Chapter 8: Making Waves

Copyright 2011, David A. Randall

8.1 The shallow-water equations

In most of this chapter we will discuss the shallow-water equations, which can be written as

$$\frac{\partial \mathbf{v}}{\partial t} + (\zeta + f)\mathbf{k} \times \mathbf{v} = -\nabla\left[g\left(h + h_S\right) + K\right], \tag{1}$$

$$\frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{v}h) = 0 . \tag{2}$$

Here $\mathbf{v}$ is the horizontal velocity vector, $\zeta \equiv \mathbf{k} \cdot (\nabla \times \mathbf{v})$ is the vertical component of the vorticity, f is the Coriolis parameter, h is the depth of the fluid, h_S is the height of the "bottom topography," g is the acceleration of gravity, and $K \equiv \frac{1}{2}\mathbf{v} \cdot \mathbf{v}$ is the kinetic energy per unit mass. In (1), all frictional effects have been neglected, for simplicity. Although the shallow water equations are highly idealized, they are extremely useful for testing numerical models that are used to simulate atmospheric dynamics.

For the special case of a one-dimensional, non-rotating small-amplitude gravity wave on a resting basic state, without topography, Eqs. (1) and (2) become

$$\frac{\partial u}{\partial t} + g\frac{\partial h}{\partial x} = 0 , \tag{3}$$

and

$$\frac{\partial h}{\partial t} + H\frac{\partial u}{\partial x} = 0 , \tag{4}$$

respectively. Here H is the mean depth of the fluid. We refer to (3) - (4) as "the gravity wave equations." Let

$$c^2 \equiv gH \,. \tag{5}$$

By combining (3) - (4) we can derive

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}\,, \tag{6}$$

and

$$\frac{\partial^2 h}{\partial t^2} = c^2 \frac{\partial^2 h}{\partial x^2}\,, \tag{7}$$

which are both examples of "the wave equation."

Assuming solutions of the form, $e^{i(kx-\omega t)}$ we obtain the dispersion equation

$$\omega^2 = c^2 k^2 \,. \tag{8}$$

The exact phase speed of pure gravity waves (without the effects of rotation) is $\pm\sqrt{gH}$, regardless of wave length. There are two waves, one propagating in the positive *x*-direction, and the other in the negative *x*-direction.

8.2 The wave equation

The solutions of the wave equation, (6), are constant along space-time lines (or surfaces) called "characteristics." A solution is fully determined if u and $\frac{\partial u}{\partial t}$ are specified somewhere on each characteristic. The characteristics can, and generally do, intersect boundaries. As with the advection equation, $f(x-ct)$ is a particular solution of the wave equation (6), but $g(x+ct)$ is a second particular solution. We can assume $c > 0$ without loss of generality. The general solution of (6) is given by

$$u(x,t) = f(x-ct) + g(x+ct)\,, \tag{9}$$

where, as shown below, the forms of *f* and *g* are determined completely by the initial conditions, which are

$$\begin{aligned} u_{t=0} &= F(x), \\ \left(\frac{\partial u}{\partial t}\right)_{t=0} &= G(x). \end{aligned} \tag{10}$$

Note that $\left(\frac{\partial u}{\partial t}\right)_{t=0} = G(x)$ contains information about $h(x,0)$, so together these two initial conditions contain information about both the mass field and the wind field at $t = 0$.

Substituting (9) into (10), we find that

$$\begin{aligned} f(x) + g(x) &= F(x), \\ -cf'(x) + cg'(x) &= G(x). \end{aligned} \tag{11}$$

Here a prime denotes differentiation. Differentiating the first of (11), and then using the second, we can solve for $f'(x)$ and $g'(x)$:

$$\begin{aligned} f'(x) &= \frac{1}{2}\left[F'(x) - \frac{G(x)}{c}\right], \\ g'(x) &= \frac{1}{2}\left[F'(x) + \frac{G(x)}{c}\right]. \end{aligned} \tag{12}$$

These can be integrated to obtain $f(x)$ and $g(x)$:

$$\begin{aligned} f(x) &= \frac{1}{2}\left[F(x) - \frac{1}{c}\int_0^x G(\xi)\,d\xi\right] + C_1, \\ g(x) &= \frac{1}{2}\left[F(x) + \frac{1}{c}\int_0^x G(\xi)\,d\xi\right] + C_2. \end{aligned} \tag{13}$$

Here C_1 and C_2 are constants of integration. Finally, we obtain $u(x,t)$ by replacing x by $x - ct$ and $x + ct$, respectively, in $f(x)$ and $g(x)$ of (13), and then substituting back into (9). This gives

$$u(x,t) = \frac{1}{2}\left[F(x-ct) + F(x+ct) + \frac{1}{c}\int_{x-ct}^{x+ct} G(\xi)\,d\xi\right]. \tag{14}$$

Here ξ is a dummy variable of integration, and we have set $C_1 + C_2 = 0$ in order to satisfy $u(x,0) = F(x)$. Note that, as mentioned above, $G(x)$ contains information about $h(x,0)$.

Obviously, that information is needed to predict $u(x,t)$, and it is in fact used on the right-hand side of (14).

