Aliasing, Non-Linear Instability, and Diffusion

6 October 2015

Note: There exist several sign errors within the course textbook in its mathematical formulation for aliasing. Any discrepancies between the notes below and the course textbook should be reconciled in favor of the notes below, which are believed to be correct.

Aliasing and Non-Linear Instability

Our consideration of numerical instability to this point has emphasized determining the stability criteria for linear forcing terms. Generally speaking, linear numerical stability is said to exist when the amplitude of the solution does not grow exponentially with time (e.g., $e^{\omega t} < 1$).

The primitive equations, however, contain non-linear forcing terms. Consider, for instance, a one-dimensional advection equation for the zonal velocity $u$:

$$\frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x}$$

The stability of this equation can be evaluated as before, from which a stability criterion may be obtained. This stability criterion is dependent upon the chosen combination of temporal and spatial finite differencing schemes and must be adhered to in order to ensure numerical stability.

However, there exists a second potential source of non-linear instability that must be considered when determining numerical stability. Aliasing occurs when two waves represented on a model grid interact, such as by obtaining the product of the two waves, and produce fictitious waves and an erroneous redistribution of energy within wave space. Aliasing can arise in any model that discretizes the primitive equations using finite difference approximations in an Eulerian framework. In the following, we will develop a framework for demonstrating aliasing that is independent of the chosen temporal and spatial finite differencing schemes.

Note that aliasing does not impact models that use semi-Lagrangian methods, wherein non-linear terms are encapsulated within the total derivative-based (i.e., flow-following) formulation of the primitive equations. Aliasing also does not impact models that use spectral methods, wherein model variables and their partial derivatives are treated analytically and potentially troublesome wave interactions (to be defined below) are not permitted. The absence of aliasing with such methods is one of several reasons why they gained widespread use in operational numerical weather prediction in recent decades.

The byproduct of aliasing is the accumulation of erroneous wave energy at short wavelengths (generally speaking, $\leq 4\Delta x$), which can lead to the model solution becoming unstable with time. Previously, we have demonstrated that short wavelengths have large truncation error (i.e., are
poorly-resolved), are among those whose amplitudes grow most rapidly if the stability criterion is violated, and are associated with significant departures of both phase speed and group velocity from the advective velocity (numerical dispersion). Thus, aliasing is yet another reason why shorter wavelengths are particularly problematic for gridpoint methods for numerical weather prediction and thus why implicit or explicit damping of such wavelengths can be beneficial.

*Analytic Framework for Aliasing*

Herein, we follow the example provided by the course text. Consider the non-linear one-dimensional advection equation above. As we have done previously, let us assume a wave-like solution for \( u \). For simplicity, assume that this wave-like solution is given by the linear superposition of cosine waves, e.g.,

\[
  u = \sum_{m=0}^{\infty} a_m \cos(k_m x)
\]

Here, wavenumber \( k_m = 2\pi m/L \), where \( m \) is a zonal wavenumber and \( L \) is the domain length. Note the slight difference in how this \( k \) is defined relative to that in our lecture on linear numerical stability, where \( k = 2\pi/L \). Here, the \( k \) is defined specific to a given wavelength through its dependence on \( m \).

The ratio of \( m \) to \( L \) is equal to the inverse wavelength, such that the ratio of \( L \) to \( m \) defines the wavelength (e.g., \( m = 1 \) defines a wave with wavelength \( L \), \( m = 2 \) defines a wave with wavelength \( L/2 \), etc.). In other words, \( m \) is the number of waves over the wavelength \( L \).

The first partial derivative of \( u \) with respect to \( x \) can be obtained analytically and is given by:

\[
  \frac{\hat{u}}{\hat{x}} = -\sum_{m=0}^{\infty} a_m k_m \sin(k_m x)
\]

Consequently,

\[
  -u \frac{\hat{u}}{\hat{x}} = \left( \sum_{m=0}^{\infty} a_m \cos(k_m x) \right) \left( \sum_{n=0}^{\infty} a_n k_n \sin(k_n x) \right)
\]

Note that the indices \( m \) and \( n \) may be switched without changing the result. The separate notation for each term is used to indicate that a wave in \( u \) of a given wavelength may interact with a wave in \( \hat{u}/\hat{x} \) of another wavelength.

