CHAPTER 10 Finite Differences on the Sphere

10.1 Introduction

There are a number of problems associated with trying to solve differential equations numerically in spherical geometry. Difficulties can arise from the use of spherical coordinate systems, and from trying to discretize the surface of the sphere itself. These difficulties include what is called the "pole problem."

Even before any numerical considerations are confronted, we run into difficulties simply in specifying a coordinate system. It is important to distinguish between true scalar-valued and vector-valued functions. The value of a scalar function, say temperature, is independent of coordinate system. That is, the temperature at some point in space is the same regardless of how we define the position of the point. On the other hand, the individual components of a vector-valued function, such as the wind velocity, obviously differ depending on the coordinate system. 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.

As a simple example, consider a jet directed over the North Pole. This is represented by the shaded arrow in Fig. 10.1. Measured at points along the prime meridian, the wind will

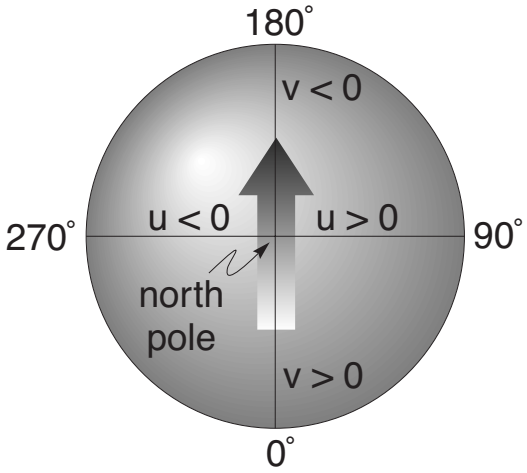


Figure 10.1: For the wind vector shown in the sketch, points along the prime meridian have a strong northward component. There is a discontinuity at the pole, and points along international date line have a strong southward component. Points near 90° longitude have a strong positive zonal component, while points near 270° longitude have a strong negative zonal component.

have a positive v component. Measured along the international date line, however, the wind

will have 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 $90°$ longitude, and is negative measured along $270°$ longitude. This problem does not occur in a Cartesian coordinate system centered on the pole. At each point along a great circle which includes the pole, the components measured in Cartesian coordinates are well defined and vary continuously.

10.2 Coordinate systems and map projections

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

$$\nabla A = \left(\frac{1}{r\cos\varphi}\frac{\partial A}{\partial \lambda}, \frac{1}{r}\frac{\partial A}{\partial \varphi}, \frac{\partial A}{\partial r} \right) , \tag{10.1}$$

$$\nabla \bullet \mathbf{V} = \frac{1}{r\cos\varphi}\frac{\partial V_\lambda}{\partial \lambda} + \frac{1}{r\cos\varphi}\frac{\partial}{\partial \varphi}(V_\varphi \cos\varphi) + \frac{1}{r^2}\frac{\partial}{\partial r}(V_r r^2) , \tag{10.2}$$

$$\begin{aligned} \nabla \times \mathbf{V} = \Bigg\{ & \frac{1}{r}\left[\frac{\partial V_r}{\partial \varphi} - \frac{\partial}{\partial r}(rV_\varphi) \right], \\ & \frac{1}{r}\frac{\partial}{\partial r}(rV_\lambda) - \frac{1}{r\cos\varphi}\frac{\partial V_r}{\partial \lambda}, \\ & \frac{1}{r\cos\varphi}\left[\frac{\partial V_\varphi}{\partial \lambda} - \frac{\partial}{\partial \varphi}(V_\lambda \cos\varphi) \right] \Bigg\}, \end{aligned} \tag{10.3}$$

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

$$\nabla A = \left(\frac{1}{a\cos\varphi}\frac{\partial A}{\partial \lambda}, \frac{1}{a}\frac{\partial A}{\partial \varphi} \right), \tag{10.4}$$

$$\nabla \bullet \mathbf{V} = \frac{1}{a\cos\varphi}\frac{\partial V_\lambda}{\partial \lambda} + \frac{1}{a\cos\varphi}\frac{\partial}{\partial \varphi}(V_\varphi \cos\varphi) , \tag{10.5}$$

$$\mathbf{k} \bullet (\nabla \times \mathbf{V}) = \frac{1}{a\cos\varphi}\left[\frac{\partial V_\varphi}{\partial \lambda} - \frac{\partial}{\partial \varphi}(V_\lambda \cos\varphi) \right] . \tag{10.6}$$

Here a is the radius of the spherical planet.

