Chapter 12: Finite Differences on the Sphere

12.1 Introduction

Before we can formulate a finite-difference model on the sphere, we must first define a grid that covers the sphere. There are many ways to do this. Perhaps the most obvious possibility is to generate a grid using lines of constant latitude and longitude, i.e., a "spherical coordinate system." A grid derived from a coordinate system is said to be "structured." On a structured grid, indexing can be defined along coordinate lines, so that the neighbors of a particular cell can be specified by defining an index for each coordinate direction, and then simply incrementing the indices to specify neighbors, as we have done many times when using cartesian grids, which are structured grids derived from cartesian coordinates.

It is also possible to define grids without reference to any coordinate system. Examples are planar hexagonal and triangular grids, and spherical grids derived from the icosahedron, the octahedron, and the cube. These are called "unstructured grids," although the name is not very descriptive. With an unstructured grid, the neighbors of each cell are listed in a table, which can be generated once and saved.

The governing equations can be written either with or without a coordinate system. When a coordinate system is used, the components of the wind vector are defined along the coordinate directions. On an unstructured grid, the orientations of the cell walls can be used to define local normal and tangent components of the wind vector. For example, a model using a unstructured C-grid can predict the normal component of the wind on each cell wall.

12.2 Spherical coordinates

We now express the shallow water equations in the spherical coordinate system. In three-dimensional spherical coordinates (λ,φ,r) , i.e., longitude, latitude, and radius, the gradient, divergence, and curl operators take the following forms:

$$\nabla A = \left(\frac{1}{r\cos\varphi}\frac{\partial A}{\partial \lambda}, \frac{1}{r}\frac{\partial A}{\partial \varphi}, \frac{\partial A}{\partial r} \right), \tag{1}$$

$$\nabla \cdot \mathbf{V} = \frac{1}{r\cos\varphi}\frac{\partial V_\lambda}{\partial \lambda} + \frac{1}{r\cos\varphi}\frac{\partial}{\partial \varphi}\left(V_\varphi \cos\varphi\right) + \frac{1}{r^2}\frac{\partial}{\partial r}\left(Vr^2\right) \tag{2}$$

$$\nabla \times \mathbf{V} = \left\{\frac{1}{r}\left[\frac{\partial V_r}{\partial \varphi} - \frac{\partial}{\partial r}\left(rV_\varphi\right)\right], \frac{1}{r}\frac{\partial}{\partial r}\left(rV_\lambda\right) - \frac{1}{r\cos\varphi}\frac{\partial V_r}{\partial \lambda}, \frac{1}{r\cos\varphi}\left[\frac{\partial V_\varphi}{\partial \lambda} - \frac{\partial}{\partial \varphi}\left(V_\lambda \cos\varphi\right)\right]\right\}. \tag{3}$$

For use with the two-dimensional shallow-water equations, we can simplify these to

$$\nabla A = \left(\frac{1}{a\cos\varphi}\frac{\partial A}{\partial \lambda}, \frac{1}{a}\frac{\partial A}{\partial \varphi}\right), \tag{4}$$

$$\nabla \cdot \mathbf{V} = \frac{1}{a\cos\varphi}\frac{\partial V_\lambda}{\partial \lambda_\lambda} + \frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}\left(V_\varphi \cos\varphi\right), \tag{5}$$

$$\mathbf{k} \cdot \left(\nabla \times \mathbf{V}\right) = \frac{1}{a\cos\varphi}\left[\frac{\partial V_\varphi}{\partial \lambda} - \frac{\partial}{\partial \varphi}\left(V_\lambda \cos\varphi\right)\right]. \tag{6}$$

Here a is the radius of the spherical planet.

The shallow water equations in spherical coordinates can be expressed as

$$\frac{\partial u}{\partial t} + \frac{u}{a\cos\varphi}\frac{\partial u}{\partial \lambda} + \frac{v}{a}\frac{\partial u}{\partial \varphi} - \left(f + \frac{u}{a}\tan\varphi\right)v + \frac{g}{a\cos\varphi}\frac{\partial}{\partial \lambda}\left(h + h_S\right) = 0\,, \tag{7}$$

$$\frac{\partial v}{\partial t} + \frac{u}{a\cos\varphi}\frac{\partial v}{\partial \lambda} + \frac{v}{a}\frac{\partial v}{\partial \varphi} + \left(f + \frac{u}{a}\tan\varphi\right)u + \frac{g}{a}\frac{\partial}{\partial \varphi}\left(h + h_S\right) = 0\,, \tag{8}$$

$$\frac{\partial h}{\partial t} + \frac{u}{a\cos\varphi}\frac{\partial h}{\partial \lambda} + \frac{v}{a}\frac{\partial h}{\partial \varphi} + \frac{h}{a\cos\varphi}\left[\frac{\partial u}{\partial \lambda} + \frac{\partial}{\partial \varphi}\left(v\cos\varphi\right)\right] = 0\,. \tag{9}$$

Here h is the depth of the fluid, and h_S is the height of the bottom topography.

In a spherical coordinate system, the lines of constant longitude converge at the poles, so longitude is multivalued at the poles. This means that the components of the wind vector are discontinuous at the poles, although the wind vector itself is perfectly well behaved at the pole. For example, consider a jet directed over the North Pole, represented by the shaded arrow in Fig. 12.1. Measured at points along the prime meridian, the wind consists entirely of a positive v component. Measured along the international date line, however, the wind consists entirely of a negative v component. A discontinuity occurs at the pole, where "north" and "south" have no

meaning. Similarly, the u component of the wind is positive measured near the pole along longitude, and is negative measured along longitude. This problem does not occur in a Cartesian coordinate system centered on the pole. At each point along a great circle that includes the pole,

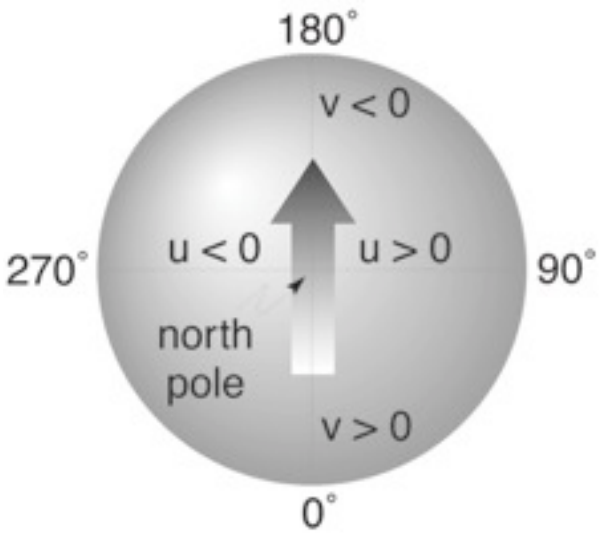

Fig. 12.1: For the wind vector shown in the sketch, points along the prime meridian have a strong northward component. There is a discontinuity at the pole, and points along international date line have a strong southward component. Points near 90° longitude have a strong positive zonal component, while points near 270° longitude have a strong negative zonal component.

the components measured in Cartesian coordinates are well defined and vary continuously.