Or, expanding the summation notation,
\[-u \frac{\partial u}{\partial x} = (a_0 + a_1 \cos(k_1x) + a_2 \cos(k_2x) + \ldots + a_\infty \cos(k_\infty x)) \times (a, k_1 \sin(k_1x) + a_2 k_2 \sin(k_2x) + \ldots + a_\infty k_\infty \sin(k_\infty x)) \]

In the above, for \( m = 0 \), \( \cos(k_0x) = 1 \), such that \( a_0 \cos(k_0x) = a_0 \). For \( n = 0 \), \( \sin(k_0x) = 0 \), such that \( a_0 k_0 \sin(k_0x) = 0 \).

Generally, the product of any two waves can be expressed as:

\[ a_m a_n k_n \sin(k_nx) \cos(k_mx) \quad \text{or, equivalently,} \quad a_n a_m k_m \sin(k_mx) \cos(k_nx) \]

We can simplify this expression further. Note that \( \sin x \cos y \) can be expressed using a trigonometric identity as follows:

\[ \sin x \cos y = \frac{\sin(x + y) + \sin(x - y)}{2} \]

For \( x = k_nx \) and \( y = k_mx \), we obtain:

\[ a_m a_n k_n \sin(k_nx) \cos(k_mx) = \frac{1}{2} a_m a_n k_n \sin((k_n + k_m)x) \sin((k_n - k_m)x) \]

Or, substituting for \( k_n \) and \( k_m \),

\[ \frac{1}{2} a_n a_m k_n \sin \left( \frac{2\pi}{L} (n + m)x \right) \sin \left( \frac{2\pi}{L} (n - m)x \right) \]

There exist two waves defined by the above – the \( n + m \) wave and the \( n - m \) wave. Note, as before, that the indices \( m \) and \( n \) may be swapped without changing the result.

In physical space, where all wavenumbers are possible, this is not a problem. However, on a model grid, only waves of wavelength \( 2\Delta x \) and larger may be represented. Recall that the ratio of \( L \) to \( m \) describes the wavelength of a wave. Consider a one-dimensional model grid with \( j_{\text{max}} \) grid points, such that \( j_{\text{max}} \Delta x = L \). Thus,

\[ \frac{j_{\text{max}} \Delta x}{m} = 2\Delta x \]

Solving for \( m \), we obtain \( j_{\text{max}}/2 \). This represents the \textit{maximum} value of \( n + m \) that may be represented on a model grid. You can prove this by considering other wavelengths longer than \( 2\Delta x \) in the above – e.g., for the \( 3\Delta x \) wave, \( m \) equals \( j_{\text{max}}/3 \), which is smaller than \( j_{\text{max}}/2 \).

Thus, the following inequality must hold in order for the \( n + m \) wave defined by the product of \( u \) and \( \partial u/\partial x \) to be represented on the model grid:
\[ n + m \leq \frac{j_{\text{max}}}{2} \]

Or, stated in the inverse, the following inequality describes the case where the \( n + m \) wave defined by the product of \( u \) and \( \partial u / \partial x \) is unable to be represented on the model grid:

\[ n + m > \frac{j_{\text{max}}}{2} \]

Let us consider this unrepresentable wave in more detail. The inequality describing this wave can alternatively be written as:

\[ n + m = j_{\text{max}} - s \]

Here, \( s \) is some generic wavenumber, where \( s < \frac{j_{\text{max}}}{2} \). Thus, all values of \( j_{\text{max}} - s \) are greater than \( j_{\text{max}} / 2 \). Considering only the \( n + m \) wave, if we substitute this relationship for \( n + m \), we obtain:

\[ \sin \left( \frac{2\pi}{L} (n + m)x \right) = \sin \left( \frac{2\pi}{L} (j_{\text{max}} - s)x \right) \]

However, because we previously defined \( L = j_{\text{max}} \Delta x \), we can also substitute for \( L \) in the above. Further, the position \( x \) along the wave is given by the product of the grid index \( j \) and the grid spacing \( \Delta x \), such that we obtain:

\[ \sin \left( \frac{2\pi}{j_{\text{max}} \Delta x} (j_{\text{max}} - s)j \Delta x \right) \]

Simplifying the terms inside of the \( \sin \) function, we obtain:

\[ \sin \left( \frac{2\pi}{j_{\text{max}}} (j_{\text{max}} - s) \right) = \sin \left( \frac{2\pi j}{j_{\text{max}}} - \frac{2\pi js}{j_{\text{max}}} \right) \]

