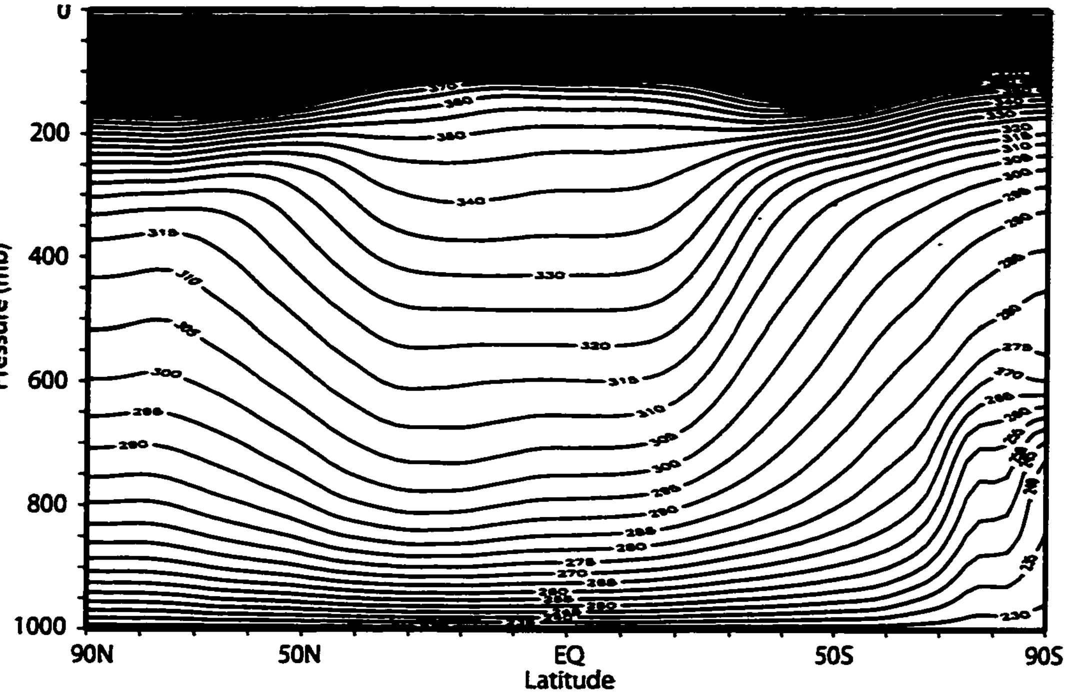


Figure 4. The observed, zonally-averaged distribution of generalized vertical coordinate surfaces for July 1999. Figure provided by Dr. Konor.

The continuity equation is given by

$$\left(\frac{\partial m}{\partial t}\right)_{\zeta} + \nabla_{\zeta} \bullet (m\mathbf{V}) + \frac{\partial}{\partial \zeta}(m\dot{\zeta}) = 0. \tag{75}$$

In the finite-difference realm, the mass is defined as

$$m_l \equiv -\frac{p_{l-1/2} - p_{l+1/2}}{(\delta\zeta)_l} \text{ where } (\delta\zeta)_l \equiv \zeta_{l-1/2} - \zeta_{l+1/2} \text{ for } l = 1, 2, \ldots, L. \tag{76}$$

With $\mathbf{V}_l = \hat{\mathbf{k}} \times \nabla\psi_l + \nabla\chi_l$, the continuity equation becomes

$$\frac{\partial m_l}{\partial t} - J(m_l, \psi_l) + \nabla \bullet (m_l \nabla \chi_l) + \frac{(m\dot{\zeta})_{l-1/2} - (m\dot{\zeta})_{l+1/2}}{(\delta\zeta)_l} = 0 \text{ for } l = 1, 2, \ldots, L \tag{77}$$

with boundary conditions $(m\dot{\zeta})_{1/2} = (m\dot{\zeta})_{L+1/2} = 0$ on the vertical mass flux. The very neat trick of consistency is done by specifying $(m\dot{\zeta})_{l+1/2}$. This is discussed in Appendix I.

3.5.4 *Momentum*

The horizontal advection of momentum is the same as in the previous two models. The pressure-gradient force can be written as

$$-\nabla_p \Phi = -\nabla_\zeta M - \Pi \nabla_\zeta \theta . \tag{78}$$

This form has the important property that when $\zeta \equiv \theta$, then $-\nabla_p \Phi = -\nabla_\theta M$. In other words, when the coordinate becomes pure isentropic, then the pressure-gradient force becomes the pressure-gradient force in the isentropic model. This term is discussed further in Appendix H. The momentum equations can be written

$$\frac{D\mathbf{V}}{Dt} = \left(\frac{\partial \mathbf{V}}{\partial t}\right)_\zeta + \mathbf{V} \bullet \nabla_\zeta \mathbf{V} + \dot{\zeta}\frac{\partial \mathbf{V}}{\partial \zeta} = -\nabla_\zeta M - \Pi \nabla_\zeta \theta - f\hat{\mathbf{k}} \times \mathbf{V} + \mathbf{F} . \tag{79}$$

In discrete form, the Montgomery potential in (79) is calculated with $M_l = \Phi_l + \Pi_l \theta_l$ where the geopotential Φ_l is obtain by integrating the hydrostatic equation in the form $\partial\Phi/\partial\Pi = -\theta$. The hydrostatic equation is discussed further in Appendix H. The vertical advection of momentum term and vertical diffusion of momentum term (not shown in (79)) are discussed in Appendix J.

3.5.5 *Thermodynamic equation*

At the internal edges of the model thermodynamic equation is given by

$$(m\Pi\delta\zeta)_{l+1/2}\frac{\partial \theta_{l+1/2}}{\partial t} + (m\Pi\mathbf{V}\delta\zeta)_{l+1/2} \bullet \nabla\theta_{l+1/2} \tag{80}$$

$$+\left(\Pi\frac{\partial\theta}{\partial\zeta}\delta\zeta\right)_{l+1/2}(m\dot{\zeta})_{l+1/2} = (mQ\delta\zeta)_{l+1/2} \quad \text{for } l = 1, 2, \ldots, L-1,$$

which is KA (3.20). This can be written as

$$\frac{\partial\theta_{l+1/2}}{\partial t} = \frac{1}{(m\Pi\delta\zeta)_{l+1/2}}\Big[(mQ\delta\zeta)_{l+1/2} - (m\Pi\mathbf{V}\delta\zeta)_{l+1/2}\bullet\nabla\theta_{l+1/2} \tag{81}$$

$$-\left(\Pi\frac{\partial\theta}{\partial\zeta}\delta\zeta\right)_{l+1/2}(m\dot{\zeta})_{l+1/2}\Big] \text{ for } l = 1, 2, \ldots, L-1,$$

where

$$(m\Pi\delta\zeta)_{l+1/2} \equiv \frac{1}{2}[m_{l+1}\Pi_{l+1}(\delta\zeta)_{l+1} + m_l\Pi_l(\delta\zeta)_l] \text{ for } l = 1, 2, \ldots, L-1, \tag{82}$$

which is KA (3.27) and

$$(m\Pi\mathbf{V}\delta\zeta)_{l+1/2} \equiv \frac{1}{2}[m_{l+1}\Pi_{l+1}\mathbf{V}_{l+1}(\delta\zeta)_{l+1} + m_l\Pi_l\mathbf{V}_l(\delta\zeta)_l] \tag{83}$$

$$\text{for } l = 1, 2, \ldots, L-1,$$

which is KA (3.30), and

$$\left(\Pi\frac{\partial\theta}{\partial\zeta}\delta\zeta\right)_{l+1/2} \equiv \Pi_{l+1/2}(\theta_l - \theta_{l+1}) \text{ for } l = 1, 2, \ldots, L-1 \tag{84}$$

which is KA (3.35) (within a sign). Finally, the heating terms are given by KA (3.36), namely,

$$(mQ\delta\zeta)_{l+1/2} \equiv \frac{1}{2}[(mQ)_{l+1}(\delta\zeta)_{l+1} + (mQ)_l(\delta\zeta)_l] \text{ for } l = 1, 2, \ldots, L-1. \tag{85}$$

KA define θ_l as $\theta_l \equiv (\theta_{l-1/2} + \theta_{l+1/2})/2$. Equation (84) can be written as follows

$$\left(\Pi\frac{\partial\theta}{\partial\zeta}\delta\zeta\right)_{l+1/2} = \Pi_{l+1/2}(\theta_l - \theta_{l+1}) \tag{86}$$

$$= \Pi_{l+1/2}\left(\frac{\theta_{l-1/2}+\theta_{l+1/2}}{2} - \frac{\theta_{l+1/2}+\theta_{l+3/2}}{2}\right) = \frac{1}{2}\Pi_{l+1/2}(\theta_{l-1/2} - \theta_{l+3/2}).$$

The sign difference between this equation and the corresponding equation in the KA paper (eq. (3.35)) is due to the definition of $(\delta\zeta)_l$. We want this number to be positive, that is, $(\delta\zeta) \equiv \zeta_{l-1/2} - \zeta_{l+1/2}$. In (82) and (83), Π_l is defined by

$$\Pi_l \equiv \frac{1}{1+\kappa} \frac{\Pi_{l+1/2} p_{l+1/2} - \Pi_{l-1/2} p_{l-1/2}}{p_{l+1/2} - p_{l-1/2}} \text{ for } l = 1, 2, \ldots, L, \tag{87}$$

where

$$\Pi_{l+1/2} = c_p \left(\frac{p_{l+1/2}}{p_0} \right)^{\kappa} \text{ for } l = 0, 1, 2, \ldots, L. \tag{88}$$

The advective term of the form $\mathbf{V} \bullet \nabla\theta$ in (80) is discussed in Appendix P. At the top edge of the model the thermodynamic energy equation is given by

$$(m\Pi\delta\zeta)_{1/2} \frac{\partial\theta_{1/2}}{\partial t} + (m\Pi\mathbf{V}\delta\zeta)_{1/2} \bullet \nabla\theta_{1/2} = (mQ\delta\zeta)_{1/2} \tag{89}$$

which is KA (3.21). This can be written as

$$\frac{\partial\theta_{1/2}}{\partial t} = \frac{1}{(m\Pi\delta\zeta)_{1/2}} [(mQ\delta\zeta)_{1/2} - (m\Pi\mathbf{V}\delta\zeta)_{1/2} \bullet \nabla\theta_{1/2}], \tag{90}$$

where

$$\begin{gathered} (mQ\delta\zeta)_{1/2} \equiv \frac{1}{2}(mQ)_1 (\delta\zeta)_1 \\ (m\Pi\delta\zeta)_{1/2} \equiv \frac{1}{2} m_1 \Pi_1 (\delta\zeta)_1 \text{ and } (m\Pi\mathbf{V}\delta\zeta)_{1/2} \equiv \frac{1}{2} m_1 \Pi_1 \mathbf{V}_1 (\delta\zeta)_1 \end{gathered} \tag{91}$$

which are KA (3.37), (3.28) and (3.31). Substituting (91) into (90) gives

$$\frac{\partial\theta_{1/2}}{\partial t} = \frac{(mQ)_1}{m_1\Pi_1} - \nabla\theta_{1/2} \bullet \mathbf{V}_1 = \frac{(mQ)_1}{m_1\Pi_1} - [\nabla\theta_{1/2} \bullet \nabla\chi_1 - J(\theta_{1/2}, \psi_1)]. \tag{92}$$

In the current version of the model, the top coordinate surface is a surface of constant potential temperature. Therefore, $\partial\theta_{1/2}/\partial t = 0$.

At the bottom edge of the model the thermodynamic energy equation is given by

$$(m\Pi\delta\zeta)_{L+1/2} \frac{\partial\theta_{L+1/2}}{\partial t} + (m\Pi\mathbf{V}\delta\zeta)_{L+1/2} \bullet \nabla\theta_{L+1/2} = (mQ\delta\zeta)_{L+1/2} \tag{93}$$

which is KA (3.22). This can be written as

$$\frac{\partial\theta_{L+1/2}}{\partial t} = \frac{1}{(m\Pi\delta\zeta)_{L+1/2}}[(mQ\delta\zeta)_{L+1/2} - (m\Pi\mathbf{V}\delta\zeta)_{L+1/2} \bullet \nabla\theta_{L+1/2}], \tag{94}$$

where

$$(mQ\delta\zeta)_{L+1/2} \equiv \frac{1}{2}(mQ)_L(\delta\zeta)_L$$
$$(m\Pi\delta\zeta)_{L+1/2} \equiv \frac{1}{2}m_L\Pi_L(\delta\zeta)_L \text{ and } (m\Pi\mathbf{V}\delta\zeta)_{L+1/2} \equiv \frac{1}{2}m_L\Pi_L\mathbf{V}_L(\delta\zeta)_L \tag{95}$$

which are KA (3.38), (3.29) and (3.32). Substituting (95) into (94) gives

$$\frac{\partial\theta_{L+1/2}}{\partial t} = \frac{(mQ)_L}{m_L\Pi_L} - \nabla\theta_{L+1/2} \bullet \mathbf{V}_L \tag{96}$$
$$= \frac{(mQ)_L}{m_L\Pi_L} - [\nabla\theta_{L+1/2} \bullet \nabla\chi_L - J(\theta_{L+1/2}, \psi_L)].$$

Note that there is no vertical advection term in (96). In a simulation in which there is no heating, then the bottom edge potential temperature is, in a sense, disconnected from the rest of model.

We found that the horizontal advection term, if left in advective form, would produce unrealistically cold temperatures. This problem was corrected by re-writing (96) in flux form. Then, by using the monotone advection scheme in Appendix L, we can prevent the formation of local extrema. We use a simple dry convective adjustment procedure to correct unstable layers in the lower layers of the model.

4. Geodesic grid

Before proceeding with the numerical results of the models, let us briefly review the horizontal structure. We use a geodesic grid similar to that of Williamson (1968) and Sadourny *et al.* (1968). This grid eliminates the pole problem and provides approximately homogeneous and isotropic resolution over the sphere. A simple geodesic grid can be constructed starting with an icosahedron inscribed within a unit sphere, as shown in Figure 5a. Each face of the icosahedron is subdivided into four new faces by bisecting the edges

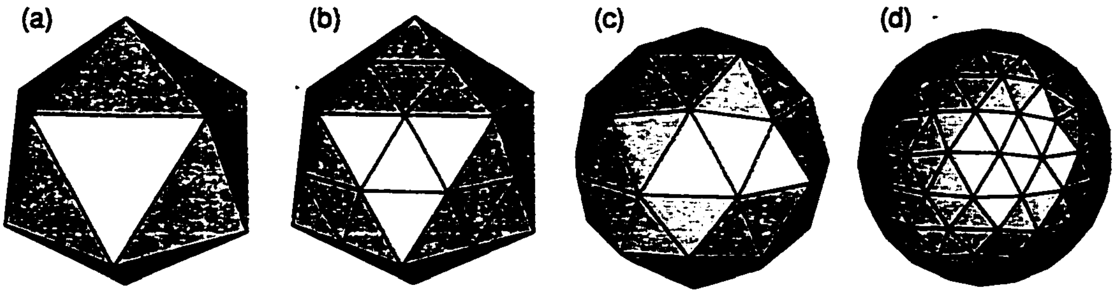
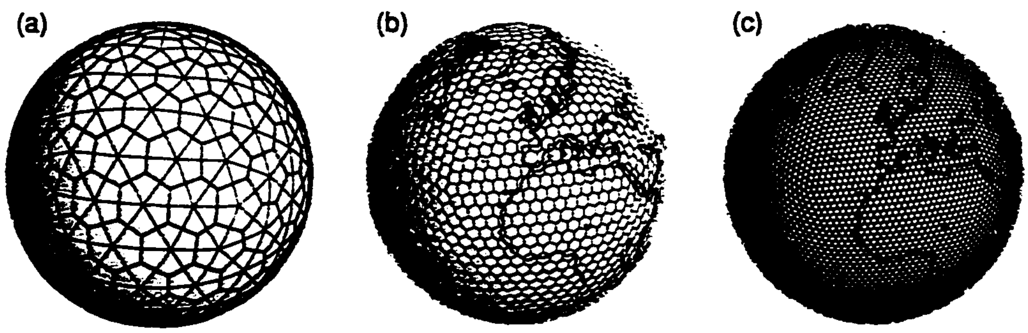


Figure 5. (a) An icosahedron. Imagine that it is inscribed within a unit sphere. (b) Bisect each forming four new faces from each old face. (c) Project the new vertices onto the unit sphere. (d) Repeating the algorithm gives a polyhedron with 162 vertices.

and projecting the new vertices onto the unit sphere. The result of this process is shown in Figure 5b. By recursively subdividing each triangular face and projecting the new vertices onto the unit sphere, we can generate polyhedra that progressively approximate a sphere. The original icosahedron has 12 vertices. The polyhedron shown in Figure 5c has 42 vertices, and in Figure 5d has 162 vertices. In general, the number of vertices goes as $2 + 5 \times 2^{2n+1}$ where n is number of applications of the subdivision algorithm. A shallow-water model based on this grid is discussed in Heikes and Randall (1995a), and refinements of the grid structure are discussed in Heikes and Randall (1995b). Recently, a dramatically improved finite-difference scheme for the geodesic grid has been introduced by Ringler and Randall (2002a) and (2002b).

We define the vertices of these polyhedra to be the model grid points, and associate a Voronoi cell with each grid point. The Voronoi cell associated with a particular grid point, say grid point p_0, consists of all the points on the sphere closer to p_0 than to any other grid point. These cells are the control volumes for our finite-volume operators. Such

Figure 6. Voronoi grids. (a) 162 cells. (b) 2562 cells. (c) 10242 cells.



a grid with 162 cells is shown in Figure 6a. This is the Voronoi grid generated from Figure 5d. In the numerical simulations we will present, the model is run with 2562 cells, 10242 cells and 40962 cells. These grids have resolutions of about 4.5°, 2.25° and 1.125°, respectively. Other properties of the grid are listed below

Table 1:

number of cells	Number of cells along a great circle	Average cell area (km^2)	Ratio of smallest cell area to largest cell area	Average distance between cell centers (km)	Ratio of smallest distance to largest distance
2562	80	1.991×10^5	0.948	481.6	0.791

Table 1:

number of cells	Number of cells along a great circle	Average cell area (km^2)	Ratio of smallest cell area to largest cell area	Average distance between cell centers (km)	Ratio of smallest distance to largest distance
10242	160	4.980×10^4	0.951	240.9	0.788
40962	320	1.245×10^4	0.952	120.5	0.787

The horizontal differencing follows Masuda (1986). The model predicts momentum in the form of vorticity, η, and divergence, δ, without dependence on horizontal staggering (Randall, 1994). Vorticity and divergence are scalar quantities, hence this formulation avoids the difficulties associated with vector valued functions near the poles. We need to diagnose a streamfunction, ψ, and velocity potential, χ, each time step from the elliptic equation $\nabla^2\psi = \eta - f$ and $\nabla^2\chi = \delta$. This is done with multigrid on the spherical geodesic grid. See Appendix Q.

5. Numerical results

5.1 Held-Suarez forcing

The Held-Suarez (1994; HS) test is a benchmark calculation for intercomparison of the statistically steady states of dry dynamical cores. Several of the numerical results in this section are run using the forcing and boundary condition prescribed in the benchmark. We begin this section by showing the forcing and boundary condition, and we will discuss the rest of the benchmark in a later section.

The forcing is a very simple parameterization of radiative processes represented with Newtonian relaxation of temperature towards a zonally symmetric state. A heating term is added to the thermodynamic equation as follows

$$Q = \frac{\partial T}{\partial t} = \ldots - k_T(\varphi, \sigma)[T - T_{eq}(\varphi, p)]. \tag{97}$$

The zonally symmetric equilibrium temperature profile is given by

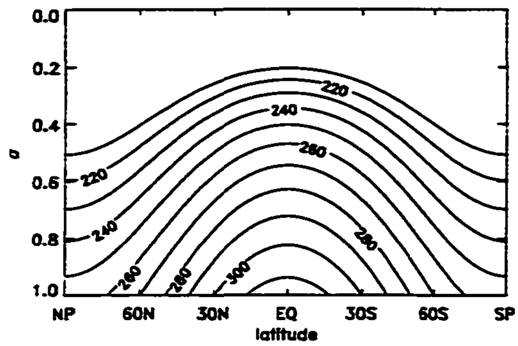
$$T_{eq}(\varphi, p) = \max\left\{200K, \left[315K - (\Delta T)_y \sin^2\varphi - (\Delta\theta)_z \log\left(\frac{p}{p_0}\right)\cos^2\varphi\right]\left(\frac{p}{p_0}\right)^{\kappa}\right\}, \tag{98}$$

where $(\Delta T)_y = 60K$ and $(\Delta\theta)_z = 10K$. The radiative relaxation coefficient is given by

$$k_T(\varphi, p) = k_a + (k_s - k_a)\max\left(0, \frac{\sigma - \sigma_b}{1 - \sigma_b}\right)\cos^4\varphi, \tag{99}$$

where $k_a = 1/40\text{day}^{-1}$, $k_s = 1/4\text{day}^{-1}$ and $\sigma_b = 0.7$. From (99) it is clear that the relaxation is much stronger near the Earth's surface and near the equator. A plot of T_{eq} is shown in Figure 7.

Figure 7. The zonally symmetric equilibrium temperature profile T_{eq} (K) used in the Newtonian relaxation of the Held-Suarez benchmark. Contour interval 10 K.



The model also uses a simple parameterization of surface friction. Dissipation is represented with a Rayleigh damping of low-level winds. A term is added to the momentum equation as follows

$$\frac{\partial \mathbf{V}}{\partial t} = \ldots - k_{\mathbf{V}}(p)\mathbf{V}, \tag{100}$$

where

$$k_{\mathbf{V}}(p) = k_f \max\left(0, \frac{\sigma - \sigma_b}{1 - \sigma_b}\right), \tag{101}$$

where $k_f = 1\,\text{day}^{-1}$. So, the friction varies linearly in intensity through the bottom 30% on the atmosphere.

5.2 A comparison of the Lorenz grid and Charney-Phillips grid

5.2.1 Description of the computational mode

One reason we like the generalized vertical coordinate is because it is based on the Charney-Phillips vertical staggering of prognostic variables. As mentioned earlier This staggering eliminates a computation mode that is present in the Lorenz grid. To demonstrate this we have run two sigma coordinate models -- one based on the L-grid and one based on the CP-grid -- and looked for evidence that the Lorenz staggering favors more $2\Delta z$ noise in the vertical than the Charney-Phillips staggering.

As mentioned earlier, Arakawa and Moorthi (1987) discuss several shortcomings of the L-grid. Suppose we consider a single model column with L layers. Then there are $L-1$ internal edges between the L layer centers. Since potential temperature is defined at layer centers, there is one more discrete potential temperature variable in a model column than there are interior layer edges. This mismatch causes problems. Consider the form of the hydrostatic equation $\partial\Phi/\partial\Pi = -\theta$ where $\Pi \equiv c_p(p/p_0)^{\kappa}$ is the Exner function. As we can see in (48), the momentum equation requires a horizontal gradient of Φ. Since $\mathbf{V}$ is defined at layer centers, Φ must also be defined at layer centers. A trapezoidal integration of the hydrostatic equation requires an approximation for θ at the layer edges. Lorenz uses a simple arithmetic average to interpolate potential temperature to the layer edges,

$$\theta_{l+1/2} \equiv (\theta_l + \theta_{l+1})/2. \tag{102}$$

Pressure is also averaged from the edges to the centers with $p_l \equiv (p_{l-1/2} + p_{l+1/2})/2$. And, the Exner function at a layer center is simply $\Pi_l = c_p(p_l/p_0)^{\kappa}$. Hence, the discrete form of the above hydrostatic equation is given by

$$\Phi_l - \Phi_{l+1} = -\theta_{l+1/2}(\Pi_l - \Pi_{l+1}) \quad \text{for } l = 1, 2, \ldots, L-1. \tag{103}$$

The horizontal gradient of the geopotential Φ_l, defined at layer centers, is now ready to be used in the pressure-gradient force. It is clear from (102) that if we superimpose a zig-zag pattern of the form $(-1)^l\Delta\theta$ on an existing vertical profile of θ at layer centers, the averaged $\theta_{l+1/2}$ will not change. Thus, this $2\Delta z$ signal in the θ profile will have no effect on the dynamics of the model. In other words, it is not visible to the momentum equations. This problem is illustrated in Figure 8. If the dynamics cannot react to the

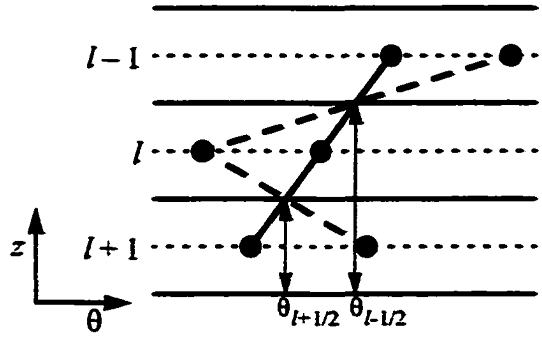


Figure 8: In the Lorenz staggering, potential temperature is defined at layer centers, denoted with gray dots. It is averaged to layer edges with $\theta_{l+1/2} \equiv (\theta_l + \theta_{l+1})/2$. Two very different vertical profiles of potential temperature, one connected with solid lines and the other with dashed lines, can give the same layer edge averages.

temperature, then the thermal wind relation cannot be satisfied.

Moorthi and Arakawa continue to describe other problems associated with the *L*-grid. These include a reduction in the effective static stability in the top and bottom layer. More importantly, they show the L-grid produces a spurious amplification of short wavelength waves due to false satisfaction of the criteria for baroclinic instability. This analysis is in the context of the linearized quasi-geostrophic equations. They also show the same amplification occurs in the non-linear primitive equations. In the quasi-geostrophic equations, the amplifying mode is trapped near the surface. In the primitive equation model the amplifying short waves can propagate vertically, and can influence longer wavelengths through non-linear interactions. These difficulties associated with the L-grid have helped motivate a revival of the CP-grid.

5.2.2 A numerical experiment to quantify the noise

We have designed a simple numerical experiment to detect noise that may result from the computational model. There are two sigma models involved in the experiment. The first is a dry dynamical core that has been teased out of the current CSU GCM. To do this radiative and surface parameterizations are replaced with the Held-Suarez *toy physics* forcing and boundary conditions described above. The vertical coordinate in this model was presented by Arakawa and Suarez (1983) and is based on a L-grid distribution of the prognostic variables. The second model in the experiment was described in section 3.2 and is based on a CP-grid. In order to compare apple with apples, the two models were made as similar as possible. The timesteps and horizontal grids are the same, as are the number and distribution of the vertical coordinate surfaces. The strength of horizontal diffusion of momentum is the same, and we have tried to make the dry convective adjustment as gentle as possible.

In the experiment, each model is run until it is spun-up or reaches a statistically equilibrated state. At this point, each has *forgotten* the initial conditions. This typically takes about 50 days of simulated time. Then the potential temperature field is recorded every four days. Four days is sufficiently long time that the samples are appreciatively different from one another. Once we have collected about 100 samples from each model, we compute a time-average potential temperature profile at each cell. This profile is then subtracted from each sample, and the result is decomposed in the vertical into Fourier modes. Since we are interested in the $2\Delta z$ noise, we focus on the shortest wavelength resolved in the vertical. We average the amplitude of the shortest mode over all the samples at each gridpoint. This is shown in Figure 9. The noise on the L-grid appears to be

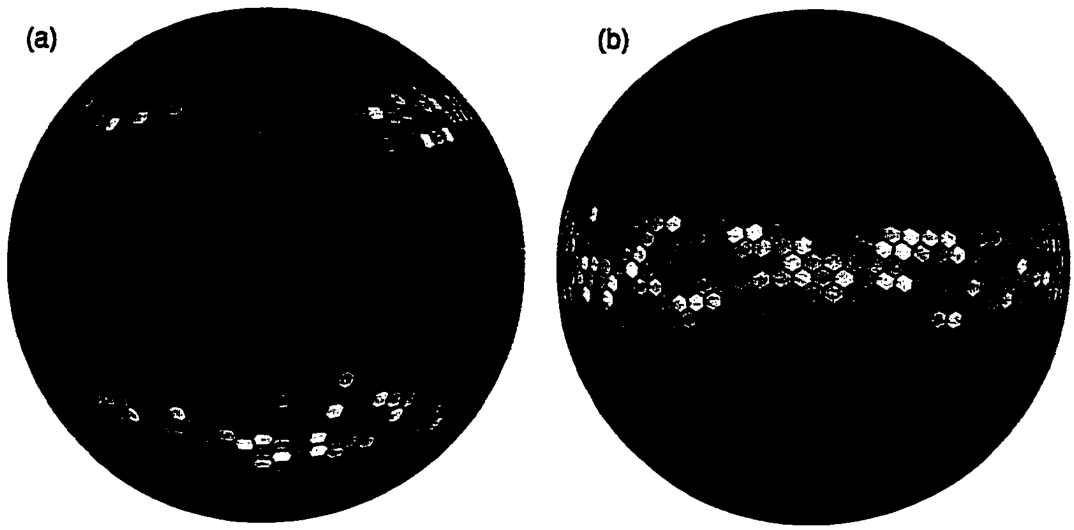


Figure 9. Time averaged amplitude of $2\Delta z$ mode. (a) L-grid. maximun=1.76, minimum=0.67 (b) CP-grid. maximum=0.99, minimum=0.18 For each plot, the color scale is linear from the maximum value to the minimum value.

largest at middle-latitudes, which the noise on the CP-grid is largest along the equator. Figure 10 shows a contour plot of the data in Figure 9. We can also sort the amplitudes of

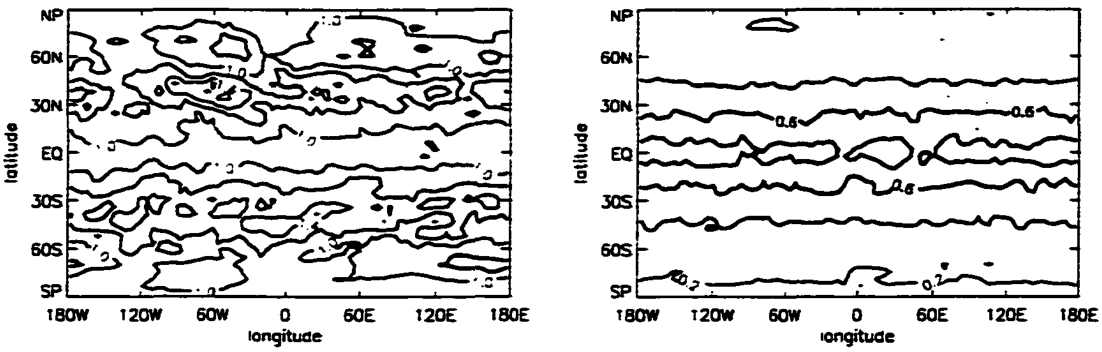
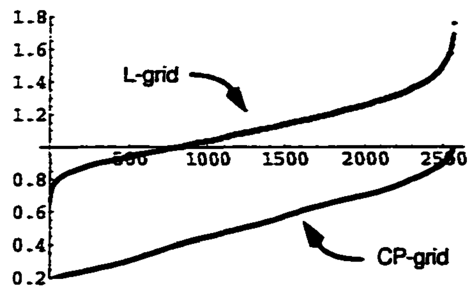


Figure 10. Time averaged amplitude of $2\Delta z$ mode. (a) L-grid, (b) CP-grid. Contour interval 0.2 K Values greater than or equal to 1.0 K are shown in red.

the $2\Delta z$ mode depicted in Figure 9 into ascending order, and display them. This is shown in Figure 11. This grid resolution has 2562 cells, so there are 2562 points in each curve.

Figure 11. Amplitude (K) of the $2\Delta z$ mode for L-grid (upper curve) and CP-grid (lower curve).



This very crude experiments seems to indicate that CP-grid is less noisy than the L-grid.

5.3 Sensitivity to horizontal and vertical resolution

Next we use the HS benchmark to investigate statistically steady behaviour of the three models. In particular, we will investigate the correlations between meridional wind and temperature to see how these correlations are affected by changes in horizontal and vertical resolutions. If the finite-difference operators are consistent with the continuous operators, then, as the spatial and temporal resolutions are refined, the discrete solution should converge the continuous solution. So, we expect the quality of the numerical results to improve with increasing resolution. Our focus will be on vertical resolution. Towards this end, each of the three models has been run at three vertical resolutions -- 16 layers, 32 layers and 64 layers. Current climate models typically run with between 16 and 32 layers. It is reasonable to believe that in the future they will run at much higher vertical resolutions. To further complicate matters, the models are run at two horizontal resolutions -- 2562 cells and 10242 cells. We are especially interested in the question of how much vertical resolution is necessary with the isentropic-coordinate and the generalize vertical coordinate relative to the sigma coordinate to accurately resolve the flow.

5.3.1 *Analysis of the mean meridional circulation in the models and in the real world*

As mentioned above, we are interested in the statistical correlations between the meridional winds and temperature. To compute these quantities in the model, we first linearly interpolate p, T, ψ and χ from the native geodesic grid to a latitude/longitude grid. The resolution of the target interpolation grid matches that of the source geodesic grid. That is, the 2562 cell geodesic grid is interpolated onto a a 80×40 lon/lat grid, and the 10242 cell grid is interpolated onto a 160×80 lon/lat grid. We compute the zonal and meridional winds using

$$u = \frac{1}{a\cos\varphi}\frac{\partial\chi}{\partial\lambda} - \frac{\partial\psi}{a\partial\varphi}, \quad \text{and} \quad v = \frac{1}{a\cos\varphi}\frac{\partial\psi}{\partial\lambda} + \frac{\partial\chi}{a\partial\varphi}. \tag{104}$$

At this point p, T, u and v are still in their original vertical coordinate systems. This is necessary for the derivatives (104) to yield the correct result. The second step is to linearly interpolate T and v onto surfaces of constant pressure. The set of pressure surfaces ranges from 50 hPa to 1000 hPa in steps of 25 hPa. There are 39 surfaces.

In order to understand the mean meridional circulation, we will have to average flows both zonally and in time. Consider an arbitrary field of the form $\alpha \equiv \alpha(t, \lambda, \varphi, z)$ which is a function of time, longitude, latitude and height. We can define a zonal average operator, $[(\ )]$, which will remove the longitudinal dependence, that is, $[\alpha] \equiv [\alpha](t, \varphi, z)$. The departure from the zonal-mean, or perturbation of α, is defined by $\alpha^* \equiv \alpha - [\alpha]$. The zonal average operator distributes across addition, so $[\alpha^*] = [\alpha - [\alpha]] = [\alpha] - [[\alpha]] = 0$.

The meridional velocity, v, and temperature, T, can be decomposed into mean and perturbation quantities. For example, the zonal-mean plus departure from zonal-mean of temperature is denoted by $T = [T] + T^*$, and the analogous expression for time mean and departure is denoted $T = \bar{T} + T'$. Since $\overline{\alpha'} = 0$, the time mean of a product can be written

$$\begin{aligned} \overline{vT} &= \overline{(\bar{v} + v')(\bar{T} + T')} \\ &= \overline{\bar{v}\bar{T}} + \overline{\bar{v}T'} + \overline{v'\bar{T}} \div \overline{v'T'} \\ &= \bar{v}\bar{T} + \overline{v'T'}. \end{aligned} \tag{105}$$

This is the temporal covariance of v and T. The term $\bar{v}\bar{T}$ in (105) can be expanded in terms of zonal-mean and perturbation and the zonal-mean applied to the result,

$$\begin{aligned} [\bar{v}\bar{T}] &= [\overline{[v] + v^*}\,\overline{[T] + T^*}] \\ &= [(\overline{[v]} + \overline{v^*})(\overline{[T]} + \overline{T^*})] \\ &= [\overline{[v]}\,\overline{[T]}] + [\overline{[v]}\,\overline{T^*}] + [\overline{v^*}\,\overline{[T]}] + [\overline{v^*}\,\overline{T^*}] \\ &= \overline{[v]}\,\overline{[T]} + [\overline{v^*}\,\overline{T^*}]. \end{aligned} \tag{106}$$

An analogous expression to (105) for the zonal average of a product is given by $[vT] = [v][T] + [v^*T^*]$. This true for any two quantities. In particular, it is true for v' and T'. Thus,

$$\overline{[v'T']} = \overline{[v']}\,\overline{[T']} + \overline{[v'^*T'^*]}. \tag{107}$$

Substituting (106) and (107) into the zonal average of (105) gives

$$\overline{[vT]} = \overline{[v]}\,\overline{[T]} + [\overline{v^*}\,\overline{T^*}] + \overline{[v'][T']} + \overline{[v'^*T'^*]}\,. \tag{108}$$

With (108) we can make observations about the mean meridional circulation in both the numerical models and observations. The zonal and time averaged, observed mean meridional circulation in isobaric coordinates consists of a strong direct cell, called the Hadley cell, in low latitudes, and a weaker indirect cell, called the Ferrel cell, in middle latitudes. These cells represent circulations in the latitude-height plane. A direct cell has its rising branch in warm air and its sinking branch in cooler air. An indirect cell is the opposite. Figure 12 shows the observed shape and intensity of the mass stream function associated with the mean meridional circulation for winter and summer. The mass stream function for the zonally averaged transport in isobaric coordinates is defined by

$$\frac{\partial\Psi_p}{\partial p} = \frac{2\pi a\cos\varphi}{g}\overline{[v]}, \quad \text{and} \quad \frac{\partial\Psi_p}{\partial\varphi} = -\frac{2\pi a^2\cos\varphi}{g}\overline{[\omega]}\,. \tag{109}$$

Qualitatively, the density of the streamlines is proportional to the speed of the winds, and the direction is indicated in the figures. The body of the Hadley cell lies in the winter hemisphere, and is stronger in the southern winter than in the northern winter. The Ferrel cell lies poleward of the Hadley cell in the winter hemisphere.