In order to relate the wave equation to the advection equation that we have already analyzed, we reduce (6) to a pair of first-order equations by defining

$$p \equiv \frac{\partial u}{\partial t} \text{ and } q \equiv -c\frac{\partial u}{\partial x}. \tag{15}$$

Substitution of (15) into the wave equation (6) gives

$$\frac{\partial p}{\partial t} + c\frac{\partial q}{\partial x} = 0, \tag{16}$$

and differentiation of the second of (15) with respect to t, with the use of the first of (15), gives

$$\frac{\partial q}{\partial t} + c\frac{\partial p}{\partial x} = 0. \tag{17}$$

If we alternately add (16) and (17), and subtract (17) from (16), we obtain

$$\frac{\partial P}{\partial t} + c\frac{\partial P}{\partial x} = 0 \text{, where } P \equiv p + q \text{, and} \tag{18}$$

$$\frac{\partial Q}{\partial t} - c\frac{\partial Q}{\partial x} = 0 \text{, where } Q \equiv p - q, \tag{19}$$

respectively. Now we have a system of two first-order equations, each in the form of the advection equation. Note, however, that the "advections" are in opposite directions! Assuming that $c > 0$, P is "advected" towards increasing x, while Q is "advected" towards decreasing x. From (18) and (19), it is clear that P is constant along the line $x - ct =$ constant, and Q is constant along the line $x + ct =$ constant. Eqs. (18) and (19) are called the *normal forms* of (16) and (17).

These concepts are applicable, with minor adjustments, to any hyperbolic system of equations. The curves $x - ct =$ constant and $x + ct =$ constant are called "characteristics." A hyperbolic equation is characterized, so to speak, by two such families of curves. In the present case they are straight, parallel lines, but in general they can have any shape so long as they do not intersect each other.

8.3 Staggered grids for the shallow water equations

Now we discuss the differential-difference equations

$$\frac{du_j}{dt} + g\left(\frac{h_{j+1} - h_{j-1}}{2\Delta x}\right) = 0\,, \tag{20}$$

$$\frac{du_j}{dt} + H\left(\frac{u_{j+1} - u_{j-1}}{2\Delta x}\right) = 0\,, \tag{21}$$

which are, of course, differential-difference analogs of the one-dimensional shallow water equations, (3) - (4). Consider a distribution of the dependent variables on the grid as shown in Fig. 8.1. Notice that from (20) and (21) *the set of red quantities will act completely independently of the set of black quantities*, if there are no boundaries. With cyclic boundary conditions, this is still true if the number of grid points in the cyclic domain is even. What this means is that we have two families of waves on the grid: "red" waves that propagate both left and right, and "black" waves that propagate both left and right. Physically there should only be one family of

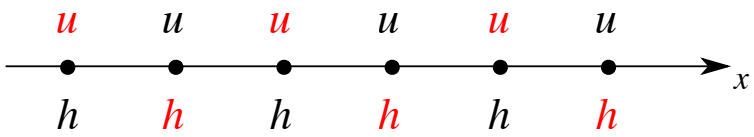

Fig. 8.1: A grid for solution of the one-dimensional shallow water equations.

waves.

Here is a mathematical way to draw the same conclusion. The wave solutions of (20) and (21) are

$$\left(u_j, h_j\right) \sim e^{i(kj\Delta x - \omega t)}\,, \tag{22}$$

giving

$$\begin{aligned} \omega u_j - g h_j \frac{\sin\left(k\Delta x\right)}{\Delta x} &= 0, \\ \omega h_j - H u_j \frac{\sin\left(k\Delta x\right)}{\Delta x} &= 0\,. \end{aligned} \tag{23}$$

Provided that u_j and h_j are not both identically zero, we obtain the dispersion relation

$$\omega^2 = gH\left(\frac{\sin k\Delta x}{\Delta x}\right)^2. \tag{24}$$

The phase speed satisfies $c^2 = gH\left(\frac{\sin k\Delta x}{k\Delta x}\right)^2$. In the exact solution, the phase speed is independent of wave number, and is given by Eq. (5). The finite-difference phase speed depends on wave number. This is computational dispersion again.

For convenience, define $p \equiv k\Delta x$. Suppose that ω is given. If $p = p_0$ satisfies (24), then $p = -p_0$, $p = \pi - p_0$ and $p = -(\pi - p_0)$ also satisfy it. This shows that there are four possible modes for the given frequency, although physically there should only be two. The "extra" pair of modes comes from the redundancy on the grid. *The extra modes are computational modes "in space."* Earlier we encountered computational modes in time.

Without loss of generality, we assume that $0 < p_0 < \frac{\pi}{2}$, so that $\sin(p_0) > 0$. Then the two solutions $p = p_0$ and $p = -p_0$ are approximations to the true solution, and therefore could be considered as physical, while the other two, $p = \pi - p_0$ and $p = -(\pi - p_0)$, could be considered as computational. This distinction is less significant than in the case of the advection equation, however. In the case of advection, the envelope of a computational mode moves toward the downstream direction. In the case of the wave equation, there is no "downstream" direction.

For a given ω, the general solution for u_j is a linear combination of the four modes, and can be written as

$$u_j = \left[Ae^{ip_0 j} + Be^{-ip_0 j} + Ce^{i(\pi - p_0)j} + De^{-i(\pi - p_0)j} \right] e^{-i\omega t} . \tag{25}$$

By substituting (25) into (21), we find that h_j satisfies

$$h_j = \frac{H \sin p_0}{\omega \Delta x} \left[Ae^{ip_0 j} - Be^{-ip_0 j} + Ce^{i(\pi - p_0)j} - De^{-i(\pi - p_0)j} \right] e^{-i\omega t} . \tag{26}$$

If we assume $\omega > 0$, so that $\sin(p_0) = \frac{\omega \Delta x}{\sqrt{gH}}$ [see (24)], then (26) reduces to

$$h_j = \sqrt{\frac{H}{g}} \left[Ae^{ip_0 j} - Be^{-ip_0 j} + Ce^{i(\pi - p_0)j} - De^{-i(\pi - p_0)j} \right] e^{-i\omega t} . \tag{27}$$

8.4 Numerical simulation of geostrophic adjustment as a guide to grid design

Winninghoff (1968) and Arakawa and Lamb (1977; hereafter AL) discussed the extent to which finite-difference approximations to the shallow water equations can simulate the process

of geostrophic adjustment, in which the dispersion of inertia-gravity waves leads to the establishment of a geostrophic balance, as the energy density of the inertia gravity waves decreases with time due to their dispersive phase speeds and non-zero group velocity. These authors considered the momentum and mass conservation equations, and defined five different staggered grids for the velocity components and mass.

