Grid-Based Methods for Numerical Weather Prediction

Structured and Unstructured Grids

A majority of modern numerical weather prediction models, particularly those used to obtain forecasts over limited-area domains, utilize grid-based methods to discretize the atmosphere into coherent points or blocks on/for which the primitive equations may be solved. Such a grid may be structured or unstructured in nature. Except where otherwise noted, we will consider grid points to be located at the center of a grid box that has a conic or other shape.

A *structured grid* is one in which the grid is arranged regularly; e.g., rectangles with sides of uniform length over the entire domain. Such sides may be defined by physical distance (e.g., each grid box has sides of length 4 km) or by latitude and longitude (e.g., each grid box has sides of length 0.04° latitude and longitude). An example of a structured grid is given in Fig. 1 (left). An *unstructured grid* is one in which the grid is arranged irregularly; e.g., triangles or hexagons with sides that may have variable length across the domain. Unstructured grids require that information about the model grid's construction be retained in memory throughout its execution, though this requires a trivial amount of memory compared to that required to store model data. An example of an unstructured grid is given in Fig. 1 (right).

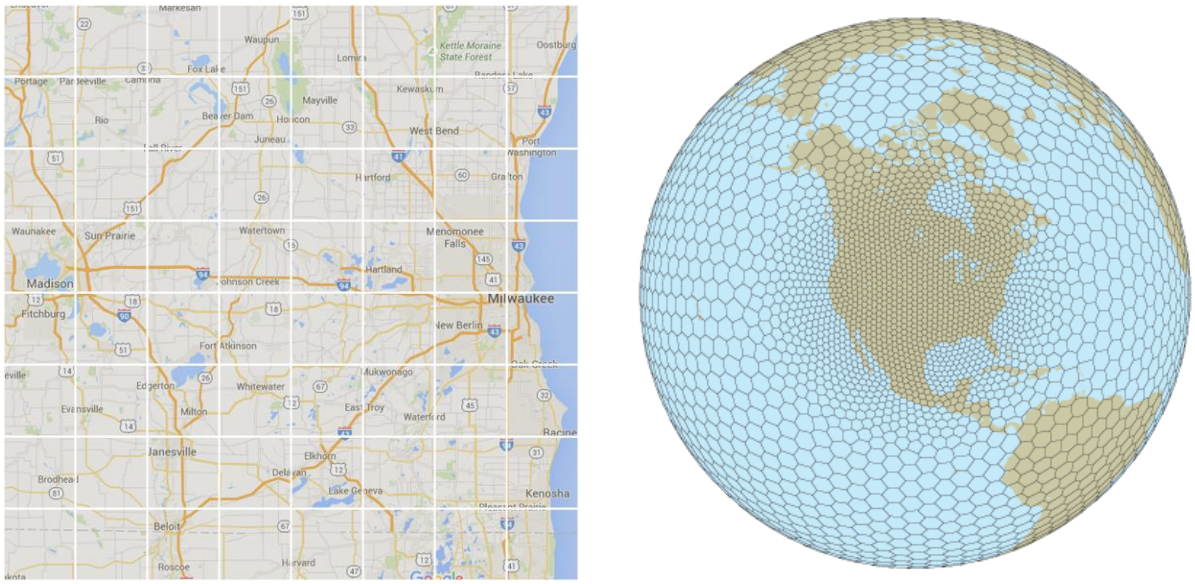

Figure 1. (left) An example of a structured grid, with sides of uniform distance, applied to southeastern Wisconsin. (right) An example of an unstructured grid, constructed of hexagons with sides of variable length, across the Western Hemisphere. Image at right obtained from https://mpas-dev.github.io/atmosphere/atmosphere.html.

Whether a structured or unstructured grid is utilized, grid points on the grid may be *staggered*, such that grid points for certain variables (momentum) are offset by some distance from grid points for other variables (mass/thermodynamic). Given the physical linkages that exist between these

variables, the primary benefit of grid staggering is to increase the effective resolution of the model by decreasing the distance over which finite difference approximations need to be computed. However, doing so often requires the use of a shorter time step to maintain computational stability. There are multiple ways by which staggering may be achieved, with the most widely used method in the horizontal being the Arakawa-*C* grid. An example of Arakawa *C*-grid staggering for a structured grid with an inner nested grid is given in Fig. 2. We will discuss grid staggering in greater detail when we cover finite difference approximations.

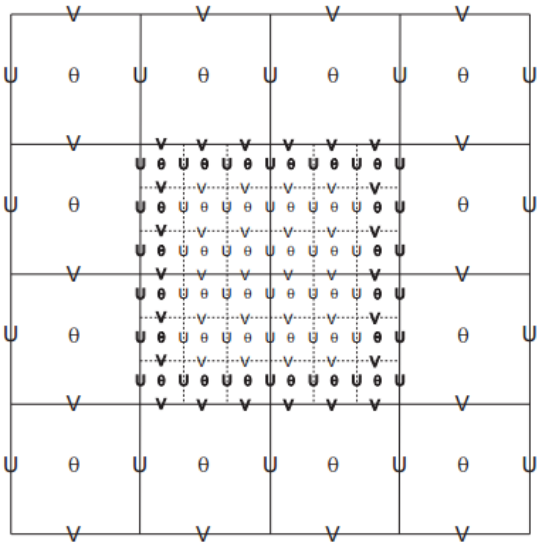

Figure 2. The Arakawa *C*-grid, representing an example of a staggered structured grid. Momentum fields are defined normal to the grid box edges whereas mass and thermodynamic fields are defined in the center of each grid box. This example also contains an inner domain, with identical grid staggering and a horizontal grid spacing one-third that of the outer domain. Image obtained from http://www2.mmm.ucar.edu/wrf/users/docs/technote/v4_technote.pdf, their Fig. 7.3.

Map Projections

Limited-area grid point models use *map projections*, or mathematical relationships that transform a portion of the spherical Earth to a horizontal grid surface. There exist several desirable traits for a chosen map projection:

- **Preservation of angles**. Angles on the grid should be equivalent to those on the Earth.
- **Preservation of areas**. Areas on the grid should be equivalent to those on the Earth. This implies that *distances* on the grid should be equivalent to those on the Earth.
- **Preservation of shapes**. Shapes on the grid should be equivalent to those on the Earth.