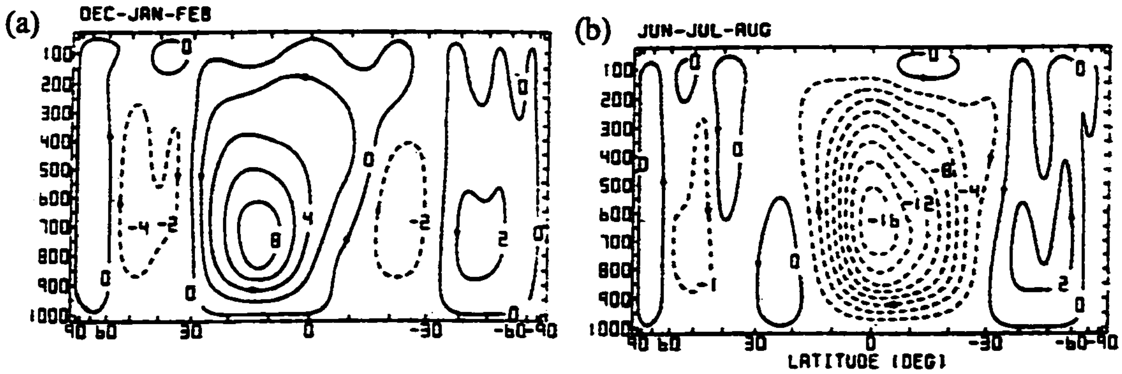


Figure 12. Mass stream function (10^{10} kg s^{-1}) for the isobaric mean meridional circulations for northern hemisphere (a) winter and (b) summer. Arrows indicate direction of circulation. From Townsend and Johnson (1985).

We can try to attach some physical significance to each of the terms in (108). Imagine looking at plots of the meridional wind on a large set of weather maps. On individual maps, the zonal average operator would reveal features such as the Hadley as shown in Figure 12. The result of applying the zonal average is called symmetric since it removes all variations in the zonal direction. On the other hand, mid-latitude eddies, which have no preferred meridional wind direction on a global scale, would average to zero and be invisible to the zonal average operator. We can use the departure from the zonal-mean to define the eddy component, $v^* = v - [v]$. On certain maps, for example winter maps, the eddy component would appear stronger due to winter storms. Instead of looking at individual maps, we can imagine looking at the same geographical location on several successive maps. At locations where the meridional wind is consistently in the same direction, the time average would be non-zero. For example, there can be persistent waves caused by topographic features. In this case, the wave is tied to a geographical location. Applying the time average operator reveals these stationary circulations. The non-stationary part of the total circulation is called transient, $v' = v - \bar{v}$. Geographic features can also cause enhanced transient flow. Examples are storm tracks off the east coasts of the continents, which result from land-sea temperature contrasts.

The $\overline{[v][T]}$ term in (108) represents the symmetric and stationary portion of the circulation. The second term, $[\overline{v}^*\overline{T}^*]$, represents the contribution to the total temperature flux from the stationary eddies. Again, these would have a preferred longitude. In our simple numerical models, which have no topography, we would expect this term to be very small since the model's surface is a perfectly uniform. The third term $\overline{[v'][T']}$ represents contributions to the total flow from transient symmetric circulations. On the real earth, this term would show the seasonal shifting of the stronger winter Hadley cell between northern and southern hemispheres. In our simple models, there is no time variation of the thermal forcing, so again, we would expect this term to be small. The fourth term, $\overline{[v'^*T'^*]}$, represents the transient eddies. This is the mechanism by which heat is transported poleward in middle-latitudes. These are the storms in the model.

In the numerical model, the zonal average is the average around a latitude circle of the data interpolated to the latitude/longitude grid. The time average is the average over numerous *snapshots* of the model simulations. These snapshots are separated by four days so that they are appreciatively different from one another. Typically, a time average is over about 200 samples. The terms of (108) from a simulation with the isentropic coordinate model are plotted in Figure 13. The many contours of $\overline{[v]}\overline{[T]}$ are concentrated mainly

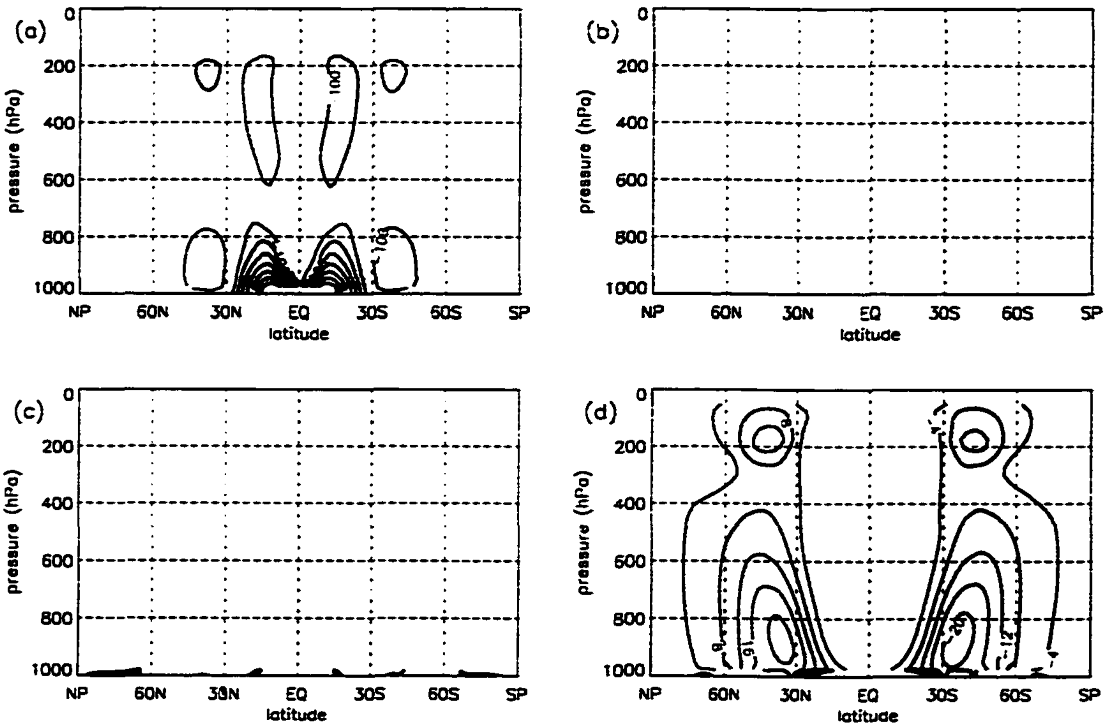


Figure 13. The terms of (108) from the isentropic coordinate model with 10242 cells and 64 layers. (a) $\overline{[v]}\overline{[T]}$, contour interval 100 m K s^{-1}, (b) $[\overline{v}^*\overline{T}^*]$ (m K s^{-1}), (c) $\overline{[v']}\overline{[T']}$ (m K s^{-1}), (d) $\overline{[v'^*T'^*]}$, contour interval 4 m K s^{-1}. Terms (b) and (c) are small compared to (a) and (d).

near the equator. On the other hand, contours of $\overline{[v'^*T'^*]}$ found at middle-latitudes. The terms $[\overline{v}^*\overline{T}^*]$ and $\overline{[v']}\overline{[T']}$ are comparatively small as expected. The stationary eddies and the transient eddies combine to give the total eddy contribution, $[\overline{v^*} \overline{T^*}] + [\overline{v'^* T'^*}] = [\overline{v^* T^*}]$.

Next we briefly discuss the mean meridional circulations in terms of energy. Energy is supplied to the atmosphere primarily in the tropics. It is then the job of the general circulation to redistribute that energy toward the poles. In the real atmosphere, the total energy of an air parcel can be partitioned into kinetic, internal, potential and latent components. The kinetic energy component can be ignored since it is small compared to the others. The latent component is like an internal energy. When water vapor condenses and precipitates out of a parcel, there is a source of energy to the parcel in the release of latent heat. We don't have to worry about these processes in the numerical models, however, since they do not include water vapor. The internal energy of an air parcel is proportional to its temperature. Looking at Figure 13 it appears as if energy is being concentrated in the tropics since low-level, warm air converges above the equator. Equatorward bound parcels do not indefinitely accumulate at the surface. They are removed by the rising branch of the Hadley cell. As a parcel rides along in the rising branch it acquires potential energy while it cools. When it finally turns and moves poleward, its potential energy has significantly increased. So, the tropical atmosphere imports warm, low potential energy parcels near the earth's surface, and exports cool, high potential energy parcels in the poleward branch. Since the net export must be greater than the import, the change in potential energy must dominate.

To make this more quantitative, we define dry static energy $s \equiv c_p T + \phi$ where $c_p = 1004 \text{ J kg}^{-1} \text{ K}^{-1}$ is the heat capacity at constant pressure, $\phi \equiv gz$ is the geopotential. Mimicking (108), we can write

$$\overline{[vs]} = \overline{[v]}\overline{[s]} + [\overline{v^*}\overline{s^*}] + \overline{[v']}\overline{[s']} + [\overline{v'^* s'^*}] \tag{110}$$

Again, the second and third terms of (110) will be small and are not plotted. The first and forth terms are plotted Figure 14.

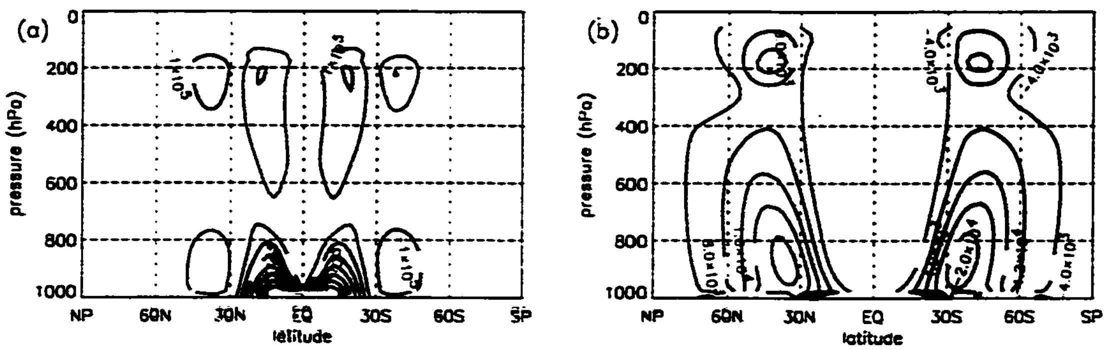


Figure 14. The terms of (110) from the isentropic coordinate model. (a) $\overline{[v]}\overline{[s]}$, contour interval 10^5 kg m^2 s^{-2}, (b) $\overline{[v'^*s'^*]}$, contour interval 4×10^3 kg m^2 s^{-1}.

5.3.2 *Spectral model*

Now we return to the resolution study. To compare simulation quality as a function of resolution, we need to know what a good simulation looks like. Isaac Held has provided us with results from a dry dynamical core based of the a spectral model. This is the model used in the Held and Suarez (1994) paper. The data are at two horizontal resolutions of T30 and T63. In Figure 15 we see can features similar to those found in the real atmosphere including mid-latitude zonal jets and low-level trade winds in the sub-tropics. We can see evidence of a Hadley circulations as well as a Ferrel cells in plots of the meridional velocity. There is a shallow layer of converging flow over the equator, with a deeper layer of diffuse divergence aloft. Increasing the resolution from T30 to T63 does not significantly change the winds. The region of easterlies directly over the equator becomes a little wider, and both the Hadley cell and the Ferrel cell become stronger. Figure 16 shows various correlation statistics for T30. Panel (a) shows that the circulation helps to cool low-latitudes, particularly near the surface, and warm high-latitudes. Panel (b) shows that strongest zonal wind anomalies occur just poleward of the jet maximum. In panel (c) we can see that the meridional wind tends to carry zonal momentum poleward,

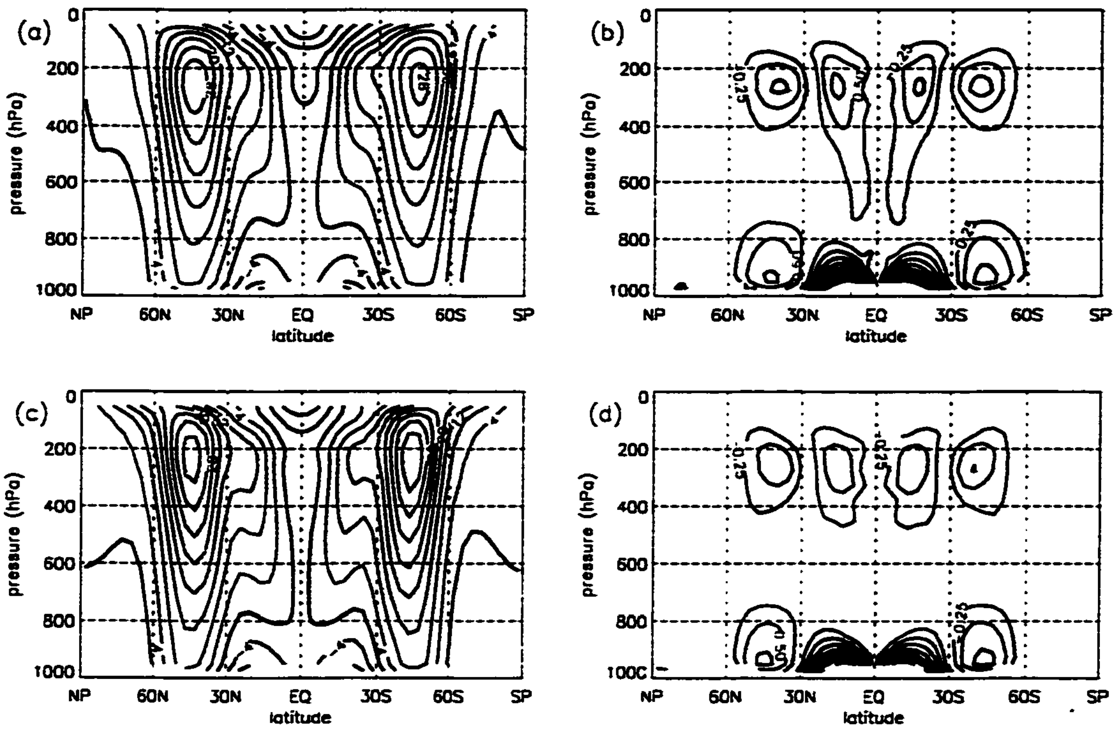


Figure 15. Spectral T63. Zonal-mean (a) zonal wind (m s^{-1}) and (b) meridional wind (m s^{-1}). Spectral T30. Zonal-mean (c) zonal wind (m s^{-1}) and (d) meridional wind (m s^{-1}). Zonal wind contour intervals 4 m s^{-1}, meridional wind contour intervals 0.25 m s^{-1}.

particularly on the equatorward side of the jet maximum. This helps to strengthen the jet. Panel (d) shows the largest meridional wind anomalies occurs pretty much within the jet maximum. Panel (e) shows that much of the meridional heat transport is done by the middle-latitude eddies. The Hadley cell is less efficient at transporting heat poleward because the horizontal temperature gradients in the tropics are weaker. If the horizontal gradients of a field are relatively small it is difficult for the meridional wind to produce very large anomalies. Looking at Figure 7, we can see that the temperature field wants to relax toward shape that is flat in the tropics, and has strong meridional temperature gradient in middle-latitudes. Increasing the horizontal resolution, as shown in Figure 17, results in significant changes with all of the correlations becoming stronger. In the next sections we show similar results for each of the three coordinate systems discussed above.

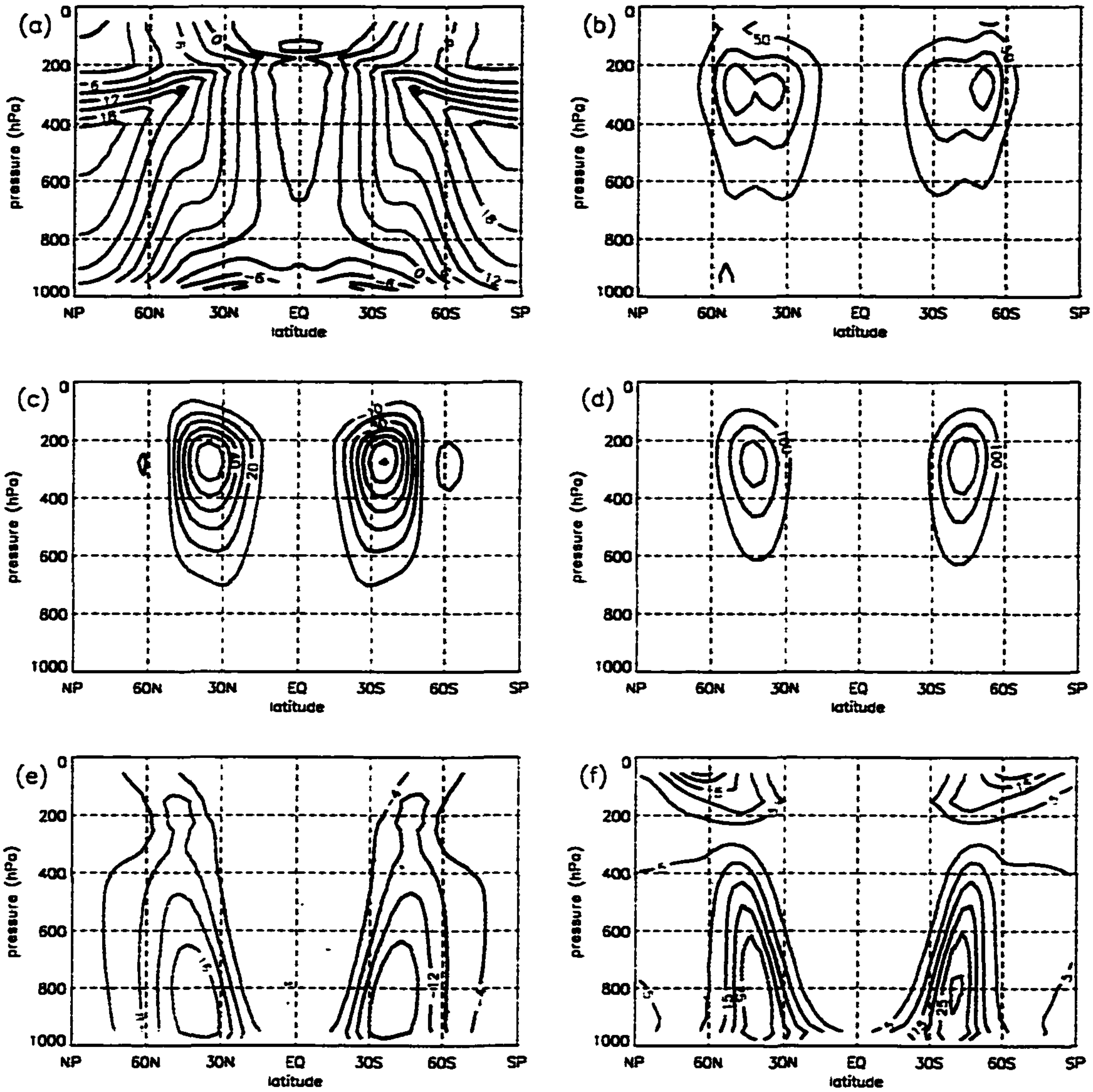


Figure 16. Spectral T30. (a) $\overline{[T-T_{eq}]}$ (K), (b) $\overline{[u^*u^*]}$ (m^2 s^{-2}), (c) $\overline{[u^*v^*]}$ (m^2 s^{-2}), (d) zonal-mean meridional wind variance, $\overline{[v^*v^*]}$ (m^2 s^{-2}), (e) zonal-mean meridional heat transport, $\overline{[v^*T^*]}$ (m K s^{-1}), (f) zonal-mean temperature variance, $\overline{[T^*T^*]}$ (K^2).

5.3.3 Charney-Phillips σ-coordinates

In this section we show results from the Charney-Phillips sigma coordinate model. A value must be assigned to each coordinate surface. In the sigma coordinate this is a very simple task. Suppose we pick pressure at the top of the model to be $p_T = 1$ hPa, and we pick a globally integrated surface pressure to be $p_S = 1000$ hPa, and a simple definition

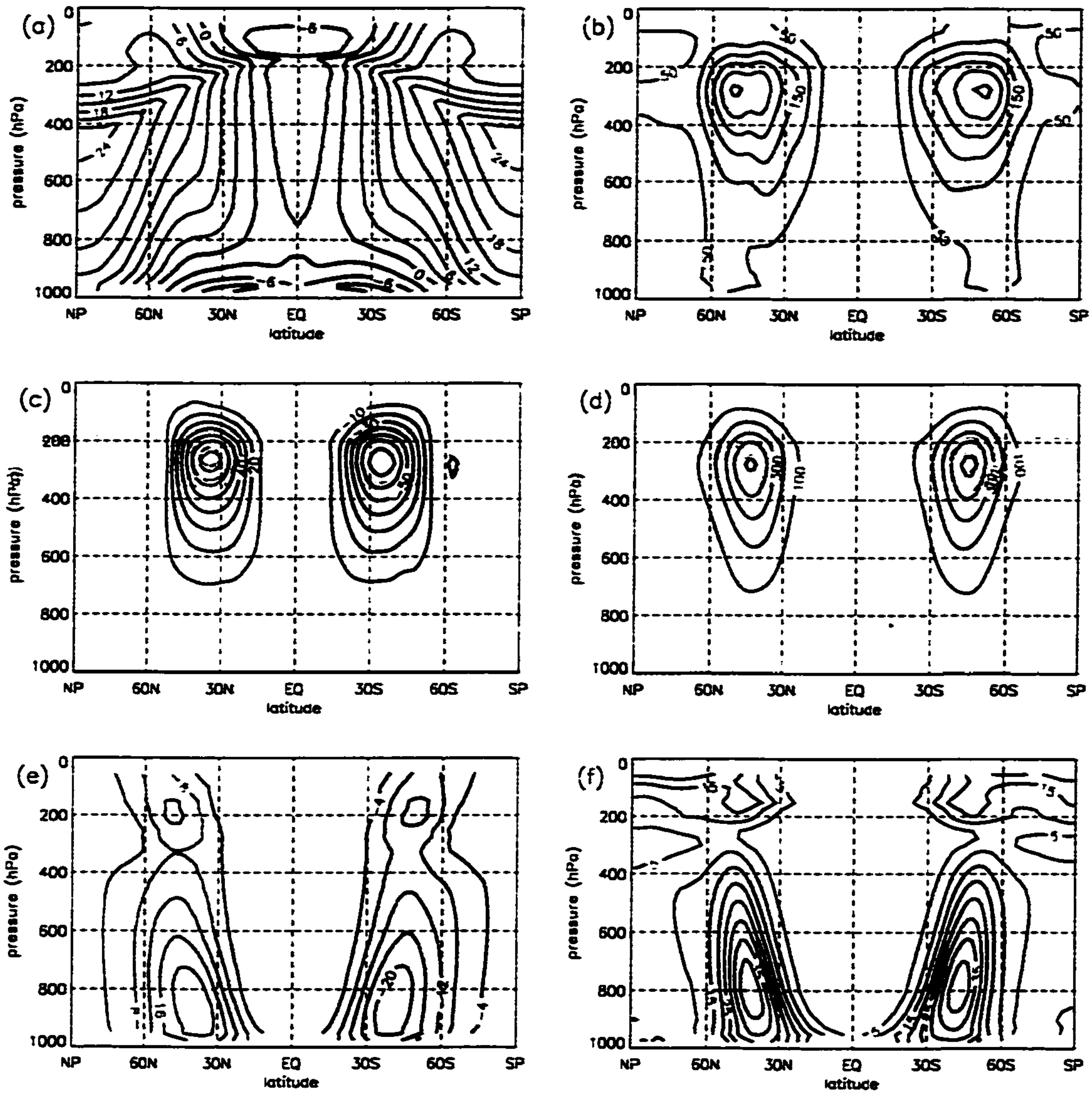


Figure 17. Spectral T63. (a) $\overline{[T-T_{eq}]}$ (K), (b) $\overline{[u^*u^*]}$ ($m^2 s^{-2}$), (c) $\overline{[u^*v^*]}$ ($m^2 s^{-2}$), (d) zonal-mean meridional wind variance, $\overline{[v^*v^*]}$ ($m^2 s^{-2}$), (e) zonal-mean meridional heat transport, $\overline{[v^*T^*]}$ (m K s^{-1}), (f) zonal-mean temperature variance, $\overline{[T^*T^*]}$ (K^2).

of σ, for example, $\sigma \equiv (p-p_T)/(p_S-p_T)$. If there are L model layers, then there are $L+1$ layer edges. We will set the value of sigma on the edges to be $0, 1/L, 2/L, \ldots, (L-1)/L, 1$. Since $p = p_T + \sigma(p_S - p_T)$, the coordinate surfaces are evenly spaced in pressure.

The zonal-mean zonal wind and zonal-mean meridional wind are shown in Figure 18 and in Figure 19.

The quality of the solution improves with increasing horizontal resolution much

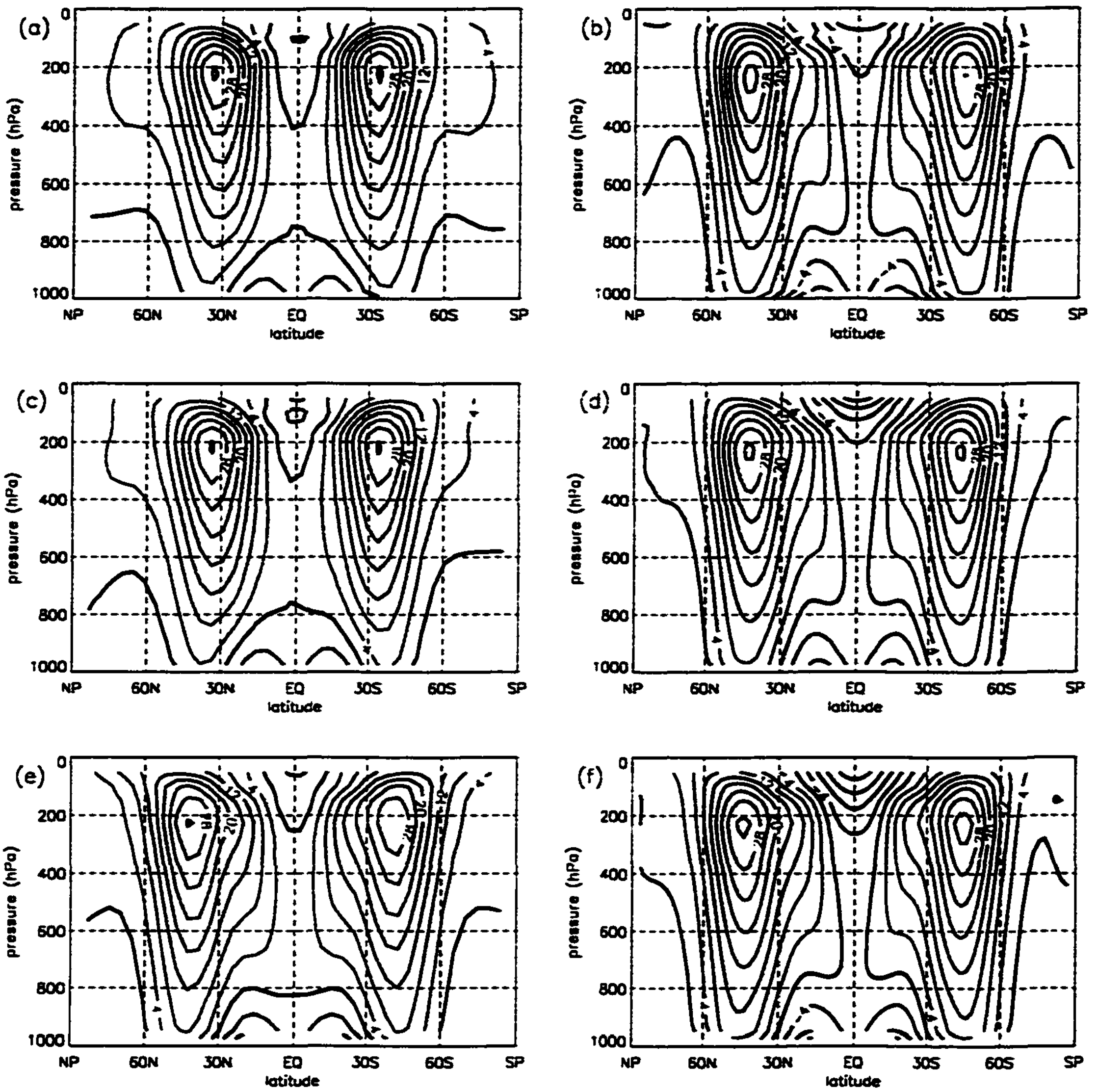


Figure 18. Sigma coordinate. Zonal-mean zonal wind $[u]$ (m s^{-1}).
(a) 2562 cells, 64 layers. (b) 10242 cells, 64 layers. (c) 2562 cells, 32 layers.
(d) 10242 cells, 32 layers. (e) 2562 cells, 16 layers. (f) 10242 cells, 16 layers.

more than by increasing vertical resolution. The change from 2562 cells to 10242 cells allows for easterlies throughout the entire vertical column in the tropics, and the jet maximum becomes better defined and positioned. Curiously, with 10242 cells, the easterlies aloft become weaker with increasing vertical resolution. Comparing Figure 18,

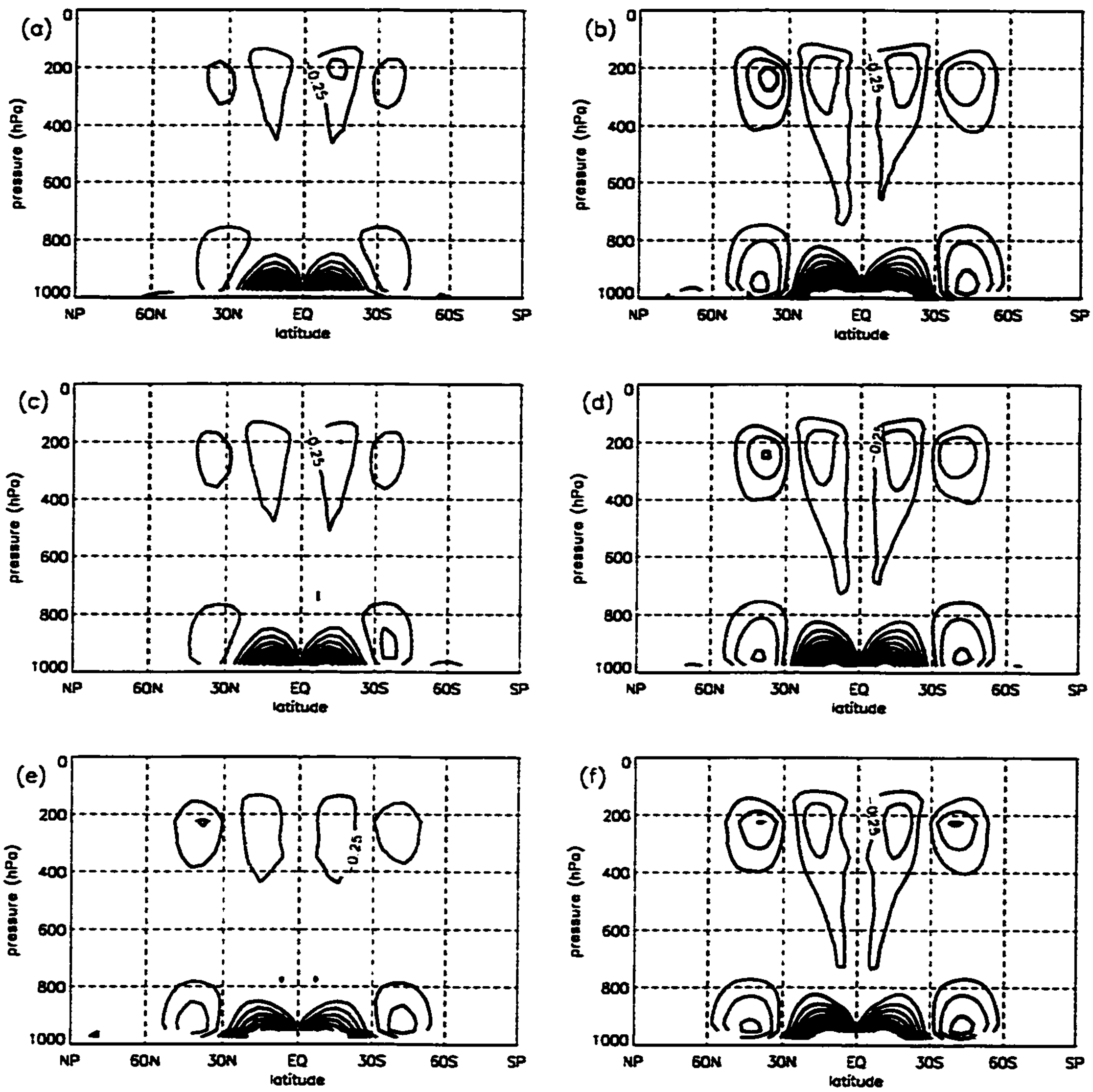


Figure 19. Sigma coordinate. Zonal-mean meridional wind [v] (m s^{-1}).
(a) 2562 cells, 64 layers. (b) 10242 cells, 64 layers. (c) 2562 cells, 32 layers.
(d) 10242 cells, 32 layers. (e) 2562 cells, 16 layers. (f) 10242 cells, 16 layers.

Figure 19 and Figure 15, we can see the T63 horizontal resolution is comparable to 10242 cells. Figure 20 shows the zonal-mean meridional heat transport. For each horizontal resolution, increasing the vertical resolution from 16 layers to 32 layers improves the solution, but increasing from 32 layers to 64 layer has little effect. However, for each vertical resolution, increasing the horizontal resolution from 2562 cells to 10242 cells improves the solution, particularly in the stratosphere.

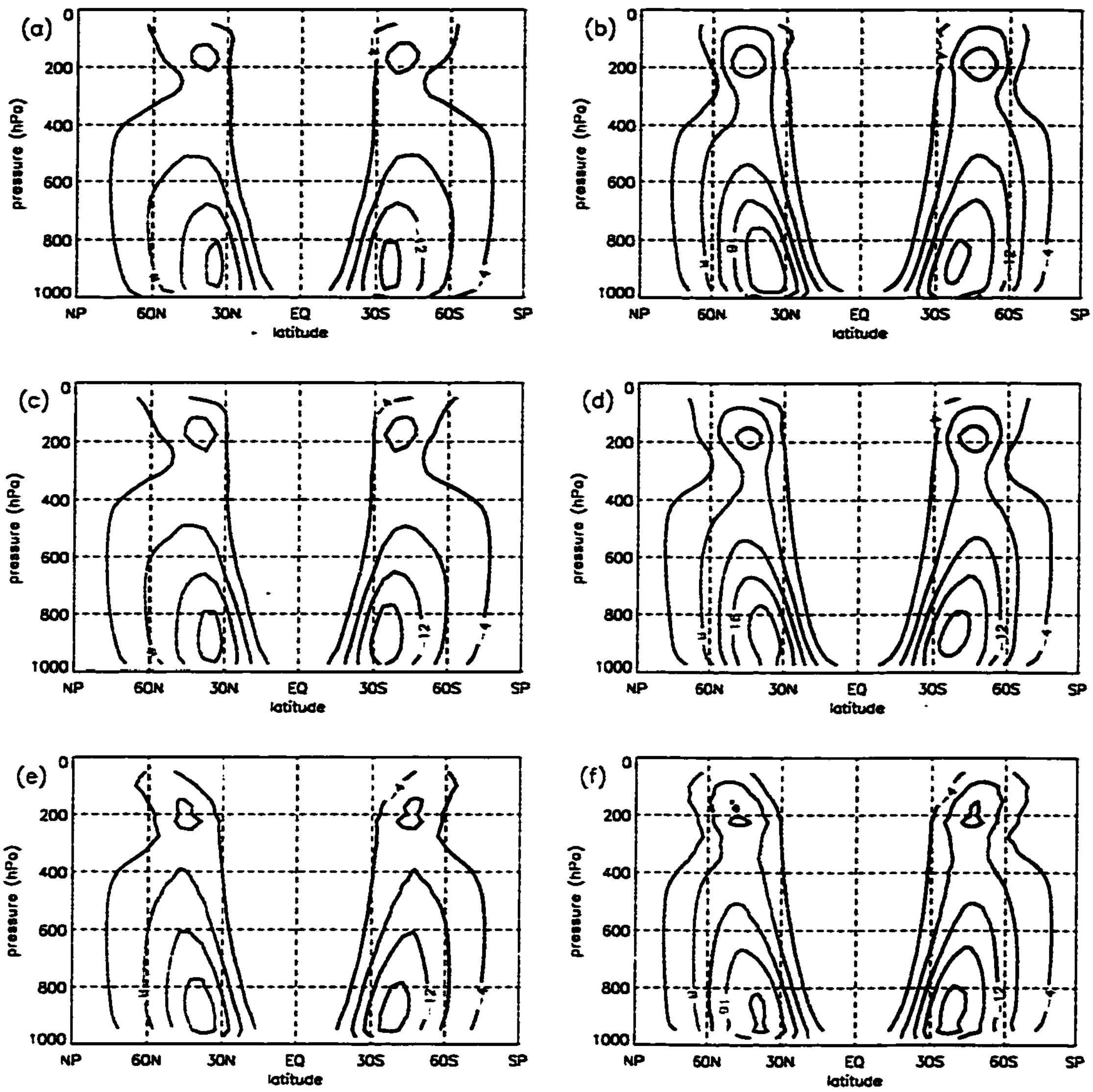


Figure 20. Sigma coordinate. Zonal-mean meridional heat transport, $[v^*T^*]$ (m K s^{-1}).
(a) 2562 cells, 64 layers. (b) 10242 cells, 64 layers. (c) 2562 cells, 32 layers.
(d) 10242 cells, 32 layers. (e) 2562 cells, 16 layers. (f) 10242 cells, 16 layers.