AL considered the shallow water equations linearized about a resting basic state, in the following form:

$$\frac{\partial u}{\partial t} - fv + g\frac{\partial h}{\partial x} = 0 \,, \tag{28}$$

$$\frac{\partial v}{\partial t} + fu + g\frac{\partial h}{\partial y} = 0 \,, \tag{29}$$

$$\frac{\partial h}{\partial t} + H\delta = 0 \,. \tag{30}$$

Here H is the constant depth of the "water" in the basic state, $\delta \equiv \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$ is the divergence, and all other symbols have their conventional meanings. From (28) -(30), we can derive an equivalent set in terms of vorticity, $\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$, and divergence:

$$\frac{\partial \delta}{\partial t} - f\zeta + g\left(\frac{\partial^2}{\partial x^2}h + \frac{\partial^2}{\partial y^2}h\right) = 0 \,, \tag{31}$$

$$\frac{\partial \zeta}{\partial t} + f\delta = 0 \,, \tag{32}$$

$$\frac{\partial h}{\partial t} + H\delta = 0 \,. \tag{33}$$

Of course, (33) is identical to (30). We can eliminate the vorticity and mass in (31) by using (32) and (33), respectively. Then by assuming wave solutions, we obtain the dispersion relation:

$$\left(\frac{\sigma}{f}\right)^2 = 1 + \lambda^2\left(k^2 + l^2\right). \tag{34}$$

Here σ is the frequency, $\lambda \equiv \frac{\sqrt{gH}}{f}$ is the radius of deformation, and k and l are the wave numbers in the x and y directions, respectively. The frequency and group speed increase monotonically with wave number and are non-zero for all wave numbers. As discussed by AL, these characteristics of (34) are important for the geostrophic adjustment process.

In their discussion of various numerical representations of (28) - (30), AL defined five

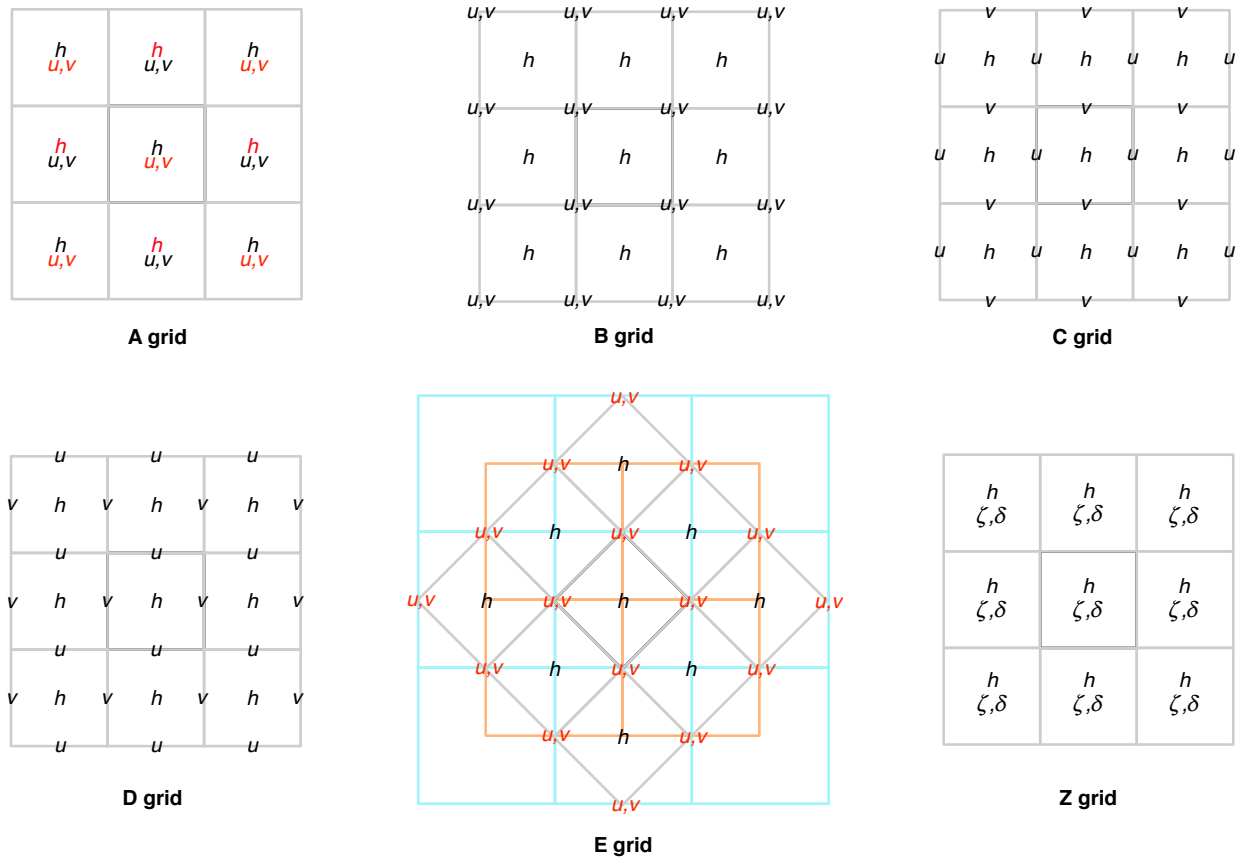

Fig. 8.2: Grids A-E and Z, on a square mesh. The E grid can be obtained by rotating the B grid by 45°. In the sketch of the E grid, the mass variables can be considered to live in the rotated grey boxes, or they can be considered to live in the overlapping blue and orange boxes.