The shallow water equations in spherical coordinates can be expressed as

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - \left(f + \frac{u}{a}\tan\varphi\right)v + \frac{g}{a\cos\varphi}\frac{\partial}{\partial\lambda}(h + h_S) = 0, \tag{10.7}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + \left(f + \frac{u}{a}\tan\varphi\right)u + \frac{g}{a}\frac{\partial}{\partial\varphi}(h + h_S) = 0, \tag{10.8}$$

$$\frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x} + v\frac{\partial h}{\partial y} + \frac{h}{a\cos\varphi}\left[\frac{\partial u}{\partial\lambda} + \frac{\partial}{\partial\varphi}(v\cos\varphi)\right] = 0. \tag{10.9}$$

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

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 $g(\lambda, \varphi) = (x, y)$. There are mappings which map almost all of S onto P; examples are given below. Unfortunately, these mappings, or projections, tend to badly distort distances and areas near the singular points of the transformation. 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). Define the metric coefficients to be h_x and h_y so that the distance increment dl satisfies

$$dl^2 = h_x^2 dx^2 + h_y^2 dy^2. \tag{10.10}$$

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

$$U = h_x\frac{dx}{dt}, \tag{10.11}$$

$$V = h_y\frac{dy}{dt}. \tag{10.12}$$

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

$$\frac{dU}{dt} - \left[f + \frac{1}{h_x h_y}\left(V\frac{\partial h_y}{\partial x} - U\frac{\partial h_x}{\partial y}\right)\right] V + \frac{g}{h_x}\frac{\partial}{\partial x}(h + h_S) = 0, \tag{10.13}$$

$$\frac{dV}{dt} + \left[f + \frac{1}{h_x h_y}\left(V\frac{\partial h_y}{\partial x} - U\frac{\partial h_x}{\partial y}\right)\right] U + \frac{g}{h_y}\frac{\partial}{\partial y}(h + h_S) = 0, \tag{10.14}$$

where

$$\frac{d}{dt}(\ ) = \frac{\partial}{\partial t}(\ ) + \frac{U}{h_x}\frac{\partial}{\partial x}(\ ) + \frac{V}{h_y}\frac{\partial}{\partial y}(\ ). \tag{10.15}$$

The continuity equation is given by

$$\frac{dh}{dt} + \frac{h}{h_x h_y}\left[\frac{\partial}{\partial x}(h_x U) + \frac{\partial}{\partial y}(h_y V)\right] = 0. \tag{10.16}$$

For example, if we set

$$x = \lambda \text{ and } y = \varphi, \tag{10.17}$$

and correspondingly set the metric coefficients to

$$h_x = a\cos\varphi \text{ and } h_y = a, \tag{10.18}$$

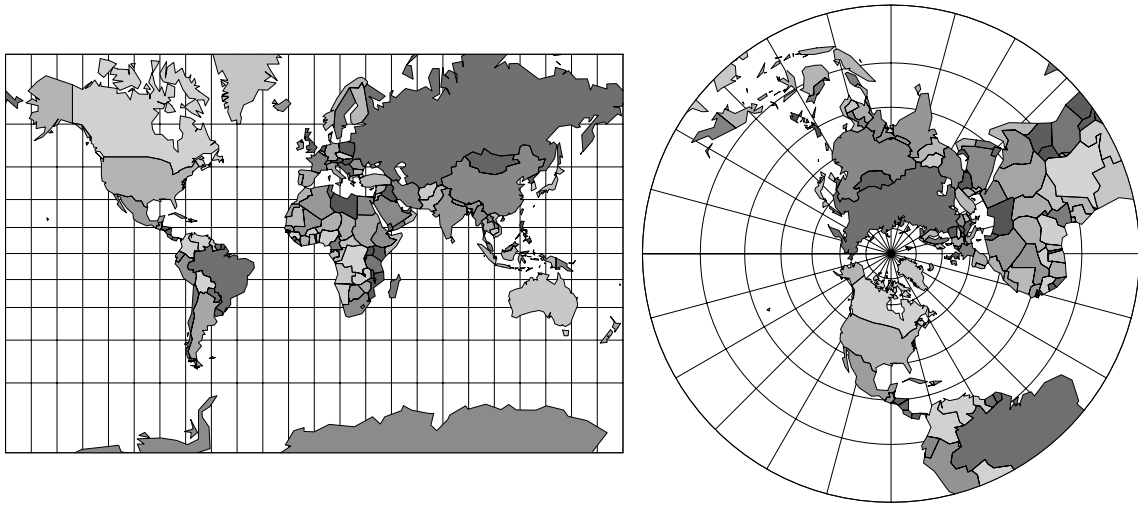
then by (10.11) and (10.12) we have

$$U = u \equiv a\cos\varphi\frac{d\lambda}{dt} \text{ and } V = v \equiv a\frac{d\varphi}{dt}. \tag{10.19}$$

Substituting (10.18) and (10.19) into (10.13), (10.14) and (10.16) gives (10.7), (10.8) and (10.9), the shallow water equations in spherical coordinates.