We can now apply another trigonometric identity,

\[ \sin(x - y) = \sin x \cos y - \cos x \sin y \]

Doing so, we obtain:

\[ \sin \left( \frac{2\pi j}{j_{\text{max}}} - \frac{2\pi js}{j_{\text{max}}} \right) = \sin(2\pi j/j_{\text{max}}) \cos \left( \frac{2\pi js}{j_{\text{max}}} \right) - \cos(2\pi j/j_{\text{max}}) \sin \left( \frac{2\pi js}{j_{\text{max}}} \right) \]
However, for all grid indices \( j \) (which are positive integers), \( \sin(2\pi j) = 0 \) and \( \cos(2\pi j) = 1 \). Thus, the above expression simplifies to the following:

\[
-\sin\left(\frac{2\pi sj}{j_{\text{max}}}\right)
\]

Noting that \( x = j\Delta x \) and \( L = j_{\text{max}}\Delta x \), this can be rewritten as:

\[
-\sin\left(\frac{2\pi s}{L}x\right)
\]

Because this is equal to the unresolvable \( m + n \) wave, we state that the unresolvable wave shows up on the model grid as one that has wavenumber equal to \( s \), where \( s = j_{\text{max}} - (n + m) \).

What does this mean in practice? Let us consider the interaction of two waves, for instance a \( 2\Delta x \) wave and a \( 4\Delta x \) wave. For the \( 2\Delta x \) wave, \( m = j_{\text{max}}/2 \). For the \( 4\Delta x \) wave, \( n = j_{\text{max}}/4 \). Please see the discussion at the bottom of page three of these notes to recall the basis for these definitions. In this case,

\[
m + n = \frac{j_{\text{max}}}{2} + \frac{j_{\text{max}}}{4} = \frac{3j_{\text{max}}}{4}
\]

This defines a wave with wavelength \( \frac{4}{3} \Delta x \), which cannot be resolved on the model grid. But,

\[
s = j_{\text{max}} - (n + m) = j_{\text{max}} - \frac{3j_{\text{max}}}{4} = \frac{j_{\text{max}}}{4}
\]

This defines a wave with wavelength \( 4\Delta x \), which can be resolved on the model grid! The unresolvable wave actually is resolved on the model grid, but in a non-physical way: it is aliased to a wavelength that is resolvable. Stated differently, the energy associated with the wave that is unresolved is folded over the shortest-resolvable wave (the \( 2\Delta x \) wave) into a wave that is resolved on the model grid.

Let us consider the idea of folding in a bit more detail. Consider a model grid with \( j_{\text{max}} = 24 \) grid points. We can obtain the values of \( m \) and \( n \) for the \( 2\Delta x \) and \( 4\Delta x \) waves on this grid as follows. Recall that \( L = j_{\text{max}}\Delta x = 24\Delta x \) and the ratio of \( L \) to \( m \) (or \( n \)) defines the wavelength of the wave. Thus, for the \( 2\Delta x \) wave,

\[
\frac{L}{m} = 2\Delta x \rightarrow \frac{24\Delta x}{m} = 2\Delta x \rightarrow m = 12
\]
And, for the $4\Delta x$ wave,

$$\frac{L}{n} = 4\Delta x \rightarrow \frac{24\Delta x}{n} = 4\Delta x \rightarrow n = 6$$

Thus, $m + n = 12 + 6 = 18$. As a result, $s = j_{max} - (n + m) = 24 - 18 = 6$. Since this is equal to $n$, the wave with wavenumber $s$ in this case is the $4\Delta x$ wave, as before. The unresolved wavenumber was 6 greater than the maximum-resolvable wavenumber ($12$, defined by the $2\Delta x$ wave), while the wavenumber to which it is aliased is 6 smaller than the maximum-resolvable wavenumber. This is the manifestation of folding over the shortest-resolvable wavelength, which is illustrated in Figure 1 below.

![Figure 1](image.png)

**Figure 1.** Conceptual illustration of how the interaction of two waves, with $m + n > j_{max}/2$, produces aliasing, manifest as the folding of wave energy across the shortest-resolvable wavelength, for $j_{max} = 24$. Adapted from Warner (2011), their Figure 3.32.