grids denoted by "A" through "E," as shown in Fig. 8.2. The figure also shows the Z grid, which will be discussed later. AL also gave the simplest centered finite-difference approximations to of (28) - (30), for each of the five grids; these equations are fairly obvious and will not be repeated here. The two-dimensional dispersion equations for the various schemes were derived but not published by AL; they are included in Fig. 8.3, which also gives a plot of the nondimensional frequency, σ / f , as a function of kd and ld , for the special case $\lambda / d = 2$. Here d is the grid size, assumed to be the same in the x and y directions. The significance of this particular choice of λ / d is discussed later. The plots show how the nondimensional frequency varies out to

$kd = \pi$ and $ld = \pi$; these wave numbers correspond to the shortest waves that can be represented on the grid.

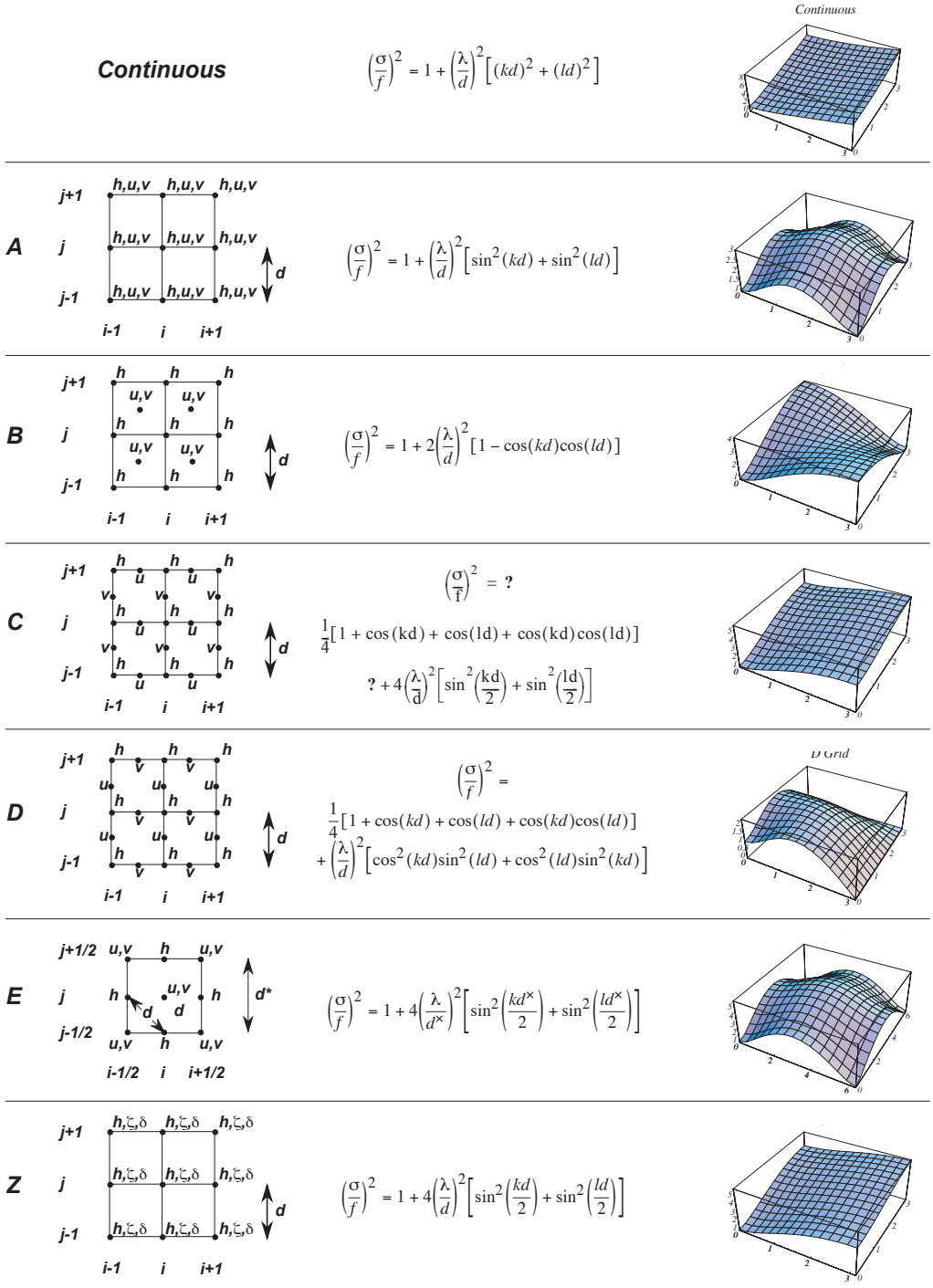

Fig. 8.3: Grids, dispersion equations, and plots of dispersion equations for grids A - E and Z. The continuous dispersion equation and its plot are also shown for comparison. For plotting, it has been assumed that $\lambda / d = 2$.