- **Correct directions**. Directions on the grid should be equivalent to those on the Earth.
- The shortest distance between two lines should be a great circle, as it is on the Earth.

No individual map projection can satisfy all of these traits; each is simply a flat approximation of the spherical Earth. For numerical weather prediction, and the atmospheric sciences in general, the most desirable of these traits is that **angles be preserved**. This is particularly important for accurate, straightforward representation of the horizontal wind components (u, v) and meteorological fields derived from them (e.g., vorticity and divergence) on the model grid.

A **conformal** map projection is one in which angles are inherently preserved. The defining characteristic of a conformal map is that the latitude and longitude lines are *locally* perpendicular to each other everywhere on the map. Conformal maps, however, do not preserve shapes, distances, or areas. Near what are known as the *standard parallels* of the map projection, where the cone, cylinder, or planar surface onto which the Earth is projected actually interest the Earth, shape, distance, and area distortion are small. Away from the standard parallels, shape, distance, and area distortion can become large. The choice of map projection for a given limited-area model domain is made with respect to the range of latitudes over which distortion is relatively small.

The three most commonly used conformal map projections in numerical weather prediction and the atmospheric sciences in general are the **Lambert conic**, **Mercator**, and **polar stereographic** map projections. Schematics of each are presented in Fig. 3 below.

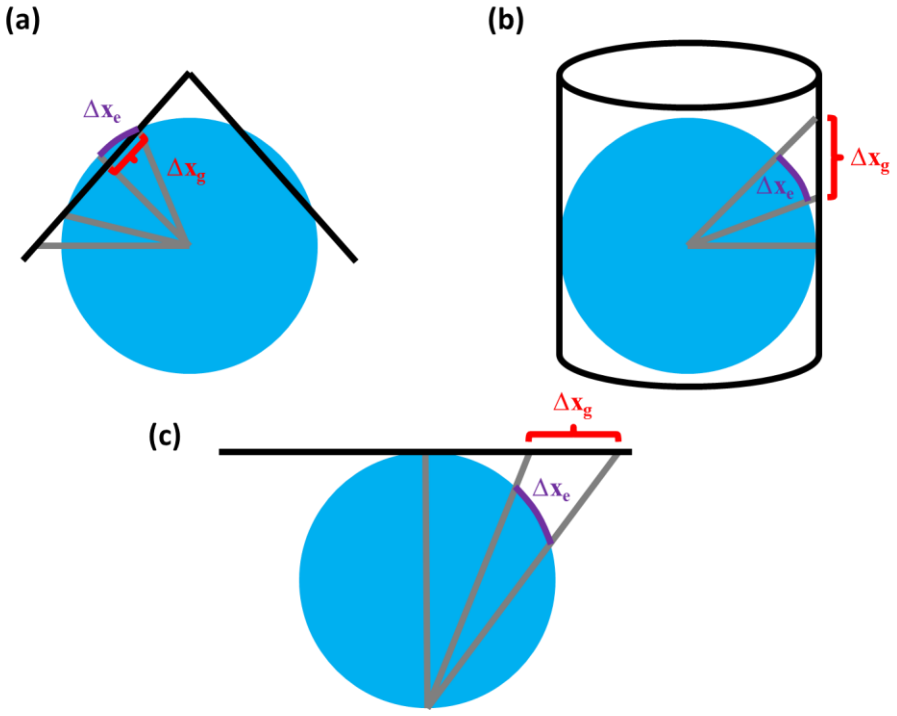

Figure 3. Graphical schematics of the (a) Lambert conic, (b) Mercator, and (c) polar stereographic conformal map projections. Please refer to the text for additional details.

The Lambert conic projection is obtained by fitting a cone, with its tip located directly above either the North or South Pole, either tangent or secant to the Earth. The schematic presented in Fig. 3 is an example of a secant fit, where the cone's surface intersects the Earth at two points, and is the more common of the two fits. Typically, the secant points, representing the standard parallels, are taken to be ~30°N and ~60°N. The Lambert conic projection is constructed by projecting outward from the center of the Earth through both the Earth's surface and the fitted cone; consequently, map distortion for this projection is smallest in the middle latitudes.

The Mercator projection is obtained by fitting a cylinder, with vertical axis located along the Earth's poles, tangent or secant to the Earth. The schematic presented in Fig. 3 is an example of a tangent fit, where the cylinder's surface intersects the Earth at one point, and is the more common of the two fits. Typically, the tangent point, representing the standard parallel, is taken to be 0°. The Mercator projection is constructed by projecting outward from the center of the Earth through both the Earth's surface and the fitted cylinder; consequently, map distortion for this projection is smallest in the tropics.

The polar stereographic projection is obtained by placing a flat planar surface parallel to the Earth's equator either tangent or secant to the Earth near the North or South Pole. The schematic presented in Fig. 3 is an example of a tangent fit where the planar surface intersects the Earth at the North Pole. Both tangent and secant fits of the planar surface to the Earth are common with polar stereographic projections, and the definition of the standard parallel(s) for each are similar to those for the Lambert conic and Mercator projections described above. Polar stereographic projections are constructed by projecting outward from the opposite pole (in Fig. 3, the South Pole) through the Earth's and planar surfaces. Map distortion for this projection is smallest near the poles.

The distance distortion can be quantified by the **map scale factor**, or *m*. For each projection, the map scale factor is generally defined as the *distance between points on the model grid* divided by the *distance between points on the Earth*, i.e.,

$$m = \frac{\Delta x_g}{\Delta x_e}$$

where Δx_g (distance on the model grid) and Δx_e (corresponding distance on the Earth) are defined graphically in Fig. 3. At the standard parallels, where the Earth's surface intersects the secant or tangent surface, $m = 1$. For secant projections, $m < 1$ ($\Delta x_g < \Delta x_e$) between the standard parallels and $m > 1$ ($\Delta x_g < \Delta x_e$) outside of the standard parallels. For tangent projections, $m > 1$ away from the standard parallel.