12.3 Map Projections

An early approach to numerically solving the shallow water equations on the sphere was to project the equations from the sphere to a plane, and solve the equations on a regular grid using a coordinate system defined in the plane. The surface of a sphere and that of a plane are not topologically equivalent, however. In other words, there does not exist a one-to-one mapping g such that for every point on the sphere $(\lambda,\varphi) \in S$ there exists $(x,y) \in P \equiv \{(x,y) | -\infty < x,y < \infty\}$ satisfying $g(\lambda,\varphi) = (x,y)$. It is possible to map almost all of S onto P; examples are given below. Unfortunately, these mappings, also called projections, tend to badly distort distances and areas near the singular points of the associated transformations. Nevertheless, we can use a projection to map the piece of the sphere where the transformation is well behaved onto a finite region of the plane. An approach to map the entire sphere is the composite mesh method, discussed later.

We can derive the equations of motion in various map projections if we first express them in a general orthogonal coordinate system (x,y) , where x and y do not necessarily have the

dimensions of length. Define the metric coefficients to be h_x and h_y so that the distance increment satisfies

$$dl^2 = h_x^2 dx^2 + h_y^2 dy^2 \,. \tag{10}$$

As a matter of notation, the metric coefficients h_x and h_y are distinguished from the depth of the fluid, *h,* by a subscript. In the (x,y) coordinate system, the horizontal velocity components are given by

$$U = h_x \frac{dx}{dt} \,, \tag{11}$$

$$V = h_y \frac{dy}{dt} \,. \tag{12}$$

Williamson (1979) gives the equations of motion for the general velocity components:

$$\frac{DU}{Dt} - \left[f + \frac{1}{h_x h_y} \left(V \frac{\partial h_y}{\partial x} - U \frac{\partial h_x}{\partial y} \right) \right] V + \frac{g}{h_x} \frac{\partial}{\partial x} (h + h_S) = 0 \,, \tag{13}$$

$$\frac{DV}{Dt} + \left[f + \frac{1}{h_x h_y} \left(V \frac{\partial h_y}{\partial x} - U \frac{\partial h_x}{\partial y} \right) \right] U + \frac{g}{h_y} \frac{\partial}{\partial y} (h + h_S) = 0 \,. \tag{14}$$

The Lagrangian time derivative is given by

$$\frac{D}{Dt}(\ ) = \frac{\partial}{\partial t}(\ ) + \frac{U}{h_x} \frac{\partial}{\partial x}(\ ) + \frac{V}{h_y} \frac{\partial}{\partial y}(\ ) \,. \tag{15}$$

The continuity equation can be written as

$$\frac{Dh}{Dt} + \frac{h}{h_x h_y} \left[\frac{\partial}{\partial x} (h_y U) + \frac{\partial}{\partial y} (h_x V) \right] = 0 \,. \tag{16}$$

As an example, with spherical coordinates we have

$$x = \lambda \text{ and } y = \varphi \,, \tag{17}$$

and correspondingly set the metric coefficients to

$$h_x = a\cos\varphi \text{ and } h_y = a \,, \tag{18}$$

then by (11) and (12) we have

$$U = u \equiv a\cos\varphi\frac{D\lambda}{Dt} \text{ and } V = v \equiv a\frac{D\varphi}{Dt} \,. \tag{19}$$

Substituting (18) and (19) into (13), (14) and (16) gives (7), (8) and (9), the shallow water equations in spherical coordinates.

The Polar Stereographic and Mercator projections are sometimes used in modeling the atmospheric circulation. Both are examples of conformal projections, that is, they preserve angles, but not distances. Also, in both of these projections the metric coefficients are independent of direction at a given point, i.e., $h_x = h_y$. The effects of these projections on the

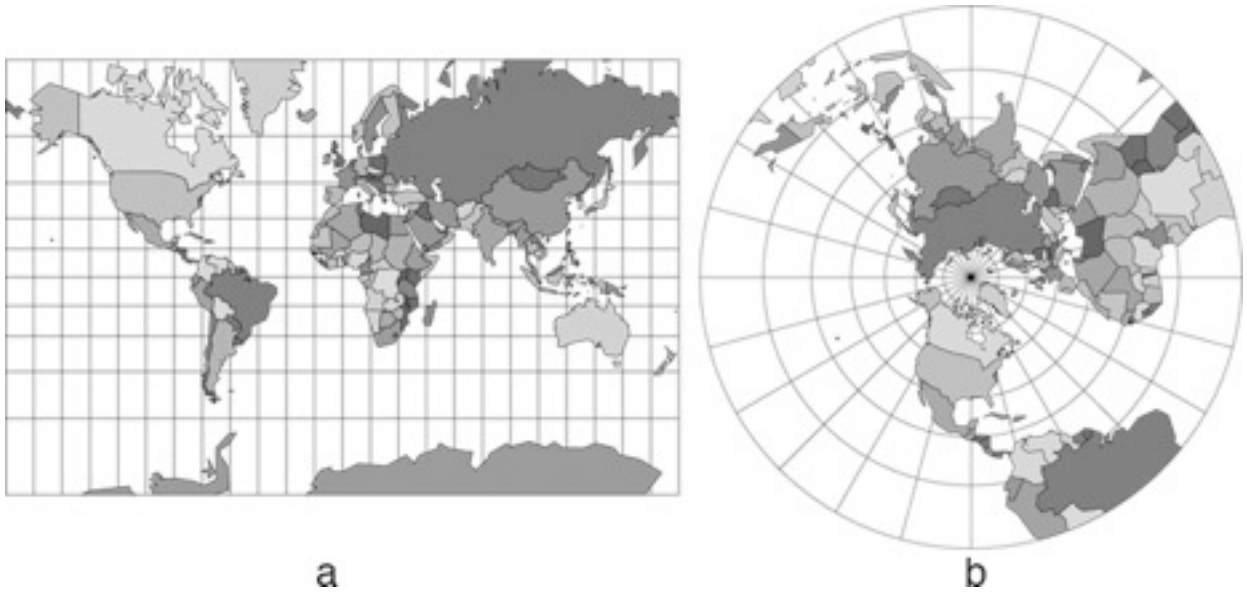

Fig. 12.2: Map projections of the continents: a.) Mercator projection. b.) North polar stereographic projection.

outlines of the continents are shown in Fig. 12.2.