The A grid may appear to be the simplest, since it is unstaggered. For example, the Coriolis terms of the momentum equations are easily evaluated, since u and v are defined at the same points. There is a serious problem, however. In the sketch of the A grid in Fig. 8.2, some of the variables are colored red, and others are colored black. The red winds are used to predict the red masses, and vice versa. Similarly, the black winds are used to predict the black masses. But, as least as far as the linearized equations are concerned, the red and black variables to not communicate. This means the the A grid can support two solutions at once -- a red one, and a black one -- and those solutions can differ considerably. In such a case, the pattern of the variables on the A grid is characterized by strong noise at the smallest scales, i.e., a checkerboard pattern. Such dynamically "invisible" noise cannot participate in the dynamics of the model, e.g., by propagating and dispersing as in the process of geostrophic adjustment. As a result, the high-wavenumber behavior of a model based on the A grid is poor. A plot of the dispersion equation for the A grid, shown in Fig. 8.3, indicates a maximum of the frequency (group speed equal to zero) for some combinations of k and l. As a result, solutions on the A grid are extremely noisy in practice and must be smoothed, e.g., through artificial diffusion or filtering (Kalnay-Rivas et al., 1977). Because of this well known problem, the A grid is rarely used today. The problem of the A grid is obviously closely analogous to that of the unstaggered one-dimensional grid discussed above.

Next, consider the B grid. Fig. 8.2 shows that the velocity vectors are defined at the corners of the mass cells. The velocity components, i.e., u and v, point along the directions of the walls that intersect at the corners. As on the A grid, the coriolis terms are easily evaluated, without averaging, since u and v are defined at the same points. On the other hand, the pressure-gradient terms must be averaged, again as on the A grid. There is an important difference, however. On the A grid, the averaging used to approximate the x-component of the pressure-gradient force, $\partial h / \partial x$, is averaging *in the x-direction*. On the B grid, the corresponding averages are in the *y-direction*. On the B grid, an oscillation in the x-direction, on the smallest represented scale, is not averaged out in the computation of $\partial h / \partial x$; it can, therefore, participate in the model's dynamics, and so is subject to geostrophic adjustment. A similar conclusion holds for the convergence / divergence terms of the continuity equation. For example, the averaging in the y–direction does no harm for solutions that are uniform in the y-direction. Nevertheless, it does do some harm, as is apparent in the plot of the B-grid dispersion equation, as shown in Fig. 8.3. The frequency does not increase monotonically with total wave number; for certain combinations of k and l, the group speed is zero. AL concluded that the B grid gives a fairly good simulation of geostrophic adjustment, but with some tendency to small-scale noise.

Now consider the C grid. The pressure gradient terms are easily evaluated, without averaging, because h is defined east and west of u points, and north and south of v points. Similarly, the mass convergence / divergence terms of the continuity equation can be evaluated without averaging the winds. On the other hand, averaging *is* needed to obtain the coriolis terms, since u and v are defined at different points. For very small-scale inertia-gravity waves, the coriolis terms are negligible; we essentially have pure gravity waves. This suggests that the C

grid will perform well if the horizontal resolution of the model is high enough so that the smallest waves that can be represented on the grid are insensitive to the Coriolis force. More precisely, AL argued that the C grid does well when the grid size is small compared to λ, the radius of deformation. A plot of the dispersion equation, given in Fig. 8.3, shows that the frequency increases monotonically with wave number, as in the exact solution, although not as rapidly. Recall, however, that this plot is for the special case $\lambda / d = 2$. We return to this point later.

Next, we turn to the D grid. Inspection of the stencil shown in Fig. 8.2 reveals that the D grid allows a simple evaluation of the geostrophic wind. In view of the importance of geostrophic balance for large-scale motions, this may appear to be an attractive property. It is also apparent, however, that considerable averaging is needed in the pressure-gradient force, mass convergence / divergence, and even in the coriolis terms. As a result, the dispersion equation for the D grid, shown in Fig. 8.3, is very badly behaved, giving zero phase speed for the shortest represented waves, and also giving a zero group speed for some modes.

Finally, consider the the E grid. As shown in Fig. 8.2, the E grid can be viewed as a modified B grid, rotated by 45˚. Because of this rotation, the grid spacing for the E grid is $d^* \equiv \sqrt{2}d$, for the same "density" of h points as in the other four grids. The mass can be considered to live inside the rotated grey cells. The velocity components, i.e., u and v, point diagonally across the walls of the grey cells surrounding the mass points; this is different from the B grid, on which, as mentioned above, the velocity components point along the directions of the walls that intersect at the corners of the mass cells.

Alternatively, the E grid can be considered to live within the *overlapping* but unrotated orange and blue boxes. From this point of view, the E grid is the superposition of two C grids, shifted with respect to each other, so that the v points on one of the C grids coincide with the u points on the other, and vice versa.

The E grid at first seems perfect; no averaging is needed for the coriolis terms, the pressure-gradient terms, or the mass convergence / divergence terms. Nevertheless there is a problem. Consider a solution that is uniform in one of the grid directions, say the *y*-direction. In that case, we effectively have a one-dimensional problem. In one dimension, the E grid "collapses" to the A grid, with a reduced grid spacing $d = d^* / \sqrt{2}$. For such one-dimensional motions, the E grid has all the problems of the A grid. These problems are apparent in the plot of the dispersion equation, given in Fig. 8.3. (For the E grid, the nondimensional frequency is plotted as a function of kd^* and ld^*, out to a value of 2π; this corresponds to the shortest "one-dimensional" mode.) The group speed is zero for some combinations of k and l.

The A-grid can be viewed as a super-position of two E-grids, in which one of the E-grids is shifted by one-half of the grid spacing. This can be seen in Fig. 8.2. The super-position of two E-grids is similar to the super-imposed red and black grids in Fig. 8.1.