Two map projections are commonly used in numerical modeling of the atmospheric circulation -- Polar Stereographic and Mercator. 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 outlines of the continents are shown in Fig. 10.2.

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



a b

Figure 10.2: Map projections of the continents: a.) Mercator projection. b.) North polar stereographic projection.

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):

$$X = \frac{2a\cos\varphi\cos\lambda}{1 + \sin\varphi}, \tag{10.20}$$

$$Y = \frac{2a\cos\varphi\sin\lambda}{1 + \sin\varphi}. \tag{10.21}$$

Note that there is a problem at the South Pole. Taking differentials of (10.20) and (10.21) gives

$$\begin{bmatrix} dX \\ dY \end{bmatrix} = \frac{2a}{1 + \sin\varphi} \begin{bmatrix} -\cos\varphi\sin\lambda & -\cos\lambda \\ \cos\varphi\cos\lambda & -\sin\lambda \end{bmatrix} \begin{bmatrix} d\lambda \\ d\varphi \end{bmatrix}. \tag{10.22}$$

Now we determine the metrics of the polar stereographic map projection. 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

$$dl^2 = (a\cos\varphi)^2 d\lambda^2 + a^2 d\varphi^2. \tag{10.23}$$

Solving (10.22) for $d\varphi$, and $d\lambda$, and substituting the results into (10.23), we obtain

$$dl^2 = \left(\frac{1+\sin\varphi}{2}\right)^2 dX^2 + \left(\frac{1+\sin\varphi}{2}\right)^2 dY^2. \tag{10.24}$$

Comparing (10.24) with (10.10), we see that metric coefficient for the polar stereographic projection is given by

$$h_x = h_y = \frac{1+\sin\varphi}{2}. \tag{10.25}$$

We define the map factor, $m(\varphi)$, as the inverse of the metric coefficient, i.e., $m(\varphi) = 2/(1+\sin\varphi)$. Using (10.13), (10.14) and (10.16), we can write the shallow water equations in north polar stereographic coordinates:

$$\frac{dU}{dt} - \left[f + \frac{UY - VX}{2a^2}\right]V + mg\frac{\partial}{\partial x}(h + h_S) = 0, \tag{10.26}$$

$$\frac{dV}{dt} + \left[f + \frac{UY - VX}{2a^2}\right]U + mg\frac{\partial}{\partial x}(h + h_S) = 0, \tag{10.27}$$

$$\frac{dh}{dt} + m^2 h\left[\frac{\partial}{\partial X}\left(\frac{U}{m}\right) + \frac{\partial}{\partial Y}\left(\frac{V}{m}\right)\right] = 0. \tag{10.28}$$

The total derivative is given by (10.15).

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.

More recently, 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 (10.26), (10.27) and (10.28). 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. 10.3. Values for

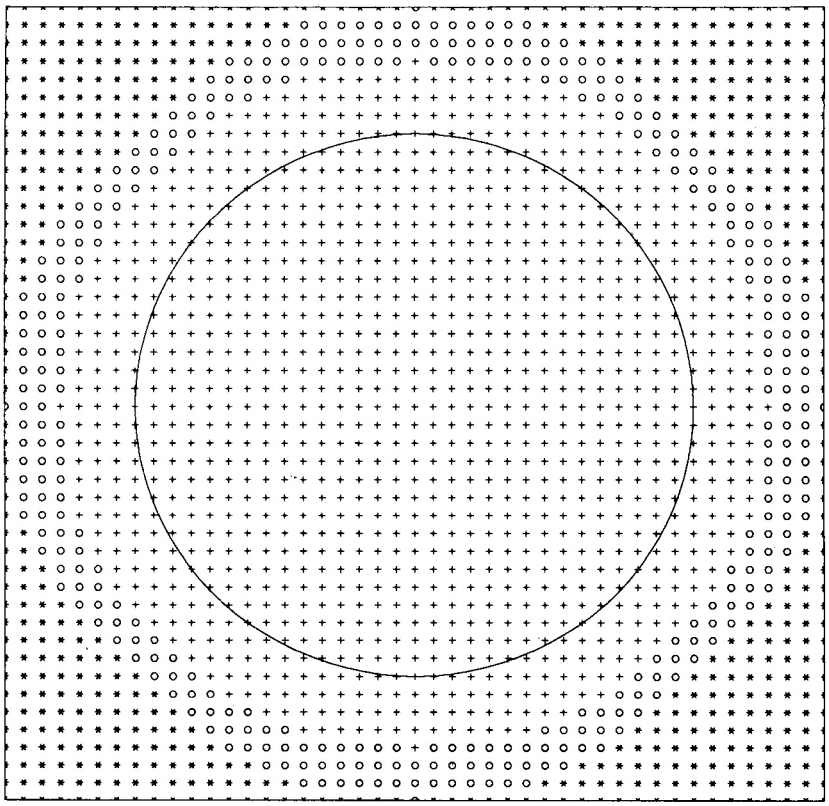


Figure 10.3: Composite grid method grid. Two such grids are used to cover the sphere. Points labeled with ○ are the boundary conditions for the points labeled with +. Values at the ○ points are obtained by interpolation from the other grid. The big circle is the image of the Equator. Points labeled * are not used.