Using spherical geometry, precise relationships for *m* for each map projection may be obtained. These are provided below, assuming a tangent fit to the cylinder for the Mercator projection and secant fits to the surface for the Lambert conic and polar stereographic projections.

$$m = \frac{\sin\phi_{ref}}{\sin\phi}\left(\frac{\tan\left(\frac{\phi}{2}\right)}{\tan\left(\frac{\phi_{ref}}{2}\right)}\right)^{n}, \text{ where } n = \frac{\ln\left(\frac{\sin\phi_{ref1}}{\sin\phi_{ref2}}\right)}{\ln\left(\frac{\tan\left(\frac{\phi_{ref1}}{2}\right)}{\tan\left(\frac{\phi_{ref2}}{2}\right)}\right)} \qquad \text{(Lambert conic)}$$

$$m = \sec\phi \qquad \text{(Mercator)}$$

$$m = \frac{1+\sin\phi_{ref}}{1+\sin\phi} \qquad \text{(Polar stereographic)}$$

For each projection, ϕ is latitude and subscripts of *ref* refer to the reference or standard latitudes. The map scale factor for these projections, and all conformal projections, is equal in the x and y directions. This is not true for non-conformal latitude-longitude projections, however.

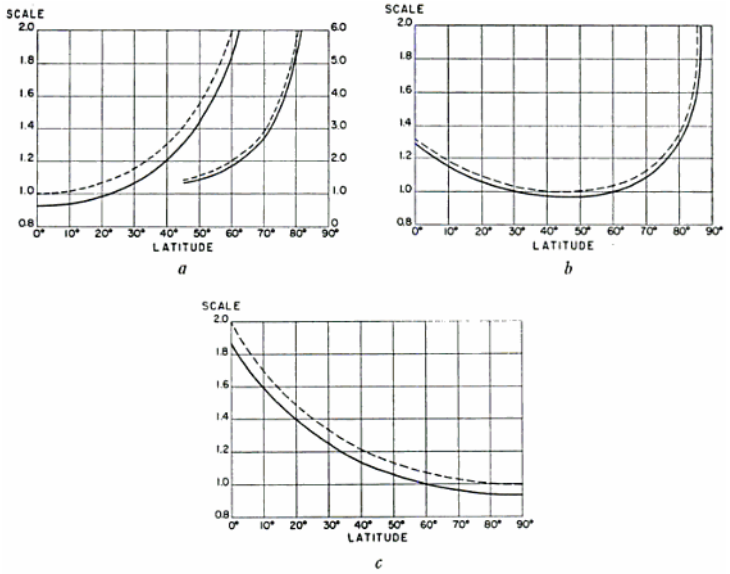

Figure 4. Map scale factors as a function of latitude for the (a) Mercator, (b) Lambert conic, and (c) polar stereographic map projections. In each panel, solid (dashed) lines refer to the map scale factor for secant (tangent) forms of each projection. The standard parallels for the secant projections are (a) 20°S and 20°N, (b) 30°N and 60°N, and (c) 60°N. The standard parallels for the tangent projections are (a) 0°, (b), 45°N, and (c) 90°N. For the Mercator projection, the second set of curves scales with the y-axis on the right side of the panel. Figure reproduced from Saucier (1955, *Principles of Meteorological Analysis*), their Fig. 2.113.

Graphs of the map scale factor for each projection as a function of latitude are presented in Fig. 4. We desire that $m \approx 1$ over the range of latitudes considered within our simulation domain. We find

that the Mercator projection is ideally suited for simulations within the tropics, the Lambert conic projection is ideally suited for simulations within the middle latitudes, and the polar stereographic projection is ideally suited for simulations at higher latitudes. Secant forms of each projection typically are applicable over a wider range of latitudes than the tangent forms.

Latitude-Longitude Grids

It is also possible to use a latitude-longitude, or *equirectangular cylindrical*, map projection. Like a Mercator projection, the latitude-longitude map projection is based upon the representation of the spherical Earth on the surface of a cylinder. Whereas the Mercator projection is obtained by literally projecting the Earth onto the cylinder's surface, however, the latitude-longitude map projection is obtained simply by unfurling the Earth's surface onto the surface of a cylinder that is tangent to the Earth at the Equator.

Whereas conformal map projections are associated with equal horizontal grid spacing in terms of *distance* (e.g., constant Δx_g), latitude-longitude grids are associated are associated with equal horizontal grid spacing in terms of *latitude and longitude*. Models that utilize latitude-longitude map projections oftentimes will pose the primitive equations in spherical coordinates, where the horizontal coordinates are latitude and longitude rather than the Cartesian coordinates x and y. While it is possible for a limited-area model grid to be constructed from the latitude-longitude map projection, this map projection is more commonly associated with global model grids.

Latitude-longitude map projections are not conformal; as a consequence, the map scale factor for latitude-longitude map projections varies between the x and y directions. As the distance between latitudes is constant with latitude (1° = 111,177 m), there exists no map distortion in the meridional (north-south) direction; so that the map scale factor m_y is 1. The map distortion in the zonal (east-west) direction is a function of latitude, where:

$$m_x = \sec \phi$$

This value is 1 at the Equator and increases slowly through the tropics, more rapidly through the middle latitudes, and exponentially approaching the poles. Thus, the least distance distortion in the zonal direction with the latitude-longitude projection is found within the tropics.

Because latitude-longitude grids are based upon discrete increments of latitude and longitude, there exists a singularity at the North and South Poles where all meridians (or lines of constant longitude) converge. As one approaches the poles, the physical distance between individual grid points along parallels (or lines of constant latitude) becomes infinitesimally small. This necessitates a small time step for computational stability. The added computational expense that results generally far exceeds the added benefit from implicitly finer model resolution.

Models utilizing latitude-longitude grids can handle this in one of two ways. One method, as is utilized by the WRF-ARW model, involves applying a Fourier filter to the model fields near the poles to truncate high wavenumber (e.g., small-scale) variability. This permits the use of a longer model time step despite a comparatively large number of grid points over a small area, though it still necessitates solving the primitive equations on the relatively dense grid near the poles.

The manifestation of this Fourier filter in the WRF-ARW model takes the form given in Section 4.1 of Skamarock et al. (2019):

$$\hat{\varphi}_{filtered}(k) = a(k)\hat{\varphi}(k)$$

Here, k is a dimensionless wavenumber, $a(k)$ is the filter function, $\hat{\varphi}(k)$ are the Fourier coefficients for a generic variable φ before filtering, and $\hat{\varphi}_{filtered}(k)$ are the Fourier coefficients for a generic variable φ after filtering. The $\hat{\varphi}(k)$ are obtained by applying a one-dimensional Fourier transform to φ on the constant latitude-longitude grid; the filtered variable φ is then obtained by applying a one-dimensional inverse transform given filtered Fourier coefficients $\hat{\varphi}_{filtered}(k)$.