Now recall the conclusion of AL, described earlier, that the C grid gives a good simulation of geostrophic adjustment provided that $\lambda / d > 1$. Large-scale modelers are never happy to choose d and λ so that λ / d can be less than one. Nevertheless, in practice modes for which $\lambda / d << 1$ can be unavoidable, at least for some situations. For example, Hansen et al. (1983) described a low-resolution atmospheric GCM, which they called Model II, designed for

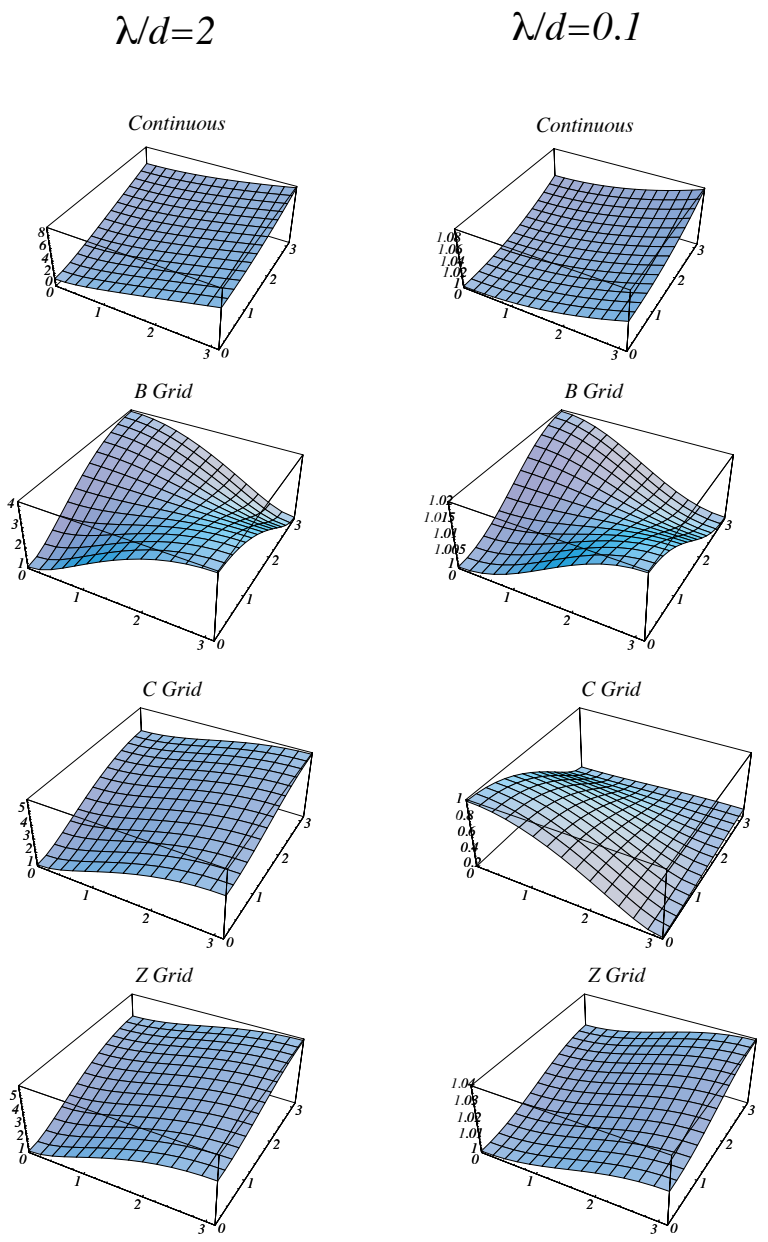

Fig. 8.4: Dispersion relations for the continuous shallow water equations, and for finite-difference approximations based on the B, C, and Z grids. The horizontal coordinates in the plots are kd and ld, respectively, except for the E grid, for which kd^* and ld^* are used. The vertical coordinate is the normalized frequency, σ / f. For the E grid, the results are meaningful only in the triangular region for which $kd^* + ld^* \le 2\pi$. The left column shows results for $\lambda / d = 2$, and the right column for $\lambda / d = 0.1$.

very long climate simulations in which low resolution was a necessity. Model II used a grid size of 10 degrees of longitude by 8 degrees of latitude; this means that the grid size was larger than the radius of deformation for many of the physically important modes that could be represented

on the grid. As shown by AL, such modes cannot be well simulated using the C grid. Having experienced these problems with the C grid, Hansen et al. (1983) chose the B grid for Model II.

Ocean models must contend with small radii of deformation, so that very fine grids are needed to ensure that $\lambda / d > 1$, even for external modes. For this reason, ocean models tend to use the B grid (e.g., Semtner and Chervin, 1992).

In addition, three-dimensional models of the atmosphere and ocean generate internal modes. With vertical structures typical of current general circulation models, the highest internal modes can have radii of deformation on the order of 50 km or less. The same model may have a horizontal grid spacing on the order of 500 km, so that λ / d can be on the order of 0.1. Fig. 8.4 demonstrates that the C grid behaves very badly for $\lambda / d = 0.1$. The phase speed actually decreases monotonically as the wave number increases, and becomes very small for the shortest waves that can be represented on the grid. Janjic and Mesinger (1989) have emphasized that, as a result, models that use the C grid have difficulty in representing the geostrophic adjustment of high internal modes. In contrast, the dispersion relation for the B grid is qualitatively insensitive to the value of λ / d. The B grid has moderate problems for $\lambda / d = 2$, but these problems do not become significantly worse for $\lambda / d = 0.1$.

In summary, the C grid does well with deep, external modes, but has serious problems with high internal modes, whereas the B grid has moderate problems with all modes. *The C grid's problem with high internal modes can be avoided by using a sufficiently fine horizontal grid spacing for a given vertical grid spacing.*

Now consider an unstaggered grid for the integration of (31) - (33), which was called the Z grid by Randall (1994). This grid is also illustrated in Fig. 8.2. Inspection shows that with the Z grid the components of the divergent part of the wind "want" to be staggered as in the C grid, while the components of the rotational part of the wind "want" to be staggered as in the D grid. This means that the Z grid does not correspond to any of the grids A through E.