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.

10.3 Latitude–longitude grids and the "pole problem"

An obvious way to discretize the shallow water equations expressed in spherical coordinates is to use a regular latitude-longitude grid in which the grid intervals $(\Delta\lambda, \Delta\varphi)$ are constants. A discretization scheme is straight forward except for the row of grid points next to the pole, where special considerations are necessary.

A portion (one eighth) of a uniform latitude-longitude grid is shown in Fig. 10.4. 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

Figure 10.4: 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.

trapezoidal in shape. The are other ways to deal with the polar regions, e.g. by defining local cartesian coordinates at the poles.

A regular latitude-longitude grid has some drawbacks. 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 5 (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 (10.7), (10.8), and (10.9) about a state of rest, as follows:

$$\frac{\partial u}{\partial t} + \frac{g}{a\cos\varphi}\frac{\partial h}{\partial \lambda} = 0, \tag{10.29}$$

$$\frac{\partial v}{\partial t} + \frac{g}{a}\frac{\partial h}{\partial \varphi} = 0, \tag{10.30}$$

$$\frac{\partial h}{\partial t} + \frac{H}{a\cos\varphi}\left[\frac{\partial u}{\partial \lambda} + \frac{\partial}{\partial \varphi}(v\cos\varphi)\right] = 0. \tag{10.31}$$

Here we have neglected rotation and bottom topography, for simplicity, and H denotes the mean depth of the fluid. We spatially discretize these as follows:

$$\frac{\partial u}{\partial t}_{i+\frac{1}{2},j} + \frac{g(h_{i+1,j} - h_{i,j})}{a\cos\varphi\Delta\lambda} = 0, \tag{10.32}$$

$$\frac{\partial v}{\partial t}_{i,j+\frac{1}{2}} + \frac{g(h_{i,j+1} - h_{i,j})}{a\Delta\varphi} = 0, \tag{10.33}$$

$$\frac{\partial h_{i,j}}{\partial t} + H\left\{\frac{\left(u_{i+\frac{1}{2},j} - u_{i-\frac{1}{2},j}\right)}{a\cos\varphi_j\Delta\lambda} + \left[\frac{(v\cos\varphi)_{i,j+\frac{1}{2}} - (v\cos\varphi)_{i,j-\frac{1}{2}}}{a\cos\varphi_j\Delta\varphi}\right]\right\} = 0. \tag{10.34}$$

Note that the C grid has been used here. We look for solutions of the form

$$u_{i+\frac{1}{2},j} = Re\left\{\hat{u}_j exp\left[\dot{i}s\left(i+\frac{1}{2}\right)\Delta\lambda + \dot{i}\sigma t\right]\right\}, \tag{10.35}$$

$$v_{i,j+\frac{1}{2}} = Re\left\{\hat{v}_{j+\frac{1}{2}} exp[\dot{i}si\Delta\lambda + \dot{i}\sigma t]\right\}, \tag{10.36}$$

$$h_{i,j} = Re\{\hat{h}_j exp[\dot{i}si\Delta\lambda + \dot{i}\sigma t]\}, \tag{10.37}$$

where $\dot{i} \equiv \sqrt{-1}$. Note that the zonal wave number, s, is defined with respect to longitude rather than distance. After substitution of these solutions into our finite-difference equations, we obtain

$$\dot{i}\sigma\hat{u}_j + \frac{\dot{i}s}{a\cos\varphi_j}\frac{\sin(s\Delta\lambda/2)}{s\Delta\lambda/2}g[S_j(s)\hat{h}_j] = 0, \tag{10.38}$$

$$\dot{i}\sigma\hat{v}_{j+\frac{1}{2}} + \frac{g(\hat{h}_{j+1} - \hat{h}_j)}{a\Delta\varphi} = 0, \tag{10.39}$$

$$-i\sigma\hat{h}_j + H\left\{\frac{-is}{a\cos\varphi_j}\frac{\sin(s\Delta\lambda/2)}{s\Delta\lambda/2}[S_j(s)\hat{u}_j] + \left[\frac{(\hat{v}\cos\varphi)_{j+\frac{1}{2}} - (\hat{v}\cos\varphi)_{j-\frac{1}{2}}}{a\cos\varphi_j\Delta\varphi}\right]\right\} = 0 \quad , (10.40)$$