The filter function $a(k)$ takes the form:

$$a(k) = \min\left[1, \max\left(0, \left(\frac{\cos\phi}{\cos\phi_0}\right)^2 \left(\frac{1}{\sin^2(\pi k / n)}\right)\right)\right]$$

Here, ϕ is latitude, ϕ_0 is the latitude above which the polar filter is applied (no filtering takes place at lower latitudes), and n is the number of zonal grid points along a parallel. The value of n is determined from the chosen horizontal grid spacing. At higher latitudes, cos ϕ approaches zero, such that the value of $a(k)$ approaches zero approaching the poles. Thus, greater filtering is applied at higher latitudes. Generally speaking, the $\sin^2$ function permits the retention of variability with wavelength of approximately 2Δ (i.e., in wavenumber space, where $k = n/2$) and larger.

Another method is to utilize a *reduced latitude-longitude grid*. In this method, the distance between latitude and longitude points is not fixed across the model grid; rather, the distance between longitude points grows larger approaching the poles. A hypothetical example of this is a latitude-longitude grid with grid spacing of 0.2° latitude at all latitudes and grid spacing of 0.2° longitude between the Equator and 60°N/S that grows to 2° at the poles.

Numerical Considerations of Conformal and Latitude-Longitude Map Projections

Fundamentally, the primitive equations apply to the Earth; in other words, ∂x and ∂y are finite displacements in the x and y directions, respectively, *on the Earth*. The horizontal velocities u and

v are related to position displacements *on the Earth*. However, with a numerical model, ∂x and ∂y are finite displacements in the *x*- and *y*-directions, respectively, *on the model grid*.

For conformal map projections, the physical distance Δx_e between grid points varies across the grid as a function of the map scale factor *m*, where $\Delta x_e = \Delta x_g / m$. For latitude-longitude map projections, the physical distance Δx_e between meridional grid points does not vary across the grid; however, the physical distance Δx_e between zonal grid points does vary across the grid as a function of the map scale factor m_x.

Consequently, the modeled form of the primitive equations must account for the departure of the horizontal grid spacing Δx_g from Δx_e. This is accomplished by (a) redefining the model kinematic variables with respect to the map scale factor and (b) coupling the primitive equations to the map scale factor. Note that the model kinematic variables are redefined exclusively so that map scale factor terms that result from coupling the primitive equations to the map scale factor cancel with those introduced into the kinematic variables. In other words, the precise definition of the redefined kinematic variables has nothing to do with the physical representation of the kinematic variables.

The model kinematic variables are redefined so that m_x acts only on ∂x and m_y acts only on ∂y in the expanded forms of the primitive equations. Thus, the combined effect of redefining the model kinematic variables and coupling the primitive equations to the map scale factor is to transform ∂x and ∂y so that they represent physical distance on the Earth and not the model grid. Doing so does, however, result in terms that reflect meridional variability in the map scale factor *m*, m_x, and/or m_y, since the map scale factor is uniform in the zonal direction for the latitude-longitude and all conformal map projections.

The coupled forms of the primitive equations, as integrated within the WRF-ARW model, are given by equations (2.23) through (2.31) of Skamarock et al. (2008). In light of these equations, the above discussion highlights two important considerations:

- When considering numerical stability – and, specifically, Δx in the context of the Courant number and CFL criterion – for models that utilize a map projection, Δx refers to Δx_e. Given that Δx_e varies across the grid, this means that the CFL criterion also can vary across the grid. Furthermore, given that $m \geq 1$ except between the standard parallels of a secant map projection, this means that $\Delta x_e \leq \Delta x_g$. This necessitates using a smaller time step Δt to maintain numerical stability than would be necessary if no map projection were used.

- Latitudinal variability in the map scale factor near the secant or tangent points of conformal map projections, or near the Equator for latitude-longitude map projections, is relatively small. It becomes large, however, away from these locations, as can be inferred from Fig. 4. Latitudinal variability, whether small or large, must be accounted for when coupling the primitive equations to the map scale factor. Doing so introduces a comparatively large non-physical correction to the primitive equations when applied on the portion(s) of model grids where latitudinal variability in the map scale factor is large.

As a result of these and other considerations, the choice of map projection for a given simulation domain is motivated by a desire to *minimize distance distortion* (i.e., departures of Δx_e from Δx_g) and to *minimize latitudinal variability in the map scale factor*. A review of Fig. 4 suggests that both may be accomplished by choosing the map projection based upon where the map scale factor is closest to 1 over the widest range of latitudes covered by the simulation domain. In other words, apply the latitude-longitude map projection globally or in the tropics, the Mercator map projection in the tropics, the Lambert conic map projection in the middle latitudes, and the polar stereographic map projection at higher latitudes.

Further, the exact forms of the curvature and Coriolis terms within the *u*-, *v*-, and *w*-momentum equations depend upon the chosen map projection. For the WRF-ARW model, these are given by equations (2.32) – (2.34) of Skamarock et al. (2008) for conformal map projections and equations (2.35) – (2.37) of Skamarock et al. (2008) for the latitude-longitude map projection.

For conformal map projections, spherical geometry allows terms involving $\tan \phi$ to be written in terms of the spatial variability of the map scale factor. The same is not true for latitude-longitude map projections, however. Note also that departures of the grid *y*-axis from the Earth's *y*-axis (i.e., meridians, or lines of constant longitude) must be accounted for in the formulation of the vertical component of the Coriolis force $e = 2\Omega\cos(\phi)$ for both types of map projections. Otherwise, the exact forms of the curvature and Coriolis terms for both types of map projections are similar to their forms applicable on the Earth or on grids where no map projection is used.

Spherical Geodesic Grids

A *geodesic* is the shortest possible line between two points on a sphere or other curved surface. Those familiar with transcontinental aircraft routing are familiar with geodesics by another name: *great circle routes*. Indeed, a geodesic is a segment of a great circle. The roughly spherical Earth can be discretized into a *spherical geodesic grid* by dividing it into a collection of spherical equilateral triangles whose sides are geodesics. Spherical geodesic grids are associated with minimal distance distortion between the Earth and model grid. Furthermore, spherical geodesic grid cells are distributed nearly homogeneously across the sphere. Both attributes make spherical geodesic grids particularly appealing for global numerical weather prediction. The NOAA Finite-volume Icosahedral Model (FIM) and NCAR/LANL Model for Prediction Across Scales (MPAS) models are two examples of modern models that utilize spherical geodesic grids.