No averaging is required with the Z grid. The only spatial differential operator appearing in (31) - (33) is the Laplacian, $\nabla^2(\ )$, which is applied to h in the divergence equation. With the usual centered finite-difference stencils, the finite-difference approximation to $\nabla^2 h$ is defined at the same point as h itself. An unstaggered grid is thus a natural choice for the numerical integration of (31) – (33).

Fig. 8.4 shows that the dispersion relation for the Z grid is very close to that of the C grid, for $\lambda / d = 2$, but is drastically different for $\lambda / d = 0.1$. Whereas the C grid behaves very badly for $\lambda / d = 0.1$, the dispersion relation obtained with the Z grid is qualitatively insensitive to the value of λ / d; it resembles the dispersion relation for the continuous equations, in that the phase speed increases monotonically with wave number and the group speed is non-zero for all wave numbers. Since the Z grid is unstaggered, collapsing it to one dimension has no effect.

The discussion presented above suggests that geostrophic adjustment in shallow water is well simulated on an unstaggered grid when the vorticity and divergence equations are used. The vorticity and divergence equations are routinely used in global spectral models, but are rarely used in global finite-difference models. The reason seems to be that it is necessary to solve elliptic equations to obtain the winds from the vorticity and divergence, e.g., to evaluate the advection terms of the nonlinear primitive equations. Experience shows that this is not a major practical problem

8.5 Time-differencing schemes for the shallow-water equations

In this section we will consider both space and time differencing for the linearized shallow water equations.

We begin our discussion with the one-dimensional shallow-water equations. The spatial coordinate is *x*, and the single velocity component is *u*. We consider the non-rotating case with $v \equiv 0$. We have divergence (i.e., $\partial u / \partial x$), but no vorticity. Linearizing about a state of rest, the continuous equations are (3) and (4).

We use a staggered one-dimensional (1D) grid, which for this simple problem can be interpreted as the 1D C grid, or the 1D B grid, or the 1D Z grid.

We can anticipate from our earlier analysis of the oscillation equation that forward time-differencing for both the momentum equation and the continuity equation is unstable, and that is actually true. We can also anticipate that a scheme that is centered in both space and time will be conditionally stable and neutral when stable. Such a scheme is given by:

$$\frac{u^{n+1}_{j+\frac{1}{2}} - u^{n-1}_{j+\frac{1}{2}}}{2\Delta t} + g\left(\frac{h^n_{j+1} - h^n_j}{\Delta x}\right) = 0 \,, \tag{35}$$

$$\frac{h^{n+1}_j - h^{n-1}_j}{2\Delta t} + H\left(\frac{u^n_{j+\frac{1}{2}} - u^n_{j-\frac{1}{2}}}{\Delta x}\right) = 0 \,. \tag{36}$$

Compare with (20) - (21). With assumed solutions of the form $u^n_j = \hat{u}^n \exp(ikj\Delta x)$, $h^n_j = \hat{h}^n \exp(ikj\Delta x)$ and the usual definition of the amplification factor, we find that

$$\left(\lambda^2 - 1\right)\hat{u}^n + \lambda \frac{g\Delta t}{\Delta x} 4i \sin\left(\frac{k\Delta x}{2}\right)\hat{h}^n = 0 \,, \tag{37}$$

$$\lambda \frac{H\Delta t}{\Delta x} 4i \sin\left(\frac{k\Delta x}{2}\right)\hat{u}^n + \left(\lambda^2 - 1\right)\hat{h}^n = 0 \,. \tag{38}$$

Non-trivial solutions occur for

$$\left(\lambda^2 - 1\right)^2 + \lambda^2 \left(\frac{4c_{GW}\Delta t}{\Delta x}\right)^2 \sin^2\left(\frac{k\Delta x}{2}\right) = 0 \,. \tag{39}$$

where $c_{GW} \equiv \sqrt{gH}$. As should be expected with the leapfrog scheme, there are four modes altogether. Two of these are physical and two are computational.

We can solve (39) as a quadratic equation for λ^2. As a first step, rewrite it as

$$\left(\lambda^2\right)^2 + \lambda^2 \left(-2 + b\right) + 1 = 0 \,, \tag{40}$$

where, for convenience, we define

$$b \equiv \left(\frac{4c_{GW}\Delta t}{\Delta x}\right)^2 \sin^2\left(\frac{k\Delta x}{2}\right) \geq 0 \,. \tag{41}$$

Obviously, for $\Delta t \to 0$ with fixed Δx we get $b \to 0$. The solution of (40) is

$$\begin{aligned} \lambda^2 &= \frac{-\left(b-2\right) \pm \sqrt{\left(b-2\right)^2 - 4}}{2} \\ &= \frac{-\left(b-2\right) \pm \sqrt{b\left(b-4\right)}}{2} \end{aligned} \tag{42}$$

Inspection of (42) shows that for $b \to 0$, we get $|\lambda| \to 1$, as expected. For $\lambda = |\lambda| e^{i\theta}$ we see that

$$\left|\lambda^2 \left[\cos\left(2\theta\right) + i \sin\left(2\theta\right)\right]\right| = \frac{-\left(b-2\right) \pm \sqrt{b\left(b-4\right)}}{2} \,. \tag{43}$$

It follows that

$$|\lambda|^2 \cos\left(2\theta\right) = -\left(\frac{b-2}{2}\right), \; |\lambda|^2 \sin\left(2\theta\right) = \sqrt{b\left(4-b\right)}, \text{ for } b \leq 4 \,, \tag{44}$$

from which we obtain

$$\tan(2\theta) = \frac{-2\sqrt{b(4-b)}}{2-b} \text{ for } b \le 4 \text{ ,} \tag{45}$$

and

$$|\lambda|^4 = \left(\frac{b-2}{2}\right)^2 + \frac{b(4-b)}{4} = 1 \text{ for } b \le 4 \text{ .} \tag{46}$$

The scheme is thus neutral for $b \le 4$, as was anticipated from our earlier analysis of the oscillation equation.