where $S_j(s)$ is a "smoothing parameter" which depends on wave number. The smoothing parameter appears in the term of (10.38) corresponding to the zonal pressure gradient force, and in the term of (10.40) corresponding to the zonal mass flux divergence. These are the key terms for zonally propagating gravity waves. We have introduced $S_j(s)$ artificially in (10.39) and (10.40); later in this discussion it 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 u and v in (10.38)-(10.40), we can obtain the "meridional structure equation" for $\hat{h}$. It is

$$\begin{aligned} &c^2\left[\frac{s}{a\cos\varphi_j}\frac{\sin(s\Delta\lambda/2)}{s\Delta\lambda/2}S_j(s)\right]^2\hat{h}_j \\ &+\frac{c^2}{(a\Delta\varphi)^2}\left[(\hat{h}_j-\hat{h}_{j-1})\frac{\cos\varphi_{j-\frac{1}{2}}}{\cos\varphi_j}-(\hat{h}_{j+1}-\hat{h}_j)\frac{\cos\varphi_{j+\frac{1}{2}}}{\cos\varphi_j}\right] = \sigma^2\hat{h}_j \ , \end{aligned} \quad (10.41)$$

where $c^2 \equiv gH$. By using the boundary condition $\hat{h}_j = 0$ at the poles, this equation can be solved as an eigenvalue problem for the frequency, σ. For high values of the zonal wave number s, the first term on the left-hand side of (10.41) dominates the second, and we obtain

$$\begin{aligned} \sigma &= |c|\left[\frac{s}{a\cos\varphi_j}\frac{\sin(s\Delta\lambda/2)}{s\Delta\lambda/2}S_j(s)\right] \\ &= 2|c|\frac{S_j(s)\sin\left(\frac{s\Delta\lambda}{2}\right)}{a\cos\varphi_j\Delta\lambda} \ . \end{aligned} \quad (10.42)$$

Although we have not introduced a specific time differencing scheme here, we know that the CFL criterion takes the form

$$\sigma\Delta t < \varepsilon \ , \quad (10.43)$$

where ε is a constant of order one. In view of (10.42) and (10.43), 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

$$\frac{|c|\Delta t}{a\cos\varphi_j \Delta\lambda} \sin\left(\frac{s\Delta\lambda}{2}\right) < \frac{1}{2} \tag{10.44}$$

The worst case is $\sin\left(\frac{s\Delta\lambda}{2}\right) = 1$, for which (10.44) reduces to

$$\frac{|c|\Delta t}{\Delta x_j} < \frac{1}{2} \tag{10.45}$$

where we define $\Delta x_j \equiv a\cos\varphi_j \Delta\lambda$. For the grid shown in Fig. 10.4, with a longitudinal grid spacing of $\Delta\lambda = 5°$ and a latitudinal grid spacing of $\Delta\varphi = 4°$ (which are the values used to draw the figure), the northernmost row of grid points where the u component of velocity is defined is at latitude 86 °N. The zonal distance between grid points on the northernmost row is then $\Delta x = 39$ km. The fast, external gravity wave has a phase speed of approximately 300 m s^{-1}. Substituting into (10.45), 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 was 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. One 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 (10.42) shows that this can be accomplished by choosing the smoothing parameter so that

$$\frac{S_j(s)\sin\left(\frac{s\Delta\lambda}{2}\right)}{a\cos\varphi_j \Delta\lambda} = \frac{1}{d^*} \ , \tag{10.46}$$

where d^* is a suitably chosen length, comparable to the zonal grid spacing at the Equator. With the use of (10.46), (10.42) reduces to

$$\sigma = \frac{2|c|}{d^*}, \tag{10.47}$$

and the CFL condition reduces to

$$\frac{2|c|}{d^*}\Delta t < \varepsilon, \tag{10.48}$$

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

$$d^* \equiv a\Delta\varphi, \tag{10.49}$$

i.e. the distance between grid points in the meridional direction, then, referring back to (10.46), we see that $S_j(s)$ must be chosen so that

$$S_j(s) = \left(\frac{\Delta\lambda}{\Delta\varphi}\right)\frac{\cos\varphi_j}{\sin\left(\frac{s\Delta\lambda}{2}\right)}. \tag{10.50}$$

Of course, at low latitudes (10.50) can give values of $S_j(s)$ which are greater than one; these should be replaced by one, so that we actually use

$$S_j(s) = Min\left\{\left(\frac{\Delta\lambda}{\Delta\varphi}\right)\frac{\cos\varphi_j}{\sin\left(\frac{s\Delta\lambda}{2}\right)}, 1\right\}. \tag{10.51}$$

A plot of (10.51) is given in Fig. 10.5, for the case of the shortest zonal mode. The plot shows that some smoothing is needed all the way down into the subtropics.