The process of discretizing the Earth on a spherical geodesic grid is illustrated in Fig. 5. To obtain a spherical geodesic grid, one may begin by fitting an icosahedron – a three-dimensional geometric solid with twenty triangular faces, twelve vertices, and thirty sides – to the spherical Earth. The twelve vertices are positioned such that they touch the spherical surface (Fig. 5a).

Next, this grid may be refined by subdividing the twenty triangular faces. For example, bisecting (or splitting in half) each side of a spherical triangle and connecting the resulting subdivision points (Fig. 5b) results in four new flat triangular faces for each original spherical triangular face. Note that this is not the only means of subdividing the twenty triangular faces; it is merely illustrative. The new vertices that result from this subdivision – the bisection points – are then positioned such that they also touch the spherical surface (Fig. 5c). Replacing the lines between subdivision points with geodesics transforms the flat triangular faces into spherical grid triangles (Fig. 5d).

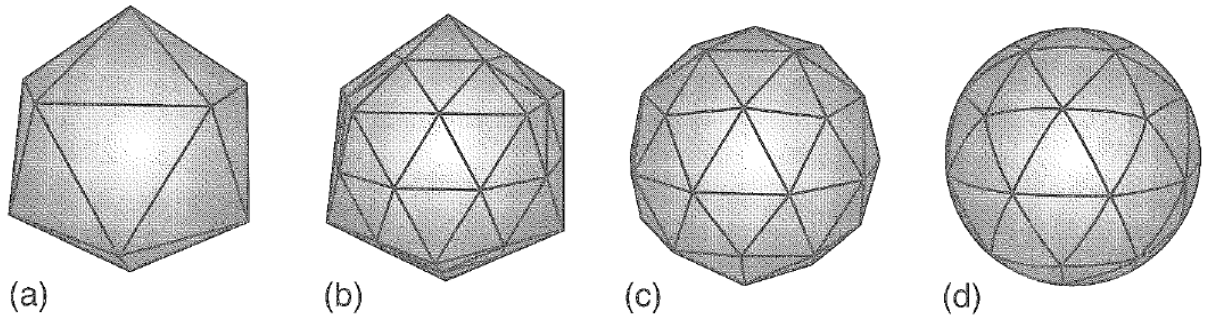

Figure 5. Process of constructing a spherical geodesic grid by fitting an icosahedron to the spherical Earth and subdividing it with spherical triangular faces. Please refer to the text for further details. Reproduced from Warner (2011), their Fig. 3.10.

Consider Fig. 5a. Each vertex is the intersection point of five triangles, or equivalently is the center point of a spherical pentagon. Now, consider Fig. 5d. The original vertices remain as the intersection points of five triangles or center points of spherical pentagons. However, the new vertices are the intersection points of six triangles, or equivalently the center points of spherical hexagons. Thus, there are spherical triangle sides that connect two vertices that are each the center of spherical hexagons, while there are also spherical triangle sides that connect a vertex that is the center of a spherical hexagon to a vertex that is the center of a spherical pentagon. The distances of these two types of sides differ slightly from one another, with the latter slightly larger than the former. This results in the minimal distance distortion previously described.

Repeating the refinement process described above, such as that outlined in Fig. 5(b-d), can further refine the grid until the desired horizontal grid spacing is achieved. For a spherical geodesic grid, the horizontal grid spacing is defined relative to the size of each grid cell; e.g., what is the great circle distance between the centers of adjacent grid cells? Furthermore, note that this refinement need not be done uniformly across the sphere. If refined in a non-uniform fashion, the refinement process allows for variable resolution across the simulation domain. Variable resolution is discussed in greater detail in the next section of these notes.

Though there may be a "method to the madness," so to speak, with respect to how the spherical triangles are obtained, spherical geodesic grids are *unstructured* grids. Thus, information regarding how the model grid is arranged – e.g., the locations of each grid center – must be carried with the

model during its integration. For structured grids such as those defined based upon conformal and latitude-longitude map projections, this is not true; adjacent points are simply one point away in the *x*/longitude or *y*/latitude dimension.

Spherical geodesic grids need not exclusively be based upon triangular refinement. For instance, it is possible to use the triangular refinement as the basis for a grid that is composed primarily of spherical hexagons. This is the approach taken by the MPAS model, as illustrated in Fig. 1 (right). The hexagonal refinement is obtained from a triangular refinement, such as illustrated in Fig. 6, as follows. Identify the bisection point of each spherical triangle's side. Draw a geodesic line through this point perpendicular to each spherical triangle's side. Terminate these geodesic lines where they intersect with each other. The resultant spherical polygons are hexagons whose centers are triangle vertices. The hexagons encompass the portions of each triangle that lies closest to the hexagon's centers.

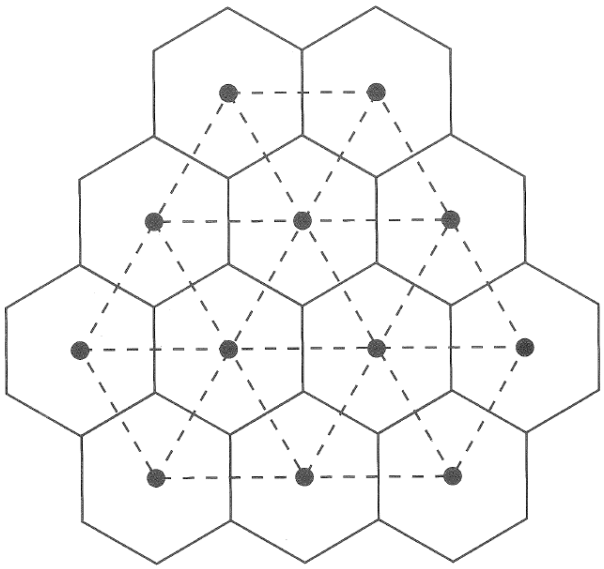

Figure 6. Relationship between triangular and hexagonal refinement of a spherical geodesic grid. Please refer to the text for further details. Reproduced from Warner (2011), their Fig. 3.12.