5.3.4 Isentropic coordinates

To use the isentropic coordinates, we must first assign a value of θ to each coordinate surface. By inspecting (98), if we set $p = 1000$ hPa then the Newtonian relaxation of temperature wants to force the surface temperature to be about 315 K at the equator and 255 K at the poles. Values of θ significantly colder than 255 K would be computationally wasteful since the model will never form air much colder than this value.

In other words, coordinate surfaces less than 255 K will result in massless layers prostrate against the earth's surface which can never be inflated. With this in mind, we pick $\theta_S = 250$ K for the bottom edge of the model. Also by looking at (98), the model will try to form an isothermal stratosphere with a static temperature at the top edge of $T_T = 200\text{K}$. If we again want the pressure on the top edge to be a constant $p_T = 1$ hPa, then $\theta_T = T_T(p_0/p_T)^{\kappa} \approx 1440.8$ K. The stratosphere is very stable, so within the stratosphere potential temperature increase rapidly with height. If we pick values of θ that evenly divide the range between θ_S and θ_T, then almost all of the coordinate surface will be packed against the top edge of the model as shown in Figure 21(a). Instead, we assign

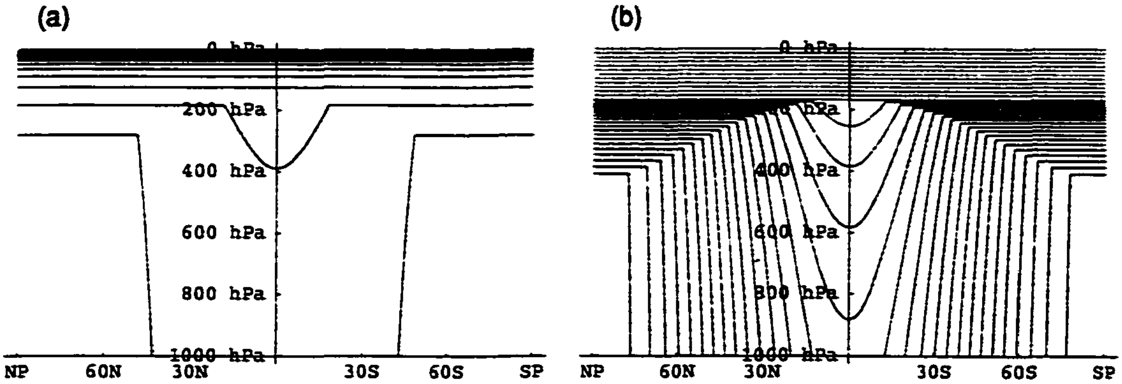


Figure 21. With 32 model layers (a) Values of θ so that they evenly divide the range between θ_S and θ_T. (b) Values of θ generated according to the algorithm in the text.

values of θ using the following recipe.

Again consider (98). Suppose we set $\varphi = 0$, and equate the two parts of the "max" operator, and solve for pressure at the tropopause, p_{trp}. So, $p_{\text{trp}} \approx 168.4$ hPa will satisfy $\left[315 - 10\ln\left(\frac{p_{\text{trp}}}{p_0}\right)\right]\left(\frac{p_{\text{trp}}}{p_0}\right)^{\kappa} = 200$. And, $\theta_{\text{trp}} = 200(p_0/p_{\text{trp}})^{\kappa} \approx 332.8$ K is the corresponding potential temperature. In the troposphere we set values of θ so that they evenly divide the potential temperature range θ_S to θ_{trp}, and stratosphere so that they evenly divide the pressure range p_{trp} to p_T. We maintain second-order accurate finite-difference operators in the troposphere. In the example shown in Figure 21(b), of the 32 model layers, 12 are entirely within the stratosphere, and 20 are in within both the stratosphere and the troposphere. In general, we place $3/8$ of the layers within the stratosphere $5/8$ within the troposphere.

It is difficult to get good resolution in the tropical troposphere since $\theta_{\text{trp}} - \theta_{\text{eq}} \approx 18$ K is a pretty small number compared to $\theta_{\text{trp}} - \theta_S \approx 83$ K where $\theta_{\text{eq}} = 315$ K the value at the surface at the equator. If the surfaces are spaced to evenly partition 83 K, then at best there will be a handful of surfaces in the tropics. In the case with 32 layers, as shown in Figure 21(b), the whole depth of a vertical column in the tropics is covered with only five layers. Attempting to correct this problem by choosing layers which accurately resolve the tropical troposphere results in very thin layers immediately below the tropopause. Unrealistically strong winds can develop in these thin layers and can cause numerical difficulties.

We have seen a plot of T_{eq} in Figure 7. It is also interesting to look at the corresponding equilibrium distribution of pressure and mass ($m \equiv -\partial p/\partial\theta$) in θ-space. The curve in (b) represents the earth's surface. The packing of contours along this curve implies that the relaxation is trying to concentrate mass in a very narrow region near the earth's surface. Mass is discontinuous right at the earth's surface. This is evident by the shape gradients in Figure 22(b). We can also see from Figure 22(b) that much of the computational domain in θ-space is almost massless.

Next we will look at the zonal and meridional wind, and meridional heat transport

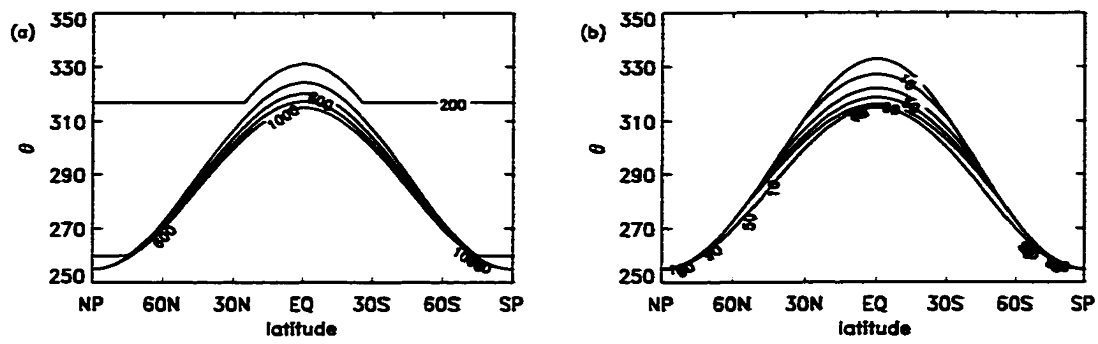


Figure 22. (a) The distribution of pressure (hPa) in θ-space. Contour interval 200 hPa. (b) The distribution of mass (hPa K^{-1}) in θ-space. Contour interval 10 hPa K^{-1}.

in the isentropic model. Looking Figure 23 and Figure 24 we can see some obvious differences between the sigma coordinate and the isentropic coordinate. With the isentropic coordinate, results change with increasing vertical resolution. Indeed, higher vertical resolution is necessary for at least an adequate simulation. In Figure 23, with 16 layers, the jet is too weak and too far poleward. The results significantly improve going to 32 layers. The increased resolution in the stratosphere allows the contours to close-off and not intersect the top edge of the model. It also allows for the formation of easterlies in the stratosphere. The shape of the contour lines through the tropics lacks the details which are present in Figure 18. This results from the lack of resolution in the tropics. The equatorward meridional wind in the tropics extends much too far away from the surface particularly with 16 layers. The depth of the low level winds is improved with 32 layers, and 64 layers gives results which agree better with Figure 19.

Figure 25 shows the zonal-mean meridional heat transport. With this coordinate, increasing the vertical resolution improves the simulation. The isentropic surfaces are tilted from the horizontal, so that a path parallel to the earth's surface in the meridional direction will pass thought coordinate surfaces. So, an increase in vertical resolution will also increase the horizontal resolution and the quality of the simulation.

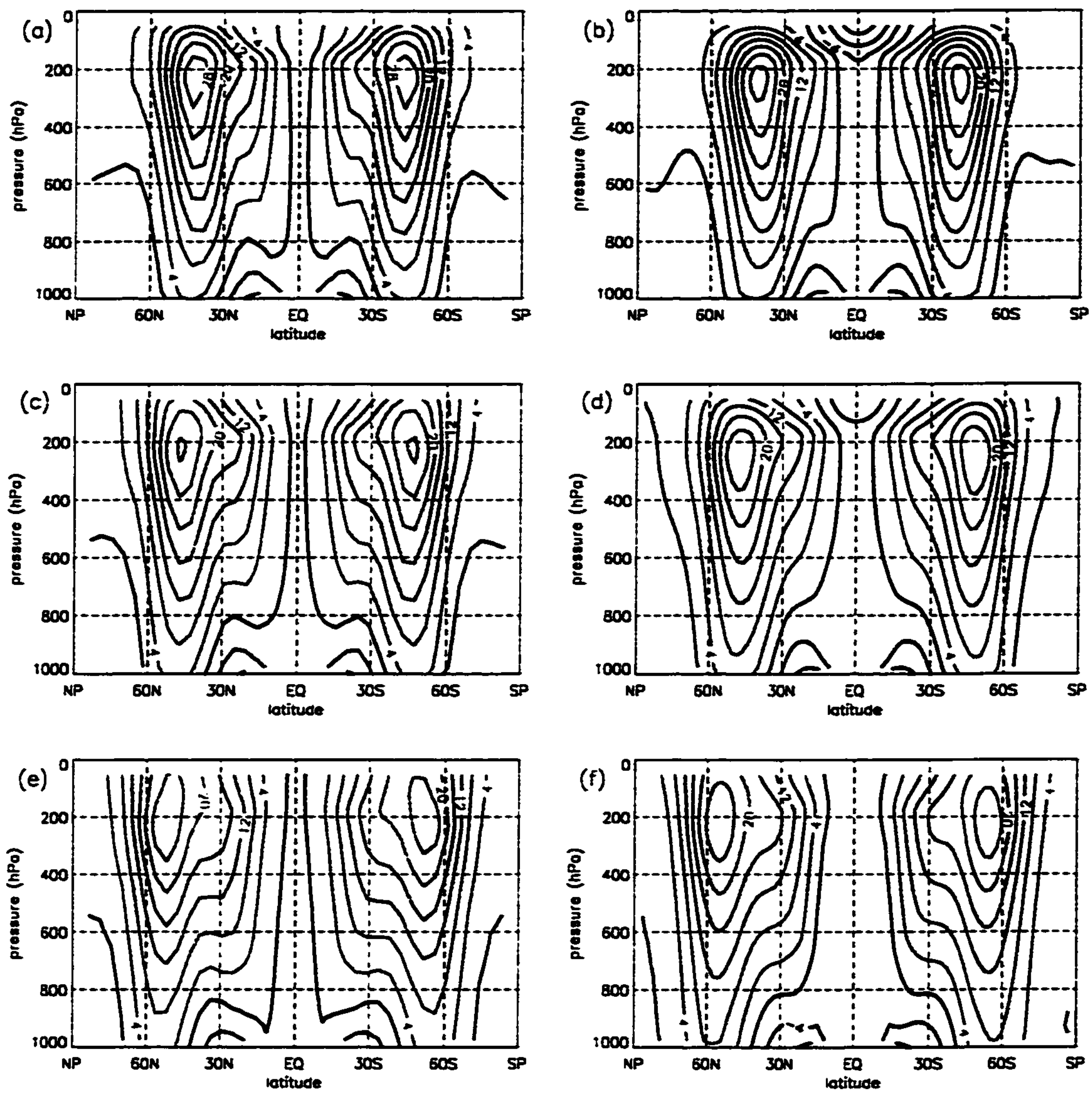


Figure 23. Isentropic coordinate. Zonal-mean zonal wind $[u]$ (m s^{-1}).
(a) 2562 cells, 64 layers. (b) 10242 cells, 64 layers. (c) 2562 cells, 32 layers.
(d) 10242 cells, 32 layers. (e) 2562 cells, 16 layers. (f) 10242 cells, 16 layers

5.3.5 Generalized vertical coordinate

We use a technique similar to the isentropic coordinates to select a value of ζ for each coordinate surface. Since we know the equilibrium pressure and potential temperature at the tropopause above the equator, p_{trp} and θ_{trp}, we can compute ζ_{trp} using the definition of ζ. This number is very close to θ_{trp} since the tropopause is in a part of the computational domain that is more-or-less pure isentropic. We pick $\zeta_S = 250$ K for

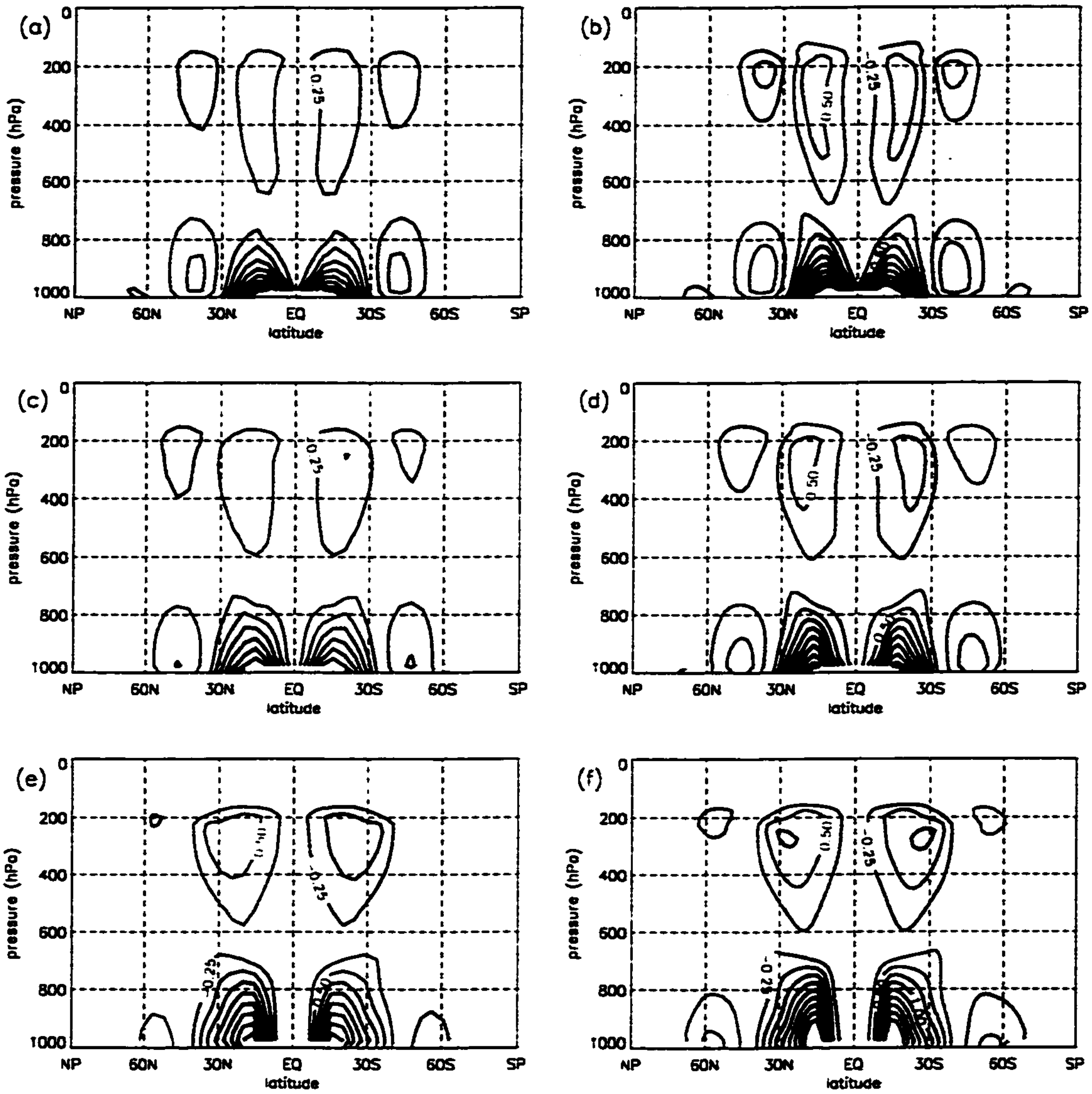


Figure 24. Isentropic coordinate. Zonal-mean meridional wind $[v]$ (m s^{-1}).
(a) 2562 cells, 64 layers. (b) 10242 cells, 64 layers. (c) 2562 cells, 32 layers.
(d) 10242 cells, 32 layers. (e) 2562 cells, 16 layers. (f) 10242 cells, 16 layers

the value of the coordinate surface coincident with the earth's surface, and linearly interpolate from ζ_S to ζ_{trp} to obtain values in the troposphere portion of the computational domain. In the stratosphere, the surfaces are again positioned so that they are linear in pressure. For the case with 32 layers, the distribution of coordinate surfaces in shown in Figure 26. Note that the surfaces look the same as the pure isentropes in the upper-half of the computational domain, but flatten out, like sigma surfaces, in the lower-half of the computational domain.

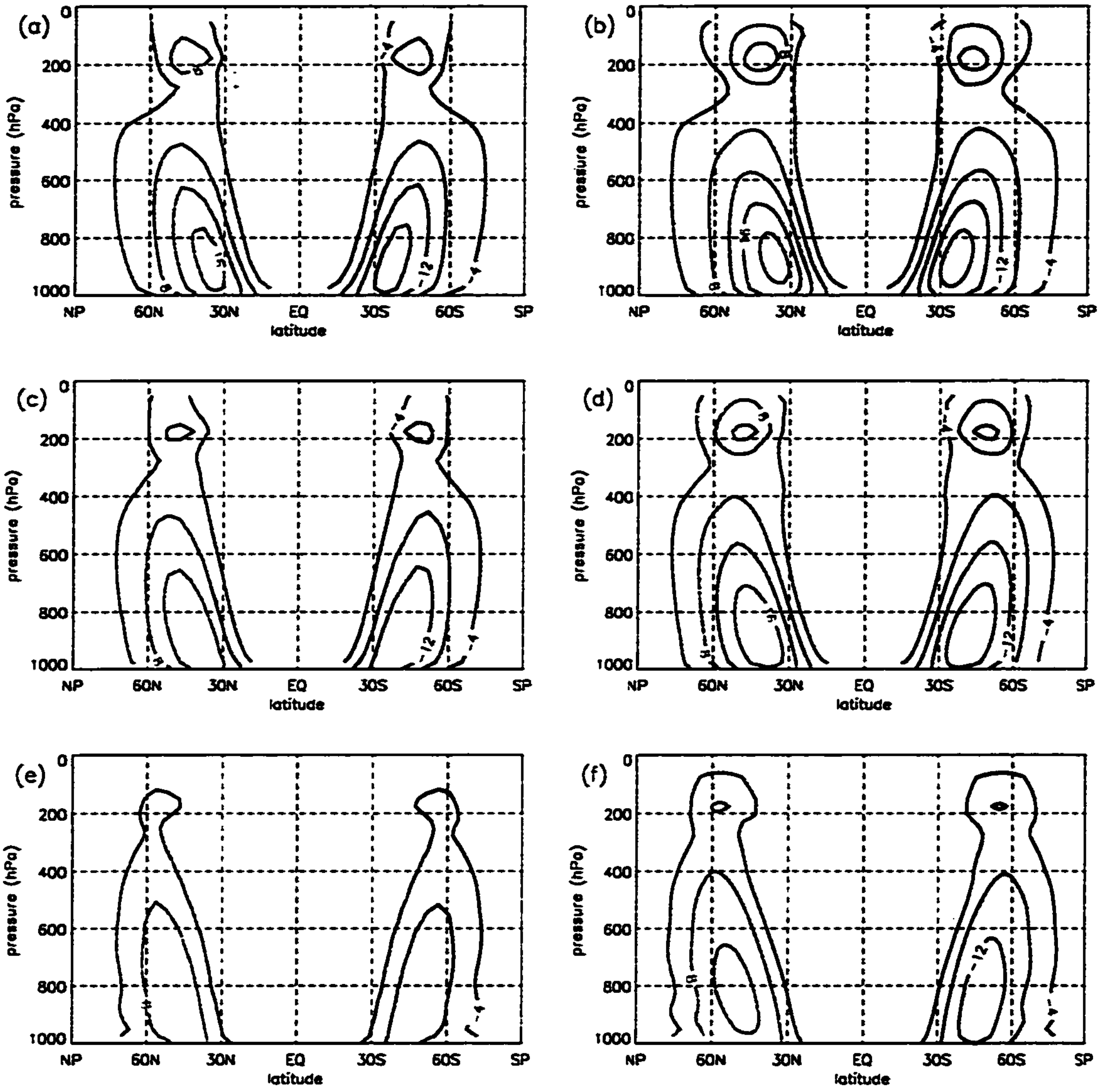
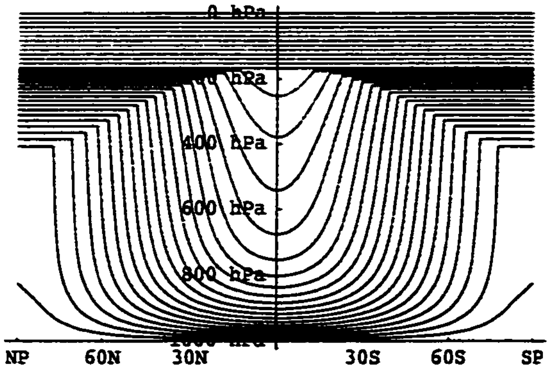


Figure 25. Isentropic coordinate. Zonal-mean meridional heat transport, $[v^*T^*]$ (m K s^{-1}).
(a) 2562 cells, 64 layers. (b) 10242 cells, 64 layers. (c) 2562 cells, 32 layers.
(d) 10242 cells, 32 layers. (e) 2562 cells, 16 layers. (f) 10242 cells, 16 layers.

Figure 26. Generalized vertical coordinate. Equilibrium distribution of coordinate surfaces with 32 model layers.



Again, we will show plots of the zonal-mean zonal wind, zonal-mean meridional wind, and zonal-mean meridional heat transport. The zonal-mean zonal wind, not too surprisingly, looks like the zonal-mean zonal wind of the pure isentropic model. This is shown in Figure 27. In both the pure isentropic model and the generalized vertical

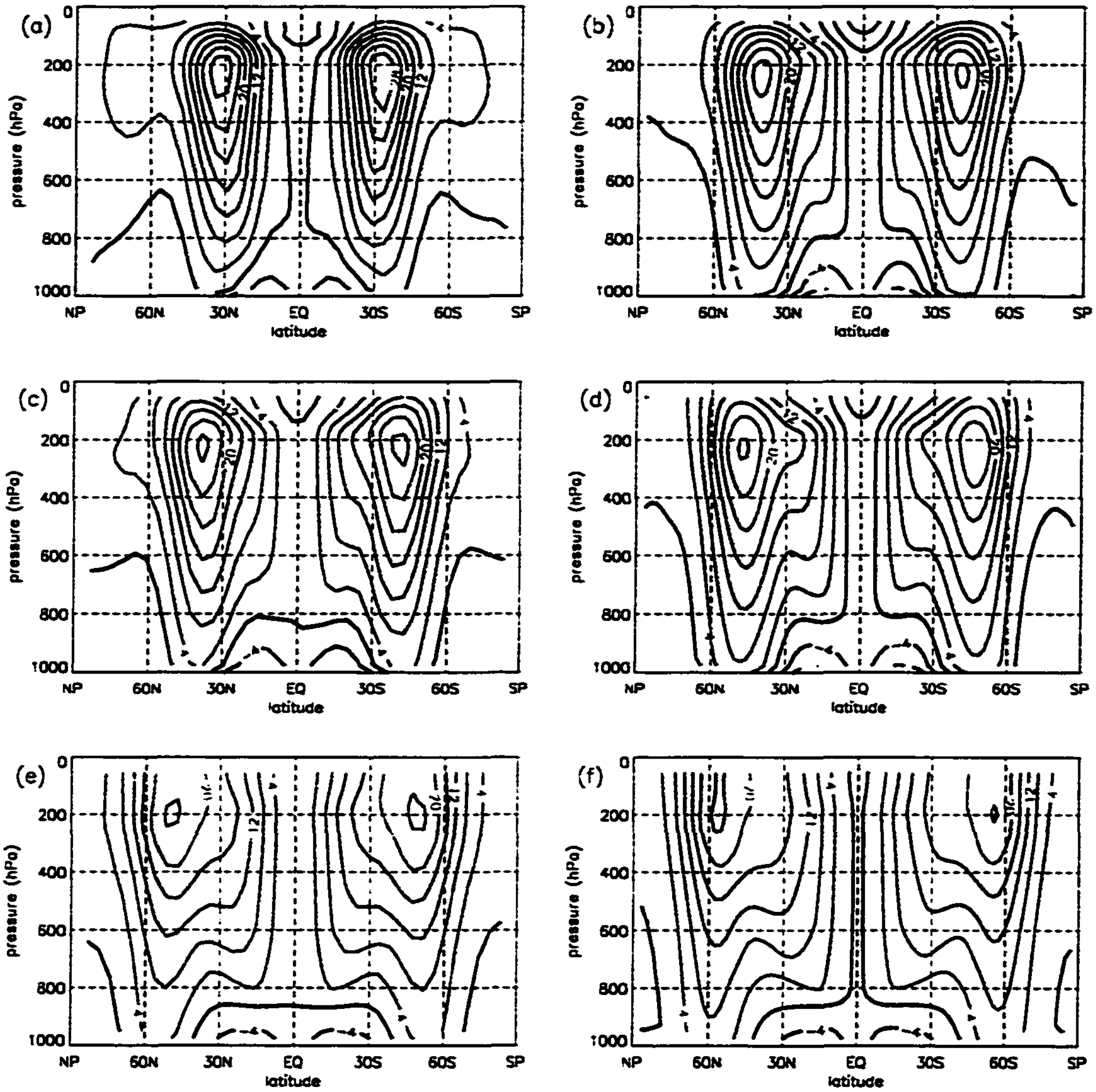


Figure 27. Generalized vertical coordinate. Zonal-mean zonal wind $[u]$ (m s^{-1}).
(a) 2562 cells, 64 layers. (b) 10242 cells, 64 layers. (c) 2562 cells, 32 layers.
(d) 10242 cells, 32 layers. (e) 2562 cells, 16 layers. (f) 10242 cells, 16 layers

coordinate model, the jet max is too far poleward at coarser vertical resolution, and moves to its proper latitude as the resolution is increased. In the 10242 cell model with 32 layers and 64 layers, the shape of the contours in the tropics is better than the corresponding pure isentropic simulations. If we compare Figure 21(b) and Figure 26, we can see that the generalized vertical coordinate model has far superior vertical resolution in the tropics near the earth's surface in the sigma domain. Figure 28 shows the zonal-mean meridional

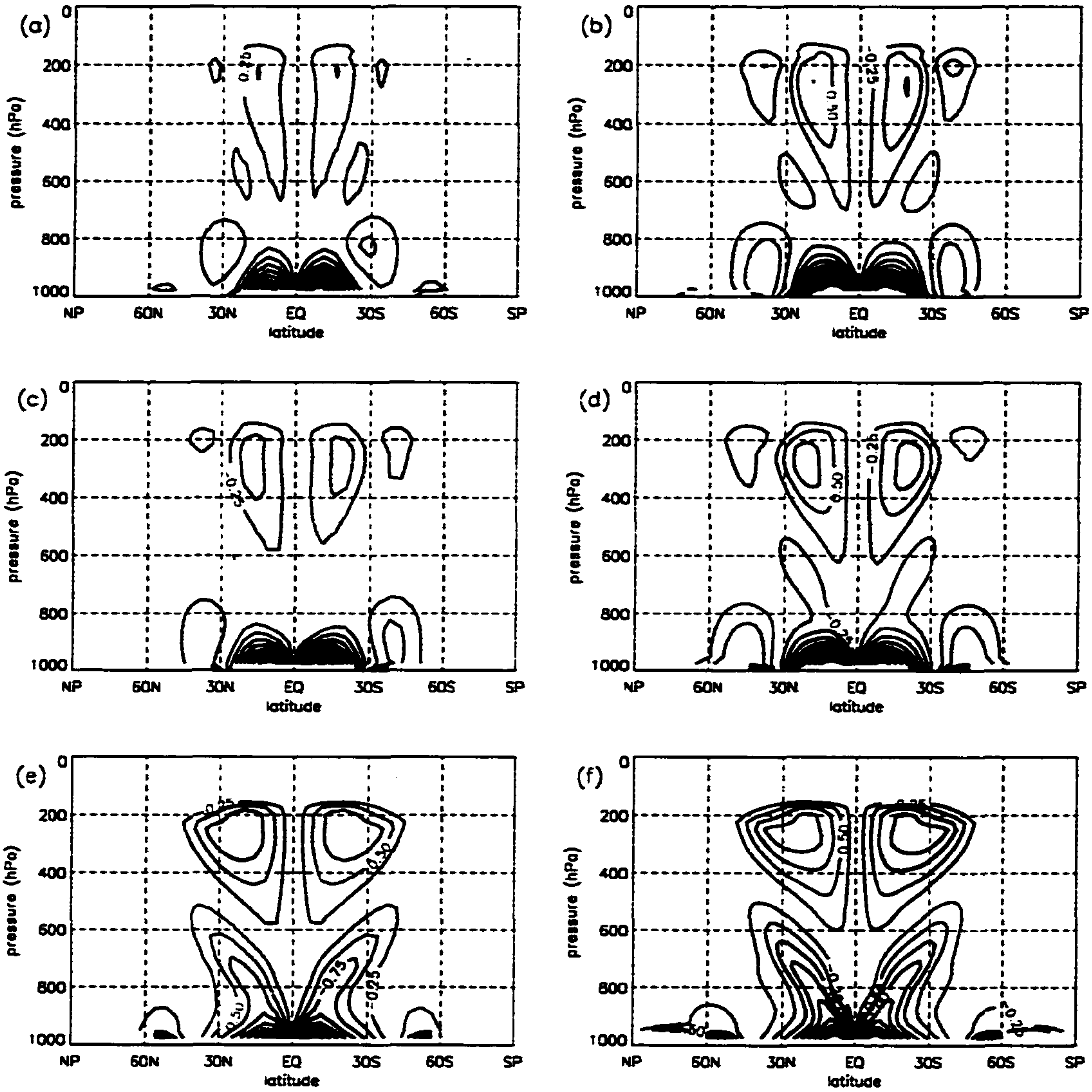


Figure 28. Generalized vertical coordinate. Zonal-mean meridional wind $[v]$ (m s^{-1}).
(a) 2562 cells, 64 layers. (b) 10242 cells, 64 layers. (c) 2562 cells, 32 layers.
(d) 10242 cells, 32 layers. (e) 2562 cells, 16 layers. (f) 10242 cells, 16 layers

wind. Here too we can see that the equatorward branch of the Hadley circulation stays below 800 hPa.

Figure 29 shows the zonal-mean meridional heat transport. Again, these results look very similar to the pure isentropic model.

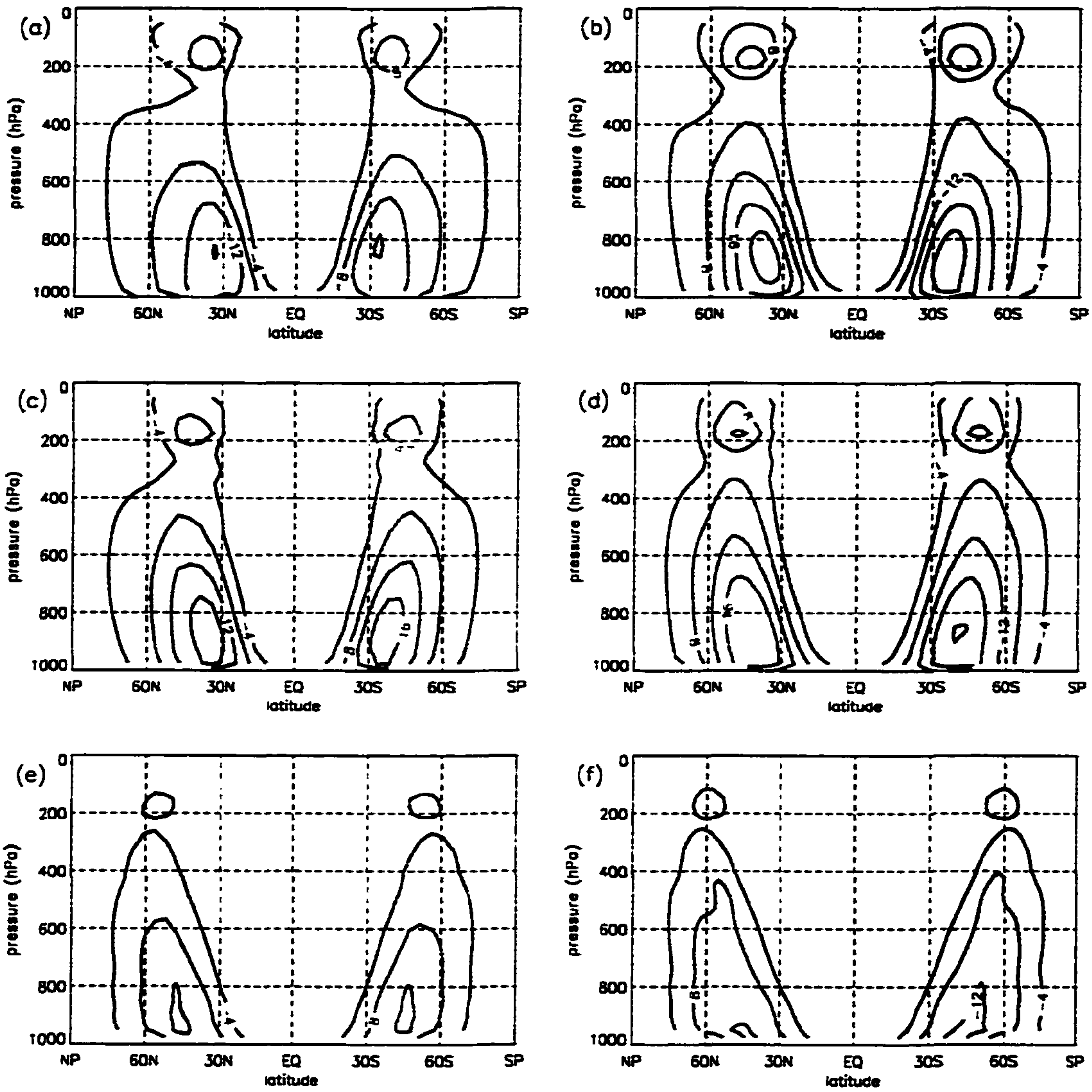


Figure 29. Generalized vertical coordinate. Zonal-mean meridional heat transport, $[v^*T^*]$ (m K s^{-1}).(a) 2562 cells, 64 layers. (b) 10242 cells, 64 layers. (c) 2562 cells, 32 layers. (d) 10242 cells, 32 layers. (e) 2562 cells, 16 layers. (f) 10242 cells, 16 layers.

The conclusions we draw from these model comparisons are heavily dependent of the Held and Suarez forcing. As indicated in Figure 1, the real atmosphere is not so close to neutral stability in the tropics, and therefore would not suffer as much from inadequate resolution. Nevertheless, these numerical results may indicate that the hybrid vertical coordinate model will require greater vertical resolution than the sigma model.

5.4 The mean meridional circulation and fluid dynamical peristalsis

5.4.1 The mean meridional circulation as viewed in isentropic coordinates

Within the atmosphere, the systematic pattern of heating in low-latitudes and cooling in high-latitudes implies that mean meridional isentropic circulations exist that span each hemisphere. An upward branch associated with net heating exists in the tropics, while a downward branch associated with net cooling exits at higher latitudes. In order to balance these vertical motions, mass continuity requires a quasi-horizontal branch from equator to pole at upper levels, and pole to equator at lower levels.