Note, however, that the interaction between two waves does not necessarily result in aliasing. Consider, for instance, the interaction of two well-resolved waves on this grid: the $12\Delta x$ wave ($m = 2$) and the $8\Delta x$ wave ($n = 3$). Here, $m + n = 2 + 3 = 5$, which defines a wave with wavelength $4.8\Delta x$ that can be resolved on the model grid. Only those interactions where $m + n > j_{max}/2$ will result in aliasing.

Let us continue to consider this model grid with $j_{max} = 24$ grid points. Allowable values of $m$ and $n$ each range from 0 to 12. There exist 42 distinct combinations of $m$ and $n$ that result in aliasing:
<table>
<thead>
<tr>
<th>Value(s) of $n$ (or $m$)</th>
<th>Value(s) of $m$ (or $n$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12</td>
</tr>
<tr>
<td>11</td>
<td>2, 3, 4, 5, 6, 7, 8, 9, 10, 11</td>
</tr>
<tr>
<td>10</td>
<td>3, 4, 5, 6, 7, 8, 9, 10</td>
</tr>
<tr>
<td>9</td>
<td>4, 5, 6, 7, 8, 9</td>
</tr>
<tr>
<td>8</td>
<td>5, 6, 7, 8</td>
</tr>
<tr>
<td>7</td>
<td>6, 7</td>
</tr>
</tbody>
</table>

This listing does not count duplicates; e.g., aliasing for $n = 11$ can also occur for $m = 12$, but this case is already accounted for by $n = 12$ and $m = 11$. One could follow a similar procedure to identify the distinct combinations (totaling 49) of $m$ and $n$ which do not result in aliasing.

Of these 42 combinations, 30 of them result in $m + n \leq 18$: six each for $n$ between 9 and 12, four for $n = 8$, and two for $n = 7$. Why are we interested in $m + n \leq 18$? Consider Figure 1. Unresolvable wavenumbers from 13 through 18 alias, or fold, to resolvable wavenumbers between 6 and 11. These identify waves with wavelengths of $2-4\Delta x$, or those that are poorly resolved on the model grid. Thus, **aliasing preferentially results in the artificial accumulation of wave energy at short, poorly-resolved wavelengths.**

![Figure 2](image-url)

**Figure 2.** Examples of modeled kinetic energy spectra relative to theory and observations (i.e., the “correct spectrum”) for (a) a case where numerical diffusion dampens short wavelengths (large wavenumber $k$, here expressed on a logarithmic axis) and (b) a case where aliasing is not controlled for by numerical diffusion, resulting in an excess of kinetic energy at short wavelengths (or large $k$). Reproduced from Warner (2011), their Figure 3.33.

When we introduced the concept of *effective resolution* earlier in the semester, we defined it as the smallest wavelength at which the modeled kinetic energy spectrum matches that from theory and observations. At smaller but still resolvable wavelengths, ideally the modeled kinetic energy spectrum is associated with less energy than that from theory and observations. Aliasing, however, can result in an excess accumulation of wave energy in short wavelengths, as the above
discussion illustrates. This can result in a modeled kinetic energy spectrum with greater energy that that from theory and observations at short wavelengths. Examples of each are provided in Figure 2 above.

This is a problematic occurrence. We know shorter wavelengths to have large truncation error, significant numerical dispersion, and to often be those that most rapidly become unstable if the numerical stability criterion is violated. Amplifying the amount of energy contained within these wavelengths only exacerbates this problem. This is yet another illustrative example of the utility of numerical diffusion, whether implicit – as considered in our lecture on linear stability, albeit manifest only for certain combinations of differencing schemes – or explicit in nature. We now wish to consider explicit numerical diffusion in more detail.

**Numerical Diffusion**

Diffusion can be defined by the spreading out, or smoothing, of atmospheric fields in all three spatial dimensions. As a result, diffusion is often characterized by *damping* given that it weakens gradients by reducing the magnitude of local maxima and minima (Figure 3). Physical diffusion, driven by turbulent eddies, is an important transporter of atmospheric fields. However, there also exist two forms of artificial, or numerical, diffusion:

- **Implicit numerical diffusion.** This is manifest through damping properties possessed by certain finite differencing schemes, in which the model solution is damped over time ($|e^{\alpha t}| < 1$). Specific characteristics of this damping, such as its scale selectivity and Courant number dependence, vary between finite differencing schemes.

- **Explicit numerical diffusion.** This is manifest by adding an explicit damping term to the predictive equations for each model variable. Specific characteristics of this term, such as its stability, its wavelength dependence, the extent to which it dampens, and so on, vary between diffusion formulations and finite differencing schemes.