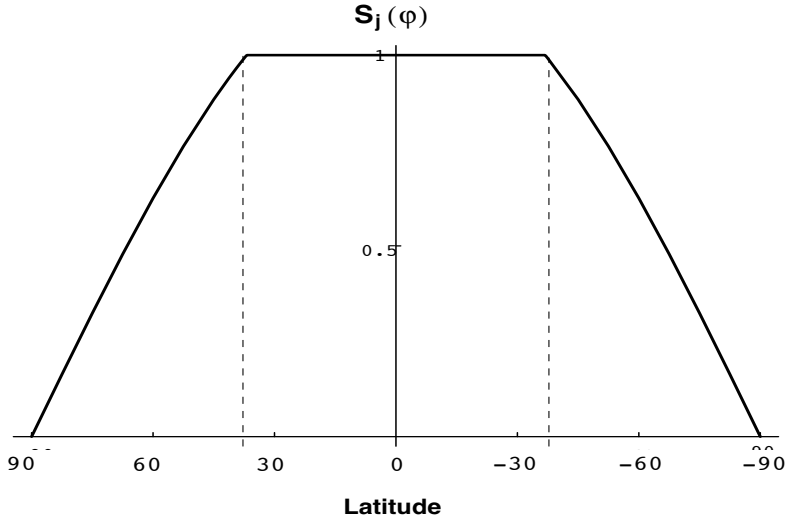


Figure 10.5: A plot of the smoothing parameter as given by (10.51), for the "worst case" of the shortest zonal mode. The dashed vertical lines demarcate the belt of latitude near the Equator for which no smoothing is needed. It has been assumed that the longitudinal grid spacing is 5/4 times the latitudinal grid spacing, as it is for the grid shown in Fig. 10.4.

10.4 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 $0°$ 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. 10.6; compare with Fig. 10.4. For a given N, the total number of grid points on the sphere is $4N^2+2$. 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 $\Delta\lambda_j \equiv \frac{\pi}{2}\frac{1}{j-1}$. At $j = N+1$, the Equator, the ratio is one, and near the pole

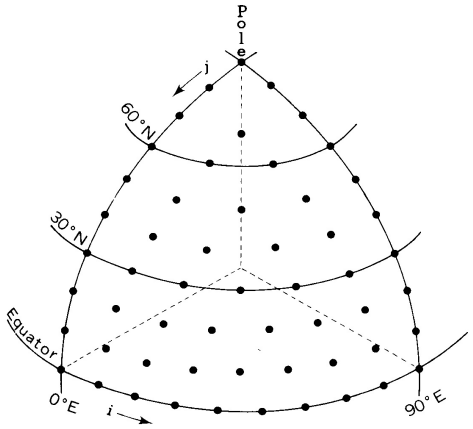


Figure 10.6: Kurihara grid on one octant of the sphere.

the ratio approaches $\pi/2 \cong 1.57$.

Kurihara built a model using this grid, based on the shallow water equations. He tested the model with the Rossby-Haurwitz wave, with wave number 4 as the initial condition. This set of initial conditions was used by Phillips (1959), and 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.

10.5 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. 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. 10.7 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.

10.6 Spherical geodesic grids

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

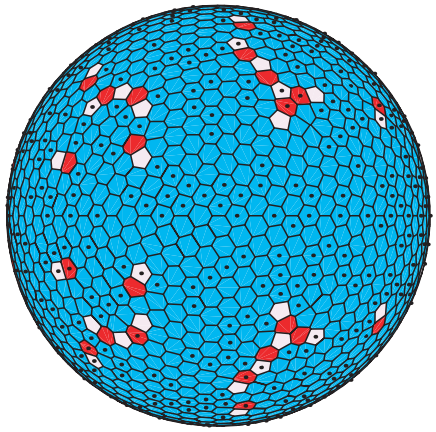


Figure 10.7: Wandering electron grid. White cells have five walls, light gray cells have six walls, and dark gray cells have seven walls.