The isentropic mass stream function for the zonally averaged mass transport (Johnson, 1989) is defined by

$$\frac{\partial \Psi_\theta}{\partial \theta} = -2\pi a \cos\varphi \overline{[mv]} \quad \text{and} \quad \frac{\partial \Psi_\theta}{\partial \varphi} = 2\pi a^2 \cos\varphi \overline{[m\dot{\theta}]}. \tag{111}$$

Plots of this quantity derived from ECMWF analysis of January and July 1979 are shown in Figure 30. Here we can clearly see that the direct Hadley circulations extend from

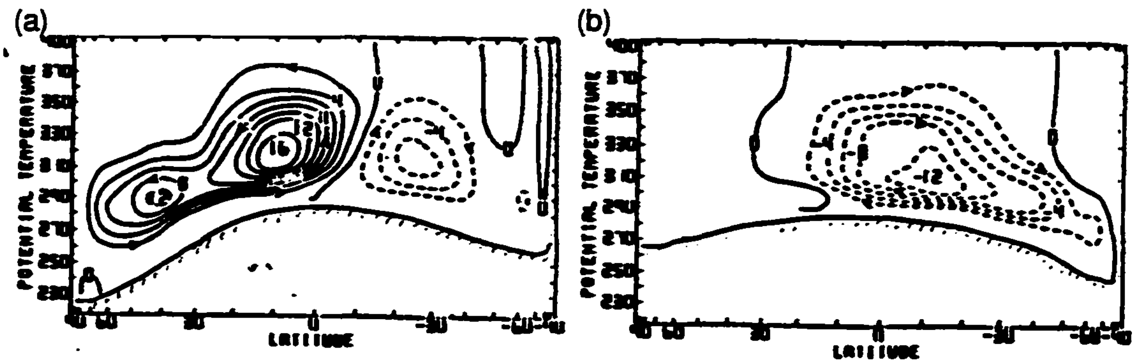


Figure 30. Mass stream function (10^{10} kg s^{-1}) for mean meridional circulations computed from ECMWF analyses for (a) January 1979 and (b) July 1979. From Johnson (1989).

equator to pole, are the Ferrel cells are absent when viewed in isentropic-coordinates. This is different than the view in isobaric coordinates as shown in Figure 12. The winter hemisphere, which experiences the greater differential heating, contains the larger cell. There is a gentle meridional slope from equator to pole indicating that a parcel gradually cools moving towards the pole, and gradually warms moving towards the equator.

5.4.2 A tracer in the isentropic coordinate model

Suppose τ is the mixing ratio of an arbitrary tracer. The tracer equation is given by

$$\frac{\partial}{\partial t}(m\tau) + \nabla \bullet (m\tau \mathbf{V}) + \frac{\partial}{\partial \theta}(m\tau\dot{\theta}) = \Delta_{src} m - \Delta_{snk} m\tau, \tag{112}$$

where $\Delta_{src} m$ is a source of tracer, and $\Delta_{snk} m\tau$ is a sink of tracer. Since we wish to simulate the poleward transport of the tracer, the source is placed at the equator and the sink at the poles. The shape of Δ_{src} and Δ_{snk} is shown in Figure 31. We designed a tracer

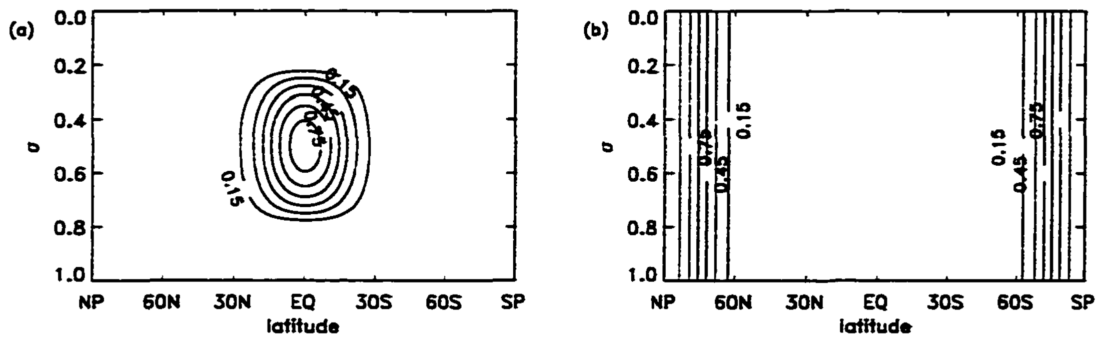


Figure 31. (a) the source term $\Delta_{src}(\varphi, \sigma)$ (kg m^{-1} (2 days)$^{-1}$) is centered above the equator and is non-zero between 200 hPa and 800 hPa, (b) the sink term $\Delta_{snk}(\varphi, \sigma)$ (kg^{-1} (2 days)$^{-1}$)

equation for the isentropic-coordinate model. This tracer equation is based on a monotone flux-corrected transport. We must contend with the possibility of massless cells. The mass in a massless cell is zero, so the actual amount of substance $m\tau$ is also zero regardless of the mixing ratio τ. So, the value of τ for a massless cell is undefined. The details of the scheme are discussed in Appendix M.

The zonal averaging operator applied to the product of mass and meridional velocity can be written

$$
\begin{aligned}
[mv] &= [([m]+m^*)([v]+v^*)] \\
&= [m][v]+[[m]v^*]+[m^*[v]]+[m^*v^*] \\
&= [m][v]+[m^*v^*].
\end{aligned}
\tag{113}
$$

The term $[m^*v^*]$ is called the bolus mass flux. To attach some physical significance to this term, imagine applying the zonal averaging operator along a latitude circle between two coordinate surfaces $\theta_{l-1/2}$ and $\theta_{l+1/2}$. Suppose in the northern hemisphere, $[m^*v^*]$ is positive. This will be true if $v>0$ at longitudes where the isentropes are far apart, $\partial p/\partial\theta$ relatively large, and $v<0$ when the isentropes are close, $\partial p/\partial\theta$ relatively small. See Figure 32. In other words, to use an analogy from plumbing, there is larger

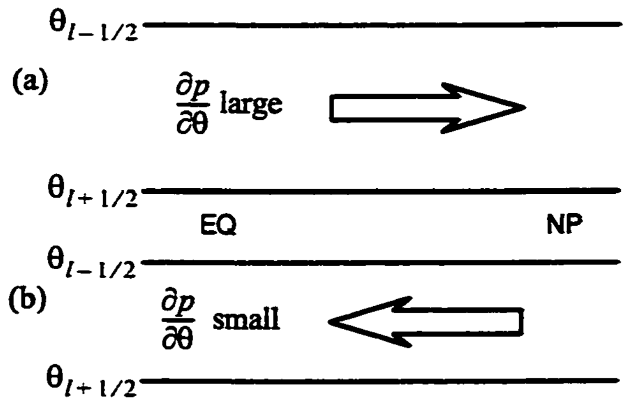


Figure 32. If $[m^*v^*]>0$ between isentropes $\theta_{l-1/2}$ and $\theta_{l+1/2}$ in the northern hemisphere, then along a latitude-circle (a) $\partial p/\partial\theta$ is relatively large for longitudes where $v>0$, and (b) $\partial p/\partial\theta$ is relatively small for longitudes where $v<0$.

pipe, larger mass transport, when the flow is toward the pole, and a smaller pipe, smaller mass transport, when the flow is toward the equator.

We can write the meridional advection of the tracer as

$$
\begin{aligned}
[(mv)\tau] &= [([(mv)]+(mv)^*)([\tau]+\tau^*)] \\
&= [(mv)][\tau]+[(mv)^*\tau^*] \\
&= [m][v][\tau]+[m^*v^*][\tau]+[(mv)^*\tau^*].
\end{aligned}
\tag{114}
$$

The last line in (114) has three terms. The first term is advection by the mean mass flux, the second term is advection by the bolus mass flux, and the third term is a kind of diffusion.

5.4.3 *Model results*

The following results are produced with the pure isentropic-coordinate model run with 10242 cells and 48 vertical layers. It is run with the Held-Suarez forcing discussed in section 5.1. The simulation begins from an isothermal atmosphere at rest and initially contains no tracer. It takes about two years of simulated time to reach a statistically equilibrated distribution of tracer, so we record data beginning at the third year. We record about 130 *snapshots* of the model each separated in time by four days. The terms of (114) are shown in Figure 33. The advection by the mean flow, $\overline{[m][v][\tau]}$, is large in the tropics while the advection by the bolus mass flux, $\overline{[m^*v^*][\tau]}$, is large in middle-latitudes. The

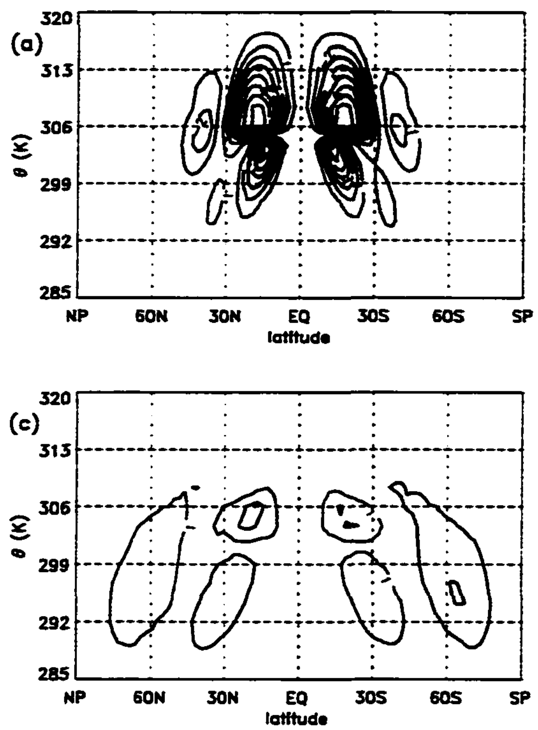


Figure 33. The terms of (114) on θ-surfaces. (a) $\overline{[m][v][\tau]}$, (b) $\overline{[m^*v^*][\tau]}$, (c) $\overline{[(mv)^*\tau^*]}$ and (d) $\overline{[(mv)\tau]} = \overline{[m][v][\tau]} + \overline{[m^*v^*][\tau]} + \overline{[(mv)^*\tau^*]}$. Contour interval 10^4 kg m^{-2} K^{-1}.

third term $\overline{[(mv)^*\tau^*]}$ is very small compared to the first two terms. The sum of all three terms looks like hemispheric cells shown in Figure 30.

If the data are first interpolated to surfaces of constant pressure and then zonally and temporally averaged, a very different view of the mean meridional circulation emerges. In this case, the pseudo-density becomes $m \equiv \partial p/\partial p = 1$, so the deviation of m from its zonally averaged value must be zero, that is, $m^* \equiv 0$. Figure 34 shows the data interpolated to pressure surfaces. So, again advection by the mean flow is large in the

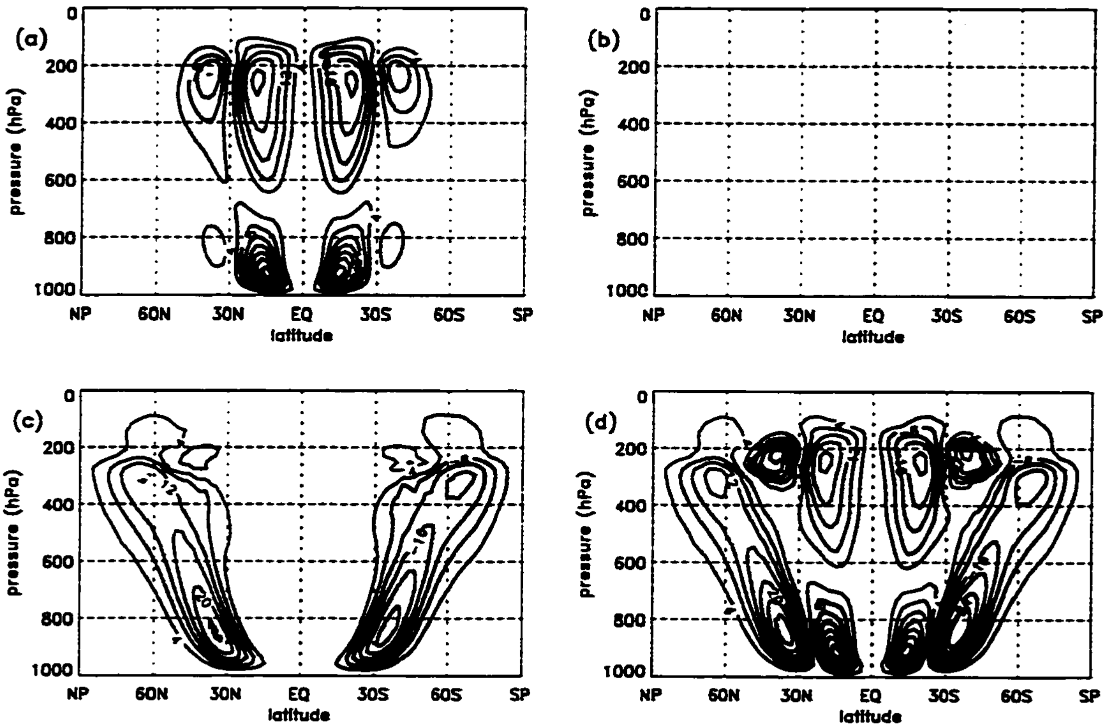


Figure 34. The terms of (114) on p-surfaces. (a) $\overline{[m][v][\tau]}$, (b) $\overline{[m^*v^*][\tau]}$, (c) $\overline{[(mv)^*\tau^*]}$ and (d) $\overline{[(mv)\tau]} = \overline{[m][v][\tau]} + \overline{[m^*v^*][\tau]} + \overline{[(mv)^*\tau^*]}$. Contour interval kg m^{-2} Pa^{-1}.

tropics. However, the advection by the bolus flux is zero. Some other mechanism must take over the advection of the trace towards the pole. This is accomplished by the $\overline{[(mv)^*\tau^*]}$ term. The sum of all three terms gives a circulation in which both a Hadley cell and a Ferrel cell exist in each hemisphere.

5.5 A simulation of stratospheric sudden warming

We have tested the isentropic vertical coordinate model by simulating the sudden warming event of February 1979. Isentropic coordinates are particularly useful for this application. During stratospheric sudden warming events diabatic effects are relatively weak. The sudden warming occurs at very high latitudes. A more traditional latitude/ longitude grid, with its converging meridians, results in grid points becoming very close at high latitudes. On the other hand, the spherical geodesic grid, with its isotropic distribution of grid points, is particularly well suited for this simulation.

5.5.1 Description of the event

During January and February 1979, the northern polar stratosphere experienced three warming events in which temperature maxima occurred on approximately 25 January, 6 February and 26 February. The third event was the most dramatic and caused a reversal of the zonal-mean winds from the climatological westerly direction to an anomalous easterly direction. See Figure 35. This change can be easily understood from thermal wind arguments since, as a result of the warming, temperature actually increased with increasing latitude. This particular event occurred between 17 February and 5 March 1979.

It is customary to divide the sudden warming event into three stages: onset, maturity and recovery stages. The first stage is characterized by increased propagation of waves from the troposphere into the stratosphere and a breakdown of the polar vortex into two cyclonic vortices. In the mature stage, there are two vortices which are separated by a filament of low potential vorticity air extending up from the tropics. And finally, in the recovery stage, the two vortices recombine to again to form one vortex.

We can describe the synoptic structure of the event using 10 hPa geopotential height and temperature fields. See Figure 36. At the beginning of the event, on 17

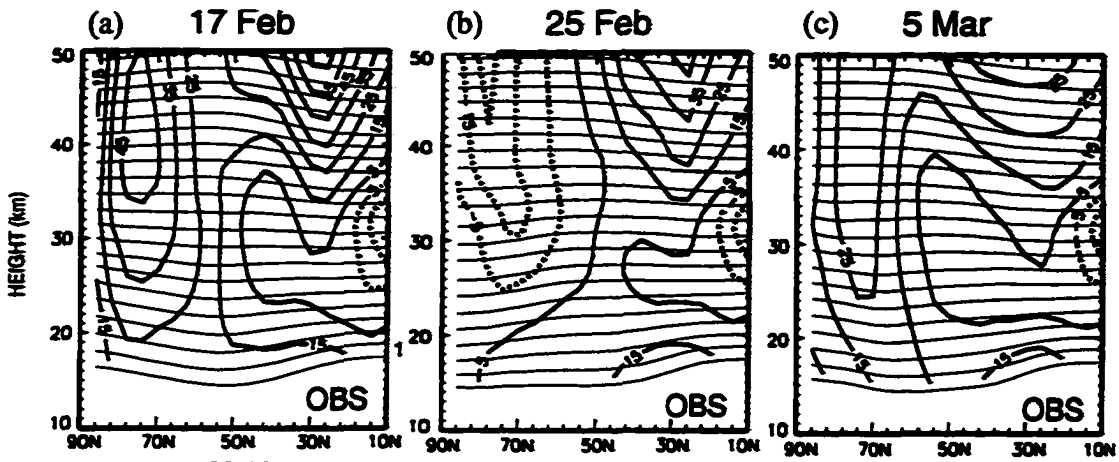
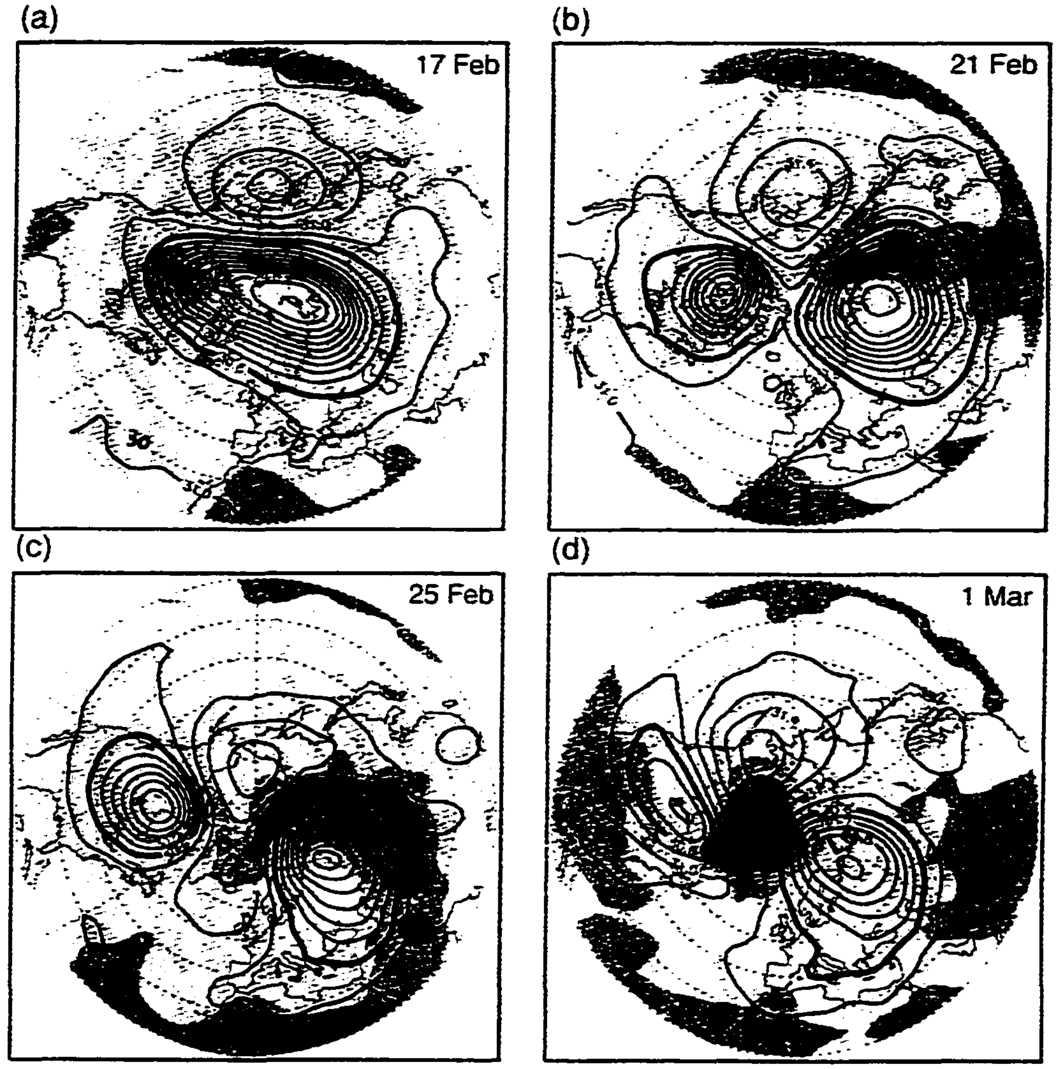
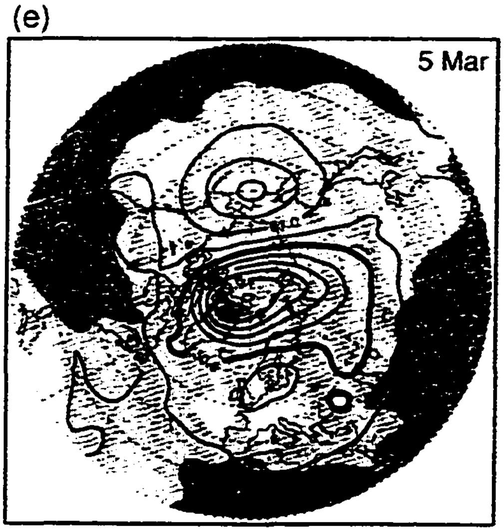


Figure 35. Observed zonal-mean zonal wind on days (a) 17 February, (b) 25 February, and (c) 5 March. Contour interval is 10 m s^{-1}. Note that the vertical axes is height. From Jung *et al.* (2001).

February, there is a large and elongated cyclone centered just off the pole near 60°E which covers the entire polar region. Accompanying the cyclone is a much smaller anticyclone in the region over the Aleutian islands. The coldest air is approximately coincident with the center of the cyclone and is about 210 K. Note that the maximum temperature anyplace in the northern hemisphere is about 240 K. From 17 to 21 February, the polar vortex splits into two cyclones. The stronger of the two is situated over Eurasia and the weaker over North America. A relatively weak anticyclone forms near 10°W over the Atlantic Ocean while the anticyclone over the Aleutians intensifies. From 21 February to 25 February, a pool of very warm air with temperature up to about 255 K forms east of the cyclone over Eurasia, and by 1 March this pool moves northward to cover the entire polar region. Over the next few days the entire pattern rotates slightly toward the west. Eventually, the two anticyclones merge to form a single anticyclone which moves directly over the pole. During this time the cyclones separate further. From 25 February to 1 March, the cyclone over North America weakens while the Eurasia cyclone remain unchanged. After 1 March, the Eurasian cyclone moves poleward where it is joined by the remains of the North American cyclone, and temperature over the polar region return to normal.





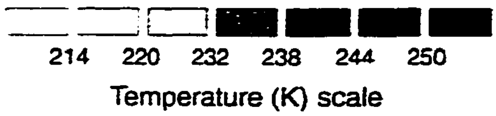


Figure 36. The observed (NCEP/NCAR reanalysis) geopotential height (contours) and temperature (gray scale) on the 10 hPa surface for days from 17 February 1979 to 5 March 1979. The contour interval is 0.2 km. The thicker contour is the 30.6 km geopotential height. From Jung *et al.* (2001).

5.5.2 Description of the simulation

The initial condition are derived from an amalgamation of two data sets. Below 10 hPa, the model is initialized with NCEP/NCAR reanalysis of temperature, geopotential height and wind. Above 10 hPa, the temperature and geopotential height are obtained from National Meteorological Center (presently NCEP) data, and the winds are diagnosed from the geostrophic relation. These data were provided by Jung and Konor on a 4° latitude × 5° longitude with momentum in terms of zonal and meridional wind. For the geodesic grid, they were numerically differentiated to obtain vorticity and divergence and linearly interpolated to the geodesic grid points.

The model domain covers the middle atmosphere above 400 K. Hence, 400 K is the value of the lowest coordinate surface, and the boundary condition at the bottom edge is included by specifying the Montgomery potential on the 400 K surface.

As the flow evolves, it is forced by a changing Montgomery potential at the lower boundary. Recall the form of the hydrostatic equation in isentropic coordinates, $\partial M/\partial\theta = \Pi$. Integration of this equation gives the Montgomery potential within each model layer. Numerically this is written

$$M_l = M_{l+1} + (\theta_l - \theta_{l+1})\Pi_{l+1/2} \quad \text{for } l = 1, 2, \ldots, L-2 \tag{115}$$

and

$$M_L = M_{l+1/2} + (\theta_L - \theta_{L+1/2})\Pi_{L+1/2} \tag{116}$$

The tendency $\partial M_{L+1/2}/\partial t$ is provided at the beginning of each simulated day, and $M_{L+1/2}$ is computed at the beginning of each timestep by linearly extrapolating the tendency forward in time.

Upward propagating waves can reflect off the upper boundary. The real atmosphere does not, of course, have an upper lid to case reflections. To reduce the effect of these reflected waves, the model has a very small Rayleigh friction in the top most layer with a coefficient of $2.8 \times 10^{-6}\ \mathrm{s}^{-1}$.

Waves from the lower boundary break the polar vortex into two vortices. We need a mechanism to restore it to one vortex. Radiative processes approximated with a simple Newtonian type heating which attempts to restore the initial zonally symmetric state. We specify a vertical mass flux as follows

$$(m\dot{\theta})_{l+1/2} = -\gamma(z_{l+1/2})(p_{l+1/2} - p^*_{l+1/2}) \quad \text{for } l = 1, 2, \ldots, L-1 \tag{117}$$

where

$$\gamma(z) = (10^{-6}\ \mathrm{s}^{-1})\left[1.5 + \tanh\left(\frac{z-35}{7}\right)\right] \tag{118}$$

where z is in kilometers, and p^* is the zonally averaged initial distribution of pressure. A plot of $\gamma(z)$ is shown in Figure 37. The form of (118) is from Holton (1976).

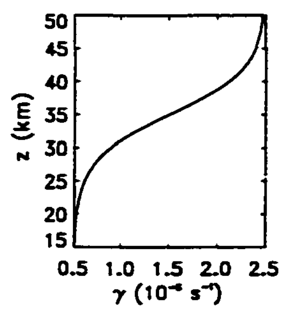


Figure 37. Relaxation coefficient γ. The time scale is about 22 days a $z = 15$ km and 5 days at $z = 50$ km, so the restoring force of the heating is felt more strongly at the top of the model. It is here that the westerlies recover first.

5.5.3 Model results

The following results were run with 10242 cells and 23 layers. The lowest coordinate surface is $\theta = 400$ K, and the highest coordinate surface is $\theta = 3163$ K. The remaining surfaces are distributed evenly in height

Figure 38 shows the simulated geopotential height on the 10 hPa surface. Comparing Figure 36 and Figure 38 we can see that the simulation captures many of the important feature of the warming event. By day 4 of the simulation, the polar vortex has split into the Eurasian and North American cyclones. While the location of both is simulated well, their intensity is slightly weak. At the same time, the anticyclone over the Aleutians is slightly strong. By day 8, the air east of Eurasian cyclone has warmed considerably. The region of the air in excess of 250 K is a little bigger than in the corresponding observation. The Aleutian anticyclone, which is still a little strong, has moved toward the pole. By day 12, the temperature maximum has been advected over the pole, and the North American cyclone has weakened considerably. By day 16, the Eurasian cyclone has combined with the remains of the North American cyclone and moved back over the pole. The resulting vortex is weaker than in observations.

Figure 39 shows the simulated, zonally-averaged zonal wind. We can see that the initial climatological westerlies give way to easterlies by day 4. Note that these data are plotted with potential temperature on the vertical axes. By day 8, the maximum easterlies occur. The westerlies do not recover as fully as in the observation. This is consistent with the weaker polar vortex.

Figure 40 shows the Rossby-Ertel potential vorticity on the 831 K surface. From the beginning of the simulation until day 4, we can clearly see the splitting of the polar vortex. Beginning on day 4, we can see a tongue of low potential vorticity air advected from low-latitudes, around the eastern edge of the Eurasian cyclone and into the polar regions. This is a characteristic of the mature stage of the sudden warming event. The low-PV air is approximately coincident with the region of significant warming. This indicated that the warming is associated with descending air.

The isentropic model is particularly well suited for stratospheric simulation since the difficulty with isentropic intersecting the is nonexistent. This model demonstrates that it can simulate all three stages of the sudden warming event.

5.6 Baroclinic Instability

The performance of the isentropic and generalized vertical coordinate model is tested by simulating the nonlinear evolution of mid-latitude disturbances on the sphere. The initial conditions consist of a zonally symmetric jet centered at 45° latitude and 250 hPa. It has a maximum velocity of 50 ms^{-1}. There are mirror imagine jets in both the northern and southern hemispheres. There are 23 model layers, and the values of ζ on each coordinate

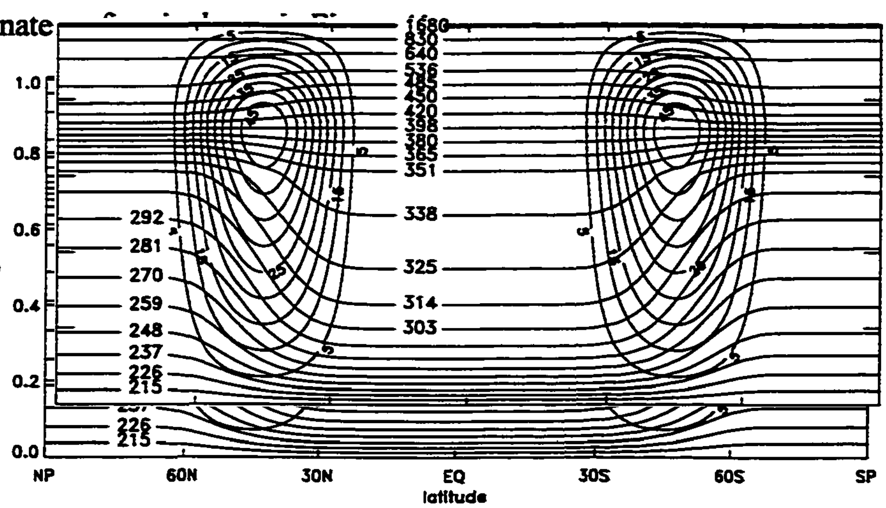


Figure 41. Distribution of ζ-surfaces in σ-space and isotachs of the initial jet (m s^{-1})

In initial jet can be written as the product two functions

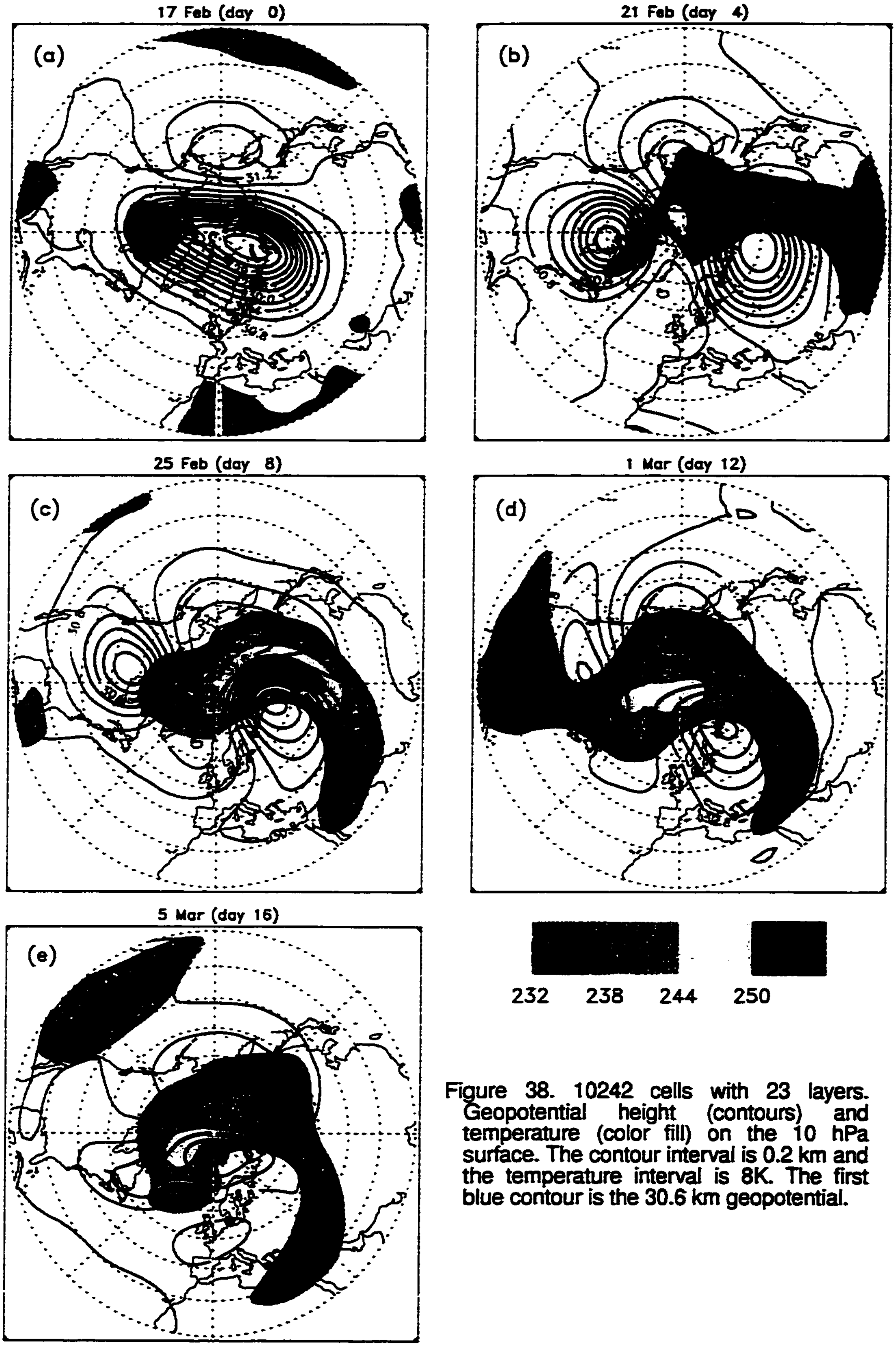
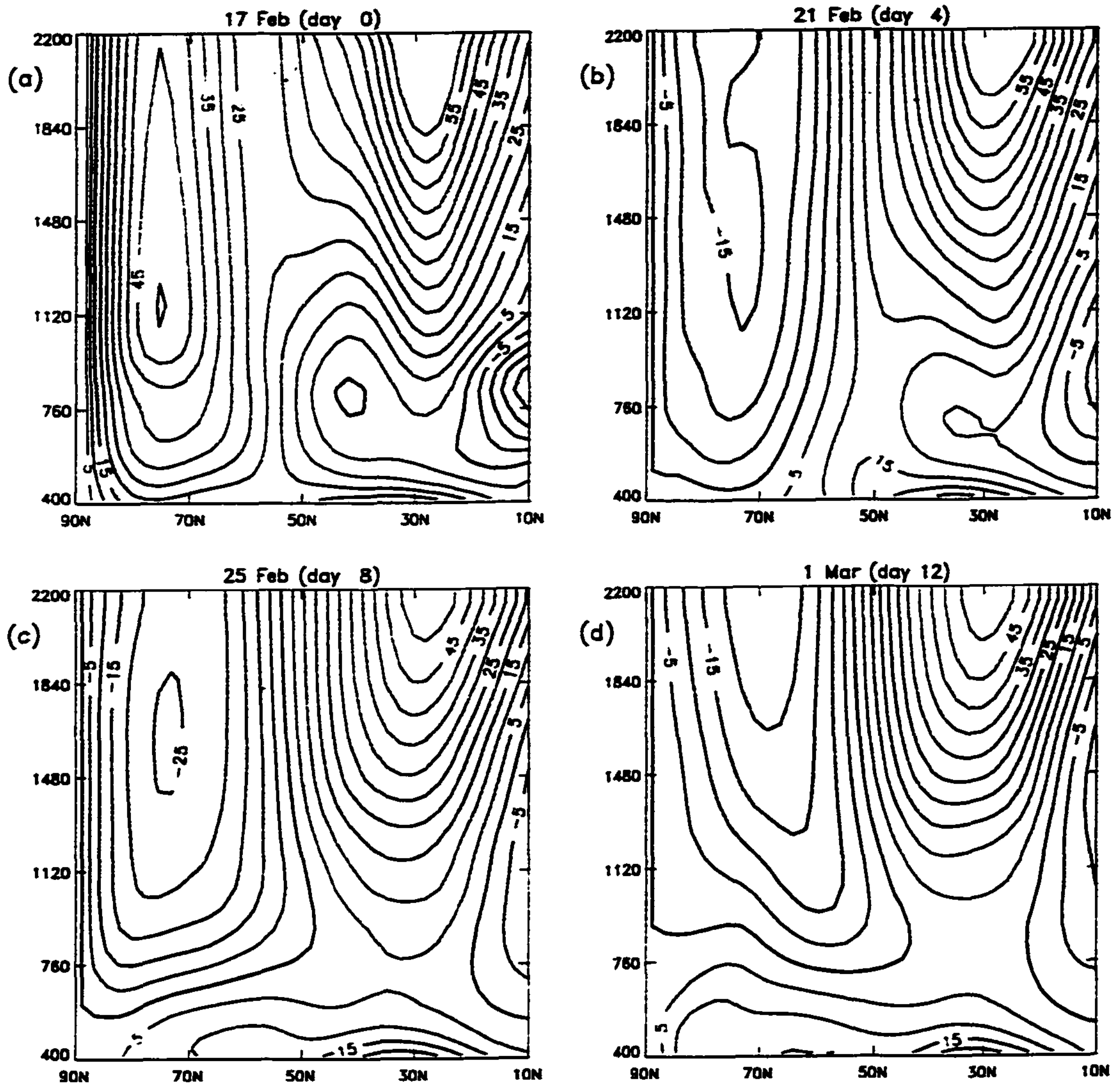


Figure 38. 10242 cells with 23 layers. Geopotential height (contours) and temperature (color fill) on the 10 hPa surface. The contour interval is 0.2 km and the temperature interval is 8K. The first blue contour is the 30.6 km geopotential.