Physical diffusion occurs primarily on scales that are not well-resolved by the model and thus must be parameterized. Given that the eddies responsible for physical diffusion occur primarily within the planetary boundary layer, physical diffusion is typically parameterized by a planetary boundary layer parameterization. Planetary boundary layer parameterizations also handle vertical diffusion where sharp gradients occur outside of the planetary boundary layer (e.g., with jets or deep, most convection). We will cover planetary boundary layer parameterizations and physical diffusion in more detail in a subsequent lecture.
Figure 3. Conceptualization of diffusion, whether physical or artificial in nature. In panel (a), a sharp horizontal gradient in temperature exists in the middle of the grid. In panel (b), diffusion has weakened this gradient and reduced the magnitude of the local temperature minima and maxima that were present in panel (a).

Implicit numerical diffusion was covered during our lecture material on linear numerical stability. As noted above, only selected finite differencing schemes are associated with implicit numerical diffusion. For example, the centered-in-time and centered-in-space differencing schemes are not associated with implicit numerical diffusion, while the forward-in-time and Runge-Kutta 3 time differencing schemes are associated with implicit numerical diffusion. Again, the specific details of such implicit diffusion are specific to the chosen differencing scheme. Please refer to the earlier lecture on linear numerical stability for specific examples.

Herein, we focus upon explicit numerical diffusion, or that which arises due to the inclusion of an explicit diffusion term within a model’s predictive equations. Though non-physical in nature,
explicit numerical diffusion may nonetheless be beneficial if it can dampen shorter wavelength, poorly-resolved phenomena while largely not affecting longer wavelength phenomena given the problems posed by shorter wavelength phenomena – truncation error, linear numerical stability, numerical dispersion, and aliasing – discussed over the course of the semester to date.

**Numerical Formulation for Explicit Diffusion**

A generalized explicit numerical diffusion term is given by:

$$\frac{\partial h}{\partial t} = (-1)^{\frac{n}{2}} K_n \nabla^n h$$

Here, $h$ is any model dependent variable, $n$ is the order of the diffusion operator ($n = 0, 2, 4, 6$), and $K_n$ is the diffusion (or damping) coefficient.

Consider the zeroth-order ($n = 0$) diffusion, i.e.,

$$\frac{\partial h}{\partial t} = -K_0 h$$

This defines a diffusion that is applied directly to $h$. It acts uniformly over all wavelengths. Such uniform damping is typically not employed within numerical models except perhaps near the lateral boundaries.

Consider the second-order ($n = 2$) diffusion, i.e.,

$$\frac{\partial h}{\partial t} = K_2 \nabla^2 h$$

This defines a diffusion that acts on the Laplacian of $h$. Recall that the Laplacian of a field has the opposite sign of the field itself. As a result, where $h$ is a maximum, $\nabla^2 h$ is a minimum and, consequently, $h$ decreases with time. Conversely, where $h$ is a minimum, $\nabla^2 h$ is a maximum and, consequently, $h$ increases with time. Where the gradient in a field is fairly uniform, the Laplacian is approximately zero, and the change in $h$ with time is also approximately zero.

This diffusion formulation is weakly scale-selective, wherein shorter wavelengths are modestly dampened more than are longer wavelengths. Higher-ordered diffusion operators are even more scale-selective, a desirable trait because longer wavelengths are not particularly problematic. As the Laplacian’s magnitude is greatest for maxima or minima in the field being diffused, it does not introduce new maxima or minima to the field, a desirable trait. The same cannot be said for sixth-order diffusion, and so-called flux limiter methods exist to mitigate this drawback.
This can be illustrated by considering the linear stability of this diffusion operator. Let us do so for the forward-in-time, second-order centered-in-space differencing scheme, i.e.,

\[
\frac{h^{t+1}_x - h^t_x}{\Delta t} = K_2 \left( \frac{h^{t+1}_{x+1} + h^{t+1}_{x-1} - 2h^{t+1}_x}{(\Delta x)^2} \right)
\]

Recall that when applied to advection terms, this finite differencing scheme is unconditionally unstable. As we will demonstrate, however, the same does not hold true for its application to this diffusion term. Consequently, the discussion of the properties of this differencing scheme for diffusion operators is primarily illustrative in nature.