The polar stereographic projection can be visualized in terms of a plane tangent to the Earth at the North Pole. A line drawn from the South Pole that intersects the Earth will also intersect the plane. This line establishes a one-to-one correspondence between all points on the plane and all points on the sphere except for the South Pole itself. In the plane, we can define a Cartesian coordinate system (X, Y), where the positive X axis is in the direction of the image of $\lambda = 0$ (the Greenwich meridian), and the positive Y axis is in the direction of the image of $\lambda = \pi/2$. Obviously, similar mappings can be obtained by placing the plane tangent to the

sphere at points other than the North Pole. Haltiner and Williams (1984) give the equations relating the projection coordinates (*X*, *Y*) and the spherical coordinates (λ,φ):

$$X = \frac{2a\cos\varphi\cos\lambda}{1+\sin\varphi}\,,$$

(20)

$$Y = \frac{2a\cos\varphi\sin\lambda}{1+\sin\varphi}\,.$$

(21)

Note that there is a problem at the South Pole, where the denominators of (20) and (21) go to zero. Taking differentials of (20) and (21) gives

$$\begin{bmatrix} dX \\ dY \end{bmatrix} = \frac{2a}{1+\sin\varphi}\begin{pmatrix} -\cos\varphi\sin\lambda & -\cos\lambda \\ \cos\varphi\cos\lambda & -\sin\lambda \end{pmatrix}\begin{bmatrix} d\lambda \\ d\varphi \end{bmatrix}.$$

(22)

The metrics of the polar stereographic map projection can be determined as follows: Substituting $x=\lambda$, $y=\varphi$, and the metrics for spherical coordinates into $dl^2 = h_x^2 dx^2 + h_y^2 dy^2$ gives

$$dl^2 = \left(a\cos\varphi\right)^2 d\lambda^2 + a^2 d\varphi^2\,.$$

(23)

Solving the linear system (22) for $d\varphi$, and $d\lambda$, and substituting the results into (23), we obtain

$$dl^2 = \left(\frac{1+\sin\varphi}{2}\right)^2 dX^2 + \left(\frac{1+\sin\varphi}{2}\right)^2 dY^2\,.$$

(24)

Comparing (24) with (10), we see that metric coefficients for the polar stereographic projection are given by

$$h_x = h_y = \frac{1+\sin\varphi}{2}\,.$$

(25)

We define the map factor, $m(\varphi)$, as the inverse of the metric coefficient, i.e., $m(\varphi) = 2/(1+\sin\varphi)$. Using (13), (14), and (16), we can write the shallow water equations in north polar stereographic coordinates:

$$\frac{dU}{dt} - \left(f + \frac{UY - VX}{2a^2}\right)V + gm(\varphi)\frac{\partial}{\partial X}(h + h_S) = 0\,,$$

(26)

$$\frac{dV}{dt}+\left(f+\frac{UY-VX}{2a^2}\right)U+gm(\varphi)\frac{\partial}{\partial Y}(h+h_S)=0\,,$$

(27)

$$\frac{dh}{dt}+m^2(\varphi)h\left\{\frac{\partial}{\partial X}\left[\frac{U}{m(\varphi)}\right]+\frac{\partial}{\partial Y}\left[\frac{V}{m(\varphi)}\right]\right\}=0\,.$$

(28)

The total derivative is given by (15).

As discussed above, a finite region of the plane will only map onto a piece of the sphere, and vice versa. One technique to map the entire sphere is to partition it, for example, into hemispheres, and project the pieces separately. Each set of projected equations then gets its boundary conditions from the solutions of the other projected equations. Phillips (1957) divided the sphere into three regions: a tropical belt, and extratropical caps to the north and south of the tropical belt. On each region, the shallow water equations are mapped to a new coordinate system. He used a Mercator coordinate system in the tropics, a polar stereographic coordinate system fixed to the sphere at the North Pole for the northern extratropical cap, and similarly, a polar stereographic coordinate system fixed to the sphere at the South Pole for the southern extratropical cap. When a computational stencil required data from outside the region covered by its coordinate system, that piece of information was obtained by interpolation within the neighboring coordinate system. The model proved to be unstable at the boundaries between the coordinate systems.

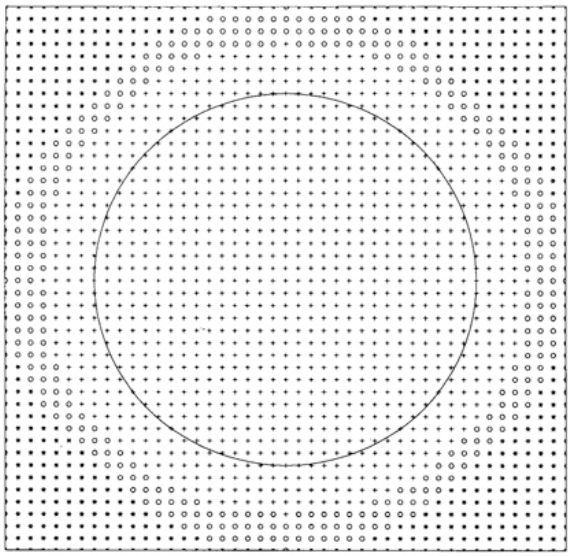

Fig. 12.3: Composite grid method grid. Two such grids are used to cover the sphere. Points labeled with are the boundary conditions for the points labeled with +. Values at the points are obtained by interpolation from the other grid. The big circle is the image of the Equator. Points labeled * are not used.

Browning (1989) discussed a different composite mesh model in which the Northern and Southern Hemispheres are mapped to the plane with a polar stereographic projection. The equations used for the northern projection are just (26), (27), and (28). The equations for the southern projection are the same as those for the northern, except for a few sign differences. This model is different from Phillips' in that the regions interior to the coordinate systems overlap a little bit as shown in Fig. 12.3. Values for dependent variables at grid points not covered by the current coordinate system are obtained by interpolation in the other coordinate system. The overlapping of the coordinate systems makes this scheme more stable than in Phillips' model, in which the coordinate systems were simply butted together at a certain latitude. This model is also easier to write computer code for because the equations are only expressed in the polar stereographic coordinate systems. Browning tested the model and reported good results.