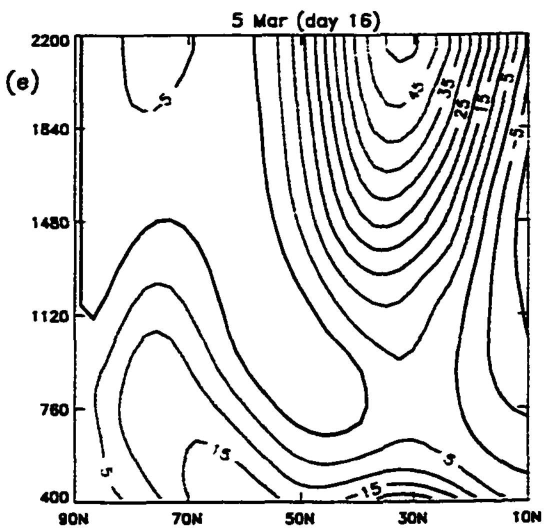


Figure 39. 10242 cells with 23 layers. Zonally averaged zonal wind (m s^{-1}). Contour interval is 5 m s^{-1}.

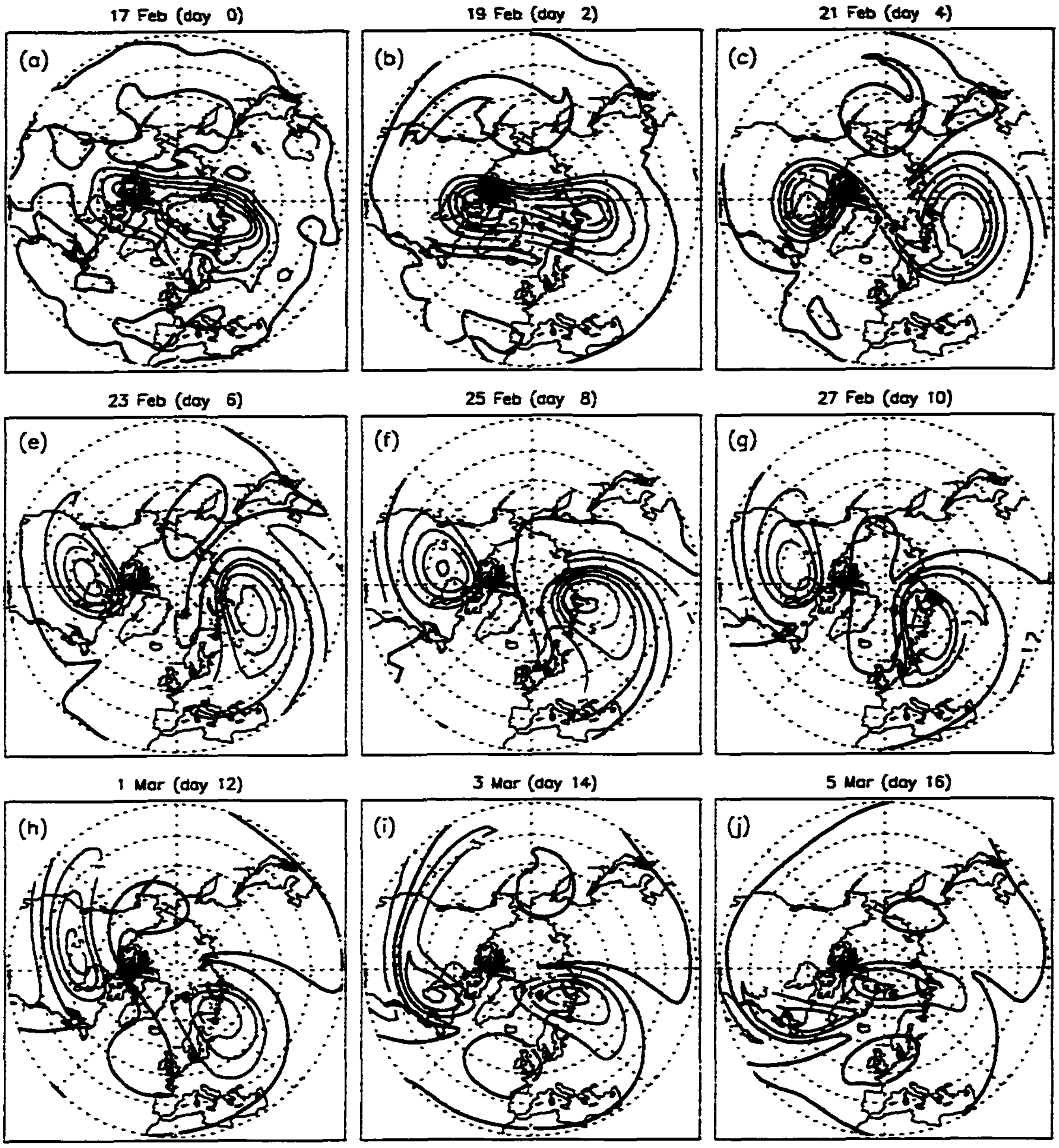


Figure 40. Rossby-Ertel potential vorticity on the 831 K isentropic surface. The contour interval is 50 potential vorticity units (PVU = 10^{-6} m^2 kg^{-1} s^{-1}) The blue contour is 200 PVU.

$$u(\varphi, p) = u_1(\varphi)u_2(p). \tag{119}$$

The meridional part is given by

$$u_1(\varphi) = \begin{cases} \left[\sin\left(\pi\dfrac{\varphi-\varphi_{min}}{\varphi_{max}-\varphi_{min}}\right)\right]^2 & \varphi_{min} < \varphi \le \varphi_{max} \\ 0 & \text{otherwise} \end{cases}, \tag{120}$$

where $\varphi_{\min} = 20°$ and $\varphi_{\max} = 70°$. The vertical part

$$u_2(p) = u_{\max}\begin{cases} \dfrac{p-p_T}{p_{\max}-p_T} & p_T \le p \le p_{\max} \\ \dfrac{p_S-p}{p_S-p_{\max}} & p_{\max} < p \le p_S \end{cases}, \tag{121}$$

where $u_{\max} = 50\ \mathrm{ms}^{-1}$, $p_T = 10\ \mathrm{hPa}$, $p_{\max} = 250\ \mathrm{hPa}$ and $p_S = 1000\ \mathrm{hPa}$. The corresponding potential temperature field is determined using the geostrophic wind relation and the hydrostatic equation. The initial conditions are discussed in more detail in Appendix N.

A slight perturbation in surface pressure is given at the lowest model edge. This perturbation provides a seed from which the instability can grow. The initial pressure of the lowest edge is given by

$$p_{L+1/2}(\lambda, \varphi) = 1000 + 0.5\sin(6\lambda)\sin^2(2\varphi)\ \mathrm{hPa} \tag{122}$$

Clearly, the perturbation will initiate a disturbance with a wavenumber 6 structure.

A gentle Newtonian heating is provided to the model to restore the thermal field to the initial zonally symmetric initial state. The heating has the form

$$Q = -\gamma c_p\left[(T - T^*) - \overline{\left(\frac{\partial T}{\partial p}\right)}(p - p^*)\right] \tag{123}$$

where p^* and T^* are the initial distributions of pressure and temperature. The coefficient in (123) is set to $\gamma = 0.1\ \mathrm{day}^{-1}$ at all layer edges except at the earth's surface. At the surface, we set $\gamma = 1.0\ \mathrm{day}^{-1}$ to approximate the effect of a surface heat flux. The value of $\overline{(\partial T/\partial p)}$ is determined from N^2 on the initial atmosphere. At pressure greater than

400 hPa, $\overline{(\partial T/\partial p)} = 1.5\times10^{-4}s^{-2}$, and at pressure less than 400 hPa, $\overline{(\partial T/\partial p)} = 3.5\times10^{-4}s^{-2}$.

We approximate surface friction with a Rayleigh damping of momentum for the layers at pressure greater than 0.9 p_S. The friction force is determined for each model layer and each time step by

$$F_l = k_{\mathbf{V}}\min\left(1, \frac{p_{l+1/2}-p_B}{p_{l+1/2}-p_{1/2}}\right)\frac{\max(0, p_{l+1/2}-p_B)}{p_{l+1/2}-p_B}\mathbf{V}_l \tag{124}$$

where $k_{\mathbf{V}} = 1$ day^{-1} and $p_B = 0.9\ p_S$.

5.6.1 Simulation results

The surface temperature is shown in Figure 42, and the surface pressure is shown in Figure 43. By day 8 of the simulation we can see that baroclinic waves are beginning develop along the meridional temperature gradient. The wavenumber 6 pattern is evident, and, as the model develops, six more-or-less equivalent weather systems evolve. By day 11 and day 12 we can see many of the features common to observed mature systems. There is a strong temperature gradient along the cold front region that follows a tongue of low pressure extending out of the cyclone. The cold front then stretches around the anticyclone along the trough line, and around the low pressure region forming a bent-back baroclinic zone. We can see that a small bubble of warm air has been advected up from low-latitudes and is trapped in the center of the low. This process is known as seclusion.

Thorncroft *et al.* describes two types of baroclinic life-cycles. Which of these two life-cycles the model demonstrates is determined by subtle differences in the initial conditions. The first type of life-cycle is triggered by the initial condition shown in this simulation. In this case the initial surface winds are zero. The second type of life-cycle has a slight barotropic enhancement with surface westerlies at latitudes below the jet maximum, and surface easterlies at latitudes above the jet maximum. The small cyclonic shear in the initial condition causes a very different evolution of the model.

For future work, it would be interesting run the generalized vertical coordinate model with the initial condition to produce the second type of life-cycle. For time being, we can see that the model produces a good simulation.

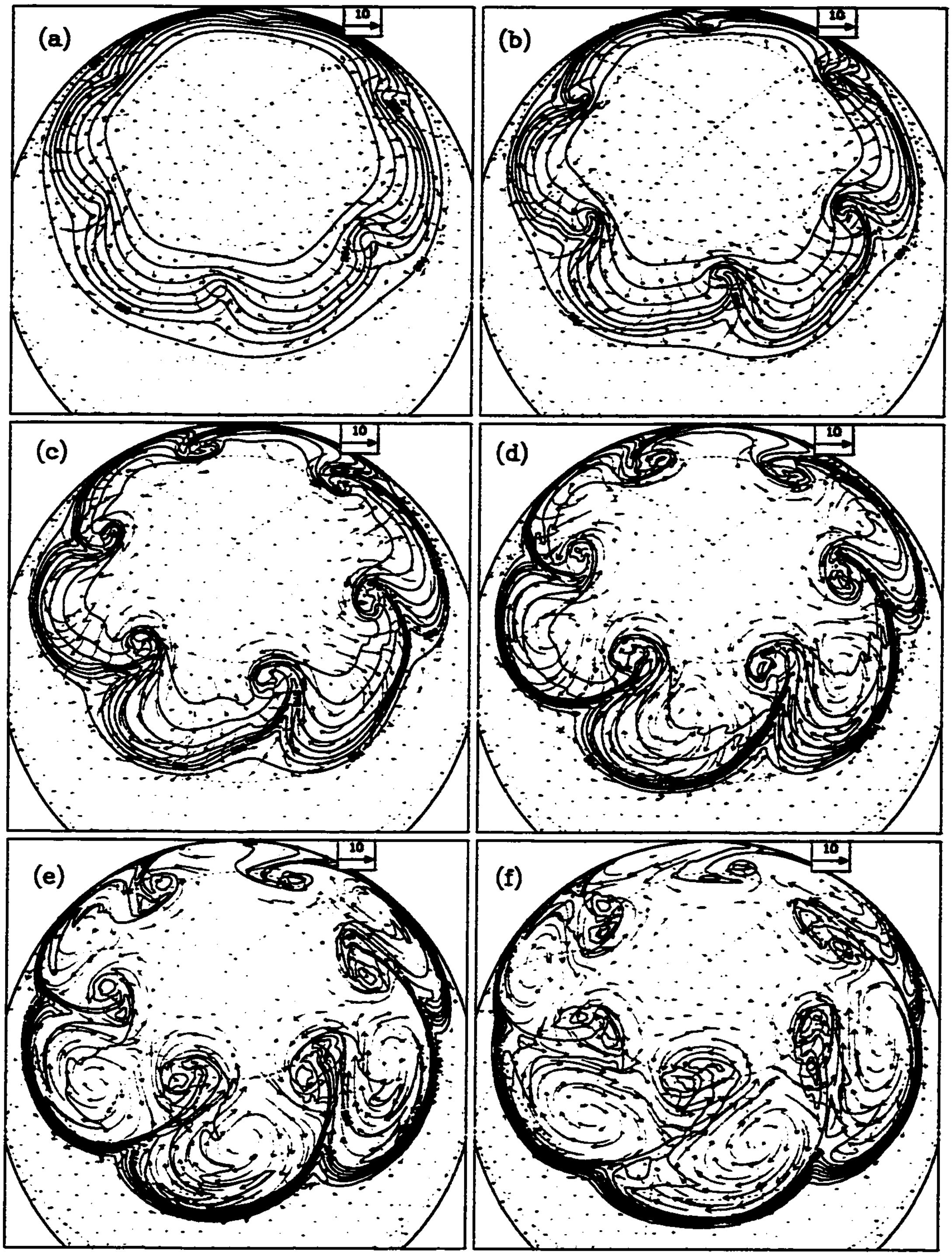


Figure 42. Surface temperature and winds. Temperature contour interval is 8 K. Wind speed indicated in figure. (a) day 8, (b) day 9, (c) day 10, (d) day 11, (e) day 12, (f) day 13.

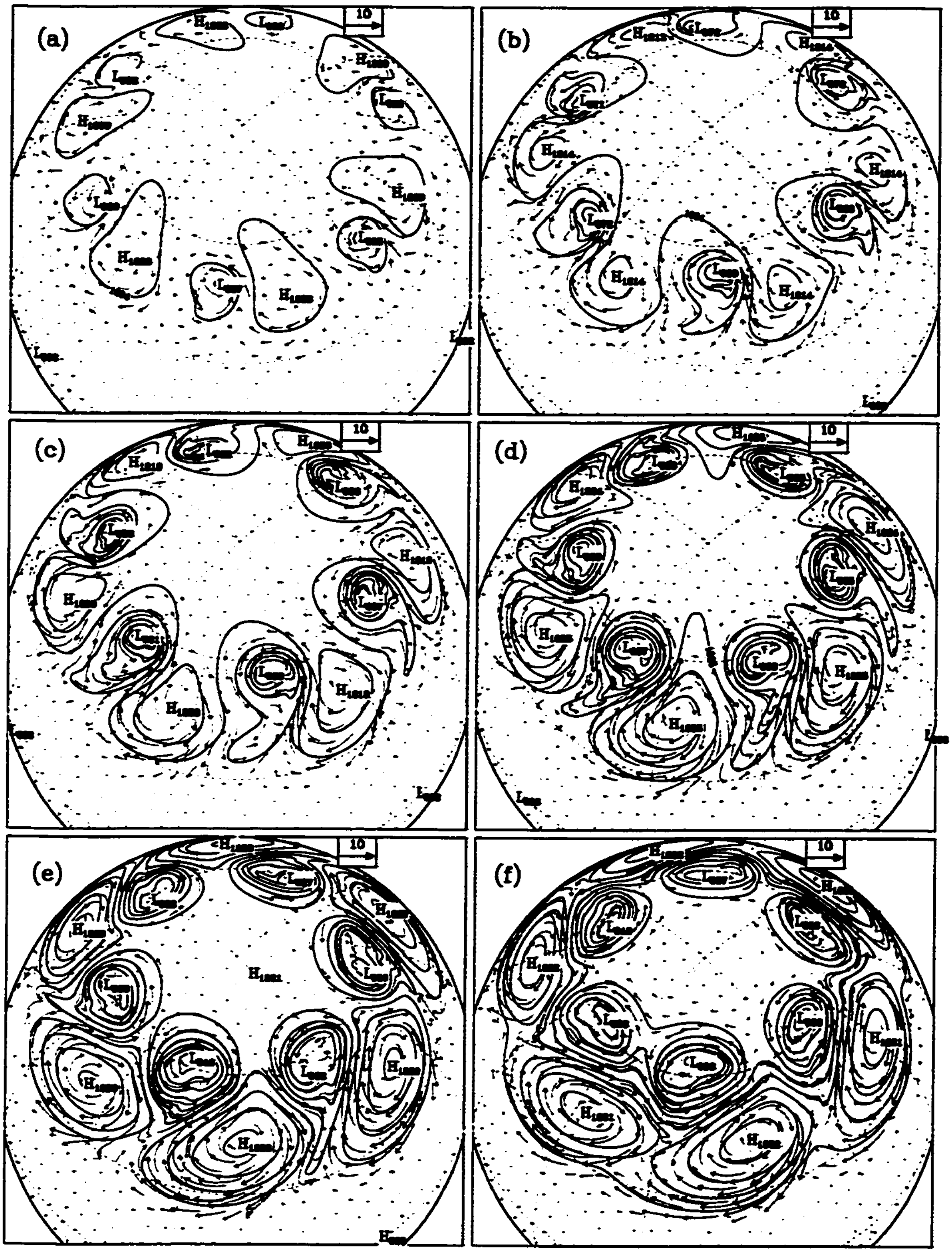


Figure 43. Surface pressure and winds. Pressure contour interval is 8 hPa. Wind speed indicated in figure. (a) day 8, (b) day 9, (c) day 10, (d) day 11, (e) day 12, (f) day 13.

6. Summary and discussion

The idea of the geodesic grid was first introduced in the late 1960's and after a period of dormancy is currently enjoying a significant revival within several modeling efforts including the CSU GCM, the Deutscher Wetterdienst, and the Japanese Frontier Project. In the same way, the isentropic coordinate was introduced in the late 1960's and after some 40 years of quiescence is now experiencing a renaissance through its coupling with sigma-coordinates. An important conclusion of this thesis is that these ideas, the geodesic grid and the isentropic/hybrid-coordinate, can be combined and work well together.

In section 2, we briefly reviewed the historical development of vertical coordinates from pressure to isentropic coordinates and finally to hybrid coordinates. In section 3, we focused on three of the models that were touched upon in section 2. We showed how the continuous equations are transformed to their discrete analogs so that the discrete equations maintain many properties of the continuous equations. These properties include conservation of mass and total energy.

These new vertical coordinates represent a fairly radical departure from the more common Lorenz-grid sigma-coordinate, and some basic questions about their use need to be addressed. The numerical results presented in section 5 are intended to answer a few of these questions. Each of the three models in section 3 are based on a Charney-Phillips staggering. In section 5.1, we showed that there is less noise generated with the Charney-Phillips staggering than with the Lorenz staggering. The amount of $2\Delta z$ noise within a model column is about twice as big with the Lorenz grid as with the Charney-Phillips grid.

Since the isentropic-coordinate surfaces behave more as Lagrangian surfaces, they are more free to follow and conform to the fluid flow than sigma-coordinate surfaces. For this reason, our ideas of model resolution for an Eulerian framework may need to be modified when we move into the realm of isentropic- and hybrid-coordinates. In section 5.2, we attempted to address very basic issues of simulation quality as a function of resolution. We attempted to show that the distribution of potential temperature within the model atmosphere, and hence the distribution of θ-coordinate surfaces, will have a significant effect on the simulation especially in regions of weak stratification. The preliminary results indicate that the hybrid-coordinate may require an increased vertical resolution relative to sigma-coordinate models. This is counter to previously held views that fewer coordinate surfaces would suffice in isentropic-coordinate models.

The use of isentropic-coordinates can be viewed a research tool to better understand the general circulation of the atmosphere. In section 5.3, we showed that the general circulation is very different viewed in isentropic coordinates than viewed in pressure coordinates. And, in isentropic-coordinates, the poleward transport of heat is simpler and more straightforward than in sigma-coordinates.

One region of the atmosphere in which isentropic coordinates are particularly useful is the stratosphere since the stratosphere is adiabatic and there are no difficulties associated coordinate surface intersections. In section 5.4, we showed that the isentropic model could simulate well a stratospheric sudden warming event. The isotropic distribution of cells in the geodesic grid is very important for this polar simulation. And, finally in section 5.5, we show the hybrid-coordinate's skill in simulating the process of baroclinic instability.

6.1 Future work

Switching from an L-grid to a CP-grid in the CSU GCM will be complicated task since the model parameterization have all been designed assuming temperature is defined at layer centers. Changing to the CP-grid will require a significant retooling of the entire model. Some small steps have been taken. Arakawa and Konor (1996) describe a scheme to define mass at layer edges which is consistent with mass defined at layer centers. We used these ideas section 5.3 to define a passive tracer in the isentropic model. This same approach can be used to define tracers like water vapor in the GCM

Parameterization in the current GCM, such as radiation and the cumulus parameterizations, expect their input to be defined at model layer centers, and they expect their input to be more-or-less evenly spaced in pressure. Initially we can *trick* the current parameterizations by feeding them input from layer edges and letting them falsely believe that the input came from layer centers. A more serious issue is the uneven spacing of input in pressure. We have seen that the layer thickness within a model column in the generalized vertical coordinate model can vary dramatically. The current parameterizations are not designed to handle very thick model layer next to very thin model layers.

References

Arakawa, A., and V. R. Lamb, 1977: Computational design of the basic dynamic processes of the UCLA general circulation model. *Methods in Computational Physics,* General circulation models of the atmosphere, **17,** J Chang, Ed., Academic Press, 173-265.

Arakawa, A., and M. J. Suarez, 1983: Vertical differencing of the primitive equations in sigma coordinates. *Mon. Wea. Rev.*, **111,** 34-45.

Arakawa, A., and S. Moorthi, 1988: Baroclinic instability in vertically discrete systems. *J. Atmos. Sci.*, **45,** 1688-1707.

Arakawa, A., and C. S Konor, 1996: Vertical differencing of the primitive equations based on a Charney-Phillips grid in hybrid sigma-theta vertical coordinates. *Mon. Wea. Rev.*, **124,** 511-528.

Arakawa, A., C. R. Mechoso and C. S. Konor, 1992: An isentropic vertical coordinate model: Design and application to atmospheric frontogenesis studies. *Meteorol. Atmos. Phys.*, **50,** 31-45.

Bleck, R., 1974: Short-range prediction in isentropic coordinates with filtered and unfiltered numerical models. *Mon. Wea. Rev.*, **102,** 813-829.

Bleck, R., 1978: On the use of hybrid vertical coordinates in numerical weather prediction models. *Mon. Wea. Rev.*, **106,** 1233-1244.

Bleck, R., 1978: Finite difference equations in generalized vertical coordinates. Part I: Total energy conservation. *Beitr. Phys. Atmos.*, **51,** 360-372.

Bleck, R., and D. B. Boudra, 1981: Initial testing of a numerical ocean circulation model using a hybrid (quasi-isopycnic) vertical coordinate. *J. Phys. Oceanogr.*, **11**, 755-770.

Bleck, R., 1984: An isentropic coordinate model suitable for lee cyclogenesis simulation. *Riv. Meteorol. Aeronaut.*, **43**, 189-194.

Bleck, R., and S Benjamin, 1993: Regional weather prediction with a model combining terrain-following and isentropic coordinates. Part I. *Mon. Wea. Rev.*, **121**, 1770-1785.

Charney, J. G., and N. A. Phillips, 1953: Numerical integration of the quasi-geostrophic equations for barotropic and simple baroclinic flows. *J. Meteor.*, **10**, 71-99.

Charney, J. G., and P. G. Drazin, 1961: Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *J. Geophy. Res.*, **55**, 83-109.

Deaven, D. G., 1976: A solution for boundary problems in isentropic coordinate models. *J. Atmos. Sci.*, **33**, 1702-1713.

Eliassen, A., and E Raustein, 1968: A numerical integration experiment with a model atmosphere based on isentropic surfaces. *Meteor. Ann.*, **5(2)**, 45-63.

Friend, A. L., D. Djuric and K. C. Brundidge, 1977: A combination of isentropic and sigma coordinates in numerical weather prediction. *Beitr. Phys. Atmos.*, **50**, 290-295.

Gary, J. M., 1973: Estimate of truncation error in transformed coordinate, primitive equation atmospheric model. *J. Atmos. Sci.*, **30**, 223-233.

Heikes, R. P., and D. A. Randall, 1995: Numerical integration of the shallow water equations on a twisted icosahedral grid. Part I: Basic design and results of tests. *Mon. Wea. Rev.*, **123**, 1862-1880.

Heikes, R. P., and D. A. Randall, 1995: Numerical integration of the shallow water equations on a twisted icosahedral grid. Part II: Grid refinement, accuracy and computational performance. *Mon. Wea. Rev.*,**123**, 1881-1887.

Held, I. M., and J. M. Suarez, 1994: A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Amer. Meteor. Soc.*, **75,** 1825-1830.

Holton, J. R., 1976: A semi-spectral numerical model for wave-mean flow interactions in the stratosphere: Application to sudden stratospheric warmings. *J. Atmos. Sci.*, **33,** 1639-1649.

Hsu, Y.-J., and A. Arakawa, 1990: Numerical modeling of the atmosphere with an isentropic vertical coordinate. *Mon. Wea. Rev.*, **118**, 1933-1959.

Johnson, D. R., 1989: The forcing and maintenance of global monsoonal circulations: An isentropic analysis. *Advances in Geophysics*, Vol. 31, Academic Press, 43-316.

Jung, J.-H., C. S. Konor, C. R. Mechoso, and A. Arakawa, 2001: A study of the stratospheric major warming and subsequent flow recovery during the winter of 1979 with an isentropic vertical coordinate model. *J. Atmos. Sci.*, **58,** 2630-2649.

Kasahara, A., and W. M. Washington, 1971: NCAR global general circulation model, including orography, cloudiness and surface temperature calculations. *J. Atmos. Sci.*, **28,** 657-701.

Kasahara, A., 1974: Various vertical coordinate systems used for numerical weather prediction. *Mon. Wea. Rev.*, **102,** 509-522.

Konor, C. S., and A. Arakawa, 1997: Design of an atmospheric model based on a generalized vertical coordinate. *Mon. Wea. Rev.*, **125**, 1649-1673.

Konor, C. S., A. Arakawa, 2000: Choice of a vertical grid in incorporating condensation heating into an isentropic vertical coordinate model. *Mon. Wea. Rev.*, **128**, 3901-3910.

Leith, C. E., 1965: Numerical simulation of the earth's atmosphere. In *Methods in Computational Physics*, **4**, Academic Press, New York, 1-28

Lorenz, E. N., 1955: Available potential energy and the maintenance of the general circulation. *Tellus*, **7**, 155-167.

Lorenz, E. N., 1960: Energy and numerical weather prediction. *Tellus*, **12**, 364-373.

Masuda, Y., and H. Ohnishi, 1986: An integration scheme of the primitive equation model with an icosahedral-hexagonal grid system and its application to the shallow-water equations. *Short- and Medium-Range Numerical Weather Prediction*. Japan Meteorological Society, Tokyo, 317-326.

Phillips, N. A., 1957: A coordinate system having some special advantages for numerical forecasting. *J. Meteor.*, **14**, 184-185.

Randall, D. A., 1994: Geostrophic adjustment and the finite-difference shallow-water equations. *Mon. Wea. Rev.*, **122**, 1371-1377.

Randall, D. A., R. Heikes, and T. D. Ringler, 2000: Global atmospheric modeling using a geodesic grid with an isentropic vertical coordinate. *General Circulation Model Development*, D. A. Randall, Ed., Academic Press, 509-538.

Ringler, T. D., and D. A. Randall, 2002a: The ZM Grid: An alternative to the Z Grid. *Mon. Wea. Rev.*, **130**, 1411-1422.

Ringler, T. D., and D. A. Randall, 2002b: A potential enstrophy and energy conserving numerical scheme for solution of the shallow-water equations on a geodesic grid. *Mon. Wea. Rev.*, **130,** 1397-1410.

Richardson, L. F., 1922: *Weather Prediction by Numerical Process*. New York, Cambridge University Press, xii+236 pp. Reprint with a new introduction by Sydney Chapman, Dover Publications, New York, xvi+236 pp. 1965.

Sadourny, R., A. Arakawa, and Y. Mintz, 1968: Integration of the nondivergent barotropic vorticity equation with an icosahedral-hexagonal grid for the sphere. *Mon. Wea. Rev.*, **96,** 351-356.

Shapiro, M. A., 1975: Simulation of upper-level frontogenesis with a 20-level isentropic coordinate primitive equation model. *Mon. Wea. Rev.*, **103,** 591-654.

Simmons, A. J., and D. M. Burridge, 1981: An energy and angular-momentum conserving vertical finite-difference scheme and hybrid vertical coordinates. *Mon. Wea. Rev.*, **109,** 758-766

Starr, V. P., 1945: A quasi-Lagrangian system of hydrodynamic equations. *J. Meteor.*, **2,** 227-237.

Sutcliffe, R. C., 1947: A contribution to the problem of development. *Quart. J. Roy. Meteor. Soc.*, **73,** 370-383.

Thorncroft, C. D., B. J. Hoskins, and M. E. McIntyre, 1993: Two paradigms of baroclinic-wave life-cycle behaviour. *Quart. J. Roy. Meteor. Soc.*, **119,** 17-55.

Townsend, R. D., and D. R. Johnson, 1985: A diagnostic study of the isentropic zonally averaged mass circulation during the first GARP global experiment. *J. Atmos. Sci.*, **42,** 1565-1579.

Uccellini, L. W., D. R. Johnson, and R. E. Schlesinger, 1979: An isentropic and sigma coordinate hybrid numerical model: Model development and some initial tests. *J. Atmos. Sci.*, **36**, 390-414.

Williamson, D. L., 1968: Integration of the barotropic model over a spherical geodesic grid. *Tellus*, **20**, 642-653.

Zalesak, S. T., 1979: Fully multidimensional flux-corrected transport algorithms for fluids. *J. Comput. Phys.* **31**, 335-350.

Zapotocny, T. H., D. R. Johnson, F. M. Reames, 1994: Development and initial test of the University of Wisconsin global isentropic-sigma model. *Mon. Wea. Rev.*, **122**, 2160-2178.

Zhu, Z., J. Thuburn, B. J. Hoskins, and P. H. Haynes, 1992: A vertical finite-difference scheme based on a hybrid σ-θ-p coordinate. *Mon. Wea. Rev.*, **120**, 851-862.

Appendix A: Potential temperature and entropy

The first law of thermodynamics is given by

$$c_v\frac{DT}{Dt}+p\frac{D\alpha}{Dt} = Q, \tag{125}$$

where $c_v = 717\ \mathrm{J\,kg^{-1}\,K^{-1}}$ is the specific heat at constant volume, α is specific volume and Q is the heating rate per unit mass. Differentiating the equation of state, $p\alpha = RT$, gives

$$p\frac{D\alpha}{Dt}+\alpha\frac{Dp}{Dt} = R\frac{DT}{Dt}. \tag{126}$$

Combining (125) and (126) to eliminate density gives

$$c_p\frac{DT}{Dt}-\alpha\frac{Dp}{Dt} = Q, \tag{127}$$

or to eliminate temperature gives

$$\frac{D\rho}{Dt}-\frac{1}{\gamma RT}\frac{Dp}{Dt} = -\frac{\rho Q}{c_pT}, \tag{128}$$

where $c_p = c_v+R = 1004\ \mathrm{J\,kg^{-1}\,K^{-1}}$ is the specific heat at constant pressure and $\gamma = c_p/c_v$. Dividing (127) by T gives the entropy form of the first law of thermodynamics

$$c_p\frac{D\ln T}{Dt}-R\frac{D\ln p}{Dt} = \frac{Q}{T} = \frac{Ds}{Dt}. \tag{129}$$

The last equality in (129) defines the entropy per unit mass denoted by s.

For an adiabatic process the heating is zero. Moving a parcel from an initial temperature T and pressure p to a standard pressure $p_0 = 1000\ \mathrm{hPa}$ will result in a temperature change which can be expressed by integrating (129), that is,

$$c_p \int_T^{\theta} \frac{1}{T} dT = R \int_p^{p_0} \frac{1}{p} dp . \tag{130}$$

Thus, the final temperature, after the parcel has been moved to p_0, is given by

$$\theta = T\left(\frac{p_0}{p}\right)^{\kappa} . \tag{131}$$

Taking the logarithm and derivative of (131) and substituting from (129) gives

$$c_p \frac{D \ln \theta}{Dt} = \frac{Ds}{Dt} . \tag{132}$$

So, as a parcel moves conserving its potential temperature it must also conserve its entropy.

Appendix B: How to get from one vertical coordinate to another

Following Kasahara (1974) let the z_1-system be the coordinate system with x, y, z_1 and t as independent variables and z_1 be the variable in the vertical direction. For the time being, we can think of z_1 as geometric height. Similarly, let the z_2-system be an alternative coordinate system with x, y, z_2 and t as independent variables. Suppose that for each fixed x, y and t, z_2 is a strictly monotonic function of z_1, that is,

$$z_2 \equiv z_2(x, y, z_1, t). \tag{133}$$

Because of the monotonicity, the relation in (133) can be inverted to give

$$z_1 \equiv z_1(x, y, z_2, t). \tag{134}$$

Suppose α is a four-dimensional field -- three spacial dimensions and time. If we write $\alpha \equiv \alpha(x, y, z_1(x, y, z_2, t), t)$ Then, the partial derivatives with respect to t, x and y in the z_2 coordinate system can be written as follows,

$$\left(\frac{\partial\alpha}{\partial t}\right)_{z_2} = \left(\frac{\partial\alpha}{\partial t}\right)_{z_1} + \left(\frac{\partial z_1}{\partial t}\right)_{z_2}\frac{\partial z_2}{\partial z_1}\frac{\partial\alpha}{\partial z_2}, \tag{135}$$

and

$$\nabla_{z_2}\alpha = \nabla_{z_1}\alpha + (\nabla_{z_2}z_1)\frac{\partial z_2}{\partial z_1}\frac{\partial\alpha}{\partial z_2}. \tag{136}$$

Using (135) and (136), we can transform the total derivative from the z_1 to the z_2 coordinate is as follows

$$\frac{D}{Dt} = \left(\frac{\partial}{\partial t}\right)_{z_1} + \mathbf{V} \bullet \nabla_{z_1} + \dot{z}_1 \frac{\partial}{\partial z_1} \tag{137}$$
$$= \left(\frac{\partial}{\partial t}\right)_{z_2} + \mathbf{V} \bullet \nabla_{z_2} + \left[\dot{z}_1 - \left(\frac{\partial z_1}{\partial t}\right)_{z_2} - \mathbf{V} \bullet \nabla_{z_2} z_1\right] \frac{\partial z_2}{\partial z_1} \frac{\partial}{\partial z_2}$$
$$= \left(\frac{\partial}{\partial t}\right)_{z_2} + \mathbf{V} \bullet \nabla_{z_2} + \dot{z}_2 \frac{\partial}{\partial z_2}.$$

The last line of (137) defines the *vertical velocity* $\dot{z}_2$ in z_2. Also from (137) we can see that the total derivative of a quantity yields the same result independent of coordinate system. With (137), we can write

$$\dot{z}_1 = \left(\frac{\partial z_1}{\partial t}\right)_{z_2} + \mathbf{V} \bullet \nabla_{z_2} z_1 + \dot{z}_2 \frac{\partial z_1}{\partial z_2}. \tag{138}$$

Taking the derivative of (142) with respect to z_2 gives

$$\frac{\partial \dot{z}_1}{\partial z_2} = \frac{D}{Dt}\frac{\partial z_1}{\partial z_2} + \frac{\partial \mathbf{V}}{\partial z_2} \bullet \nabla_{z_2} z_1 + \frac{\partial \dot{z}_2}{\partial z_2}\frac{\partial z_1}{\partial z_2}. \tag{139}$$

So, $\partial \dot{z}_1 / \partial z_1$ and $\partial \dot{z}_2 / \partial z_2$ are related with

$$\frac{\partial \dot{z}_1}{\partial z_1} = \frac{\partial \dot{z}_1}{\partial z_2}\frac{\partial z_2}{\partial z_1} = \frac{\partial z_2}{\partial z_1}\left[\frac{D}{Dt}\frac{\partial z_1}{\partial z_2} + \frac{\partial \mathbf{V}}{\partial z_2} \bullet \nabla_{z_2} z_1\right] + \frac{\partial \dot{z}_2}{\partial z_2}. \tag{140}$$

We can now transform from the standard geometric height to any coordinate z_2. Towards this end, define z_1 to be geometric height, then $z_1 \equiv z$ and $\dot{z}_1 \equiv w$. The continuity equation is given by

$$\frac{D}{Dt}\ln\rho + \nabla_z \bullet \mathbf{V} + \frac{\partial w}{\partial z} = 0. \tag{141}$$

With (136) and (140), we can write

$$\frac{D}{Dt}\ln\rho + \nabla_{z_2}\bullet\mathbf{V} - \left(\frac{\partial z_2}{\partial z}\right)(\nabla_{z_2}z)\frac{\partial\mathbf{V}}{\partial z_2} + \frac{\partial z_2}{\partial z}\left[\frac{D}{Dt}\frac{\partial z}{\partial z_2} + \frac{\partial\mathbf{V}}{\partial z_2}\bullet\nabla_{z_2}z\right] + \frac{\partial\dot{z}_2}{\partial z_2} \tag{142}$$
$$= \frac{D}{Dt}\ln\rho + \nabla_{z_2}\bullet\mathbf{V} + \frac{D}{Dt}\ln\left(\frac{\partial z}{\partial z_2}\right) + \frac{\partial\dot{z}_2}{\partial z_2}$$
$$= \frac{D}{Dt}\ln\left(\rho\frac{\partial z}{\partial z_2}\right) + \nabla_{z_2}\bullet\mathbf{V} + \frac{\partial\dot{z}_2}{\partial z_2} = 0.$$

Performing the logarithmic differentiation and expanding the total derivative in (142) yields

$$\rho\frac{\partial z}{\partial z_2}\left\{\left(\rho\frac{\partial z}{\partial z_2}\right)^{-1}\left[\left(\frac{\partial}{\partial t}\right)_{z_2}\left(\rho\frac{\partial z}{\partial z_2}\right) + \mathbf{V}\bullet\nabla_{z_2}\left(\rho\frac{\partial z}{\partial z_2}\right) + \dot{z}_2\frac{\partial}{\partial z_2}\left(\rho\frac{\partial z}{\partial z_2}\right)\right] + \nabla_{z_2}\bullet\mathbf{V} + \frac{\partial\dot{z}_2}{\partial z_2}\right\} \tag{143}$$
$$= \left(\frac{\partial}{\partial t}\right)_{z_2}\left(\rho\frac{\partial z}{\partial z_2}\right) + \nabla_{z_2}\bullet\left(\rho\frac{\partial z}{\partial z_2}\mathbf{V}\right) + \frac{\partial}{\partial z_2}\left(\rho\frac{\partial z}{\partial z_2}\dot{z}_2\right) = 0.$$

The hydrostatic equation written in the z_2-coordinate is given by

$$\rho\frac{\partial z}{\partial z_2} = -\frac{1}{g}\frac{\partial p}{\partial z_2}. \tag{144}$$

Thus (143) becomes

$$\left(\frac{\partial}{\partial t}\right)_{z_2}\left(\frac{\partial p}{\partial z_2}\right) + \nabla_{z_2}\bullet\left(\frac{\partial p}{\partial z_2}\mathbf{V}\right) + \frac{\partial}{\partial z_2}\left(\frac{\partial p}{\partial z_2}\dot{z}_2\right) = 0. \tag{145}$$

This is a useful equation.