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. 10.8. One can

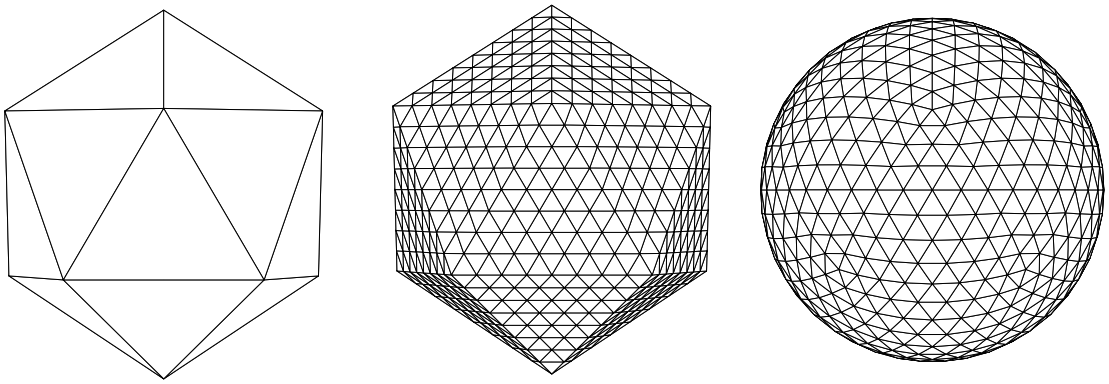
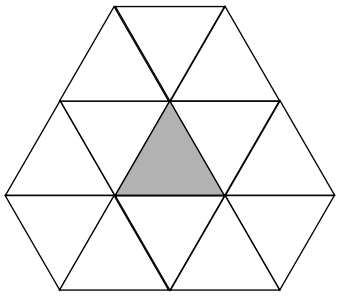


Figure 10.8: a.) Icosahedron. b.) Partition each face into 64 smaller triangles. c.) Project onto the sphere.

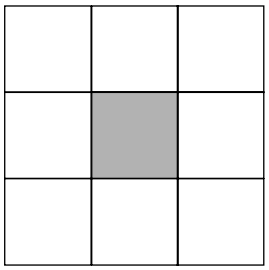
construct a more homogeneous grid by partitioning the spherical equilateral triangles instead. Williamson (1968) and Sadourny (1968) use slightly different techniques to construct their grids. However, both begin by partitioning the spherical icosahedral triangle.

On these geodesic grids, all but twelve of the cells are hexagons. Hexagonal grids are

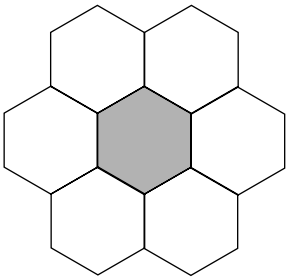
quasi-isotropic. As is well known, only three regular polygons tile the plane: equilateral triangles, squares, and hexagons. Fig. 10.9 shows planar grids made up of each of these three



12 neighbors,
3 wall neighbors



8 neighbors,
4 wall neighbors



6 neighbors,
6 wall neighbors

Figure 10.9: Cells neighboring a given cell (shaded) on triangular, square, and hexagonal grids. A "wall neighbor" is a neighbor which lies directly across a cell wall.

possible polygonal elements. On the triangular grid and the square grid, some of the neighbors of a given cell lie directly across cell walls, while others lie across cell vertices. As a result, finite-difference operators constructed on these grids tend to use "wall neighbors" and "vertex neighbors" in different ways. For example, the simplest second-order finite-difference approximation to the gradient, on a square grid, uses only "wall neighbors;" vertex neighbors are ignored. Although it is certainly possible to construct finite-difference operators on square grids (and triangular grids) in which information from all neighboring cells is used [e.g. the Arakawa Jacobian, as discussed by Arakawa (1966)], the essential anisotropies of these grids remain, and are unavoidably manifested in the forms of the finite-difference operators. Hexagonal grids, in contrast, have the property that all neighbors of a

given cell lie across cell walls; there are no "vertex neighbors." As a result, finite-difference operators constructed on hexagonal grids treat all neighboring cells in the same way; in this sense, the operators are as symmetrical and isotropic as possible. The twelve pentagonal cells also have only wall neighbors; there are no vertex neighbors anywhere on the sphere.

Williamson (1968) chose the nondivergent shallow water equations to test the new grid, i.e. he used

$$\frac{\partial \zeta}{\partial t} = -J(\psi, \eta), \tag{10.52}$$

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

$$\nabla^2 \psi = \zeta. \tag{10.53}$$

Recall that, for arbitrary functions α and β, the Jacobian in differential form has the property that

$$\int_A J(\alpha,\beta)\, dA = \oint_S \alpha \frac{\partial \beta}{\partial s} ds. \tag{10.54}$$

So, integrating (10.52) over the area A, we get

$$\frac{\partial}{\partial t}\int_A \zeta\, dA = -\oint_S u_n \eta\, ds, \tag{10.55}$$

where $u_n = -\dfrac{\partial \psi}{\partial s}$ is the velocity normal to the boundary S.