12.4 The "pole problem," and polar filters

One eighth of a uniform latitude-longitude grid is shown in Fig. 12.4. The zonal rows of grid points nearest the two poles consist of "pizza slices" which come together at a point at each pole. The other zonal rows consist of grid points which are roughly trapezoidal in shape. There are other ways to deal with the polar regions, e.g., by defining local Cartesian coordinates at the

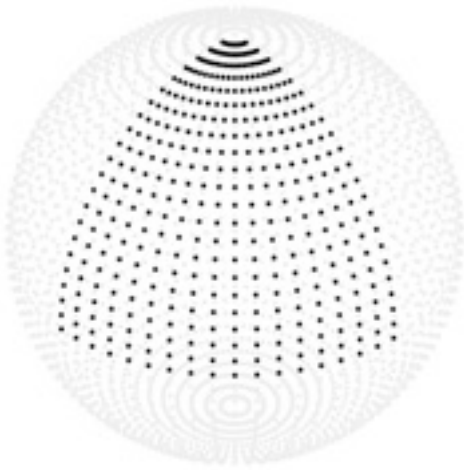

Fig. 12.4: One octant of the latitude-longitude grid used by Arakawa and Lamb (1981). In the example shown, there are 72 grid points around a latitude circle and 44 latitude bands from pole to pole. The longitudinal grid spacing is globally uniform, and in this example is 5 °. The latitudinal grid spacing is globally uniform except for "pizza slices" ringing each pole, which are 1.5 times as "tall" as the other grid cells. The reason for this is explained by Arakawa and Lamb (1981). In the example shown here, the latitudinal grid spacing is 4° except that the pizza slices are 6° tall.

poles.

The scales of meteorological action do not vary dramatically from place to place, nor do the meridional and zonal scales of the circulations of interest differ very much. This suggests that average distance between neighboring grid points should not depend on location, and also that the distances between grid points in the zonal direction should not be substantially different from the distances in the meridional direction. Latitude-longitude grids lack these two desirable properties.

In addition, the convergence of the meridians at the poles demands a short time step in order to satisfy the Courant-Friedrichs-Lewy (CFL) requirement for computational stability, as

discussed in Chapters 4 (for advection) and 7 (for wave propagation). This is often referred to as "the pole problem." To derive the stability criterion for the shallow water equations on the sphere, following Arakawa and Lamb (1977), we begin by linearizing (7), (8), and (9) about a state of rest, as follows:

$$\frac{\partial u}{\partial t} + \frac{g}{a\cos\varphi}\frac{\partial h}{\partial \lambda} = 0\,,$$

(29)

$$\frac{\partial v}{\partial t} + \frac{g}{a}\frac{\partial h}{\partial \varphi} = 0\,,$$

(30)

$$\frac{\partial h}{\partial t} + \frac{H}{a\cos\varphi}\left[\frac{\partial u}{\partial \lambda} + \frac{\partial}{\partial \varphi}\left(v\cos\varphi\right)\right] = 0\,.$$

(31)

Here we have neglected rotation and bottom topography, for simplicity, and H denotes the mean depth of the fluid. We spatially discretize (29) - (31) suing the C-grid, as follows:

$$\frac{du_{i+\frac{1}{2},j}}{dt} + \frac{g\left(h_{i+1,j} - h_{i,j}\right)}{a\cos\varphi\Delta\lambda} = 0\,,$$

(32)

$$\frac{dv_{i,j+\frac{1}{2}}}{dt} + \frac{g\left(h_{i,j+1} - h_{i,j}\right)}{a\Delta\varphi} = 0\,,$$

(33)

$$\frac{dh_{i,j}}{dt} + H\left\{\frac{\left(u_{i+\frac{1}{2},j} - u_{i-\frac{1}{2},j}\right)}{a\cos\varphi_j\Delta\lambda} + \left[\frac{\left(v\cos\varphi\right)_{i,j+\frac{1}{2}} - \left(v\cos\varphi\right)_{i,j-\frac{1}{2}}}{a\cos\varphi_j\Delta\varphi}\right]\right\} = 0\,.$$

(34)

We look for solutions of the form

$$u_{i+\frac{1}{2},j} = \mathrm{Re}\left\{\hat{u}_j \exp\left[\bar{i}s\left(i+\frac{1}{2}\right)\Delta\lambda + \bar{i}\,\sigma t\right]\right\},$$

(35)

$$v_{i,j+\frac{1}{2}} = \mathrm{Re}\left\{\hat{v}_{j+\frac{1}{2}} \exp\left[\bar{i}si\Delta\lambda + \bar{i}\,\sigma t\right]\right\},$$

(36)

$$h_{i,j} = \mathrm{Re}\left\{\hat{h}_j \exp\left[\bar{i}si\Delta\lambda + \bar{i}\,\sigma t\right]\right\},$$

(37)

where $\bar{i} \equiv \sqrt{-1}$. Note that the zonal wave number, *s*, is defined with respect to longitude rather than distance, and that the "hat" variables depend on latitude. Bu substitution of (35) - (37) into (32) - (34), we obtain

$$\bar{i}\,\sigma\hat{u}_j + \frac{\bar{i}s}{a\cos\varphi_j}\frac{\sin\left(s\Delta\lambda/2\right)}{s\Delta\lambda/2}g\left[S_j(s)\hat{h}_j\right] = 0\,,$$

(38)

$$\bar{i}\,\sigma\hat{v}_{j+\frac{1}{2}} + g\left(\frac{\hat{h}_{j+1} - \hat{h}_j}{a\Delta\varphi}\right) = 0\,,$$

(39)

$$\bar{i}\,\sigma\hat{h}_j + H\left\{\frac{\bar{i}s}{a\cos\varphi_j}\frac{\sin\left(s\Delta\lambda/2\right)}{s\Delta\lambda/2}S_j(s)\hat{u}_j + \left[\frac{\left(\hat{v}\cos\varphi\right)_{j+\frac{1}{2}} - \left(\hat{v}\cos\varphi\right)_{j-\frac{1}{2}}}{a\cos\varphi_j\Delta\varphi}\right]\right\} = 0\,,$$

(40)

where $S_j(s)$ *is an artificially inserted "smoothing parameter" that depends on wave number.* The smoothing parameter appears in the term of (38) corresponding to the *zonal* pressure gradient force, and also in the term of (40) corresponding to the *zonal* mass flux divergence. These are the key terms for zonally propagating gravity waves. Later in this discussion, $S_j(s)$ will be set to values less than unity, in order to allow computational stability with a "large" time step near the pole. For now, just consider it to be equal to one.