The momentum equations written in geometric height are given by

$$\frac{D\mathbf{V}}{Dt} + f\hat{\mathbf{k}}\times\mathbf{V} = -\frac{1}{\rho}\nabla p. \tag{146}$$

Using (136) and (144), we can transform the pressure-gradient term to z_2

$$\frac{D\mathbf{V}}{Dt} + f\hat{\mathbf{k}}\times\mathbf{V} = -\frac{1}{\rho}\nabla_{z_2}p + \frac{1}{\rho}\frac{\partial z_2}{\partial z}(\nabla_{z_2}z)\frac{\partial p}{\partial z_2} = -\frac{1}{\rho}\nabla_{z_2}p - g\nabla_{z_2}z. \tag{147}$$

Appendix C: Horizontal velocity

The following is taken from Bleck (1978). The horizontal velocities remain unchanged when changing vertical coordinates. If the horizontal coordinate is a standard Cartesian coordinate, and the vertical coordinate is pressure or potential temperature, then the coordinate system is non-orthogonal. We can avoid confusion by considering coordinate surfaces instead of coordinate axis. Figure 44 shows coordinate surfaces along

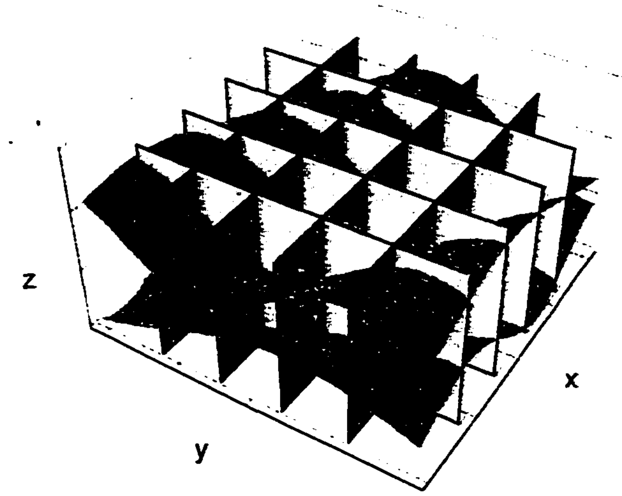


Figure 44 Surfaces of constant x, y and ζ.

which x and y are constant. A coordinate axis in the vertical direction is then defined to be the intersection of surfaces of constant x and constant y. Regardless of the orientation of the of the coordinate surfaces, the vertical axis is perpendicular to the earth's surface, not to the vertical coordinate surfaces. Without loss of generality, let's suppose the vertical coordinate is ζ. The x-axes, on the other hand, which is defined as the intersection of surfaces of constant y and ζ, are not necessarily parallel to the earth's surface as can be seen in Figure 44. Nevertheless, horizontal distance is measured along a path parallel to the earth's surface and not in the direction of the coordinate axes.

Horizontal velocity remains unchanged. For example, the velocity in the x direction, $\dot{x} \equiv dx/dt$, is defined as the rate at which a parcel passes though surfaces of constant x in a unit amount of time. So, again, this velocity would be unchanged by the definition of the vertical coordinate system. So, in the momentum equations,

$$\frac{D\mathbf{V}}{Dt} + f\hat{\mathbf{k}} \times \mathbf{V} = -\frac{1}{\rho}\nabla p, \tag{148}$$

the horizontal wind $\mathbf{V} \equiv \hat{\mathbf{i}}u + \hat{\mathbf{j}}v$ is such that $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ point in the usual cartesian directions. Then the individual components of the pressure-gradient must also point in the same directions. So, a change in vertical coordinate does not change the direction of the components of the horizontal gradient, but it does change the values of individual components according to (136).

Appendix D: Derivation of vertical mass flux CP sigma-coordinate

The discrete continuity equation is given by

$$\frac{\partial m_l}{\partial t} + \nabla \bullet (m\mathbf{V})_l + \frac{1}{(\delta\sigma)_l}[(m\dot{\sigma})_{l+1/2} - (m\dot{\sigma})_{l-1/2}] = 0 \text{ for } l = 1, 2, \ldots, L, \quad (149)$$

where

$$m_l \equiv \frac{p_{l+1/2} - p_{l-1/2}}{(\delta\sigma)_l} \text{ for } l = 1, 2, \ldots, L \quad (150)$$

and $(\delta\sigma)_l = \sigma_{l+1/2} - \sigma_{l-1/2}$ for $l = 1, 2, \ldots, L$. We wish to determine the form of $(m\dot{\sigma})_{l+1/2}$. Substitute (150) into (149) and summing throughout the column gives

$$\frac{\partial p_{l+1/2}}{\partial t} + \sum_{k=1}^{l} \nabla \bullet (m\mathbf{V})_k (\delta\sigma)_k + (m\dot{\sigma})_{l+1/2} = 0 \text{ for } l = 1, 2, \ldots, L-1, \quad (151)$$

where we have used the upper boundary condition $(m\dot{\sigma})_{1/2} = 0$. Since the lower boundary condition is given by $(m\dot{\sigma})_{L+1/2} = 0$ then the surface pressure tendency equation is given by

$$\frac{\partial p_S}{\partial t} + \sum_{k=1}^{L} \nabla \bullet (m\mathbf{V})_k (\delta\sigma)_k = 0. \quad (152)$$

Let

$$\sigma \equiv F(p, p_S). \quad (153)$$

Then if we differentiate (153) with respect to time and hold σ fixed we get

$$\left(\frac{\partial F}{\partial p}\right)_{p_S}\left(\frac{\partial p}{\partial t}\right)_{\sigma} + \left(\frac{\partial F}{\partial p_S}\right)_{p}\frac{\partial p_S}{\partial t} = 0. \quad (154)$$

The value of σ on a σ-surface cannot change. In discrete form (154) becomes

$$\left(\frac{\partial F}{\partial p}\right)_{l+1/2}\frac{\partial p_{l+1/2}}{\partial t}+\left(\frac{\partial F}{\partial p_S}\right)_{l+1/2}\frac{\partial p_S}{\partial t} = 0 \text{ for } l = 1, 2, \ldots, L-1. \tag{155}$$

Substitute (151) and (152) into (155) gives

$$\left(\frac{\partial F}{\partial p}\right)_{l+1/2}\left[-\sum_{k=1}^{l}\nabla\bullet(m\mathbf{V})_k(\delta\sigma)_k-(m\dot{\sigma})_{l+1/2}\right] \tag{156}$$
$$+\left(\frac{\partial F}{\partial p_S}\right)_{l+1/2}\left[-\sum_{k=1}^{L}\nabla\bullet(m\mathbf{V})_k(\delta\sigma)_k\right] = 0 \text{for } l = 1, 2, \ldots, L-1,$$

or

$$(m\dot{\sigma})_{l+1/2} = -\sum_{k=1}^{l}\nabla\bullet(m\mathbf{V})_k(\delta\sigma)_k \tag{157}$$
$$-\left[\left(\frac{\partial F}{\partial p}\right)_{l+1/2}\right]^{-1}\left(\frac{\partial F}{\partial p_S}\right)_{l+1/2}\sum_{k=1}^{L}\nabla\bullet(m\mathbf{V})_k(\delta\sigma)_k \text{for } l = 1, 2, \ldots, L-1.$$

For example, suppose $\sigma = F(p, p_S) = (p-p_T)/(p_S-p_T)$, then

$$\left(\frac{\partial F}{\partial p_S}\right)_{l+1/2} = -\frac{p_{l+1/2}-p_T}{(p_S-p_T)^2} \text{ and } \left(\frac{\partial F}{\partial p}\right)_{l+1/2} = \frac{1}{p_S-p_T} \tag{158}$$

and

$$\left[\left(\frac{\partial F}{\partial p}\right)_{l+1/2}\right]^{-1}\left(\frac{\partial F}{\partial p_S}\right)_{l+1/2} = -\frac{p_{l+1/2}-p_T}{p_S-p_T} = -\sigma_{l+1/2} \tag{159}$$

Appendix E: Energy conversion term in sigma-coordinates

The form of the energy conversion term in (51), (53) and (54) is determined by conservation of total energy. Therefore, we first derive the energy conversion term for the kinetic energy equation. In Arakawa and Lamb (1977) the second criteria in the design of the vertical differencing scheme for finite difference models is that the energy conversion term in the thermodynamic and the kinetic energy equations have the same form with opposite signs so that the total energy in conserved. This is in the realm of adiabatic and frictionless flow. To determine the form of the energy conversion term, we first form the kinetic energy equation. In continuous form we can multiply the momentum equation by $m\mathbf{V}$ to give

$$m\mathbf{V}\bullet\left[\frac{\partial\mathbf{V}}{\partial t}+\mathbf{V}\bullet\nabla\mathbf{V}+\dot{\sigma}\frac{\partial v}{\partial\sigma}\right] = m\mathbf{V}\bullet[-(\nabla_p\Phi)-f\hat{k}\times\mathbf{V}+\boldsymbol{F}] \tag{160}$$

$$= m\mathbf{V}\bullet\left[-\nabla\Phi+\frac{1}{m}\frac{\partial\Phi}{\partial\sigma}\nabla p\right]+m\mathbf{V}\bullet\boldsymbol{F}.$$

And the continuity equation times $\mathbf{V}^2/2$ gives

$$\frac{\mathbf{V}^2}{2}\left[\frac{\partial m}{\partial t}+\nabla\bullet(m\mathbf{V})+\frac{\partial}{\partial\sigma}(m\dot{\sigma})\right] = 0. \tag{161}$$

Combine eqs. (160) and (161) gives

$$\frac{\partial}{\partial t}\left(m\frac{\mathbf{V}^2}{2}\right)+\nabla\bullet\left(m\frac{\mathbf{V}^2}{2}\mathbf{V}\right)+\frac{\partial}{\partial\sigma}\left(m\frac{\mathbf{V}^2}{2}\dot{\sigma}\right) \tag{162}$$

$$= -\nabla\bullet(m\Phi\mathbf{V})-\frac{\partial}{\partial\sigma}\left[\Phi\left(\frac{\partial p}{\partial t}+m\dot{\sigma}\right)\right]-m\alpha\omega+m\mathbf{V}\bullet\boldsymbol{F}.$$

The details of this derivation are shown in grinding detail below

$$\frac{\partial}{\partial t}\left(m\frac{\mathbf{V}^2}{2}\right)+\nabla\bullet\left(m\frac{\mathbf{V}^2}{2}\mathbf{V}\right)+\frac{\partial}{\partial\sigma}\left(m\frac{\mathbf{V}^2}{2}\dot{\sigma}\right) = m\mathbf{V}\bullet\left[-\nabla\Phi+\frac{1}{m}\frac{\partial\Phi}{\partial\sigma}\nabla p\right]+m\mathbf{V}\bullet \boldsymbol{F} \tag{163}$$

$$= -m\mathbf{V}\bullet\nabla\Phi+\frac{\partial\Phi}{\partial\sigma}\mathbf{V}\bullet\nabla p+m\mathbf{V}\bullet \boldsymbol{F}$$

$$= -\nabla\bullet(m\Phi\mathbf{V})+\Phi\nabla\bullet(m\mathbf{V})-\alpha m\mathbf{V}\bullet\nabla p+m\mathbf{V}\bullet \boldsymbol{F}$$

$$= -\nabla\bullet(m\Phi\mathbf{V})+\Phi\left[-\frac{\partial m}{\partial t}-\frac{\partial}{\partial\sigma}(m\dot{\sigma})\right]-\alpha m\mathbf{V}\bullet\nabla p+m\mathbf{V}\bullet \boldsymbol{F}$$

$$= -\nabla\bullet(m\Phi\mathbf{V})+\Phi\left[-\frac{\partial}{\partial\sigma}\left(\frac{\partial p}{\partial t}+m\dot{\sigma}\right)\right]-\alpha m\mathbf{V}\bullet\nabla p+m\mathbf{V}\bullet \boldsymbol{F}$$

$$= -\nabla\bullet(m\Phi\mathbf{V})-\frac{\partial}{\partial\sigma}\left[\Phi\left(\frac{\partial p}{\partial t}+m\dot{\sigma}\right)\right]+\frac{\partial\Phi}{\partial\sigma}\left(\frac{\partial p}{\partial t}+m\dot{\sigma}\right)-\alpha m\mathbf{V}\bullet\nabla p+m\mathbf{V}\bullet \boldsymbol{F}$$

$$= -\nabla\bullet(m\Phi\mathbf{V})-\frac{\partial}{\partial\sigma}\left[\Phi\left(\frac{\partial p}{\partial t}+m\dot{\sigma}\right)\right]-\alpha m\left(\frac{\partial p}{\partial t}+m\dot{\sigma}+\mathbf{V}\bullet\nabla p\right)+m\mathbf{V}\bullet \boldsymbol{F}$$

$$= -\nabla\bullet(m\Phi\mathbf{V})-\frac{\partial}{\partial\sigma}\left[\Phi\left(\frac{\partial p}{\partial t}+m\dot{\sigma}\right)\right]-m\alpha\omega+m\mathbf{V}\bullet \boldsymbol{F}.$$

In the discrete realm, Arakawa and Konor, (AK) mimic the steps from the continuous realm. We multiply the pressure-gradient force (AK (3.22)) (which is derived in Appendix F) by $(m\mathbf{V})_l \equiv m_l\mathbf{V}_l$, that is,

$$-m_l\mathbf{V}_l\bullet(\nabla_p p) \tag{164}$$

$$= m_l\mathbf{V}_l\bullet\left[-\frac{1}{m_l}\nabla(m_l\Phi_l)+\frac{1}{m_l(\delta\sigma)_l}(\Phi_{l+1/2}\nabla p_{l+1/2}-\Phi_{l-1/2}\nabla p_{l-1/2})\right]$$

$$= -\nabla\bullet(m_l\mathbf{V}_l\Phi_l)+m_l\Phi_l\nabla\bullet\mathbf{V}_l+\frac{1}{(\delta\sigma)_l}[\Phi_{l+1/2}\mathbf{V}_l\bullet\nabla p_{l+1/2}-\Phi_{l-1/2}\mathbf{V}_l\bullet\nabla p_{l-1/2}]$$

$$= -\nabla\bullet(m_l\mathbf{V}_l\Phi_l)+\Phi_l\nabla\bullet(m_l\mathbf{V}_l)-\Phi_l\mathbf{V}_l\bullet\nabla m_l$$

$$+\frac{1}{(\delta\sigma)_l}[\Phi_{l+1/2}\mathbf{V}_l\bullet\nabla p_{l+1/2}-\Phi_{l-1/2}\mathbf{V}_l\bullet\nabla p_{l-1/2}] \quad \text{for } l = 1, 2, \ldots, L.$$

With the continuity equation and the definition $m_l \equiv (p_{l+1/2}-p_{l-1/2})/(\delta\sigma)_l$, we can write (164) as

$$-m_l\mathbf{V}_l \bullet (\nabla_p p) = -\nabla\bullet(m_l\mathbf{V}_l\Phi_l) \tag{165}$$

$$-\frac{1}{(\delta\sigma)_l}\left\{\Phi_{l+1/2}\left[\frac{\partial p_{l+1/2}}{\partial t}+(m\dot{\sigma})_{l+1/2}\right]-\Phi_{l-1/2}\left[\frac{\partial p_{l-1/2}}{\partial t}+(m\dot{\sigma})_{l-1/2}\right]\right\}$$

$$+\frac{1}{(\delta\sigma)_l}(\Phi_{l+1/2}-\Phi_l)\left[\frac{\partial p_{l+1/2}}{\partial t}+\mathbf{V}_l\bullet\nabla p_{l+1/2}+(m\dot{\sigma})_{l+1/2}\right]$$

$$+\frac{1}{(\delta\sigma)_l}(\Phi_l-\Phi_{l-1/2})\left[\frac{\partial p_{l-1/2}}{\partial t}+\mathbf{V}_l\bullet\nabla p_{l-1/2}+(m\dot{\sigma})_{l-1/2}\right] \text{ for } l = 1,2,...,L.$$

Comparing the right-hand side of the continuous equation (162) and the corresponding discrete finite-difference equation (165), we can see the energy conversion term $(m\alpha\omega)_l$ in the discrete equation must be the following terms

$$(m\alpha\omega)_l = -\frac{1}{(\delta\sigma)_l}(\Phi_{l+1/2}-\Phi_l)\left[\frac{\partial p_{l+1/2}}{\partial t}+\mathbf{V}_l\bullet\nabla p_{l+1/2}+(m\dot{\sigma})_{l+1/2}\right] \tag{166}$$

$$-\frac{1}{(\delta\sigma)_l}(\Phi_l-\Phi_{l-1/2})\left[\frac{\partial p_{l-1/2}}{\partial t}+\mathbf{V}_l\bullet\nabla p_{l-1/2}+(m\dot{\sigma})_{l-1/2}\right] \text{ for } l = 1,2,...L.$$

The form (166) involves differences in geopotential. The appearance of geopotential is a clue that the corresponding energy conversion term in the thermodynamic equation will be determined by specifying a form for the hydrostatic equation. This is the task of Appendix F. Jumping ahead a little bit, we can substitute the unspecified parts of the hydrostatic equation, (175) and (176), back into (166) to give

$$(m\alpha\omega)_l = -\frac{1}{(\delta\sigma)_l}A_{l+1/2}\theta_{l+1/2}\left[\frac{\partial p_{l+1/2}}{\partial t}+\mathbf{V}_l\bullet\nabla p_{l+1/2}+(m\dot{\sigma})_{l+1/2}\right] \tag{167}$$

$$-\frac{1}{(\delta\sigma)_l}B_{l-1/2}\theta_{l-1/2}\left[\frac{\partial p_{l-1/2}}{\partial t}+\mathbf{V}_l\bullet\nabla p_{l-1/2}+(m\dot{\sigma})_{l-1/2}\right] \text{ for } l = 1,2,...L.$$

To reiterate, the constraint (ii) (Arakawa and Lamb (1977)) requires that the vertical sum of the energy conversion term in the kinetic energy equation $(m\alpha\omega)_l(\delta\sigma)_l$ for $l = 1,2,...,L$ (layer centers) and the energy conversion term in the thermodynamic energy equation $(m\alpha\omega)_{l+1/2}(\delta\sigma)_{l+1/2}$ for $l = 0, 1, ..., L$ (layer edges) be equal and opposite sign, that is,

$$\sum_{l=1}^{L}(m\alpha\omega)_l(\delta\sigma)_l + \sum_{l=0}^{L}(m\alpha\omega)_{l+1/2}(\delta\sigma)_{l+1/2} = 0. \quad (168)$$

(Note: there appears to be a sign error in AK (3.34).) Now, this is a very neat trick. From (167), we can see there contribution from $\theta_{l+1/2}$ to the kinetic energy conversion term in the layer (l). The same potential temperature at this edge will also contribute to the kinetic energy conversion term of layer $(l+1)$. So, we can write the energy conversion terms in the thermodynamic energy equation as a telescoping sum which will satisfy (168), namely,

$$(m\alpha\omega)_{l+1/2}(\delta\sigma)_{l+1/2} = \theta_{l+1/2}\left\{(A_{l+1/2}+B_{l+1/2})\left[\frac{\partial p_{l+1/2}}{\partial t}+(m\dot{\sigma})_{l+1/2}\right]\right. \quad (169)$$

$$\left. +(A_{l+1/2}\mathbf{V}_l + B_{l+1/2}\mathbf{V}_{l+1})\bullet\nabla p_{l+1/2}\right\} \text{ for } l = 1, 2, ..., L,$$

and

$$(m\alpha\omega)_{1/2}(\delta\sigma)_{1/2} = \theta_{1/2}B_{1/2}\left[\frac{\partial p_{1/2}}{\partial t}+\mathbf{V}_1\bullet\nabla p_{1/2}\right] = 0, \quad (170)$$

and

$$(m\alpha\omega)_{L+1/2}(\delta\sigma)_{L+1/2} = \theta_{L+1/2}A_{L+1/2}\left[\frac{\partial p_{L+1/2}}{\partial t}+\mathbf{V}_L\bullet\nabla p_{L+1/2}\right], \quad (171)$$

where we have used the boundary conditions, $(m\dot{\sigma})_{1/2} = (m\dot{\sigma})_{L+1/2} = 0$. With the energy conversion term in the thermodynamic equation in the form (169)-(171) the equality in (168) will be true regardless of the form of $A_{l+1/2}$ and $B_{l+1/2}$. It still remains to determine the form of $A_{l+1/2}$ and $B_{l+1/2}$ in Appendix F.

Appendix F: The pressure-gradient force and hydrostatic equation in sigma-coordinates

Using the gradient operator relation in (136) and applying it to the geopotential Φ gives the relationship of the pressure-gradient force in pressure and sigma coordinates

$$
\begin{aligned}
-\nabla_p\Phi &= -\nabla_\sigma\Phi + (\nabla_\sigma p)\frac{\partial\Phi}{\partial p} \\
&= -\nabla_\sigma\Phi + \frac{1}{m}\frac{\partial\Phi}{\partial\sigma}\nabla_\sigma p \\
&= -\left(\frac{1}{m}m\right)\nabla_\sigma\Phi + \left(\frac{1}{m}\Phi\nabla_\sigma m - \frac{1}{m}\Phi\nabla_\sigma m\right) + \frac{1}{m}\frac{\partial\Phi}{\partial\sigma}\nabla_\sigma p \\
&= -\frac{1}{m}\nabla_\sigma(m\Phi) + \frac{1}{m}\Phi\frac{\partial}{\partial\sigma}\nabla_\sigma p + \frac{1}{m}\frac{\partial\Phi}{\partial\sigma}\nabla_\sigma p \\
&= -\frac{1}{m}\nabla_\sigma(m\Phi) + \frac{1}{m}\frac{\partial}{\partial\sigma}(\Phi\nabla_\sigma p),
\end{aligned}
\tag{172}
$$

where we have used $m = \partial p/\partial\sigma$. The vertically discrete form of the pressure-gradient force in (172) can be written

$$
-(\nabla_p\Phi)_l = -\frac{1}{m_l}\nabla_\sigma(m_l\Phi_l) + \frac{1}{m_l(\delta\sigma)_l}(\Phi_{l+1/2}\nabla_\sigma p_{l+1/2} - \Phi_{l-1/2}\nabla_\sigma p_{l-1/2}) \tag{173}
$$

$$\text{for } l = 1, 2, \ldots, L.$$

To compute the discrete geopotential we need to vertically integrate the hydrostatic equation in finite-difference form. The continuous form of the hydrostatic equation for the σ-coordinate is

$$
\frac{\partial\Phi}{\partial\Pi} = -\theta. \tag{174}
$$

In finite-difference form (169) can be discretize in two pieces:

$$
\Phi_l - \Phi_{l+1/2} = A_{l+1/2}(\Pi_{l+1/2})\theta_{l+1/2} \quad \text{for } l = 1, 2, \ldots, L \tag{175}
$$

and

$$\Phi_{l+1/2}-\Phi_{l+1} = B_{l+1/2}(\Pi_{l+1/2})\theta_{l+1/2} \text{ for } l = 0, 1, \ldots, L-1. \tag{176}$$

From (174), we can see that $A_{l+1/2}$ and $B_{l+1/2}$ are going to be functions of the Exner function $\Pi_{l+1/2}$ at a layer edge. Combining (175) and (176), the thickness between two layers is given by

$$\Phi_{l}-\Phi_{l+1} = (A_{l+1/2}+B_{l+1/2})\theta_{l+1/2} \text{ for } l = 1, 2, \ldots, L-1. \tag{177}$$

The form of (177) is important. It implies a hydrostatic equations in which the thickness between two adjacent layers depends only on the potential temperature at their common boundary. This is in contrast to the Lorenz grid. In the Lorenz grid the potential temperature at the edge of a layer is averaged from the two adjacent layer centers. Recall that this deficiency allows for the existence of the computational mode. Next, we can rewrite (51), (53) and (54) using

$$c_pT_{i+1/2} = \theta_{l+1/2}\Pi_{l+1/2} \text{ where } \Pi_{l+1/2} \equiv c_p\left(\frac{p_{l+1/2}}{p_0}\right)^{\kappa} \text{ for } l = 0, 1, \ldots, L \tag{178}$$

and the continuity equation applied to a layer edge and (169)-(171). The result is

$$\Pi_{l+1/2}\left[m_{l+1/2}\frac{\partial\theta_{l+1/2}}{\partial t}+(m\mathbf{V})_{l+1/2}\bullet\nabla\theta_{l+1/2}\right] \tag{179}$$

$$+\theta_{l+1/2}\left(\frac{\partial\Pi_{l+1/2}}{\partial p_{l+1/2}}\right)\left[m_{l+1/2}\frac{\partial p_{l+1/2}}{\partial t}+(m\mathbf{V})_{l+1/2}\bullet\nabla p_{l+1/2}\right]$$

$$=\frac{c_pT_{l+1/2}}{(\delta\sigma)_{l+1/2}}[(m\dot{\sigma})_{l+1}-(m\dot{\sigma})_l]-\frac{c_p}{(\delta\sigma)_{l+1/2}}[(Tm\dot{\sigma})_{l+1}-(Tm\dot{\sigma})_l]$$

$$+\frac{\theta_{l+1/2}}{(\delta\sigma)_{l+1/2}}\left\{(A_{l+1/2}+B_{l+1/2})\left[\frac{\partial p_{l+1/2}}{\partial t}+(m\dot{\sigma})_{l+1/2}\right]\right.$$

$$\left.+(A_{l+1/2}\mathbf{V}_l+B_{l+1/2}\mathbf{V}_{l+1})\bullet\nabla p_{l+1/2}\right\}+(mQ)_{l+1/2}\quad\text{for } l=1,2,\ldots,L-1,$$

and

$$\Pi_{1/2}\left[m_{1/2}\frac{\partial\theta_{1/2}}{\partial t}+(m\mathbf{V})_{1/2}\bullet\nabla\theta_{1/2}\right] \tag{180}$$

$$+\theta_{1/2}\left(\frac{\partial\Pi_{1/2}}{\partial p_{1/2}}\right)\left[m_{1/2}\frac{\partial p_{1/2}}{\partial t}+(m\mathbf{V})_{1/2}\bullet\nabla p_{1/2}\right]$$

$$=\frac{c_pT_{1/2}}{(\delta\sigma)_{1/2}}(m\dot{\sigma})_1-\frac{c_p}{(\delta\sigma)_{l+1/2}}(Tm\dot{\sigma})_1$$

and

$$\Pi_{L+1/2}\left[m_{L+1/2}\frac{\partial\theta_{L+1/2}}{\partial t}+(m\mathbf{V})_{L+1/2}\bullet\nabla\theta_{L+1/2}\right] \tag{181}$$

$$+\theta_{L+1/2}\left(\frac{\partial\Pi_{L+1/2}}{\partial p_{L+1/2}}\right)\left[m_{L+1/2}\frac{\partial p_{L+1/2}}{\partial t}+(m\mathbf{V})_{L+1/2}\bullet\nabla p_{L+1/2}\right]$$

$$=-\frac{c_pT_{L+1/2}}{(\delta\sigma)_{l+1/2}}(m\dot{\sigma})_L+\frac{c_p}{(\delta\sigma)_{L+1/2}}(Tm\dot{\sigma})_L$$

$$+\frac{\theta_{L+1/2}}{(\delta\sigma)_{L+1/2}}(A_{L+1/2})\left[\frac{\partial p_{L+1/2}}{\partial t}+\mathbf{V}_L\bullet\nabla p_{L+1/2}\right].$$

If you write the thermodynamic energy equation in terms of potential temperature, namely,

$$\frac{D\theta}{Dt} = \frac{\partial\theta}{\partial t} + \mathbf{V} \bullet \nabla\theta + \dot{\sigma}\frac{\partial\theta}{\partial\sigma} = \frac{Q}{\Pi}, \tag{182}$$

and compare with (179), then by inspection terms involving $\partial p_{l+1/2}/\partial t$ and $\nabla p_{l+1/2}$ in (179)-(181) need to be eliminated. This will happen if

$$A_{l+1/2} \equiv \frac{(\delta\sigma)_l m_l}{2}\left(\frac{\partial\Pi_{l+1/2}}{\partial p_{l+1/2}}\right) \text{ for } l = 1, 2, \ldots, L \tag{183}$$

$$B_{l+1/2} \equiv \frac{(\delta\sigma)_{l+1} m_{l+1}}{2}\left(\frac{\partial\Pi_{l+1/2}}{\partial p_{l+1/2}}\right) \text{ for } l = 0, 1, \ldots, L-1.$$

With $\Pi(p) = c_p(p/p_0)^{\kappa}$ we have

$$\frac{\partial\Pi}{\partial p} = c_p\kappa\left(\frac{p}{p_0}\right)^{\kappa-1}\frac{1}{p_0} = R\left(\frac{p}{p_0}\right)^{\kappa}\left(\frac{p}{p_0}\right)^{-1}\frac{1}{p_0} = \frac{R}{p}\left(\frac{p}{p_0}\right)^{\kappa} = \frac{RT}{p\theta}. \tag{184}$$

With (39), (183)and (184), (175)and (176) become

$$\Phi_l - \Phi_{l+1/2} = \frac{p_{l+1/2} - p_{l-1/2}}{2}\frac{R}{p_{l+1/2}}T_{l+1/2} \text{ for } l = 1, 2, \ldots, L \tag{185}$$

and

$$\Phi_{l+1/2} - \Phi_{l+1} = \frac{p_{l+3/2} - p_{l+1/2}}{2}\frac{R}{p_{l+1/2}}T_{l+1/2} \text{ for } l = 0, 1, \ldots, L-1. \tag{186}$$

This is the form of the hydrostatic equation used in the model.

Appendix G: The pressure-gradient force and hydrostatic equation in isentropic-coordinates

With (136) we can write

$$-\frac{1}{\rho}\nabla_z p = -\frac{1}{\rho}\nabla_\theta p + \frac{1}{\rho}\frac{\partial\theta}{\partial z}(\nabla_\theta z)\frac{\partial p}{\partial\theta}. \tag{187}$$

With the chain rule and the hydrostatic equation we can write

$$\frac{\partial p}{\partial z} = \frac{\partial\theta}{\partial z}\frac{\partial p}{\partial\theta} = -\rho g. \tag{188}$$

Substitute (188) into (187) gives

$$-\frac{1}{\rho}\nabla_z p = -\frac{1}{\rho}\nabla_\theta p - g(\nabla_\theta z). \tag{189}$$

With Poisons equation $\theta = T(p_0/p)^\kappa$, we have $\ln\theta = \ln T + \kappa[\ln p_0 - \ln p]$. Taking the gradient on a surface of constant θ

$$\nabla_\theta \ln\theta = \nabla_\theta \ln T - \frac{R}{c_p}\nabla_\theta \ln p = 0. \tag{190}$$

With the chain rule and the equation of state (190) becomes

$$\frac{1}{\rho}\nabla_\theta p = c_p\nabla_\theta T. \tag{191}$$

Substituting (191) into (189) gives

$$-\frac{1}{\rho}\nabla_z p = -c_p\nabla_\theta T - g\nabla_\theta z = -\nabla_\theta M, \tag{192}$$

where $M \equiv c_p T + gz$ is the Montgomery potential.

The continuous hydrostatic equation of interest is this

$$\frac{\partial M}{\partial \theta} = \Pi, \tag{193}$$

which can be conveniently discretized as

$$M_l - M_{l+1} = (\theta_l - \theta_{l+1})\Pi_{l+1/2}, \tag{194}$$

where $\theta_l = \sqrt{\theta_{l-1/2}\theta_{l+1/2}}$ is the harmonic mean.

Appendix H: The pressure-gradient force and hydrostatic equation in the generalized vertical coordinates

With (136), we can write the pressure-gradient force as

$$-\nabla_p\Phi = -\nabla_\zeta\Phi - \frac{1}{m}(\nabla_\zeta p)\frac{\partial\Phi}{\partial\zeta}, \tag{195}$$

where we have used $m = -\partial p/\partial\zeta$. Equation (195) can be re-written as follows

$$\begin{aligned}
-\nabla_p\Phi &= -\nabla_\zeta\Phi - \frac{1}{m}\frac{\partial\Phi}{\partial\zeta}\nabla_\zeta p \\
&= -\nabla_\zeta(M - c_pT) - \alpha\nabla_\zeta p \\
&= -\nabla_\zeta M + c_p\nabla_\zeta T - \frac{RT}{p}\nabla_\zeta p \\
&= -\nabla_\zeta M + \nabla_\zeta(\Pi\theta) - \frac{RT}{p}\nabla_\zeta p \\
&= -\nabla_\zeta M + \Pi\nabla_\zeta\theta + \theta\nabla_\zeta\Pi - \frac{RT}{p}\nabla_\zeta p \\
&= -\nabla_\zeta M + \Pi\nabla_\zeta\theta + \theta\left[c_p\kappa\left(\frac{p}{p_0}\right)^{\kappa}\left(\frac{p}{p_0}\right)^{-1}\frac{1}{p_0}\nabla_\zeta p\right] - \frac{RT}{p}\nabla_\zeta p \\
&= -\nabla_\zeta M + \Pi\nabla_\zeta\theta,
\end{aligned} \tag{196}$$

where we have used $M = \Phi + c_pT$, $p\alpha = RT$ and $\Pi\theta = c_pT$. Note that when $\zeta = \theta$ then $-\nabla_p\Phi = -\nabla_\theta M$. Hence, in regions of pure isentropic coordinate, the pressure-gradient force is the same as in (192).

To compute the pressure-gradient force in the generalized vertical coordinate, we need to know both the Montgomery potential M (like in θ-coordinates) and the geopotential Φ (like in σ-coordinates) at the middles of layers. The hydrostatic equation for isentropic coordinates is given by

$$\frac{\partial M}{\partial \theta} = \Pi \tag{197}$$

and can be discretize most directly as follows

$$M_l - M_{l+1} = \Pi_{l+1/2}(\theta_l - \theta_{l+1}) \quad \text{for } l = 1, 2, \ldots, L-1, \tag{198}$$

where $\theta_l \equiv f(\theta_{l-1/2}, \theta_{l+1/2})$. While the hydrostatic equation in the form

$$\frac{\partial \Phi}{\partial \Pi} = -\theta \tag{199}$$

can be discretize as follows

$$\Phi_l - \Phi_{l+1} = -\theta_{l+1/2}(\Pi_l - \Pi_{l+1}) \quad \text{for } l = 1, 2, \ldots, L-1, \tag{200}$$

where $\Pi_l \equiv f(p_{l-1/2}, p_{l+1/2})$. Substituting $M_l = \theta_l \Pi_l + \Phi_l$ into (198) we get

$$(\theta_l \Pi_l + \Phi_l) - (\theta_{l+1}\Pi_{l+1} + \Phi_{l+1}) = \Pi_{l+1/2}(\theta_l - \theta_{l+1}) \quad \text{for } l = 1, 2, \ldots, L-1 \tag{201}$$

or

$$\begin{aligned} \Phi_l - \Phi_{l+1} &= \theta_l(\Pi_{l+1/2} - \Pi_l) + \theta_{l+1}(\Pi_{l+1} - \Pi_{l+1/2}) \\ &= (-[\theta_l(\Pi_l - \Pi_{l+1/2}) + \theta_{l+1}(\Pi_{l+1/2} - \Pi_{l+1})]) \quad \text{for } l = 1, 2, \ldots, L-1, \end{aligned} \tag{202}$$

where (202) is written to look like (200). At the lower boundary, we set

$$\Phi_L = \Phi_{L+1/2} + \theta_L(\Pi_{L+1/2} - \Pi_L), \tag{203}$$

where

$$\theta_l = \frac{1}{2}(\theta_{l-1/2} + \theta_{l+1/2}) \quad \text{for } l = 1, 2, \ldots, L. \tag{204}$$

The vertical distribution of θ that satisfies (202) may include a computational mode because of the averaging in (204). However, the advantages of a hydrostatic equation more akin to that found in the isentropic model (the goal is to make the model as θ like as possible) outweighs the disadvantage of the possible computational mode.