Spherical geodesic grids are often utilized in *finite-volume* models, although not all finite-volume models use spherical geodesic grids (the Finite-Volume 3, or FV3, model being one example of a finite-volume model that does not use a spherical geodesic grid). Traditional grid-point models consider model variables only at grid points, which are assumed to be representative of the entire grid cell. Finite-volume models, on the other hand, consider the grid-cell-averaged values of model variables.

Methods for Variable Grid Resolution

Independent of the grid and map projection used, it is possible to construct a model configuration with variable grid resolution. The primary benefit to doing so is to allow for finer discretization over an area or for a feature of interest while minimizing the computational expense that would result from finer discretization over the entire domain. Nominally, variable grid resolution is most commonly considered in the horizontal, given that nearly all modern models utilize variable grid resolution in the vertical. The precise means by which variable grid resolution is achieved depends upon whether an unstructured or structured grid is utilized, and there exist multiple options for doing so in the latter case.

Unstructured grids generally permit variable grid resolution simply by adding or removing grid elements within a region of interest. The example provided in Fig. 1 (right) is an example of a variable resolution unstructured grid, with horizontal grid spacing of ~60 km over much of the globe gradually reducing to ~15 km over North America. Other grid discretizations are possible.

It is not as straightforward to utilize variable grid resolution with structured grids. There exist two primary approaches to doing so: nested domains and grid stretching. *Nested domains* reflect the case where the model is integrated over multiple, often telescoping domains. Examples of nested domains are given in Figs. 7 and 8. A nested grid is often referred to as a child domain, whereas the grid that surrounds the nested grid is often referred to as a parent domain. In the example below, the outermost grid is the parent of the intermediate grid, which is the child of the outermost grid and the parent of the innermost grid.

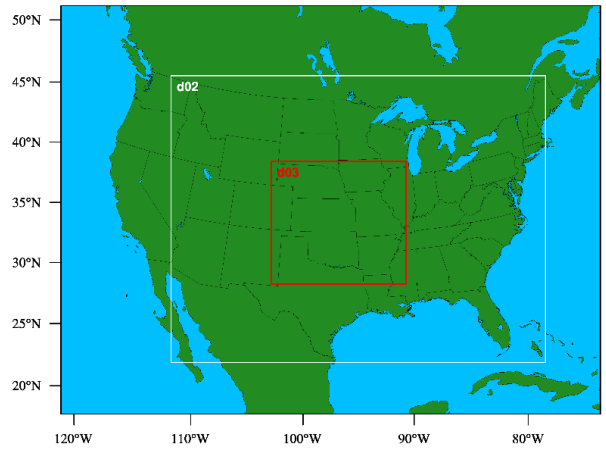

Figure 7. An example of nested limited-area model domains. In this example, the outermost nest (d01) contains 148 x 112 grid points and has a horizontal grid spacing of 36 km. A global or larger-area model provides the lateral boundary conditions for this domain. The intermediate nest (d02) contains 307 x 235 grid points and has a horizontal grid spacing of 12 km. The outermost nest provides the lateral boundary conditions for this nest. The innermost nest (d03) contains 331 x 301 grid points and has a horizontal grid spacing of 4 km. The intermediate nest provides the lateral boundary conditions for this nest.

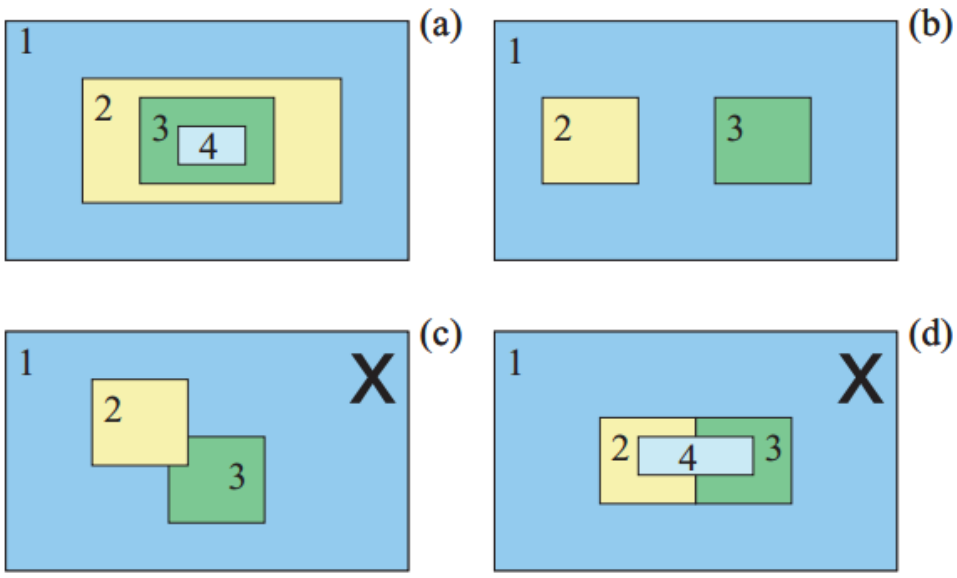

Figure 8. An example of four nested grid configurations, two of which are typically allowable and two of which are not. In (a), a total of four telescoping domains are depicted. In (b), a total of three domains are depicted, with both domains 2 and 3 having domain 1 as their parent. In (c), three domains are again depicted, with domains 2 and 3 overlapping; this is typically not allowable. Finally, in (d), four domains are depicted, with the innermost domain having multiple intermediate domains as its parents; this is also typically not allowable. Figure obtained from http://www2.mmm.ucar.edu/wrf/users/docs/technote/v4_technote.pdf, their Fig. 7.2.

Nested domain simulations can feature either one-way or two-way feedback. In the case of one-way feedback, the model is first integrated over the outermost domain, the output from which provides initial and lateral boundary conditions to a nested domain over which the model is subsequently integrated. In the case of two-way feedback, the model is integrated concurrently for all nested domains. Here, the outer domain(s) also provide initial and lateral boundary conditions to the inner domain(s), but what happens on the inner domain(s) is now able to feed back to the outer domains(s). For one-way feedback, the outer domain solution(s) was/were not affected by that on the inner domain(s).