(Conversely, as shown in the course text, the centered-in-time, second-order centered-in-space differencing scheme is unconditionally unstable when applied to a second-order diffusion term despite it being conditionally stable for linear advection.)

As before, assume a wave-like solution for \( h \) of the form:

\[
h = \hat{h} e^{i(kx - \omega t)} = \hat{h} e^{\omega_R t} e^{i(kx - \omega_I t)}
\]

where \( \omega = \omega_R + i\omega_I \). If we substitute this solution for \( h \) into our finite difference discretization of the second-order diffusion operator, expand the resulting exponential functions, and divide through by a common factor of \( \hat{h} e^{\omega_I t} e^{i(kx - \omega_I t)} \), we obtain:

\[
e^{\omega_I \Delta t} e^{-i\omega_I \Delta t} - 1 = \frac{K\Delta t}{(\Delta x)^2} \left( e^{i\Delta x} + e^{-i\Delta x} - 2 \right)
\]

On the left-hand side of this equation, \( e^{-i\omega_I \Delta t} \) can be rewritten using Euler’s relations. For the right-hand side of this equation, Euler’s relations allow us to show that \( e^{i\Delta x} + e^{-i\Delta x} = 2\cos(k\Delta x) \).

Substituting into the above, we obtain:

\[
e^{\omega_I \Delta t} \left( \cos(\omega_R \Delta t) - i \sin(\omega_R \Delta t) \right) - 1 = \frac{K\Delta t}{(\Delta x)^2} \left( 2\cos(k\Delta x) - 2 \right)
\]

Splitting this into its real and imaginary components, we obtain:

\[
e^{\omega_I \Delta t} \cos(\omega_R \Delta t) - 1 = \frac{K\Delta t}{(\Delta x)^2} \left( 2\cos(k\Delta x) - 2 \right) \quad \text{(real)}
\]

\[
-i \sin(\omega_R \Delta t)e^{\omega_I \Delta t} = 0 \quad \text{(imaginary)}
\]

Recall that to evaluate the linear stability of this equation, we eliminate \( \omega_R \) from this system of equations, leaving only \( \omega_I \) or, more specifically, \( e^{\omega_I \Delta t} \).
Because the exponential function in the imaginary equation cannot be equal to zero, \( \sin(\omega_R \Delta t) \) must equal zero in order for the equality in that equation to hold. The only values of \( \omega_R \) that result in \( \sin(\omega_R \Delta t) = 0 \) are 0 (such that \( \omega_R \Delta t = 0 \)) and \( \Delta t/\pi \) (such that \( \omega_R \Delta t = \pi \)). It can be shown that \( \omega_R = \Delta t/\pi \) is just a special form of the \( \omega_R = 0 \) case. Thus, we focus upon the \( \omega_R = 0 \) case.

For \( \omega_R = 0 \), \( \cos(\omega_R \Delta t) = 1 \) and the real component of the equation becomes:

\[
e^{\omega_R \Delta t} = 1 + 2 \frac{K \Delta t}{(\Delta x)^2} \left( \cos(k \Delta x) - 1 \right)
\]

This defines the multiplicative change in amplitude in \( h \) that occurs with each time step during the model integration for the forward-in-time, second-order centered-in-space differencing scheme applied to a second-order diffusion operator.

From this result, the stability criteria for this equation may be obtained, as is done in the “Linear Numerical Stability” lecture notes. The \( 2\Delta x \) wave is that which limits the linear stability of this diffusion operator. Two stability criteria, one ensuring that \( e^{\omega_R \Delta t} \geq -1 \) (numerically stable) and one ensuring that \( e^{\omega_R \Delta t} \geq 0 \) (numerically stable with no change in wave phase), exist. These are obtained by setting \( e^{\omega_R \Delta t} = -1 \) and \( e^{\omega_R \Delta t} = 0 \), respectively, for the \( 2\Delta x \) wave.