By eliminating $\hat{u}_j$ and $\hat{v}_{j+\frac{1}{2}}$ in (38) - (40), we can obtain the "meridional structure equation" for $\hat{h}_j$:

$$c^2\left[\frac{s}{a\cos\varphi_j}\frac{\sin\left(s\Delta\lambda/2\right)}{s\Delta\lambda/2}S_j(s)\right]^2\hat{h}_j + \frac{c^2}{\left(a\Delta\varphi\right)^2}\left[\left(\hat{h}_j - \hat{h}_{j-1}\right)\frac{\cos\varphi_{j-\frac{1}{2}}}{\cos\varphi_j} - \left(\hat{h}_{j+1} - \hat{h}_j\right)\frac{\cos\varphi_{j+\frac{1}{2}}}{\cos\varphi_j}\right] = \sigma^2\hat{h}_j\,.$$

(41)

Here $c^2 \equiv gH$ is the square of the phase speed of a pure gravity wave. For high values of the zonal wave number *s*, the first term on the left-hand side of (41) dominates the second, and we obtain

$$\sigma \cong |c| \left[\frac{s}{a\cos\varphi_j} \frac{\sin(s\Delta\lambda/2)}{s\Delta\lambda/2} S_j(s) \right]$$
$$= 2|c| \frac{S_j(s)\sin\left(\frac{s\Delta\lambda}{2}\right)}{a\cos\varphi_j \Delta\lambda}.$$

(42)

Although we have not used a specific time differencing scheme here, we know that for a conditionally stable scheme the CFL criterion takes the form

$$\sigma\Delta t < \varepsilon ,$$

(43)

where ε is a constant of order one. In view of (42) and (43), the CFL criterion will place more stringent conditions on Δt as $\cos\varphi_j$ decreases, i.e., near the poles. In addition, the criterion becomes more stringent as s increases, at a given latitude. Putting $S_j(s) = 1$ temporarily, and assuming $\varepsilon = 1$, we can write the stability condition for zonal wave number s as

$$\frac{|c|\Delta t}{a\cos\varphi_j \Delta\lambda} \sin\left(\frac{s\Delta\lambda}{2}\right) < \frac{1}{2} .$$

(44)

The worst case is $\sin\left(\frac{s\Delta\lambda}{2}\right) = 1$, for which (44) reduces to

$$\frac{|c|\Delta t}{\Delta x_j} < \frac{1}{2} ,$$

(45)

where we define $\Delta x_j \equiv a\cos\varphi_j \Delta\lambda$. For the grid shown in Fig. 12.4, with a longitudinal grid spacing of $\Delta\lambda = 5^\circ$ and a latitudinal grid spacing of $\Delta\varphi = 4^\circ$ (the values used to draw the figure), the northernmost row of grid points where the zonal component of velocity is defined is at latitude 86 ˚N. The zonal distance between grid points there is $\Delta x \cong 39$ km, which is less than one-tenth the zonal grid spacing at the Equator. Recall that the fast, external gravity wave has a phase speed of approximately 300 m s^{-1} . Substituting into (45), we find that the largest permissible time step near the pole is about 70 seconds. This is roughly one tenth of the largest permissible time step at the Equator.

It would be nice if the CFL criterion were the same at all latitudes, permitting time steps near the pole as large as those near the Equator. In order to make this possible, models that use latitude-longitude grids typically employ “polar filters” that prevent computational instability, so that a longer time step can be used.

The simplest method is to use a Fourier filter to remove the high-wave number components of the prognostic fields, near the poles. This can prevent a model from blowing up, but it leads to drastic violations of mass conservation (and many other conservation principles). The cure is almost as bad as the disease.

A better approach is to longitudinally smooth the longitudinal pressure gradient in the zonal momentum equation and the longitudinal contribution to the mass flux divergence in the continuity equation. This has the effect of reducing the zonal phase speeds of the gravity waves sufficiently so that the CFL criterion is not violated.

Inspection of (42) shows that this can be accomplished by choosing the smoothing parameter $S_j(s)$ so that

$$\frac{S_j(s)\sin\left(\frac{s\Delta\lambda}{2}\right)}{a\cos\varphi_j\Delta\lambda} = \frac{1}{d^*},$$

(46)

where d^* is a suitably chosen length, comparable to the zonal grid spacing at the Equator. When $S_j(s)$ satisfies (46), Eq. (42) reduces to

$$\sigma = \frac{2|c|}{d^*},$$

(47)

and the CFL condition reduces to

$$\frac{|c|\Delta t}{d^*} < \frac{\varepsilon}{2},$$

(48)

so that the time step required is independent of latitude, as desired. If we choose

$$d^* \equiv a\Delta\varphi,$$

(49)

i.e., the distance between grid points in the meridional direction, then, referring back to (46), we see that $S_j(s)$ must satisfy

$$S_j(s) = \left(\frac{\Delta\lambda}{\Delta\varphi}\right)\frac{\cos\varphi_j}{\sin\left(\frac{s\Delta\lambda}{2}\right)}.$$

(50)

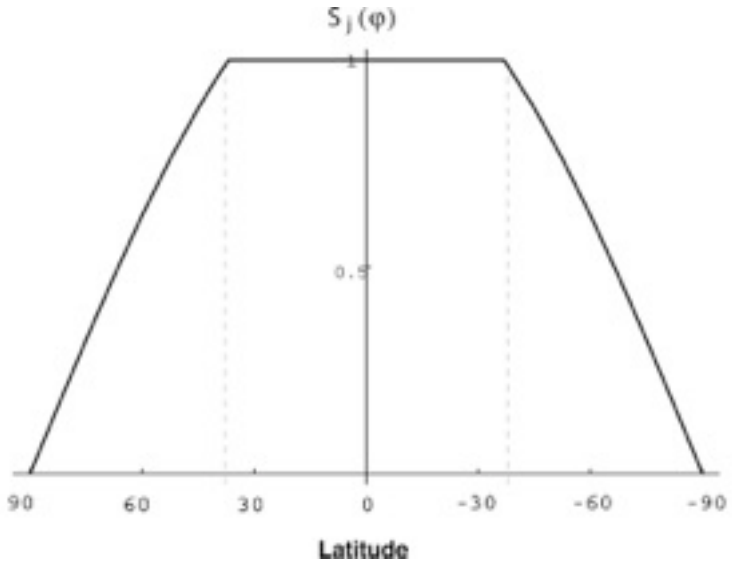

Fig. 12.5: A plot of the smoothing parameter as given by (12.51), for the "worst case" of the shortest zonal mode. The dashed vertical lines demarcate the belt of latitude centered on the Equator for which no smoothing is needed. It has been assumed that $\Delta\lambda = (5/4)\Delta\varphi$, which is true for the grid shown in Fig. 12.4.