Nested domain(s) need not necessarily start and end at the same time as their parent domain(s), nor must they necessarily remain in a fixed location throughout the duration of a numerical simulation. For instance, both research and operational numerical simulations of tropical cyclones will often utilize a nested domain configuration in which the innermost nest moves with the tropical cyclone. Not all models allow such flexibility in nested domain configuration, however. Furthermore, a nested grid will typically have a horizontal grid spacing that is finer by a factor of a small odd integer – oftentimes 3:1, but sometimes 5:1 or 7:1 – compared to its parent. As a result, there is an abrupt change in horizontal grid spacing along the lateral boundary between a parent and a child domain. If not treated appropriately, this can compromise forecast quality. This and other considerations related to lateral boundary conditions will be discussed in a later lecture.

Grid stretching enables the horizontal grid spacing of a limited-area model simulation to decrease monotonically from that of the larger-area or global model that provides it with initial and lateral boundary conditions to a specified smaller value over the domain interior. An example, as applied within the Regional version of the Canadian Global Environmental Multiscale (R-GEM) model, is given in Fig. 9. As compared to nested grids, there is no abrupt change in horizontal grid spacing between a parent and child domain with grid stretching.

Figure 9. An example of a stretched horizontal grid. In this example, the full limited-area model domain is 353 x 415 grid points, with the centermost 240 x 323 grid points having uniform horizontal resolution. The horizontal grid spacing decreases by 10% per each model grid length as one moves from the periphery of the limited-area domain, where it is identical to that of the global model used to initialize and provide lateral boundary conditions to the simulation (2°), to the center of the limited-area domain, where it is 0.04°. Figure obtained from Yeh et al. (2002, *Mon. Wea. Rev.*), their Fig. 6.

As noted above, most modern models use variable grid resolution in the vertical. Typically, this is achieved with smaller vertical grid spacing near the ground (and sometimes near the tropopause) and larger vertical grid spacing elsewhere. An example is given below in Fig. 10. However, for those simulations that use nested grids, some numerical models permit the use of variable grid resolution between domains, with more vertical levels and finer vertical discretization utilized on the inner nests. This helps maintain consistency between horizontal and vertical grid increments where other means of variable grid resolution would not. The consistency between horizontal and vertical grid increments is discussed further below.

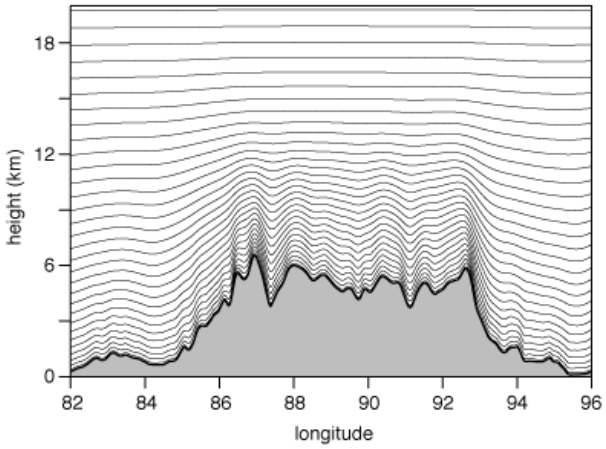

Figure 10. An example of a hybrid terrain-following vertical coordinate over a region of sloped topography. In this example, representative of the vertical coordinate employed in the Model for Prediction Across Scales, vertical levels follow the terrain below approximately 12 km above sea level. At higher altitudes, vertical levels closely resemble constant height surfaces. Image obtained from https://mpas-dev.github.io/.

Consistency between Horizontal and Vertical Grid Spacing

Historically, greater attention has been given to the horizontal grid spacing and how well it is able to faithfully represent a feature or features of interest. However, we know that many atmospheric features are not purely vertical, but rather slope horizontally with height: cold and warm fronts on synoptic scales; cold pools, density currents, and inversions on the mesoscale, gravity and inertia-gravity waves on meso- to microscales, and so on. We want these sloping features to be represented smoothly on the model's grid. In other words, we desire that features that are well-resolved by the chosen horizontal grid spacing for a numerical simulation also be well-resolved in the vertical. If a feature is insufficiently resolved in the vertical relative to its horizontal representation, spurious gravity waves can occur, thereby compromising simulation quality.

Methods for determining the appropriate vertical grid spacing for a given feature or horizontal grid spacing stem from the underlying atmospheric dynamics. For example, at midlatitudes, the vertical grid spacing needed to vertically resolve a feature can be related to the horizontal grid spacing by the following expression (Pecnick and Keyser 1989, *Meteor. Atmos. Phys.*):

$$\frac{\Delta z}{\Delta x} \leq s$$

where s is a measure of the vertical slope of an atmospheric phenomenon to be studied, such as a front. For the example of a front, as well as many other phenomena, s is ~0.005 to ~0.02 (e.g., rise 1 km for every 50 km to 200 km in the horizontal). This equation states that the slope of the model

grid (given by $\Delta z/\Delta x$) should be less than or equal to the slope of the feature the model grid needs to resolve – which is rather intuitive.

The slope *s* of a given feature can be estimated from observations (e.g., vertical cross-sections) or may be estimated from dynamical principles. For example, *s* can be estimated from the ratio of the atmospheric scale height ($\Delta z \sim H_0$) to the Rossby radius of deformation ($\Delta x \sim \Delta L$; see also Lindzen and Fox-Rabinovitz 1989, *Mon. Wea. Rev.*):

$$\frac{\Delta z}{\Delta x} \leq \frac{f}{N}$$

In the above, *f* is the Coriolis parameter and *N* is the Brunt-Väisälä frequency, a measure of static stability.

Consider a simulation that contains a cold front somewhere in its domain. Let us assume that the slope *s* is equal to 0.01 for this front and that Δx is 20 km. Per the above equation, Δz then should be equal to 0.2 km, or 200 m. For a model domain with a model top at 20 km above ground level, this suggests that 100 vertical levels are required to appropriately resolve the vertical variability in model fields associated with the cold front. By contrast, most numerical models with $\Delta x \sim 20$ km typically utilize between 30 and 80 vertical levels. Thus, one might conclude that the vertical grid spacing used in most simulations is too coarse, and this is true to a point. However, because most models use a variable vertical grid spacing, with finer spacing in regions where the sharpest vertical gradients typically exist, the discrepancy between the chosen horizontal and vertical grid spacings is not as large as the initial evaluation would suggest.