In the momentum equation the pressure-gradient term has the form

$$\frac{\partial \mathbf{V}}{\partial t} = -\nabla M + \Pi \nabla \theta + \ldots, \tag{205}$$

where we assume all derivatives are on surfaces of constant ζ.

$$\begin{aligned} \hat{\mathbf{k}} \bullet \nabla \times \frac{\partial \mathbf{V}}{\partial t} &= \frac{\partial \eta}{\partial t} = \hat{\mathbf{k}} \bullet \nabla \times (-\nabla M + \Pi \nabla \theta) + \ldots \\ &= \hat{\mathbf{k}} \bullet \nabla \times \Pi \nabla \theta + \ldots = -J(\theta, \Pi) + \ldots. \end{aligned} \tag{206}$$

There is a contribution from the pressure-gradient term in the vorticity equation. Taking the divergence of (205) we have

$$\nabla \bullet \frac{\partial \mathbf{V}}{\partial t} = \frac{\partial \delta}{\partial t} = \nabla \bullet (-\nabla M + \Pi \nabla \theta) + \ldots = -\nabla^2 M + \nabla \bullet (\Pi \nabla \theta) + \ldots. \tag{207}$$

Appendix I: Derivation of vertical mass flux in generalized vertical coordinate

The value of ζ on a ζ-surface cannot change. In continuous form we have

$$\left(\frac{\partial}{\partial t}\right)_{\zeta} F(\theta, p, p_s) = 0. \tag{208}$$

With the chain rule (208) becomes

$$\left(\frac{\partial F}{\partial \theta}\right)_{p,\, p_S}\left(\frac{\partial \theta}{\partial t}\right)_{\zeta} + \left(\frac{\partial F}{\partial p}\right)_{\theta,\, p_S}\left(\frac{\partial p}{\partial t}\right)_{\zeta} + \left(\frac{\partial F}{\partial p_S}\right)\frac{\partial p_S}{\partial t} = 0. \tag{209}$$

The thermodynamic energy equation is given by

$$\frac{D\theta}{Dt} = \left(\frac{\partial \theta}{\partial t}\right)_{\zeta} + \mathbf{V} \bullet \nabla_{\zeta}\theta + \dot{\zeta}\frac{\partial \theta}{\partial \zeta} = \frac{Q}{\Pi} \tag{210}$$

and pressure and surface pressure tendency equations are given by

$$\left(\frac{\partial p}{\partial t}\right)_{\zeta} = \frac{\partial p_T}{\partial t} + \nabla_{\zeta} \bullet \int_{\zeta_T}^{\zeta} m\mathbf{V}d\zeta + (m\dot{\zeta})_{\zeta} \quad \text{and} \quad \frac{\partial p_S}{\partial t} = \frac{\partial p_T}{\partial t} + \nabla_{\zeta} \bullet \int_{\zeta_T}^{\zeta_S} m\mathbf{V}d\zeta. \tag{211}$$

Substitution of (210) and (211) into (209) gives

$$\begin{aligned}\left\{\left(\frac{\partial F}{\partial \theta}\right)_{p,\, p_S}\left(\frac{\partial \theta}{\partial \zeta}\right) - m\left(\frac{\partial F}{\partial p}\right)_{\theta,\, p_S}\right\}\dot{\zeta} &= \left(\frac{\partial F}{\partial \theta}\right)_{p,\, p_S}\left\{\frac{Q}{\Pi} - \mathbf{V} \bullet \nabla_{\zeta}\theta\right\} \\ &+ \left(\frac{\partial F}{\partial p}\right)_{\theta,\, p_S}\left\{\frac{\partial p_T}{\partial t} + \nabla_{\zeta} \bullet \int_{\zeta_T}^{\zeta} m\mathbf{V}d\zeta\right\} + \left(\frac{\partial F}{\partial p_S}\right)\left\{\frac{\partial p_T}{\partial t} + \nabla_{\zeta} \bullet \int_{\zeta_T}^{\zeta_S} m\mathbf{V}d\zeta\right\}.\end{aligned} \tag{212}$$

Eq. (212) is known as the generalized vertical mass flux equation. In finite-difference form, we can write an equation analogous to (208), namely,

$$\frac{\partial}{\partial t}F(\theta_{l+1/2}, p_{l+1/2}, p_S) = 0. \tag{213}$$

The discrete continuity equation is given by

$$\frac{\partial m_l}{\partial t} + \nabla \bullet (m\mathbf{V})_l + \frac{1}{(\delta\zeta)_l}[(m\dot{\zeta})_{l+1/2} - (m\dot{\zeta})_{l-1/2}] = 0 \quad \text{for } l = 1, 2, \ldots, L. \tag{214}$$

The discrete thermodynamic equation is given by

$$(m\Pi\delta\zeta)_{l+1/2}\frac{\partial\theta_{l+1/2}}{\partial t} + (m\Pi\mathbf{V}\delta\zeta)_{l+1/2} \bullet \nabla\theta_{l+1/2} \tag{215}$$
$$+ \left(\Pi\frac{\partial\theta}{\partial\zeta}\delta\zeta\right)_{l+1/2}(m\dot{\zeta})_{l+1/2} = (mQ\delta\zeta)_{l+1/2} \quad \text{for } l = 1, 2, \ldots, L-1.$$

The algorithm to satisfy (213) has two steps. First, compute a provisional mass retaining only the horizontal advection terms, and compute a provisional potential temperature retaining only the diabatic heating term and the horizontal advection term. That is,

$$\hat{m}_l^n = m_l^n - \Delta t[\nabla \bullet (m\mathbf{V})_l]^n \quad \text{for } l = 1, 2, \ldots, L, \tag{216}$$

where the tendency on the right-hand side is the flux-corrected Adams-Bashforth weighted tendency. The flux correction prevents the pressure thickness of a cell from becoming less than a prescribed threshold value, typically on the order of 1mb. And,

$$\hat{\theta}_{l+1/2}^n = \theta_{l+1/2}^n \tag{217}$$
$$+ \Delta t\left[\frac{1}{(m\Pi\delta\zeta)_{l+1/2}}((mQ\delta\zeta)_{l+1/2} - (m\Pi\mathbf{V}\delta\zeta)_{l+1/2} \bullet \nabla\theta_{l+1/2})\right]^n$$
$$\text{for } l = 1, 2, \ldots, L-1,$$

where the tendency on the right-hand side is the Adams-Bashforth weighted tendency. From $\hat{m}_l$ we can integrate to get $\hat{p}_{l+1/2}$ and $\hat{p}_S$. The deviation of ζ from is correct value due to horizontal advection and diabatic heating can then be determined from

$$(\Delta\zeta)_{l+1/2} = F(\hat{\theta}_{l+1/2}, \hat{p}_{l+1/2}, \hat{p}_S) - \zeta_{l+1/2} \quad \text{for } l = 1, 2, \ldots, L-1. \tag{218}$$

In the second step, we require that $(m\dot{\zeta})_{l+1/2}$ reduce $(\Delta\zeta)_{l+1/2}$. Discretizing the left-hand side of (212) gives

$$\left[\left(\frac{\partial\hat{F}}{\partial\hat{\theta}}\right)_{l+1/2}\left(\Pi\frac{\partial\theta}{\partial\zeta}\delta\zeta\right)_{l+1/2}(m\Pi\delta\zeta)^{-1}_{l+1/2}-\left(\frac{\partial\hat{F}}{\partial\bar{p}}\right)_{l+1/2}\right](m\dot{\zeta})_{l+1/2} = \frac{(\Delta\zeta)_{l+1/2}}{\Delta t} \quad (219)$$

$$\text{for } l = 1, 2, \ldots, L-1$$

Appendix J: Vertical advection and diffusion of momentum in the generalized vertical coordinate model.

Vertical advection of momentum

The momentum equation is given by

$$\frac{\partial \mathbf{V}}{\partial t} = -\mathbf{V} \bullet \nabla \mathbf{V} - \dot{\zeta}\frac{\partial \mathbf{V}}{\partial \zeta} - (\nabla_p \Phi) - f\hat{\mathbf{k}} \times \mathbf{V} + \boldsymbol{F}. \tag{220}$$

Following Hsu and Arakawa (1990) we can write the vertical advection of momentum term in (220) as follows

$$\dot{\zeta}\frac{\partial \mathbf{V}_l}{\partial \zeta} = \frac{-1}{p_{l-1/2} - p_{l+1/2}}[(m\dot{\zeta})_{l-1/2}(\mathbf{V}_{l-1/2} - \mathbf{V}_l) + (m\dot{\zeta})_{l+1/2}(\mathbf{V}_l - \mathbf{V}_{l+1/2})] \tag{221}$$

for $l = 1, 2, \ldots, L$.

It is clear from (221) that $p_{l-1/2} - p_{l+1/2} < 0$ is desirable. If massless layers are not allowed to form, the difference $p_{l-1/2} - p_{l+1/2} < 0$ bounded away from zero is guaranteed. We run the model with a FCT of mass which prevents layers from becoming less than 50Pa thick. With $\mathbf{V}_{l+1/2} \equiv (\mathbf{V}_l + \mathbf{V}_{l+1})/2$ and $\mathbf{V}_l = \hat{k} \times \nabla\psi_l + \nabla\chi_l$, (221) becomes

$$\dot{\zeta}\frac{\partial \mathbf{V}_l}{\partial \zeta} = \frac{-1}{p_{l-1/2} - p_{l+1/2}}\left\{\frac{(m\dot{\zeta})_{l-1/2}}{2}[(\hat{\mathbf{k}} \times \nabla\psi_{l-1} + \nabla\chi_{l-1}) - (\hat{\mathbf{k}} \times \nabla\psi_l + \nabla\chi_l)] \right. \tag{222}$$

$$\left. + \frac{(m\dot{\zeta})_{l+1/2}}{2}[(\hat{\mathbf{k}} \times \nabla\psi_l + \nabla\chi_l) - (\hat{\mathbf{k}} \times \nabla\psi_{l+1} + \nabla\chi_{l+1})]\right\}.$$

Define

$$\alpha_{l-1/2} \equiv \frac{-1}{p_{l-1/2} - p_{l+1/2}}\frac{(m\dot{\zeta})_{l-1/2}}{2} \text{ and } \alpha_{l+1/2} \equiv \frac{-1}{p_{l-1/2} - p_{l+1/2}}\frac{(m\dot{\zeta})_{l+1/2}}{2}. \tag{223}$$

Substitute (223) into (222)

$$\zeta\frac{\partial \mathbf{V}_l}{\partial \zeta} = [\alpha_{l-1/2}(\hat{\mathbf{k}}\times\nabla\psi_{l-1}+\nabla\chi_{l-1})-\alpha_{l-1/2}(\hat{\mathbf{k}}\times\nabla\psi_l+\nabla\chi_l)] \tag{224}$$

$$+[\alpha_{l+1/2}(\hat{\mathbf{k}}\times\nabla\psi_l+\nabla\chi_l)-\alpha_{l+1/2}(\hat{\mathbf{k}}\times\nabla\psi_{l+1}+\nabla\chi_{l+1})].$$

Taking the curl of (224) and using (292), we have

$$\hat{\mathbf{k}}\bullet\left[\nabla\times\frac{\partial \mathbf{V}_l}{\partial t}\right] = \frac{\partial \eta_l}{\partial t} = \ldots-\hat{\mathbf{k}}\bullet\left[\nabla\times\left(\zeta\frac{\partial \mathbf{V}_l}{\partial \zeta}\right)\right] \tag{225}$$

$$= \ldots-\{[(\nabla\bullet(\alpha_{l-1/2}\nabla\psi_{l-1})+J(\alpha_{l-1/2},\chi_{l-1}))-(\nabla\bullet(\alpha_{l-1/2}\nabla\psi_l)+J(\alpha_{l-1/2},\chi_l))]$$

$$[(\nabla\bullet(\alpha_{l+1/2}\nabla\psi_l)+J(\alpha_{l+1/2},\chi_l))-(\nabla\bullet(\alpha_{l+1/2}\nabla\psi_{l+1})+J(\alpha_{l+1/2},\chi_{l+1}))]\}$$

$$= \ldots-\{[\nabla\bullet[\alpha_{l-1/2}\nabla(\psi_{l-1}-\psi_l)]+J(\alpha_{l-1/2},\chi_{l-1}-\chi_l)]$$

$$([\nabla\bullet[\alpha_{l+1/2}\nabla(\psi_l-\psi_{l+1})]+J(\alpha_{l+1/2},\chi_l-\chi_{l+1})]\}.)$$

Taking the divergence of (224) and using (293), we have

$$\nabla\bullet\frac{\partial \mathbf{V}_l}{\partial t} = \frac{\partial \eta_l}{\partial t} = \ldots-\nabla\bullet\left(\zeta\frac{\partial \mathbf{V}_l}{\partial \zeta}\right) \tag{226}$$

$$= \ldots-\{[(\nabla\bullet(\alpha_{l-1/2}\nabla\chi_{l-1})-J(\alpha_{l-1/2},\psi_{l-1}))-(\nabla\bullet(\alpha_{l-1/2}\nabla\chi_l)-J(\alpha_{l-1/2},\psi_l))]$$

$$[(\nabla\bullet(\alpha_{l+1/2}\nabla\chi_l)-J(\alpha_{l+1/2},\psi_l))-(\nabla\bullet(\alpha_{l+1/2}\nabla\chi_{l+1})-J(\alpha_{l+1/2},\psi_{l+1}))]\}$$

$$= \ldots-\{[\nabla\bullet[\alpha_{l-1/2}\nabla(\chi_{l-1}-\chi_l)]-J(\alpha_{l-1/2},\psi_{l-1}-\psi_l)]$$

$$([\nabla\bullet[\alpha_{l+1/2}\nabla(\chi_l-\chi_{l+1})]-J(\alpha_{l+1/2},\psi_l-\psi_{l+1})]\}.)$$

Appendix P shows that the continuous relations $\nabla\bullet(\alpha\nabla\beta)-\nabla\bullet(\alpha\nabla\gamma) = \nabla\bullet[\alpha\nabla(\beta-\gamma)]$ and $J(\alpha,\beta)-J(\alpha,\gamma) = J(\alpha,\beta-\gamma)$ are also true for the finite-difference operators.

Vertical diffusion of momentum

To avoid generating artificially strong winds in layers with negligible mass, a vertical momentum diffusion is included in the model. The diffusion acts only when the thickness of the layers becomes small. In the continuous case, the change of momentum due to the vertical diffusion is

$$\partial \mathbf{V}/\partial t = (1/\rho)\partial \underset{\sim}{\tau}/\partial z, \tag{227}$$

where $\underset{\sim}{\tau}$ is the horizontal stress. The horizontal stress is defined by

$$\begin{aligned} \underset{\sim}{\tau}_{l+1/2} &\equiv \rho_{l+1/2}K^*\sqrt{(\delta z)_l(\delta z)_{l+1}}(\mathbf{V}_l - \mathbf{V}_{l+1}) \\ &= \frac{\rho_{l+1/2}}{g^2}K^*\sqrt{\frac{(\delta p)_l}{\rho_l}\frac{(\delta p)_{l+1}}{\rho_{l+1}}}(\mathbf{V}_l - \mathbf{V}_{l+1}) \text{ for } l = 1, 2, \ldots, L-1, \end{aligned} \tag{228}$$

where K^* is a constant diffusion coefficient. We have used $(\delta z)_l = -(\delta p)_l/g\rho_l$. The horizontal stress at the top and bottom of the model are set $\underset{\sim}{\tau}_{1/2} = \underset{\sim}{\tau}_{L+1/2} = 0$. The density at a layer edge is given by

$$\rho_{l+1/2} = \frac{p_{l+1/2}}{RT_{l+1/2}} = \frac{p_{l+1/2}}{\kappa\theta_{l+1/2}\Pi_{l+1/2}} \tag{229}$$

and at a layer center $\rho_l \equiv (\rho_{l-1/2} + \rho_{l+1/2})/2$. So, (227) becomes

$$\begin{aligned} \frac{\partial \mathbf{V}_l}{\partial t} &= \frac{1}{\rho_l(\delta z)_l}[\underset{\sim}{\tau}_{l-1/2} - \underset{\sim}{\tau}_{l+1/2}] = -g\frac{1}{(\delta p)_l}[\underset{\sim}{\tau}_{l-1/2} - \underset{\sim}{\tau}_{l+1/2}] \\ &= -K^*\frac{1}{(\delta p)_l}\left[\rho_{l-1/2}\sqrt{\frac{(\delta p)_{l-1}}{\rho_{l-1}}\frac{(\delta p)_l}{\rho_l}}(\mathbf{V}_{l-1} - \mathbf{V}_l) - \rho_{l+1/2}\sqrt{\frac{(\delta p)_l}{\rho_l}\frac{(\delta p)_{l+1}}{\rho_{l+1}}}(\mathbf{V}_l - \mathbf{V}_{l+1})\right] \\ &= K^*[\alpha_{l-1/2}(\mathbf{V}_{l-1} - \mathbf{V}_l) - \alpha_{l+1/2}(\mathbf{V}_l - \mathbf{V}_{l+1})] \text{ for } L = 1, 2, \ldots, L, \end{aligned} \tag{230}$$

where

$$\alpha_{l-1/2} = \frac{-\rho_{l-1/2}}{(\delta p)_l}\sqrt{\frac{(\delta p)_{l-1}}{\rho_{l-1}}\frac{(\delta p)_l}{\rho_l}} \text{ and } \alpha_{l+1/2} = \frac{-\rho_{l+1/2}}{(\delta p)_l}\sqrt{\frac{(\delta p)_l}{\rho_l}\frac{(\delta p)_{l+1}}{\rho_{l+1}}} \quad (231)$$

$$\text{for } l = 1, 2, \ldots, L$$

and $\alpha_{1/2} = \alpha_{L+1/2} = 0$. The contribution of vertical diffusion to the momentum equation can then written

$$\hat{k} \bullet \left[\nabla\times\frac{\partial \mathbf{V}_l}{\partial t}\right] = \frac{\partial \eta_l}{\partial t} = \ldots + K^* \hat{k} \bullet [\nabla\times[\alpha_{l-1/2}(\mathbf{V}_{l-1} - \mathbf{V}_l) - \alpha_{l+1/2}(\mathbf{V}_l - \mathbf{V}_{l+1})]] \quad (232)$$

$$= \ldots + K^*\{[\nabla\bullet[\alpha_{l-1/2}\nabla(\psi_{l-1} - \psi_l)] + J(\alpha_{l-1/2}, \chi_{l-1} - \chi_l)]$$

$$-[\nabla\bullet[\alpha_{l+1/2}\nabla(\psi_l - \psi_{l+1})] + J(\alpha_{l+1/2}, \chi_l - \chi_{l+1})]\} \text{ for } l = 1, 2, \ldots, L$$

and

$$\nabla\bullet\frac{\partial \mathbf{V}_l}{\partial t} = \frac{\partial \eta_l}{\partial t} = \ldots K^* \nabla\bullet[\alpha_{l-1/2}(\mathbf{V}_{l-1} - \mathbf{V}_l) - \alpha_{l+1/2}(\mathbf{V}_l - \mathbf{V}_{l+1})]$$

$$= \ldots + K^*\{[\nabla\bullet[\alpha_{l-1/2}\nabla(\chi_{l-1} - \chi_l)] - J(\alpha_{l-1/2}, \psi_{l-1} - \psi_l)]$$

$$-[\nabla\bullet[\alpha_{l+1/2}\nabla(\chi_l - \chi_{l+1})] - J(\alpha_{l+1/2}, \psi_l - \psi_{l+1})]\} \text{ for } l = 1, 2, \ldots, L.$$

Appendix K: Satisfaction of the first two Arakawa constraints in the generalized vertical coordinate

Continuous Equations:

Flux form of the continuity equation:

$$\frac{\partial m}{\partial t} + \nabla \bullet (m\mathbf{V}) + \frac{\partial}{\partial \zeta}(m\dot{\zeta}) = 0. \tag{233}$$

Advective form of the potential temperature equation:

$$\frac{\partial \theta}{\partial t} + \mathbf{V} \bullet \nabla\theta + \dot{\zeta}\frac{\partial \theta}{\partial \zeta} = \frac{Q}{\Pi}. \tag{234}$$

Flux form of the potential temperature equation:

$$\begin{aligned} \frac{\partial}{\partial t}(m\theta) &= m\frac{\partial \theta}{\partial t} + \theta\frac{\partial m}{\partial t} \\ &= m\left[\frac{Q}{\Pi} - \mathbf{V} \bullet \nabla\theta - \dot{\zeta}\frac{\partial \theta}{\partial \zeta}\right] - \theta\left[\nabla \bullet (m\mathbf{V}) + \frac{\partial}{\partial \zeta}(m\dot{\zeta})\right] \\ &= m\frac{Q}{\Pi} - \nabla \bullet (m\theta\mathbf{V}) - \frac{\partial}{\partial \theta}(m\theta\dot{\zeta}). \end{aligned} \tag{235}$$

Derivation of the Continuous Kinetic Energy Equation:

Momentum equations:

$$\frac{\partial \mathbf{V}}{\partial t} + \mathbf{V} \bullet \nabla\mathbf{V} + \dot{\zeta}\frac{\partial \mathbf{V}}{\partial \zeta} = -\nabla_p \Phi - f\hat{\mathbf{k}} \times \mathbf{V}. \tag{236}$$

This is KA (2.12) without the friction term. The horizontal pressure-gradient force can be written

$$
\begin{aligned}
-\nabla_p \Phi &= -\nabla_\zeta \Phi - \frac{1}{m}\frac{\partial \Phi}{\partial \zeta}\nabla_\zeta p \qquad (237)\\
&= -\nabla_\zeta \Phi - \frac{\Phi}{m}\nabla_\zeta m + \frac{\Phi}{m}\nabla_\zeta m - \frac{1}{m}\frac{\partial \Phi}{\partial \zeta}\nabla_\zeta p\\
&= -\frac{1}{m}\nabla_\zeta (m\Phi) + \frac{\Phi}{m}\nabla_\zeta\left(-\frac{\partial p}{\partial \zeta}\right) - \frac{1}{m}\frac{\partial \Phi}{\partial \zeta}\nabla_\zeta p\\
&= -\frac{1}{m}\nabla_\zeta (m\Phi) - \frac{1}{m}\frac{\partial}{\partial \zeta}(\Phi \nabla_\zeta p).
\end{aligned}
$$

This is KA (2.18). Multiplying $m\mathbf{V}$ with the momentum equations

$$m\mathbf{V} \bullet \frac{\partial \mathbf{V}}{\partial t} + m\mathbf{V} \bullet (\mathbf{V} \bullet \nabla\mathbf{V}) + m\mathbf{V} \bullet \left(\dot{\zeta}\frac{\partial \mathbf{V}}{\partial \zeta}\right) = -m\mathbf{V} \bullet \nabla_p \Phi - m\mathbf{V} \bullet (f\hat{\mathbf{k}} \times \mathbf{V}) \qquad (238)$$

or

$$m\frac{\partial}{\partial t}\left(\frac{\mathbf{V}^2}{2}\right) + m\mathbf{V} \bullet \nabla_\zeta\left(\frac{\mathbf{V}^2}{2}\right) + m\dot{\zeta}\frac{\partial}{\partial \zeta}\left(\frac{\mathbf{V}^2}{2}\right) = -m\mathbf{V} \bullet \nabla_p \Phi. \qquad (239)$$

Multiply $\mathbf{V}^2/2$ with the continuity equation

$$\frac{\mathbf{V}^2}{2}\frac{\partial m}{\partial t} + \frac{\mathbf{V}^2}{2}\nabla_\zeta(m\mathbf{V}) + \frac{\mathbf{V}^2}{2}\frac{\partial}{\partial \zeta}(m\dot{\zeta}) = 0. \qquad (240)$$

Combine (239) and (240) gives

$$\frac{\partial}{\partial t}\left(m\frac{\mathbf{V}^2}{2}\right)+\nabla_\zeta\bullet\left(m\frac{\mathbf{V}^2}{2}\mathbf{V}\right)+\frac{\partial}{\partial\zeta}\left(m\frac{\mathbf{V}^2}{2}\dot{\zeta}\right) = -m\mathbf{V}\bullet\nabla_p\Phi \tag{241}$$

$$= -m\mathbf{V}\bullet\left[\frac{1}{m}\nabla_\zeta(m\Phi)+\frac{1}{m}\frac{\partial}{\partial\zeta}(\Phi\nabla_\zeta p)\right]$$

$$= -\mathbf{V}\bullet\nabla_\zeta(m\Phi)-\mathbf{V}\bullet\left[\frac{\partial\Phi}{\partial\zeta}\nabla_\zeta p+\Phi\nabla_\zeta\frac{\partial p}{\partial\zeta}\right]$$

$$= -\mathbf{V}\bullet\nabla_\zeta(m\Phi)-\frac{\partial\Phi}{\partial\zeta}\mathbf{V}\bullet\nabla_\zeta p+\mathbf{V}\bullet\Phi\nabla_\zeta m$$

$$= -\mathbf{V}\bullet m\nabla_\zeta\Phi-\mathbf{V}\bullet\Phi\nabla_\zeta m-\frac{\partial\Phi}{\partial\zeta}\left[\omega-\frac{\partial p}{\partial t}+m\dot{\zeta}\right]+\mathbf{V}\bullet\Phi\nabla_\zeta m$$

$$= -\mathbf{V}\bullet m\nabla_\zeta\Phi-\frac{\partial\Phi}{\partial\zeta}\left[\omega-\frac{\partial p}{\partial t}+m\dot{\zeta}\right]-\Phi\frac{\partial m}{\partial t}-\Phi\nabla_\zeta\bullet(m\mathbf{V})-\Phi\frac{\partial}{\partial\zeta}(m\dot{\zeta})$$

$$= -\nabla_\zeta\bullet(m\Phi\mathbf{V})-\frac{\partial}{\partial\zeta}(\Phi m\dot{\zeta})+\frac{\partial\Phi}{\partial\zeta}\frac{\partial p}{\partial t}+\Phi\frac{\partial}{\partial\zeta}\frac{\partial p}{\partial t}-\frac{\partial\Phi}{\partial\zeta}\omega$$

$$= -\nabla_\zeta\bullet(m\Phi\mathbf{V})-\frac{\partial}{\partial\zeta}(\Phi m\dot{\zeta})+\frac{\partial}{\partial\zeta}\left(\Phi\frac{\partial p}{\partial t}\right)-m\alpha\omega$$

$$= -\nabla_\zeta\bullet(m\Phi\mathbf{V})+\frac{\partial}{\partial\zeta}\left[\Phi\left(\frac{\partial p}{\partial t}-m\dot{\zeta}\right)\right]-m\alpha\omega,$$

where we have used $\omega\equiv\left(\frac{\partial p}{\partial t}\right)_\zeta+\mathbf{V}\bullet\nabla_\zeta p-m\dot{\zeta}$ and the hydrostatic equation of the form

$$\frac{\partial\Phi}{\partial\zeta} = \alpha m.$$

The Continuous Thermodynamic Equation:

Multiply thermodynamic equation (KA (2.10)) with mass:

$$m\frac{\partial}{\partial t}(c_pT)+m\mathbf{V}\bullet\nabla_\zeta T+m\dot{\zeta}\frac{\partial T}{\partial\zeta} = m\alpha\omega+mQ. \tag{242}$$

Combined with continuity gives

$$\frac{\partial}{\partial t}(mc_pT)+\nabla_\zeta\bullet(c_pTm\mathbf{V})+\frac{\partial}{\partial\zeta}(c_pTm\dot{\zeta}) = ma\omega+mQ. \tag{243}$$

Arakawa Constrain 2:

The pressure-gradient force defined at the center of a layer is given by

$$-(\nabla_p \Phi) = -\nabla M_l - \Pi_l \nabla \theta_l \quad \text{for} \quad l = 1, 2, ..., L. \tag{244}$$

Multiply (244) with $m_l \mathbf{V}_i$

$$-m_l \mathbf{V}_l \bullet (\nabla_p \Phi)_l = -m_l \mathbf{V}_l \bullet \nabla M_l + m_l \mathbf{V}_l \bullet (\Pi_l \nabla \theta_l) \tag{245}$$

$$= -m_l \mathbf{V}_l \bullet \nabla(\Pi_l \theta_l + \Phi_l) + m_l \mathbf{V}_l \bullet (\Pi_l \nabla \theta_l)$$

$$= -m_l \mathbf{V}_l \bullet \nabla(\Pi_l \theta_l) - m_l \mathbf{V}_i \bullet \nabla \Phi_l + m_l \mathbf{V}_l \bullet (\Pi_l \nabla \theta_l)$$

$$-\Phi_l \frac{\partial m_l}{\partial t} - \Phi_l \nabla \bullet (m_l \mathbf{V}_l) - \Phi_l \frac{1}{(\delta\zeta)_l}[(m\dot{\zeta})_{l-1/2} - (m\dot{\zeta})_{l+1/2}]$$

$$= -m_l \mathbf{V}_i \bullet (\Pi_l \nabla \theta_l) - m_l \mathbf{V}_i \bullet (\theta_l \nabla \Pi_l) - \nabla \bullet (\Phi_l m_l \mathbf{V}_i) + m_l \mathbf{V}_l \bullet (\Pi_l \nabla \theta_l)$$

$$+\Phi_l \frac{1}{(\delta\zeta)_l}\left[\frac{\partial p_{l-1/2}}{\partial t} - \frac{\partial p_{l+1/2}}{\partial t}\right] - \Phi_l \frac{1}{(\delta\zeta)_l}[(m\dot{\zeta})_{l-1/2} - (m\dot{\zeta})_{l+1/2}]$$

$$= -\nabla \bullet (\Phi_l m_l \mathbf{V}_i) - \theta_l m_l \mathbf{V}_l \bullet \nabla \Pi_l$$

$$+\frac{1}{(\delta\zeta)_l}\left\{\Phi_l\left[\frac{\partial p_{l-1/2}}{\partial t} - (m\dot{\zeta})_{l-1/2}\right] - \Phi_l\left[\frac{\partial p_{l+1/2}}{\partial t} - (m\dot{\zeta})_{l+1/2}\right]\right\} \quad \text{for} \quad l = 1, 2, ..., L.$$

Add and subtract a term of the form

$$\frac{1}{(\delta\zeta)_l}\left\{\Phi_{l-1/2}\left[\frac{\partial p_{l-1/2}}{\partial t} - (m\dot{\zeta})_{l-1/2}\right] - \Phi_{l+1/2}\left[\frac{\partial p_{l+1/2}}{\partial t} - (m\dot{\zeta})_{l+1/2}\right]\right\} \tag{246}$$

from (245) gives

$$-m_l\mathbf{V}_l \bullet (\nabla_p\Phi)_l = -\nabla \bullet (\Phi_l m_l \mathbf{V}_l) \tag{247}$$

$$+\frac{1}{(\delta\zeta)_l}\left\{\Phi_{l-1/2}\left[\frac{\partial p_{l-1/2}}{\partial t}-(m\dot{\zeta})_{l-1/2}\right]-\Phi_{l+1/2}\left[\frac{\partial p_{l+1/2}}{\partial t}-(m\dot{\zeta})_{l+1/2}\right]\right\}-(m\alpha\omega)_l$$

for $l = 1, 2, \ldots, L,$

where

$$-(m\alpha\omega)_l = -\theta_l m_l \mathbf{V}_l \bullet \nabla\Pi_l \tag{248}$$

$$+\frac{1}{(\delta\zeta)_l}\left\{(\Phi_l-\Phi_{l-1/2})\left[\frac{\partial p_{l-1/2}}{\partial t}-(m\dot{\zeta})_{l-1/2}\right]+(\Phi_{l+1/2}-\Phi_l)\left[\frac{\partial p_{l+1/2}}{\partial t}-(m\dot{\zeta})_{l+1/2}\right]\right\}$$

for $l = 1, 2, \ldots, L$

is the discrete form of the energy conversion term. The first two term on the right-hand-side of (247) and analogous to the first two terms on the right-hand-side of (241). The discrete form of the thermodynamic equation is given by

$$\frac{\partial}{\partial t}(m_l c_p T_l) + \nabla \bullet (c_p T_l m_l \mathbf{V}_l) + \frac{1}{(\delta\zeta)_l}[(c_p T m\dot{\zeta})_{l-1/2} - (c_p T m\dot{\zeta})_{l+1/2}] \tag{249}$$

$$= (m\alpha\omega)_l + (mQ)_l \quad \text{for } l = 1, 2, \ldots, L.$$

The Arakawa constrain 2 is satisfied if the energy conversion term in (249) is set to (248).

Arakawa Constraint 1:

The pressure-gradient force defined at the center of a layer is given by

$$-(\nabla_p\Phi) = -\nabla M_l - \Pi_l \nabla\theta_l \quad \text{for} \quad l = 1, 2, \ldots, L. \tag{250}$$

The mass weighed vertical sum

$$-\sum_{l=1}^{L} m_l(\nabla_p\Phi)_l(\delta\zeta)_l \tag{251}$$

$$= -\sum_{l=1}^{L}\nabla(m_lM_l)(\delta\zeta)_l - \sum_{l=1}^{L} M_l\nabla m_l(\delta\zeta)_l - \sum_{l=1}^{L}\nabla(m_l\Pi_l\theta_l)(\delta\zeta)_l + \sum_{l=1}^{L}\theta_l\nabla[\Pi_l m_l]$$

$$= -\sum_{l=1}^{L}\nabla(m_lM_l)(\delta\zeta)_l - \sum_{l=1}^{L} M_l\nabla(p_{l-1/2}-p_{l+1/2})$$

$$-\sum_{l=1}^{L}\nabla(m_l\Pi_l\theta_l)(\delta\zeta)_l + \sum_{l=1}^{L}\theta_l\nabla[\Pi_l(p_{l-1/2}-p_{l+1/2})].$$

Appendix L: A general monotone FCT procedure for tracer transport

Suppose we wish to solve the tracer transport equation in M-dimensional space,

$$\frac{\partial}{\partial t}(m\tau) + \sum_{I=1}^{M} \frac{\partial}{\partial x^I}[(m\tau)u^I] = 0, \tag{252}$$

where $(m\tau) \equiv (m\tau)(t, x^1, \ldots, x^M)$ is the nondiffusive scalar quantity; $u^I \equiv u^I(t, x^1, \ldots, x^M)$ is the I^{th} velocity component for each $I = 1, \ldots, M$. The independent variables are time t and space $x \equiv (x^1, \ldots, x^M)$. Define e_I to be the unit vector in the I^{th} direction. For the geodesic grid, we can think of these fluxes as fluxes through six (or five) cell walls and through the top and bottom edge of the cell. Consider an arbitrary, higher-order-accurate advection algorithm for the integration of (252):

$$(m\tau)_i^{n+1} = (m\tau)_i^n - \sum_{I=1}^{M} (FH_{i+1/2e_I} - FH_{i-1/2e_I}). \tag{253}$$

The high-order FH-flux may be written as the sum of a low-order flux from a nonoscillatory scheme and a residual, or so-called antidiffusive flux, i.e.,

$$FH_{i+1/2e_I} \equiv FL_{i+1/2e_I} + A_{i+1/2e_I}, \tag{254}$$

where (254) defines the residual A-flux. We can think of the A-flux as correcting at least the first-order truncation error terms in the transport of the low-order scheme. The A-flux is traditionally referred to as the "antidiffusive" flux. Using (254) in (253) results in

$$(m\tau)_i^{n+1} = {}_L(m\tau)_i^{n+1} - \sum_{I=1}^{M} (A_{i+1/2e_I} - A_{i-1/2e_I}), \tag{255}$$

where ${}_L(m\tau)_i^{n+1}$ is the solution at time level $n+1$ which results from the advection by the low-order fluxes. Dividing (255) by m_i^{n+1} gives

$$\tau_i^{n+1} \equiv \frac{(m\tau)_i^{n+1}}{m_i^{n+1}} = \frac{{}_L(m\tau)_i^{n+1}}{m_i^{n+1}} - \frac{1}{m_i^{n+1}} \sum_{I=1}^{M} (A_{i+1/2e_I} - A_{i-1/2e_I}). \tag{256}$$

The A-fluxes can be partitioned into fluxes out of a cell and fluxes into a cell as follows

$$A_i^{\text{out}} = \sum_{I=1}^{M} ([A_{i+1/2e_I}]^+ - [A_{i-1/2e_I}]^-) \tag{257}$$
$$A_i^{\text{in}} = -\sum_{I=1}^{M} ([A_{i+1/2e_I}]^- - [A_{i-1/2e_I}]^+),$$

where $[\ ]^+$ denotes the nonnegative operator, and $[\ ]^-$ denotes the nonpositive operator. Note that A_i^{out} and A_i^{in} are both positive quantities. We can write (256) as

$$\tau_i^{n+1} = \frac{(m\tau)_i^{n+1}}{m_i^{n+1}} = \frac{{}_L(m\tau)_i^{n+1}}{m_i^{n+1}} - \frac{1}{m_i^{n+1}}(A_i^{\text{out}} - A_i^{\text{in}}). \tag{258}$$

The amount by which τ is reduced during a given timestep is determined by the magnitude of A_i^{out}. Montonicity implies that

$$\tau^{\min} \le \tau_i^{n+1} \tag{259}$$

for some as yet undetermined $\tau^{\min}$. If A_i^{out} is too large, then (259) may be violated. We wish to determine the largest fraction of A_i^{out} which can be used and still maintain (259). Using (258) and (259) we can write

$$\tau^{\min} = \frac{{}_L(m\tau)_i^{n+1}}{m_i^{n+1}} - \frac{\beta_i^{\mathrm{dn}} A_i^{\mathrm{out}}}{m_i^{n+1}}, \tag{260}$$

where $0 \le \beta_i^{\mathrm{out}} \le 1$. Solving for β_i^{out} gives

$$\beta_i^{\mathrm{out}} = \frac{m_i^{n+1}}{A_i^{\mathrm{out}}}\left[\frac{{}_L(m\tau)_i^{n+1}}{m_i^{n+1}} - \tau^{\min}\right] = \frac{{}_L(m\tau)_i^{n+1} - m_i^{n+1}\tau^{\min}}{A_i^{\mathrm{out}}}. \tag{261}$$

Similarly, we wish to guarantee that $\tau_i^{n+1} \le \tau_i^{\max}$ for some $\tau_i^{\max}$. Using (258), we can write an equation analogous to (260) with this inequality

$$\tau^{\max} = \frac{{}_L(m\tau)_i^{n+1}}{m_i^{n+1}} + \frac{\beta_i^{\mathrm{up}} A_i^{\mathrm{in}}}{m_i^{n+1}} \tag{262}$$

which implies

$$\beta_i^{\mathrm{up}} = \frac{m_i^{n+1}\tau^{\max} - {}_L(m\tau)_i^{n+1}}{A_i^{\mathrm{in}}} \tag{263}$$

and

$$m_i^{\min} \le {}_L(m\tau)_i^{n+1} \le m_i^{\max}. \tag{264}$$

To determine β_i^{up} and β_i^{dn}, one must specify the limiters $m_i^{\min}$ and $m_i^{\max}$. The standard limiter (Zalesak, 1979) is

$$\begin{aligned} m_i^{\min} &= \min(m_{i-e_I}^n, m_i^n, m_{i+e_I}^n, {}_Lm_{i-e_I}^{n+1}, {}_Lm_i^{n+1}, {}_Lm_{i+e_I}^{n+1}) \text{ for } I = 1, \ldots, M \\ m_i^{\max} &= \max(m_{i-e_I}^n, m_i^n, m_{i+e_I}^n, {}_Lm_{i-e_I}^{n+1}, {}_Lm_i^{n+1}, {}_Lm_{i+e_I}^{n+1}) \text{ for } I = 1, \ldots, M. \end{aligned} \tag{265}$$

Appendix M: A tracer equation in the isentropic model

For the tracer equation, we will use the following guidelines:

A) A cell is either massless or not massless, and there exists a test to determine if a cell is massless.