Because of how it is obtained, the criterion ensuring numerical stability with no change in wave phase can be used to define the value of \( K \) that dampens the \( 2\Delta x \) wave entirely at each time step. This criterion is given by:

\[
\frac{K \Delta t}{(\Delta x)^2} \leq \frac{1}{4}
\]

If the inequality is replaced with an equal sign and the equation solved for \( K \), we obtain:

\[
K = \frac{(\Delta x)^2}{4 \Delta t}
\]

If we use this as our value for \( K \), the stability equation becomes:

\[
e^{\omega_R \Delta t} = 1 + \frac{1}{2} \left( \cos(k \Delta x) - 1 \right)
\]

If we plug in to this equation with values for wavelength \( L \), we can determine the wavelength-dependence, or scale-selectivity, of the damping function. This is depicted by the blue line in Figure 4.
The same process as followed above can be used to determine the multiplicative change in \( h \) that occurs with each time step during the model integration for the forward-in-time, second-order centered-in-space differencing scheme applied to fourth- and sixth-order diffusion operators. Second-order centered-in-space finite difference approximations for the fourth and sixth partial derivatives, respectively, are given by the following:

\[
\frac{\partial^4 h}{\partial x^4} = \frac{(h_{x+2} + h_{x-2}) - 4(h_{x+1} + h_{x-1}) + 6h_x}{(\Delta x)^4}
\]

\[
\frac{\partial^6 h}{\partial x^6} = \frac{(h_{x+3} + h_{x-3}) - 6(h_{x+2} + h_{x-2}) + 15(h_{x+1} + h_{x-1}) - 20h_x}{(\Delta x)^6}
\]

The resulting equations for \( e^{\omega \Delta t} \) for the fourth- and sixth-order diffusion operators are given by:

\[
e^{\omega \Delta t} = 1 + 2 \frac{K\Delta t}{(\Delta x)^4} (- \cos(2k\Delta x) + 4 \cos(k\Delta x) - 3)
\]

\[
e^{\omega \Delta t} = 1 + 2 \frac{K\Delta t}{(\Delta x)^6} (\cos(3k\Delta x) - 6 \cos(2k\Delta x) + 15 \cos(k\Delta x) - 10)
\]

Using these equations, values of \( K \) that results in \( e^{\omega \Delta t} = 0 \) for the \( L = 2\Delta x \) wave may be obtained. These are given by:

\[
K = \frac{(\Delta x)^4}{16\Delta t}
\]

\[
K = \frac{(\Delta x)^6}{64\Delta t}
\]

If we use these as our values for \( K \), the stability equations become:

\[
e^{\omega \Delta t} = 1 + \frac{1}{8} (- \cos(2k\Delta x) + 4 \cos(k\Delta x) - 3)
\]

\[
e^{\omega \Delta t} = 1 + \frac{1}{32} (\cos(3k\Delta x) - 6 \cos(2k\Delta x) + 15 \cos(k\Delta x) - 10)
\]

If we plug in to these equations with values for wavelength \( L \), we can determine the wavelength-dependence, or scale-selectivity, of the fourth- and sixth-order damping functions. These are depicted by the red and green lines, respectively, in Figure 4.
Figure 4. Depiction of damping magnitude per time step, $e^{\omega \Delta t}$, as a function of wavelength for second-, fourth-, and sixth-order diffusion operators using the forward-in-time, second-order centered-in-space finite differencing scheme. Note that for each diffusion operator, the value of $K$ is chosen such that the $2\Delta x$ wave is entirely dampened at each time step. Adapted from Warner (2011), their Figure 3.34.

Figure 5. In each panel, the influence of diffusion – second-order in panel (a), sixth-order with a flux limited in panel (b) – upon an initial square wave over 100 model time steps is depicted. Note that the value of $K$ for each diffusion operator is chosen such that the $2\Delta x$ wave is entirely dampened at each time step. As the second-order diffusion operator is less scale-selective than is the sixth-order diffusion operator, its effects upon the square wave extend across the wave rather than being localized to its sharp edges. Reproduced from Warner (2011), their Figure 3.35.
For all damping functions, the magnitude of damping decreases as the wavelength increases. As the order of the diffusion operator increases, the extent to which longer wavelengths are dampened decreases. As a result, higher-order diffusion operators are preferable, so long as their weaknesses (e.g., creating new maxima or minima) can be mitigated by some means.

The scale-selectivity of the second-order and sixth-order diffusion operators is again demonstrated in Figure 5. A one-dimensional (in $x$) model, and thus diffusion formulation, is again utilized. There is no advection within this model; only diffusion acts upon the initial wave. In this example, the initial wave is given by a square wave, which generally is not present within atmospheric fields but possesses sharp horizontal gradients that help to illustrate the scale-selectivity of the applied diffusion operators. The initial square wave extends over twenty-five grid points. The value of the diffusion coefficient $K$ is again chosen so that the $2\Delta x$ wave is entirely dampened at each time step, and the model is integrated forward for 100 time steps.