Consider K triangles surrounding the grid point P_0 in Fig. 10.10. 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. First, we make the approximation $A\zeta_0 \cong \int_A \zeta\, dA$. Let η_i and η_{i+1} be the absolute vorticity defined at points P_i and P_{i+1}, respectively. We approximate the absolute vorticity along the edge between P_i and P_{i+1} by $(\eta_i + \eta_{i+1})/2$. Similarly, $\partial\psi/\partial s \approx (\psi_{i+1} - \psi_i)/\Delta s$, where Δs is the distance from P_i to P_{i+1}. Hence, we can approximate (10.55) by

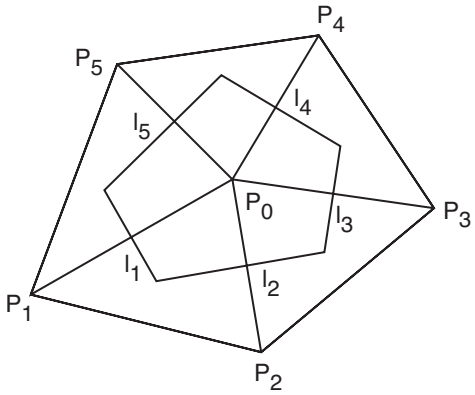


Figure 10.10:Configuration of grid triangles for the case $K = 5$.

$$\frac{\partial \zeta_0}{\partial t} = \frac{1}{A}\sum_{i=1}^{K}\left(\frac{\psi_{i+1}-\psi_i}{\Delta s}\right)\left(\frac{\eta_{i+1}+\eta_i}{2}\right)\Delta s. \tag{10.56}$$

To solve equation (10.53), we must discretize the Laplacian operator. Consider the smaller, inner polygon in Fig. 10.10. The walls of the polygon are formed from the perpendicular bisectors of the line segments $\overline{P_0P_i}$. For any arbitrary scalar function α, we can use Gauss' Theorem to write

$$\int_a \nabla^2\alpha\, da = \oint_{s'}\frac{\partial\alpha}{\partial n}ds, \tag{10.57}$$

where a is the area of the small polygon, and s' is its boundary. Using (10.53) and (10.57), we get

$$\Gamma = \oint_{s'} u_s ds, \tag{10.58}$$

where Γ is the circulation around the boundary and u_s is the counterclockwise tangential velocity. We now set $\Gamma \cong a\zeta_0$. We assume that the tangential velocity $u_s = \partial\psi/\partial n$ is approximately constant along each wall of s', and can be approximated by $(\psi_i - \psi_0)/|P_0P_i|$, where $|P_0P_i|$ is the distance from P_0 to P_i. Define $\omega_i \equiv l_i/|P_0P_i|$, where l_i is the length of wall i. The resulting finite-difference approximation is

$$\zeta_0 = \sum_{i=1}^{K} \omega_i(\psi_i - \psi_0). \tag{10.59}$$

This is a finite-difference approximation to (10.53). It can be solved for the ψ_i by relaxation, as discussed in earlier.

Williamson showed that his scheme conserves kinetic energy and enstrophy as the exact equations do. When applied to an equilateral triangular 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. In a later study, Williamson (1971) described a finite-difference scheme that is second-order accurate on the spherical geodesic grid, but lacks certain desirable conservation properties.

Sadourny (1968) discussed a nondivergent model very similar to Williamson's. Also, Sadourny (1969) developed a geodesic-grid model based on the 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 6). Like Williamson, Masuda chose the Rossby-Haurwitz wave with wave number 4 as his initial condition. Fig. 10.11 shows the

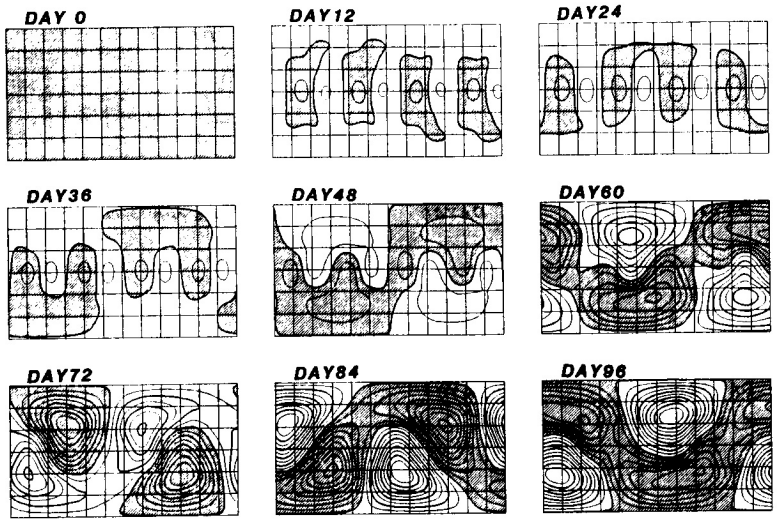


Figure 10.11:Masuda's velocity potential field.

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 (whether twisted or not) has to be slightly altered to permit consistent finite-difference approximation 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) have constructed a full-physics global atmospheric model using this approach.

There have also been recent attempts to use grids based on cubes and octahedrons (e.g. McGregor, 1996; Purser and Rancic, 1998).

10.7 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.