Although these formulations are intuitive even independent of the underlying dynamics, formally, these formulations are based on midlatitude dynamics and are not *directly* applicable in the tropics. Alternatively, so long as the model is capable of faithfully representing the underlying dynamics, one may evaluate a model's kinetic energy spectrum – a measure of the variation in kinetic energy as a function of wavelength/wavenumber tied to the underlying dynamics – for a given horizontal grid spacing to identify the vertical grid spacing below which kinetic-energy spectra are similarly represented across simulations with varying Δz. In the literature, this is referred to as *convergence*. A representative example is given by Skamarock et al. (2019, *Mon. Wea. Rev.*) in Fig. 11. For the model used (Δx = 15 km), a vertical grid spacing of 200 m was identified to be the spacing below which convergence occurs. At coarser Δz, sloping atmospheric features are insufficiently resolved, resulting in the generation of spurious gravity waves that transfer kinetic energy from larger scales (e.g., the red and black lines lying below the green and blue lines on the synoptic scale) to smaller scales whereupon the kinetic energy is dampened.

In the case of insufficient vertical resolution relative to the horizontal grid spacing, spurious gravity waves are generated. These superpose on the physical solution, degrading the quality of the model solution. Consider Fig. 12. In Fig. 12a, where an appropriate vertical grid spacing for the horizontal grid spacing and slope of the meteorological phenomena is used, the model solution is smooth. In

Fig. 12b, which differs from Fig. 12a only in that a vertical grid spacing that is three times as large is used, the model solution is contaminated by spurious gravity waves. When the horizontal grid spacing is coarsened, as in Fig. 12c, a smooth solution is obtained, but with reduced amplitude compared to that in Fig. 12a.

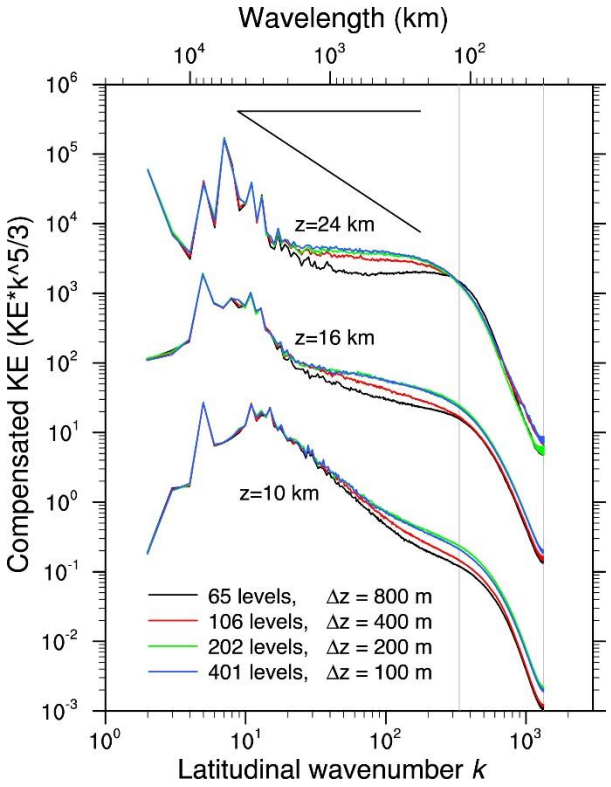

Figure 11. Kinetic energy times the latitudinal wavenumber *k* raised to the 5/3rd power (*y*-axis) as a function of latitudinal wavenumber *k* (inversely related to wavelength; smaller *k* indicates larger wavelength) at three altitudes (z = 10 km, 16 km, 24 km) from Δx = 15-km, day 6-7 global forecasts from the Model for Prediction Across Scales. Outputs from forecasts with vertical grid spacing of 800 m, 400 m, 200 m, and 100 m are given by the black, red, green, and blue curves, respectively. Note the similarity between the green and blue curves at each altitude, indicating convergence of the kinetic-energy spectra. Figure reproduced from Skamarock et al. (2019, *Mon. Wea. Rev.*), their Fig. 2.

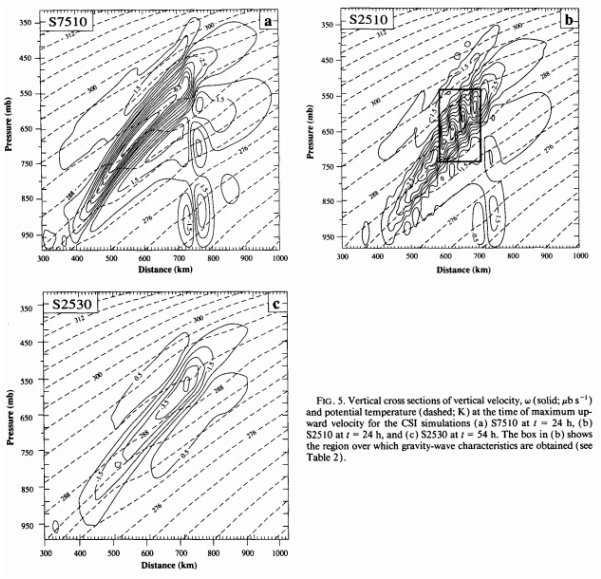

FIG. 5. Vertical cross sections of vertical velocity, ω (solid; μb s^{-1}) and potential temperature (dashed; K) at the time of maximum upward velocity for the CSI simulations (a) S7510 at t = 24 h, (b) S2510 at t = 24 h, and (c) S2530 at t = 54 h. The box in (b) shows the region over which gravity-wave characteristics are obtained (see Table 2).

Figure 12. Vertical cross-sections (x, p) of vertical velocity ω (solid lines, μbar s^{-1}) and potential temperature θ (dashed lines, K) at the time of maximum upward velocity from three numerical simulations of a case of conditional symmetric instability (a slantwise instability): (a) Δx = 10 km, 75 vertical levels; (b) Δx = 10 km, 25 vertical levels; and (c) Δx = 30 km, 25 vertical levels. In panel (b), note the high-frequency wave structure in the vertical velocity field, indicative of the presence of gravity waves, where no such structure is evident in panels (a) and (c). In panel (c), the reduced resolution of the model contributes to weaker vertical velocities compared to panels (a) and (b). Figure obtained from Persson and Warner (1991, *Mon. Wea. Rev.*), their Fig. 5.