Of course, at low latitudes (50) can give values of $S_j(s)$ which are greater than one; these should be replaced by one, so that we actually use

$$S_j(s) = Min\left\{\left(\frac{\Delta\lambda}{\Delta\varphi}\right)\frac{\cos\varphi_j}{\sin\left(\frac{s\Delta\lambda}{2}\right)},1\right\}. \tag{51}$$

A plot of (51) is given in Fig. 12.5, for the case of the shortest zonal mode. The plot shows that some smoothing is needed all the way down into the subtropics.

12.5 Kurihara's grid

Many authors have sought alternatives to the latitude-longitude grid, hoping to make the grid spacing more uniform, still within the "latitude-longitude" framework.

For example, Kurihara (1965) proposed a grid in which the number of grid points along a latitude circle varies with latitude. By placing fewer points at higher latitudes, he was able to more homogeneously cover the sphere. The grid is constructed by evenly placing $N+1$ grid points along the longitude meridian, from the North Pole to the Equator. The point at the North Pole is given the label $j = 1$, the next latitude circle south is given the label $j = 2$, and so on until the Equator is labeled $j = N+1$. Along latitude circle j there are $4(j-1)$ equally spaced grid

points, except at each pole, where there is a single point. One octant of the sphere is shown in Fig. 12.6; compare with Fig. 12.4. For a given *N*, the total number of grid points on the sphere is

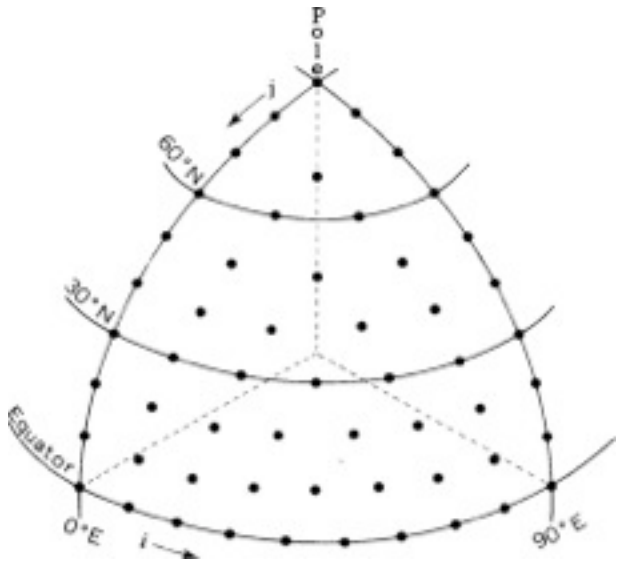

Fig. 12.6: Kurihara grid on one octant of the sphere. Compare with Fig. 12.4.

$4N^2 + 2$. The Southern Hemisphere grid is a mirror image of the Northern Hemisphere grid.

We can measure the homogeneity of the grid by examining the ratio of the zonal distance, $a\cos\varphi_j \Delta\lambda_j$, and the meridional distance $a\Delta\varphi$, for a grid point at latitude φ_j. Here, $\Delta\varphi \equiv \frac{\pi}{2}\frac{1}{N}$ and $\Delta\lambda_j \equiv \frac{1}{j-1}$. At $j = N+1$, the Equator, the ratio is one, and near the pole the ratio approaches $\pi/2 \cong 1.57$.

Kurihara built a model using this grid, based on the shallow water equations. He tested it in a simulation of the Rossby-Haurwitz wave, with zonal wave number 4 as the initial condition. This set of initial conditions was also used by Phillips (1959), and later in the suite of seven test cases for shallow water models proposed by Williamson et al. (1992). The model was run with a variety of time-stepping schemes and with varying amounts of viscosity. Each simulation covered 16 simulated days, with $N = 20$. The Rossby-Haurwitz wave should move from west to east, without distortion. In several of Kurihara's runs, however, the wave degenerated to higher wave numbers.

12.6 The Wandering Electron Grid

An approach to constructing a mesh of grid points that homogeneously covers a sphere is to model the equilibrium distribution of a set of electrons confined to the surface of the sphere.

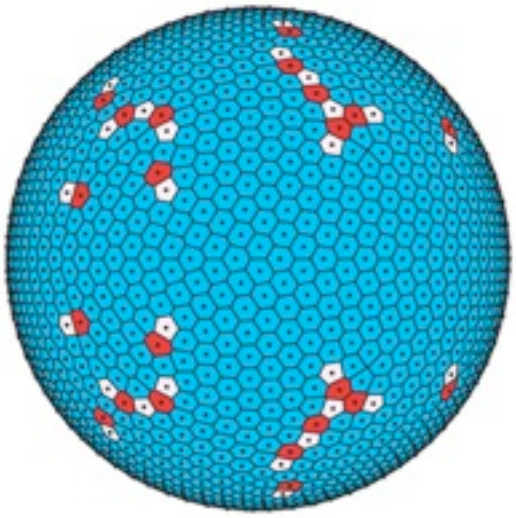

Fig. 12.7: Wandering electron grid. White cells have five walls, light gray cells have six walls, and dark gray cells have seven walls.

Because each electron is repelled by every other electron, it will move to maximize the distance between it and its closest neighbors. In this way, the electrons will distribute themselves as evenly as possible over the sphere. We associate a grid point with each electron. It seems advantageous to constrain the grid so that it is symmetric across the Equator. An Equator can be defined by restricting the movement of a subset of the electrons to a great circle. The remaining electrons can be paired so that each has a mirror image in the opposite hemisphere. We can also fix an electron at each of the poles. Experience shows that unless we fix the positions of some of the electrons, their positions tend to wander indefinitely. Fig. 12.7. shows a grid constructed using the wandering electron algorithm. Most cells have six walls, but some have five or seven walls. While this approach more or less homogeneously covers the sphere, it is not very satisfactory.

12.7 Unstructured spherical grids

Fig. 12.8 shows several discretizations of the sphere. The left-most panel shows the structured latitude-longitude grid. The second and third panels show triangular and hexagonal-pentagonal grids, respectively, generated from the icosahedron.

The fourth panel shows a “cubed sphere” grid, generated from the sphere (e.g., Ronchi et al., 1996; Nair et al., 2005; Putman and iLn, 2007; Lauritzen and Nair, 2008; Ulrich et al., 2009). The cells of the cubed sphere grid are quadrilaterals.