B) The mixing ratio τ of a massless cell is undefined. If a cell is massless, then its mixing ratio cannot be used in any calculation.

The tracer equation will live at layer edges. Consider the flux form of the tracer equation in continuous form,

$$\frac{\partial}{\partial t}(m\tau) = -\nabla \bullet (m\tau\mathbf{V}) - \frac{\partial}{\partial \theta}(m\tau\dot{\theta}). \tag{266}$$

In discrete form we can write (266) as

$$m^{(n+1)}\tau^{(n+1)} = m^{(n)}\tau^{(n)} - \Delta t\left[\nabla \bullet (m\tau\mathbf{V}) + \frac{\partial}{\partial \theta}(m\tau\dot{\theta})\right]^{(n)}, \tag{267}$$

where mass $m_{l+1/2}$ and mixing ratio $\tau_{l+1/2}$ are defined at layer edges, and the superscript (n) denotes timestep. This can be written in terms of discrete horizontal $(m\mathbf{V})_{l+1/2}$ and vertical $(m\dot{\theta})_l$ mass fluxes as follows

$$m^{(n+1)}\tau^{(n+1)} = m^{(n)}\tau^{(n)} - \Delta t\left\{\frac{1}{A}\sum_i \hat{\tau}_i((m\mathbf{V})_{l+1/2})_i + \frac{\hat{\tau}_l(m\dot{\theta})_l - \hat{\tau}_{l+1}(m\dot{\theta})_{l+1}}{(\delta\theta)_{l+1/2}}\right\}, \tag{268}$$

where the summation is over five or six cell walls.

Next, we discuss tracer fluxes for the tracer equation. The horizontal and vertical mass fluxes for the timestep $(n) \rightarrow (n+1)$ come from the FCT of mass. This ensures that the advection of mass and tracer are consistent with one another. However, in the continuity equation, the horizontal mass flux is defined at layer centers. These fluxes must be interpolated to layer edges for the tracer equation. Arakawa and Konor (AK) (1996) write the horizontal mass flux defined at a layer edge as follows:

$$(m\mathbf{V})_{l+1/2} \tag{269}$$
$$= \frac{1}{(\delta\theta)_l + (\delta\theta)_{l+1}}[(\delta\theta)_l(mV)_l + (\delta\theta)_{l+1}(m\mathbf{V})_{l+1}] \text{ for } l = 1, 2, \ldots, L-1$$

which is AK (3.13), and at the top and bottom edges:

$$(m\mathbf{V})_{1/2} = (m\mathbf{V})_1 \quad \text{and} \quad (m\mathbf{V})_{L+1/2} = (m\mathbf{V})_L \tag{270}$$

which is AK (3.18). Similarly, the vertical mass flux which is defined at the layer edges in the continuity equation, must be interpolated to the layer centers in the tracer equation. AK write the vertical mass flux defined at a layer center as follows:

$$(m\dot{\theta})_l = \frac{1}{2}[(m\dot{\theta})_{l-1/2} + (m\dot{\theta})_{l+1/2}] \quad \text{for} \quad l = 2, 3, \ldots, L-1 \tag{271}$$

which is AK (3.14) and

$$(m\dot{\theta})_1 = \frac{1}{2}(m\dot{\theta})_{3/2} \quad \text{and} \quad (m\dot{\theta})_L = \frac{1}{2}(m\dot{\theta})_{L-1/2} \tag{272}$$

since $(m\dot{\theta})_{1/2} = (m\dot{\theta})_{L+1/2} = 0$.

The cell center values can be Adams-Bashforth weighted in time

$$\tau^{AB} = w_3\tau^{(n)} + w_2\tau^{(n-1)} + w_1\tau^{(n-2)}. \tag{273}$$

In (266), $\hat{\tau}_i \equiv f(\tau_0^{(n)}, \tau_0^{AB}, \tau_i^{(n)}, \tau_i^{AB})$ denotes the value of a tracer horizontally interpolated to a cell wall, and $\hat{\tau}_l \equiv f(\tau_{l-1/2}^{(n)}, \tau_{l-1/2}^{AB}, \tau_{l+1/2}^{(n)}, \tau_{l+1/2}^{AB})$ denotes the value of a tracer vertically interpolated to a cell center. So, $\hat{\tau}_i$ and $\hat{\tau}_l$ are functions of center values from the two cells which share the wall or edge. The second guideline above determines the form of the interpolation. Suppose two cells, $cell_0$ and $cell_i$, share a cell wall. We wish to determine the flux of tracer at the wall. There are four cases to consider:

1) Suppose $cell_0$ and $cell_i$ are both not massless, then we can write

$$\hat{\tau}_i = \frac{m_0^{(n)}\tau_0^{AB} + m_i^{(n)}\tau_i^{AB}}{m_0^{(n)} + m_i^{(n)}}. \qquad (274)$$

2) Suppose $cell_0$ and $cell_i$ are both massless, then τ is undefined for both cells. We conclude that the flux of tracer through their common wall is zero.
3) Suppose $cell_0$ is not massless and $cell_i$ is massless. The mass flux $(m\mathbf{V})_i$ is directed from $cell_0$ into $cell_i$, and the flux of tracer can be written $\tau_0^{(n)}(m\mathbf{V})_i$ using only the upstream value of τ. The mass flux cannot be directed in the opposite direction since $cell_i$ is massless.
4) Suppose $cell_0$ is massless and $cell_i$ is not massless. The flux of tracer can be written $\tau_i^{(n)}(m\mathbf{V})_i$, again, using the upstream value.

The above four cases were written in terms of interpolation to a cell wall. Obviously, the same idea will work for interpolation to a cell ceiling or floor.

There are a few more ugly details to consider that pertain to the evolution of τ. Suppose that an arbitrary cell $cell_0$ is massless at time level (n) and not massless at time level $(n+1)$. Then $m^{(n)}\tau^{(n)}$ is zero and $\tau^{(n)}$ undefined. The mixing ratio becomes defined sometime during the timestep, and acquires a value from the surrounding cells. This value is carried within the advected mass. At the end of the timestep, $m^{(n+1)}\tau^{(n+1)} > 0$, or $m^{(n+1)}\tau^{(n+1)} = 0$ if $\tau^{(n+1)} = 0$. We can determine $\tau^{(n+1)}$ by dividing the product $m^{(n+1)}\tau^{(n+1)}$ by $m^{(n+1)}$. Since $\tau^{(n)}$ did not *exist*, we set $\tau^{(n+1)} \to \tau^{(n)}$ and $\tau^{(n+1)} \to \tau^{(n-1)}$. Hence, on the next timestep $\tau^{AB} = \tau^{(n+1)}$. This does not violate any conservation properties, since the τ^{Ab} values are only interpolated to cell walls to determine fluxes. Fluxes can be picked in any arbitrary fashion and still conserve tracer as long as the two cells which share a flux at a wall use equal and opposite values.

Suppose $cell_0$ is not massless at (n) and has become massless at $(n+1)$. Ideally, we want $m^{(n+1)}\tau^{(n+1)} = 0$, however this is numerically unlikely to happen. Therefore, at the end of the timestep we equally distribute any remaining tracer substance, $A(\delta\theta)_{l+1/2}(m\tau)$, to all other not massless cells. Here $A(\delta\theta)_{l+1/2}$ is the volume of the cell and $(m\tau)$ is the tracer per unit volume.

Appendix N: Baroclinic instability initial conditions

Suppose the zonal velocity $u(\varphi, p)$ can be written as the product two functions

$$u(\varphi, p) = u_1(\varphi)u_2(p), \tag{275}$$

where

$$u_1(\varphi) = \begin{cases} \left[\sin\left(\pi\dfrac{\varphi-\varphi_{\min}}{\varphi_{\max}-\varphi_{\min}}\right)\right]^2 & \varphi_{\min} < \varphi \le \varphi_{\max} \\ 0 & \text{otherwise} \end{cases}, \tag{276}$$

where $\varphi_{\min} = 20°$ and $\varphi_{\max} = 70°$, and

$$u_2(p) = u_{\max}\begin{cases} \dfrac{p-p_T}{p_{\max}-p_T} & p_T \le p \le p_{\max} \\ \dfrac{p_S-p}{p_S-p_{\max}} & p_{\max} < p \le p_S \end{cases}, \tag{277}$$

where $u_{\max} = 50\ \mathrm{ms}^{-1}$, $p_T = 10\ \mathrm{hPa}$, $p_{\max} = 250\ \mathrm{hPa}$ and $p_S = 1000\ \mathrm{hPa}$. The profile u_2 is numerically smoothed a little bit to make it more continuous.

The geostrophic wind is given by $f\mathbf{V}_g = \hat{\boldsymbol{k}} \times \nabla_p\Phi$. The zonal component of the geostrophic wind can be manipulated with the thermal wind relation as follows

$$fu = -\frac{\partial\Phi}{(a\partial\varphi)} \tag{278}$$

$$f\frac{\partial u}{\partial p} = -\frac{\partial}{(a\partial\varphi)}\frac{\partial\Phi}{\partial p} = -\frac{\partial}{(a\partial\varphi)}\frac{\partial\Phi}{\partial\Pi}\frac{\partial\Pi}{\partial p} = -\frac{\partial}{(a\partial\varphi)}(-\theta)\frac{\partial\Pi}{\partial p}$$

$$\left(\frac{\partial\Pi}{\partial p}\right)^{-1} f(\varphi)\frac{\partial}{\partial p}[u_1(\varphi)u_2(p)] = \left(\frac{\partial\Pi}{\partial p}\right)^{-1}\frac{\partial u_2}{\partial p}(f(\varphi)u_1(\varphi)) = \frac{\partial\theta}{(a\partial\varphi)},$$

where the hydrostatic equation of the form $\partial\Phi/\partial\Pi = -\theta$ has been used. (and the "g" subscript for geostrophic has been neglected.) Integrating (278) from $\varphi_{\min}$ to φ

$$\theta(\varphi, p) - \theta(\varphi_{min}, p) = \left(\frac{\partial\Pi}{\partial p}\right)^{-1}\frac{\partial u_2}{\partial p}\int_{\varphi_{\min}}^{\varphi} f(\varphi)u_1(\varphi)(ad\varphi). \qquad (279)$$

To determine $\theta(\varphi_{\min}, p)$, the profile of θ above latitude $\varphi_{\min}$, we integrate the static stability

$$N^2 = \frac{g d\theta}{\theta dz} = \frac{g d\theta dp}{\theta dp dz} = \frac{g d\theta}{\theta dp}(-\rho g) = -\frac{g^2 d\theta}{\theta}\frac{p}{dp RT} = -\frac{g^2 d\theta}{\theta}\frac{p}{dp \kappa\Pi\theta} \qquad (280)$$

which can be re-written

$$\frac{\kappa N^2}{g^2 p}\Pi dp = \frac{\kappa N^2}{g^2 p}c_p\left(\frac{p}{p_0}\right)^{\kappa} dp = \frac{\kappa N^2}{g^2}c_p\frac{p^{\kappa-1}}{p_0^{\kappa}}dp = -\frac{1}{\theta^2}d\theta. \qquad (281)$$

Integrating through the column starting from surface values (p_1, θ_1) to values (p, θ) within the atmosphere, we have

$$\frac{N^2 c_p}{g^2 p_0^{\kappa}}\int_{p_1}^{p}\kappa p^{\kappa-1}dp = -\int_{\theta_1}^{\theta}\frac{1}{\theta^2}d\theta \qquad (282)$$

$$\frac{N^2 c_p}{g^2 p_0^{\kappa}}(p^{\kappa} - p_1^{\kappa}) = \frac{1}{\theta} - \frac{1}{\theta_1}.$$

Solve for θ gives

$$\theta(\varphi_{min}, p) = \left[\frac{1}{\theta_1} + \frac{N^2 c_p}{g^2 p_0^k}(p^k - p_1^k)\right]^{-1}. \qquad (283)$$

At pressure greater than 400 hPa, $N^2 = 1.5\times10^{-4}s^{-2}$, and at pressure less than

400 hPa, $N^2 = 3.5\times10^{-4}s^{-2}$ for the increased static stability of the stratosphere. We have set $p_1 = 1000$ hPa and $\theta_1 = 310$ K. With (283) and (279), we can compute $\theta(\varphi, p)$ as a function of p, and, hence, $\zeta = F(p, p_S, \theta)$ as a function of p. Computations of u, θ and ζ are carried-out on a very high resolution latitude-pressure grid. This ensures that interpolation of u, p and θ to our model coordinate surfaces (ζ-surfaces) will be done accurately. Finally, with u on ζ surfaces, we can numerically differentiate to get η on ζ-surfaces.

Appendix O: Continuous properties of the operators

Details of the continuity equation: Consider an arbitrary scalar functions m and suppose there exist scalar functions ψ and χ such that $\mathbf{V} = \hat{\mathbf{k}} \times \nabla\psi + \nabla\chi$, then

$$\begin{aligned} \nabla\bullet(m\mathbf{V}) &= \nabla\bullet[m(\hat{\mathbf{k}} \times \nabla\psi + \nabla\chi)] \\ &= \nabla\bullet(\hat{\mathbf{k}} \times m\nabla\psi) + \nabla\bullet(m\nabla\chi) \\ &= m\nabla\psi \bullet \nabla\times\hat{\mathbf{k}} - \hat{\mathbf{k}} \bullet \nabla\times m\nabla\psi + \nabla\bullet(m\nabla\chi) \\ &= -J(m, \psi) + \nabla\bullet(m\nabla\chi). \end{aligned} \tag{284}$$

Details of the momentum equation: Consider an arbitrary scalar functions η and K, and arbitrary vector function $\mathbf{V}$. Suppose there exist scalar functions ψ and χ such that $\mathbf{V} = \hat{\mathbf{k}} \times \nabla\psi + \nabla\chi$.

The curl of a gradient is zero.

$$\hat{\mathbf{k}} \bullet (\nabla\times\nabla K) = 0. \tag{285}$$

The curl of the remaining vorticity term in the momentum equation is given by

$$\begin{aligned} \hat{\mathbf{k}} \bullet [\nabla\times(\eta\hat{\mathbf{k}} \times \mathbf{V})] &= \hat{\mathbf{k}} \bullet \{\nabla\times[\eta\hat{\mathbf{k}} \times (\nabla\chi + \hat{\mathbf{k}} \times \nabla\psi)]\} \\ &= \hat{\mathbf{k}} \bullet [\nabla\times(\eta\hat{\mathbf{k}} \times \nabla\chi)] + \hat{\mathbf{k}} \bullet [\nabla\times(\eta\hat{\mathbf{k}} \times (\hat{\mathbf{k}} \times \nabla\psi))] \\ &= \nabla\bullet(\eta\nabla\chi) - J(\eta, \psi). \end{aligned} \tag{286}$$

where we have used the following two derivations:

$$\begin{aligned} &\hat{\mathbf{k}} \bullet [\nabla\times(\eta\hat{\mathbf{k}} \times \nabla\chi)] \\ &= \hat{\mathbf{k}} \bullet [\eta\hat{\mathbf{k}}\nabla\bullet\nabla\chi - \nabla\chi(\nabla\bullet\eta\hat{\mathbf{k}}) + (\nabla\chi \bullet \nabla)\eta\hat{\mathbf{k}} - (\eta\hat{\mathbf{k}} \bullet \nabla)\nabla\chi] \\ &= \hat{\mathbf{k}} \bullet [\eta\hat{\mathbf{k}}\nabla^2\chi + (\nabla\chi \bullet \nabla)\eta\hat{\mathbf{k}}] = \eta\nabla^2\chi + \nabla\eta \bullet \nabla\chi \\ &= \nabla\bullet(\eta\nabla\chi) \end{aligned} \tag{287}$$

and

$$\hat{\mathbf{k}} \bullet [\nabla\times(\eta\hat{\mathbf{k}} \times (\hat{\mathbf{k}} \times \nabla\psi))] \tag{288}$$
$$= \hat{\mathbf{k}} \bullet [\nabla\times(\eta\hat{\mathbf{k}} \bullet \nabla\psi - (\eta\hat{\mathbf{k}} \bullet \hat{\mathbf{k}})\nabla\psi)]$$
$$= -\hat{\mathbf{k}} \bullet [\nabla\times(\eta\nabla\psi)]$$
$$= -J(\eta, \psi).$$

The divergence of the vorticity term is given by

$$\nabla\bullet(\eta\hat{\mathbf{k}} \times \mathbf{V}) \tag{289}$$
$$= \nabla\bullet[\eta\hat{\mathbf{k}} \times (\nabla\chi + \hat{\mathbf{k}} \times \nabla\psi)]$$
$$= \nabla\bullet(\eta\hat{\mathbf{k}} \times \nabla\chi) + \nabla\bullet[\eta\hat{\mathbf{k}} \times (\hat{\mathbf{k}} \times \nabla\psi)]$$
$$= -J(\eta, \chi) - \nabla\bullet(\eta\nabla\psi),$$

where we have used the following two derivations:

$$\nabla\bullet(\eta\hat{\mathbf{k}} \times \nabla\chi) = \nabla\chi \bullet \nabla\times\eta\hat{\mathbf{k}} - \eta\hat{\mathbf{k}} \bullet \nabla\times\nabla\chi = \nabla\chi \bullet \nabla\times\eta\hat{\mathbf{k}} \tag{290}$$
$$= \nabla\chi \bullet [\eta\nabla\times\hat{\mathbf{k}} - \hat{\mathbf{k}} \times \nabla\eta] = -\nabla\chi \bullet (\hat{\mathbf{k}} \times \nabla\eta) = -\hat{\mathbf{k}} \bullet (\nabla\eta \times \nabla\chi)$$
$$= -J(\eta, \chi)$$

and

$$\nabla\bullet[\eta\hat{\mathbf{k}}\times(\hat{\mathbf{k}}\times\nabla\psi)] \tag{291}$$

$$= (\hat{\mathbf{k}}\times\nabla\psi)\bullet\nabla\times\eta\hat{\mathbf{k}} - \eta\hat{\mathbf{k}}\bullet\nabla\times(\hat{\mathbf{k}}\times\nabla\psi)$$

$$= (\hat{\mathbf{k}}\times\nabla\psi)\bullet\nabla\times\eta\hat{\mathbf{k}} - \eta\hat{\mathbf{k}}\bullet[\hat{\mathbf{k}}\nabla\bullet\nabla\psi - \nabla\psi\nabla\bullet\hat{\mathbf{k}} + (\nabla\psi\bullet\nabla)\hat{\mathbf{k}} - (\hat{\mathbf{k}}\bullet\nabla)\nabla\psi]$$

$$= (\hat{\mathbf{k}}\times\nabla\psi)\bullet\nabla\times\eta\hat{\mathbf{k}} - \eta\nabla^2\psi$$

$$= (\hat{\mathbf{k}}\times\nabla\psi)\bullet[\eta\nabla\times\hat{\mathbf{k}} - \hat{\mathbf{k}}\times\nabla\eta] - \eta\nabla^2\psi$$

$$= -(\hat{\mathbf{k}}\times\nabla\psi)\bullet(\hat{\mathbf{k}}\times\nabla\eta) - \eta\nabla^2\psi = -\nabla\eta\bullet\nabla\psi - \eta\nabla^2\psi$$

$$= -\nabla\bullet(\eta\nabla\psi).$$

Derivations of curl and divergence of a scalar times a vector: Let α be an arbitrary scale function and $\mathbf{V} = \hat{\mathbf{k}}\times\nabla\psi + \nabla\chi$

$$\hat{\mathbf{k}}\bullet[\nabla\times(\alpha\mathbf{V})] = \hat{\mathbf{k}}\bullet\{\nabla\times[\alpha(\hat{\mathbf{k}}\times\nabla\psi + \nabla\chi)]\} \tag{292}$$

$$= \hat{\mathbf{k}}\bullet[\nabla\times\alpha(\hat{\mathbf{k}}\times\nabla\psi)] + \hat{\mathbf{k}}\bullet\nabla\times(\alpha\nabla\chi)$$

$$= \hat{\mathbf{k}}\bullet[\alpha\nabla\times(\hat{\mathbf{k}}\times\nabla\psi) + \nabla\alpha\times(\hat{\mathbf{k}}\times\nabla\psi)] + J(\alpha,\chi)$$

$$= \hat{\mathbf{k}}\bullet[\alpha\hat{\mathbf{k}}(\nabla\bullet\nabla\psi) - \alpha\nabla\psi(\nabla\bullet\hat{\mathbf{k}}) + \alpha(\nabla\psi\bullet\nabla)\hat{\mathbf{k}}$$

$$-\alpha(\hat{\mathbf{k}}\bullet\nabla)\nabla\psi + (\nabla\alpha\bullet\nabla\psi)\hat{\mathbf{k}} - (\nabla\alpha\bullet\hat{\mathbf{k}})\nabla\psi] + J(\alpha,\chi)$$

$$= \alpha\nabla^2\psi + \nabla\alpha\bullet\nabla\psi + J(\alpha,\chi)$$

$$= \nabla\bullet(\alpha\nabla\psi) + J(\alpha,\chi)$$

and

$$\nabla\bullet(\alpha\mathbf{V}) = \nabla\bullet[\alpha(\hat{\mathbf{k}}\times\nabla\psi + \nabla\chi)] \tag{293}$$

$$= \nabla\bullet[\alpha(\hat{\mathbf{k}}\times\nabla\psi)] + \nabla\bullet(\alpha\nabla\chi)$$

$$= \alpha\nabla\bullet(\hat{\mathbf{k}}\times\nabla\psi) + (\hat{\mathbf{k}}\times\nabla\psi)\bullet\nabla\alpha + \nabla\bullet(\alpha\nabla\chi)$$

$$= \alpha[\nabla\psi\bullet\nabla\times\hat{\mathbf{k}} - \hat{\mathbf{k}}\bullet(\nabla\times\nabla\psi)] + \hat{\mathbf{k}}\bullet(\nabla\psi\times\nabla\alpha) + \nabla\bullet(\alpha\nabla\chi)$$

$$= -J(\alpha,\psi) + \nabla\bullet(\alpha\nabla\chi).$$

Appendix P: Linear properties of the finite-difference operators

Flux Form: Consider arbitrary scalar functions α, β and γ defined at cell centers. Then

$$\nabla\bullet(\alpha\nabla\beta)+\nabla\bullet(\alpha\nabla\gamma) = \frac{1}{A_0}\sum_i\frac{\alpha_0+\alpha_i}{2}\frac{\beta_i-\beta_0}{L_i}l_i+\frac{1}{A_0}\sum_i\frac{\alpha_0+\alpha_i}{2}\frac{\gamma_i-\gamma_0}{L_i}l_i \tag{294}$$
$$= \frac{1}{A_0}\sum_i\frac{\alpha_0+\alpha_i}{2}\frac{(\beta_i+\gamma_i)-(\beta_0+\gamma_0)}{L_i}l_i = \nabla\bullet[\alpha\nabla(\beta+\gamma)]$$

and

$$J(\alpha,\beta)+J(\alpha,\gamma) = \frac{1}{A_0}\sum_i\frac{\alpha_0+\alpha_i}{2}\frac{\beta_{i+1}-\beta_{i-1}}{L_i}l_i+\frac{1}{A_0}\sum_i\frac{\alpha_0+\alpha_i}{2}\frac{\gamma_{i+1}-\gamma_{i-1}}{L_i}l_i \tag{295}$$
$$= \frac{1}{A_0}\sum_i\frac{\alpha_0+\alpha_i}{2}\frac{(\beta_{i+1}+\gamma_{i+1})-(\beta_{i-1}+\gamma_{i-1})}{L_i}l_i = J(\alpha,\beta+\gamma).$$

Advective form: Consider the Eulerian time derivative and horizontal advection terms of the thermodynamic equation

$$\frac{\partial\theta}{\partial t}+\mathbf{V}\bullet\nabla\theta = 0. \tag{296}$$

In continuous form the horizontal advection term can be written in terms of streamfunction and velocity potential as follows

$$\begin{aligned}\mathbf{V}\bullet\nabla\theta &= (\nabla\chi+\hat{\mathbf{k}}\times\nabla\psi)\bullet\nabla\theta \\ &= \nabla\chi\bullet\nabla\theta+(\hat{\mathbf{k}}\times\nabla\psi)\bullet\nabla\theta \\ &= \nabla\chi\bullet\nabla\theta-\hat{\mathbf{k}}\bullet(\nabla\theta\times\nabla\psi) \\ &= \nabla\chi\bullet\nabla\theta-J(\theta,\psi).\end{aligned} \tag{297}$$

In finite-difference form the Jacobian operator is written as

$$J(\theta,\psi) = \frac{1}{6A_0}\sum_i(\theta_0+\theta_i)(\psi_{i+1}-\psi_{i-1}). \tag{298}$$

If θ is a constant at all cells, say $\theta \equiv \theta_{\text{const}}$, then (298) becomes

$$J(\theta_{\text{const}}, \psi) = \frac{\theta_{\text{const}}}{3A_0}\sum_i(\psi_{i+1} - \psi_{i-1}) = 0 \tag{299}$$

for any distribution of ψ. We have a finite-difference operator for the flux-divergence and the Laplacian operators, namely,

$$\nabla \bullet (\theta \nabla \chi) = \frac{1}{A_0}\sum_i \frac{\theta_0 + \theta_i}{2}\frac{\chi_i - \chi_0}{L_i}l_i \text{ and } \nabla^2\chi = \frac{1}{A_0}\sum_i \frac{\chi_i - \chi_0}{L_i}l_i, \tag{300}$$

where L_i is the distance between cell centers and l_i is the length of a cell wall. See (35) and (36) in the 665 blue book. The $\nabla(\ ) \bullet \nabla(\ )$ operator in (297) can be written as follows

$$\nabla\theta \bullet \nabla\chi = \nabla \bullet (\theta\nabla\chi) - \theta\nabla^2\chi \tag{301}$$

$$= \frac{1}{A_0}\sum_i \frac{\theta_0 + \theta_i}{2}\frac{\chi_i - \chi_0}{L_i}l_i - \theta_0\frac{1}{A_0}\sum_i \frac{\chi_i - \chi_0}{L_i}l_i = \frac{1}{A_0}\sum_i \frac{\theta_i - \theta_0}{2}\frac{\chi_i - \chi_0}{L_i}l_i.$$

Again, if θ is a constant at all cells, then (301) says that

$$\nabla\theta_{\text{const}} \bullet \nabla\chi = \frac{1}{A_0}\sum_i \frac{\theta_{\text{const}} - \theta_{\text{const}}}{2}\frac{\chi_i - \chi_0}{L_i}l_i = 0 \tag{302}$$

for any distribution of χ. Note also the (301) has the property $\nabla\theta \bullet \nabla\chi = \nabla\chi \bullet \nabla\theta$. So, if θ is initially uniform at every grid point, then (299) and (302) say that it will remain uniform as the model is integrated in time.

Appendix Q: Multigrid

The model must solve a pair of elliptic equations, $\nabla^2\psi = \eta - f$ and $\nabla^2\chi = \delta$, each timestep. The multigrid algorithm provides an efficient means to do this. Suppose we wish to solve the continuous equation $\nabla^2\alpha = \beta$. In the discrete case, we can denote this problem as

$$\frac{1}{A_i}\sum_{j=1}^{6}\frac{\alpha_{i;j}-\alpha_i}{L_{i;j}}l_{i;j} = \beta_i \quad \text{for} \quad i = 1, 2, \ldots, N, \tag{303}$$

where there are N cells on the grid. For each cell $i = 1, 2, \ldots, N$, there is a linear equation involving unknowns α_i, and six (or five) neighbors $\alpha_{i;1}, \alpha_{i;2}, \ldots, \alpha_{i;6}$ of grid point i. The symbols $L_{i;j}$ and $l_{i;j}$ denote distance between cell centers and the length the cell wall shared by grid point i and its neighbor j, respectively, and A_i the area of cell i. Our goal is to solve the system of linear equations in (303). For a given cell i, (303) can be rearranged as follows:

$$\alpha_i = \frac{\sum_{j=1}^{6}\alpha_{i;j}\frac{l_{i;j}}{L_{i;j}} - A_i\beta_i}{\sum_{j=1}^{6}\frac{l_{i;j}}{L_{i;j}}}. \tag{304}$$

We can define a pointwise iteration as follows: (1) begin with an initial guess at the solution of (303), (2) form a new approximate solution consisting of the right-hand-sides of (304), and (3) replace the initial guess with this new approximation. Repeating the last two steps gives an algorithm analogous to the familiar Jacobi method. Let $\alpha \equiv \{\alpha_i | i=1, 2\ldots, N\}$ denote the true solution of the linear system (303) and $\alpha^m \equiv \{\alpha_i^m | i=1, 2\ldots, N\}$ denote the mth element in the sequence of approximate solutions which result from the pointwise iteration. Let $v^m \equiv \{\alpha_i - \alpha_i^m | i = 1, 2 \ldots, N\}$ be the error associated with the mth approximation. So, hopefully, $|v^m| \to 0$ as $m \to \infty$ in some norm. If we knew v^m then the problem would be solved. To simplify the notation, let $r_{i;j} = A_i / \sum_{j=1}^{6} \frac{l_{i;j}}{L_{i;j}}$ and $r_{i;j} = \frac{l_{i;j}}{L_{i;j}} / \sum_{j=1}^{6} \frac{l_{i;j}}{L_{i;j}}$. Then (303) and (304) can be written more compactly as

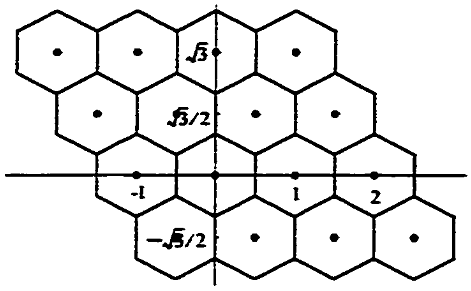
$$\alpha_i = \sum_{j=1}^{6} r_{i;j}\alpha_{i;j} - r_i\beta_i \quad \text{and} \quad \alpha_i^{m+1} = \sum_{j=1}^{6} r_{i;j}\alpha_{i;j}^{m} - r_i\beta_i \tag{305}$$

and error can written

$$v_i^{m+1} = \sum_{j=1}^{6} r_{i;j} v^m{}_{i;j}. \tag{306}$$

It is clear that $\sum_{j=1}^{6} r_{i;j} = 1$. So, (306) is an interpolation or smoothing of the error. Each application of the pointwise iteration will reduce $|v^m|$. If we imagine v^m decomposed into modes, then components of the error which vary smoothly relative to the gridpoint spacing will be reduced only slightly in amplitude for each application of (306). On the other hand, components of v^m which vary rapidly relative to the gridpoint spacing -- in particular, locally change sign -- will be substantially reduced in amplitude. Without loss of generality, suppose the distance between gridpoints is one. See Figure 45 This implies that

Figure 45. Grid of perfect hexagons with a uniform spacing of one.



$r_{i;j} = 1/6$ for all i. Then we can write these modes as

$$E_\mu(x, y) = e^{i(\mu_1 x + \mu_2 y)} \tag{307}$$

where $\mu = (\mu_1, \mu_2)$ is the vector wavenumber. We want to limit μ such that $-\pi \le \mu_1 \le \pi$ and $-2\pi/\sqrt{3} \le \mu_2 \le 2\pi/\sqrt{3}$, so that $E_\mu(x, y)$ is restricted to waves which *fit* on the grid without aliasing error. Suppose that the error v^{m+1} before one iteration of (306) has a component $\gamma_\mu E_\mu$ where γ_μ is the amplitude. If we apply one iteration of (306) we hope that the amplitude of this component will be reduced, and we can define a quantity call convergence factor which is a function of wavenumber as follows

$$\rho(\mu) = \left|\frac{\gamma_\mu^{m=1}}{\gamma_\mu^m}\right|. \tag{308}$$

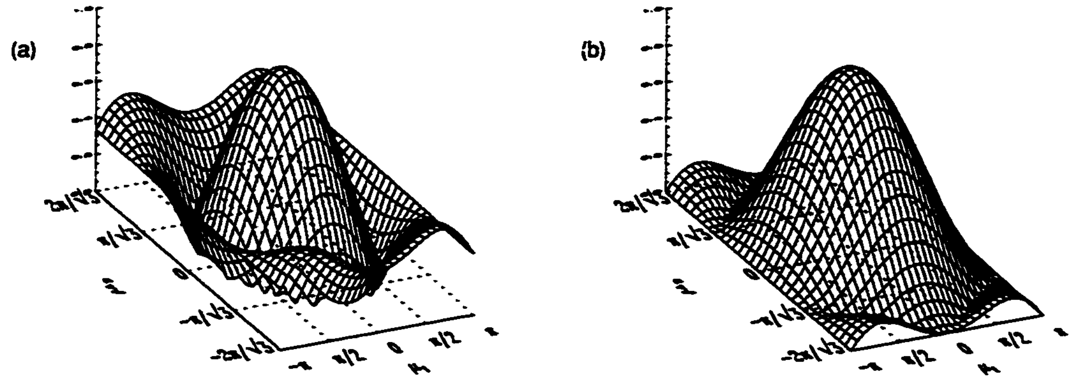


Figure 46(a) Convergence factor $\rho(\mu)$ without under-relaxation. (b) Convergence $\rho(\mu)$ factor with under-relation. The under-relaxed convergence factor will more effectively damp the high wavenumber component.

A plot of $\rho(\mu)$ is shown in Figure 46(a). This makes concrete what was qualitatively described above. For small wavenumber, the component of the error be only slightly reduced. As μ increases, $\rho(\mu)$ decreases, but again increases. So, the shortest wavelength modes are not being reduced as effectively as we might hope.

We might want to under-relax the error smoothing as follows

$$\tilde{v}_i^{m+1} = \omega v_i^m + (1-\omega)\sum_{j=1}^{6} r_{i;j} v^m_{i;j}. \tag{309}$$

If $\omega = 1/3$ in (309), then the resulting convergence factor is shown in Figure 46(b). Hence, the shortest wavelength modes are effectively reduced.