There are two primary wavelengths manifest through the initial square wave: that of the wave itself ($25\Delta x$ given no corresponding negative portion of the wave) and that of the discontinuities along the edge of the wave ($\sim 2\Delta x$). The less-scale-selective second-order diffusion operator dampens each of these wavelengths, resulting in both a weakening of the gradient across the edge of the square wave and a reduction in the amplitude of the wave itself. The more-scale-selective sixth-order diffusion operator – which includes a flux-limiter correction term as noted above – also weakens the gradient across the edge of the square wave, albeit to less extent than did the second-order diffusion operator. However, the change in amplitude of the wave is negligible with this formulation, a desirable trait.

We can also consider the scale-selectivity of the second-, fourth-, and sixth-order diffusion operators in light of the one-dimensional advection example considered during our lecture on numerical dispersion. Here, an initial Gaussian wave is advected at a constant $U = 10$ m s$^{-1}$ over a domain containing 100 grid points ($\Delta x = 1$ km) until it returns to its original location. The time step is 10 s, such that the Courant number is 0.1. The centered-in-time, second-order centered-in-space finite differencing scheme is used to discretize the advection terms within this example.

The influence of second-, fourth-, and sixth-order diffusion upon this solution is demonstrated in Figure 6. As the order of the diffusion operator increases, its impact upon longer wavelength phenomena decreases; as the order of the diffusion operator decreases, its scale-selectivity also decreases. Though numerical dispersion is less evident in the second-order diffusion example, the amplitude of the primary wave is substantially reduced relative to both the physical wave and to that in the cases with higher-ordered diffusion operators. Thus, a trade-off exists, which motivates the use of more accurate finite differencing schemes that are less affected by numerical dispersion (among other attributes)!
Figure 6. Fluid height $h$ (m) after integrating the one-dimensional advection equation for 10,000 s on the model grid described in the text above, with a Courant number of 0.1, for integrations utilizing no explicit numerical diffusion (top), a sixth-order diffusion operator (middle-top), a fourth-order diffusion operator (middle-bottom), and a second-order diffusion operator (bottom). Reproduced from Warner (2011), their Figure 3.26.

Figure 2 above demonstrates the utility of numerical diffusion – primarily explicit, but in some applications also implicit – in mitigating the impacts of aliasing upon the model solution. Where aliasing results in an artificial buildup of wave energy at short, poorly-resolved wavelengths, numerical diffusion dampens this accumulation. Though the kinetic energy spectrum deviates from that given by theory and observations at wavelengths below the effective model resolution when damping is applied, the model is less likely to become numerically unstable. As this is desirable, we choose the horizontal grid spacing of our model simulations in light of the resulting effective resolution. For example, to resolve thunderstorms, we require $\Delta x < 4$ km with an effective resolution of $< 28$ km that is barely crude enough to encompass the area covered by larger thunderstorms.

Formally, diffusion operators should be evaluated on horizontal surfaces (e.g., constant height surfaces) rather than on model surfaces, which for most modern models follow the terrain. Let us consider the example of a mountain, where the temperature at mountain top is often larger than that at the bottom of the mountain. Along a terrain-following surface, temperature would be a minimum at the top of the mountain and a maximum at the bottom of the mountain. Diffusion acting on the terrain-following surface would decrease the temperature at the bottom of the mountain and increase the temperature at the top of the mountain. This will locally increase the
thickness of the column at mountain top, resulting in relatively low pressure at mountain top, into which air will converge due to friction. Diffusion calculated on horizontal surfaces would not lead to the development of this non-physical circulation.

In WRF-ARW, diffusion may be computed on either model coordinate or horizontal surfaces, as controlled by the diff_opt option. Where terrain gradients are small, diffusion along coordinate surfaces may not be overly problematic. However, where terrain gradients are large, such as near mountains, diffusion should be computed along horizontal surfaces. The WRF-ARW default is second-order diffusion along model coordinate surfaces. However, a sixth-order diffusion operator is also available. For model simulations where $\Delta x \geq O(100 \text{ m})$, the diffusion coefficient $K_n$ in the horizontal is determined from horizontal deformation; recall that vertical diffusion is handled, and thus $K_n$ in the vertical is specified, by the boundary layer parameterization. More detail regarding explicit numerical diffusion within WRF-ARW may be found in Section 4.2 of the WRF-ARW Technical Document.