The last panel shows the “Ying-Yang” grid proposed by Kageyama and Satoh (2004), and Kagayama (2005). The grid is composed of two “sleeves” that wrap together like the leather patches that are stitched together to cover the outside of a baseball. The sleeves overlap slightly, and an interpolation is used to patch them together, something like the methods used to patch together two polar stereographic grids. Overlapping grids of this type are sometimes called

"overset grids." There have also been recent attempts to use grids based on octahedrons (e.g., McGregor, 1996; Purser and Rancic, 1998). A "Fibonacci grid" has also been suggested (Swinbank and Purser, 2006).

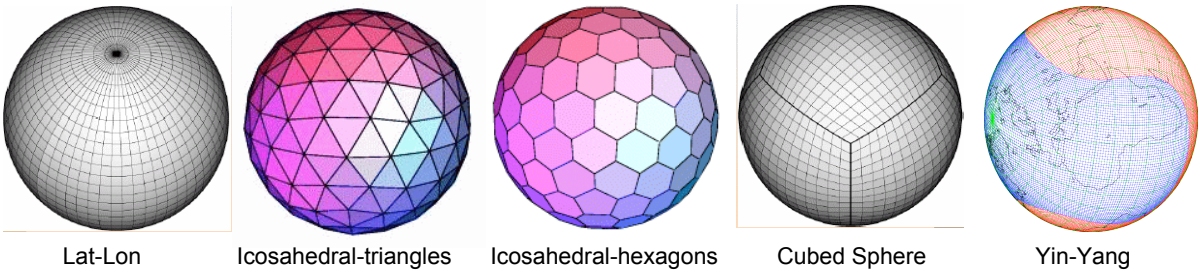

Fig. 12.8: Various ways of discretizing the sphere. Figure made by Bill Skamarock of NCAR.

Grids based on icosahedra offer an attractive framework for simulation of the global circulation of the atmosphere; their advantages include almost uniform and quasi-isotropic resolution over the sphere. Such grids are termed "geodesic," because they resemble the geodesic domes designed by Buckminster Fuller. Williamson (1968) and Sadourny (1968) simultaneously introduced a new approach to more homogeneously discretize the sphere. They constructed grids using spherical triangles which are equilateral and nearly equal in area. Because the grid points are not regularly spaced and do not lie in orthogonal rows and columns, alternative finite-difference schemes are used to discretize the equations. Initial tests using the grid proved encouraging, and further studies were carried out. These were reported by Sadourny et al. (1968), Sadourny and Morel (1969), Sadourny (1969), Williamson (1970), and Masuda (1986).

The grids are constructed from an icosahedron (20 faces and 12 vertices), which is one of the five Platonic solids. A conceptually simple scheme for constructing a spherical geodesic grid is to divide the edges of the icosahedral faces into equal lengths, create new smaller equilateral triangles in the plane, and then project onto the sphere. See Fig. 12.9. One can construct a more homogeneous grid by partitioning the spherical equilateral triangles instead. Williamson (1968) and Sadourny (1968) use slightly different techniques to construct their grids. However, both begin by partitioning the spherical icosahedral triangle. On these geodesic grids, all but twelve of the cells are hexagons. The remaining twelve are pentagons. They are associated with the twelve vertices of the original icosahedron.

Williamson (1968) chose the nondivergent shallow water equations to test the new grid. He solved the two-dimensional nondivergent vorticity equation

$$\frac{\partial \zeta}{\partial t} = -J(\psi, \eta), \tag{52}$$

where ζ is relative vorticity, $\eta = \zeta + f$ is absolute vorticity and ψ is the stream function, such that

$$\zeta = \nabla^2 \psi .$$

(53)

For arbitrary functions α and β, it follows from the form $J(\alpha,\beta) = \mathbf{k} \cdot \nabla \times (\alpha \nabla \beta)$ that the Jacobian satisfies

$$J(\alpha,\beta) = \lim_{A \to 0} \left\{ \frac{1}{A} \oint_S \alpha \frac{\partial \beta}{\partial s} ds \right\},$$

(54)

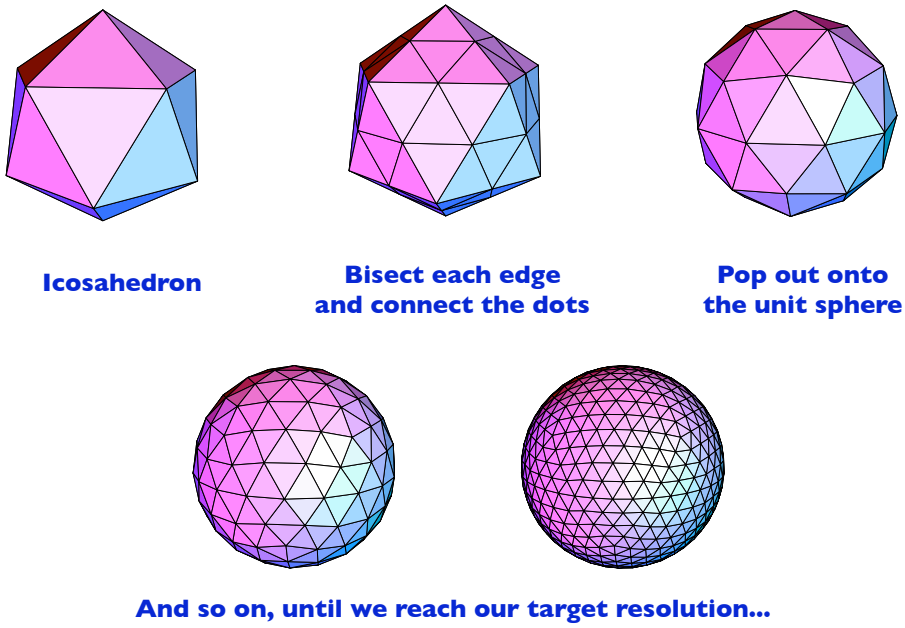

Fig. 12.9: A spherical geodesic grid is generated recursively by starting from an icosahedron.

where A is a small area, and s measures distance along the curve bounding A. Integrating (52) over the area A, and using (54), we get

$$\frac{d}{dt} \int_A \zeta \, dA = -\oint_S \frac{\partial \psi}{\partial s} \eta \, ds .$$

(55)

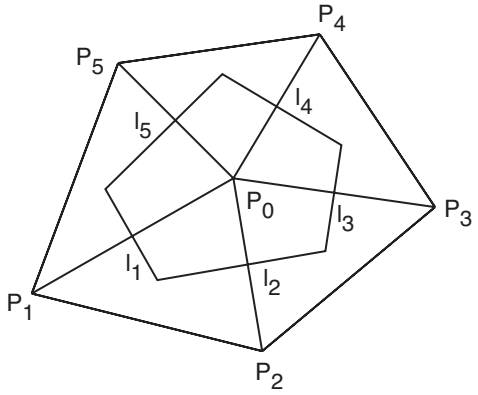

Fig. 12.10: Configuration of grid triangles for the case of a pentagon.

This can be discretized with reference to Fig. 12.10. We approximate the line integral along the polygon defined by the path $P_1, P_2, \ldots P_5, P_1$. Let ζ_0 be the relative vorticity defined at the point P_0, and let η_i be the absolute vorticity defined at the point P_i. We can approximate (55) by

$$\frac{d\zeta_0}{dt} = \frac{1}{A}\sum_{i=1}^{K}\left(\frac{\psi_{i+1}-\psi_{i-1}}{\Delta s}\right)\left(\frac{\eta_0+\eta_i}{2}\right)\Delta s\,. \tag{56}$$

We must also discretize the Laplacian. Consider the smaller, inner polygon in Fig. 12.10. Its walls are formed from the perpendicular bisectors of the line segments $\overline{P_0P_i}$. We can use Gauss's Theorem to write

$$\int_a \zeta\, dA = -\oint_{s'} (\nabla\psi)\cdot\mathbf{n}\, ds'\,, \tag{57}$$

where a is the area of the small polygon, s' is its boundary, and $\mathbf{n}$ is the outward-normal unit vector on the boundary. Eq. (57) is approximated by

$$a\zeta_0 = \sum_{i=1}^{K}\frac{l_i}{|P_0P_i|}(\psi_i-\psi_0)\,, \tag{58}$$

where $|P_0P_i|$ is the distance from P_0 to P_i, and l_i is the length of wall i. Eq. (58) can be solved for ψ_i by relaxation, using the methods discussed in Chapter 6.

Williamson showed that his scheme conserves kinetic energy and enstrophy, as the exact equations do. When applied to regular grid on a plane, the scheme is second-order accurate.

Williamson performed a numerical experiment, using a Rossby-Haurwitz wave as the initial condition. A run of 12 simulated days produced good results. Sadourny et al. (1968) discussed a nondivergent model very similar to Williamson's. Also, Sadourny and Morel (1969) developed a geodesic-grid model based on the free-surface shallow water equations.

Masuda (1986) developed an elegant algorithm for solving the shallow water equations on the sphere. He used the Z-grid (see Chapter 7). Like Williamson, Masuda chose the Rossby-

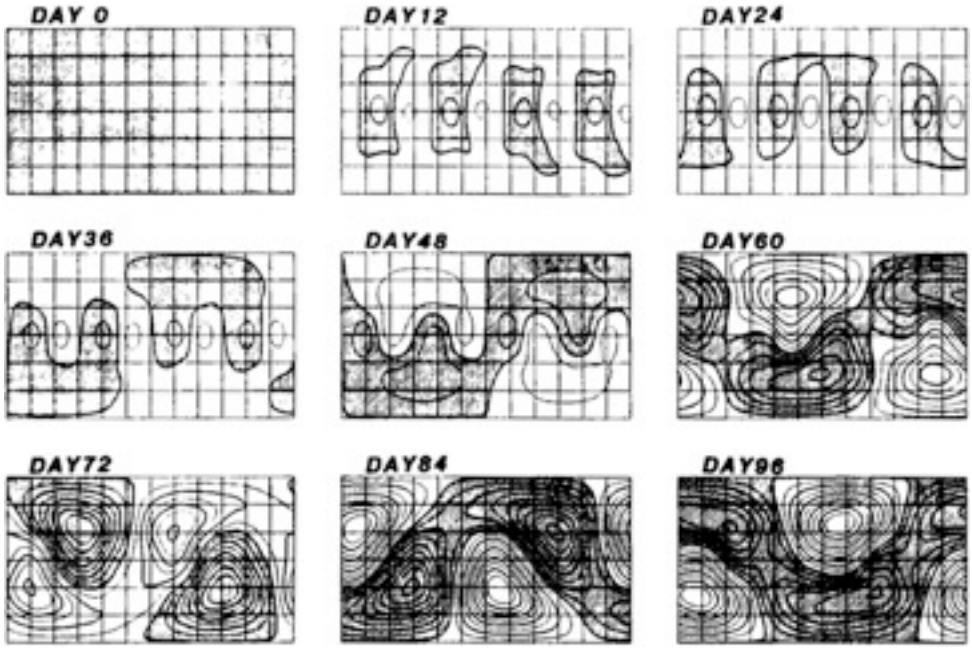

Fig. 12.11: Masuda's velocity potential field.

Haurwitz wave with wave number 4 as his initial condition. Fig. 12.11 shows the evolution of the velocity potential field in a 96-simulated-day run using Masuda's model. The initial conditions are nondivergent, so initially the velocity potential is zero. As time progresses, a wave number 4 pattern develops. As time progresses further, the pattern at higher latitudes begins to break down, forming a wave number 1 pattern. Significantly, the wave number 1 pattern is antisymmetric across the Equator, even though the initial condition is symmetric across the Equator. Masuda suggested that this is due to the antisymmetry of the grid across the Equator.

Heikes and Randall (1995 a, b) extended Masuda's work, through the use of a "twisted icosahedral grid" that has symmetry across the equator. They used a multi-grid method to compute the stream function and velocity potential from the vorticity and divergence, respectively. Heikes and Randall (1995 b) also showed that the grid (twisted or not) has to be slightly altered, or "optimized," to permit consistent finite-difference approximations to the divergence, Jacobian, and Laplacian operators that are used in the construction of the model. They tested their model using certain standard test cases for shallow water on the sphere

(Williamson et al. 1992), and obtained good results. Ringler et al. (1999) constructed a full-physics global atmospheric model using this approach.

12.8 Summary

In order to construct a numerical model on the sphere, it is necessary to map the sphere onto a computational domain. There are various ways of doing this. The most straightforward is to use latitude-longitude coordinates, but this leads to the pole problem. The pole problem can be dealt with by using filters, but these approaches suffer from some problems of their own. Semi-implicit differencing could be used to avoid the need for filtering.

Another approach is to use a regular grid on the sphere. A perfectly regular grid is mathematically impossible, but geodesic grids can come close.

A third approach, discussed in the next chapter, is to use the spectral method, with spherical harmonics as the basis functions.

Problems

1. Program the *one-dimensional* linearized C-grid shallow-water equations with a leapfrog time step. Use a grid spacing corresponding to a longitudinal grid spacing of 5 ˚ and a latitude of 70 ˚N. Construct and apply a "polar filter" designed to allow computational stability in a model with a meridional grid spacing of 4 degrees.