Returning to (43), we find that

$$\sin(2\theta) = 0,\ \cos(2\theta) = \pm 1 \quad \text{and} \pm|\lambda|^2 = \frac{-(b-2) \pm \sqrt{b(b-4)}}{2}. \text{ for } b > 4 \text{ .} \tag{47}$$

You should be able to see that for $b > 4$ there are always unstable modes.

We conclude that the scheme is stable and neutral for $b \le 4$. This condition can also be written as $\left(\frac{c_{GW}\Delta t}{\Delta x}\right)\left|\sin(k\Delta x)\right| \le \frac{1}{2}$. The worst case occurs for $\left|\sin(k\Delta x)\right| = 1$, which corresponds to $k\Delta x = \pi$, i.e., the $2\Delta x$-wave. It follows that

$$\frac{c_{GW}\Delta t}{\Delta x} < \frac{1}{2} \text{ is required for stability.} \tag{48}$$

and that the $2\Delta x$-wave will be the first to become unstable.

In atmospheric models, the fastest gravity waves, i.e., the external-gravity or "Lamb" waves, have speeds on the order of 300 m s^{-1} , which is about equal to the speed of sound in the Earth's atmosphere. The stability criterion for the leapfrog scheme as applied to the wave problem, i.e., (48), can therefore be painful. In models that do not permit vertically propagating sound waves (i.e., quasi-static models, or anelastic models, or shallow-water models), the external gravity wave is almost always the primary factor limiting the size of the time step. This is unfortunate, because the external gravity modes are believed to play only a minor role in weather and climate dynamics.

With this in mind, the gravity-wave terms of the governing equations are often approximated using implicit differencing. For the simple case of first-order backward-implicit differencing, we replace (35) - (36) by

$$\frac{u^{n+1}_{j+\frac{1}{2}} - u^{n}_{j+\frac{1}{2}}}{\Delta t} + g\left(\frac{h^{n+1}_{j+1} - h^{n+1}_{j}}{\Delta x}\right) = 0\,,$$

(49)

$$\frac{h^{n+1}_{j} - h^{n}_{j}}{\Delta t} + H\left(\frac{u^{n+1}_{j+\frac{1}{2}} - u^{n+1}_{j+\frac{1}{2}}}{\Delta x}\right) = 0\,.$$

(50)

This leads to

$$(\lambda - 1)\hat{u}^{n} + \lambda\frac{g\Delta t}{\Delta x}2i\sin\left(\frac{k\Delta x}{2}\right)\hat{h}^{n} = 0\,,$$

(51)

$$\lambda\frac{H\Delta t}{\Delta x}2i\sin\left(\frac{k\Delta x}{2}\right)\hat{u}^{n} + (\lambda - 1)\hat{h}^{n} = 0\,.$$

(52)

The condition for non-trivial solutions is

$$(\lambda - 1)^2 + \lambda^2 4\left(\frac{c_{GW}\Delta t}{\Delta x}\right)^2 \sin^2\left(\frac{k\Delta x}{2}\right) = 0\,,$$

(53)

which, using (53), is equivalent to

$$\lambda^2\left(1 + \frac{b}{4}\right) - 2\lambda + 1 = 0\,.$$

(54)

This time there are no computational modes; the two physical modes satisfy

$$\lambda^2 = \frac{2 \pm \sqrt{4 - 4\left(1 + \frac{b}{4}\right)}}{2\left(1 + \frac{b}{4}\right)} = \frac{1 \pm i\sqrt{\frac{b}{4}}}{1 + \frac{b}{4}}\,.$$

(55)

The solutions are always oscillatory, and

$$|\lambda|^2 = \frac{1+\frac{b}{4}}{\left(1+\frac{b}{4}\right)^2} = \frac{4}{4+b} \le 1 ,$$

(56)

i.e., the scheme is unconditionally stable, and in fact it damps all modes.

The trapezoidal implicit scheme gives superior results; it is more accurate, and unconditionally neutral. We replace (49) - (50) by

$$\frac{u^{n+1}_{j+\frac{1}{2}} - u^{n}_{j+\frac{1}{2}}}{\Delta t} + \frac{g}{\Delta x}\left[\left(\frac{h^n_{j+1} + h^{n+1}_{j+1}}{2}\right) - \left(\frac{h^n_j + h^{n+1}_j}{2}\right)\right] = 0 ,$$

(57)

$$\frac{h^{n+1}_j - h^n_j}{\Delta t} + \frac{H}{\Delta x}\left[\left(\frac{u^n_{j+\frac{1}{2}} + u^{n+1}_{j+\frac{1}{2}}}{2}\right) - \left(\frac{u^n_{j-\frac{1}{2}} + u^{n+1}_{j-\frac{1}{2}}}{2}\right)\right] = 0 .$$

(58)

This leads to

$$(\lambda - 1)\hat{u}^n + \left(\frac{1+\lambda}{2}\right)\frac{g\Delta t}{\Delta x} 2i \sin\left(\frac{k\Delta x}{2}\right)\hat{h}^n = 0 ,$$

(59)

$$\left(\frac{1+\lambda}{2}\right)\frac{H\Delta t}{\Delta x} 2i \sin\left(\frac{k\Delta x}{2}\right)\hat{u}^n + (\lambda - 1)\hat{h}^n = 0 .$$

(60)

For non-trivial solutions, we need

$$(\lambda - 1)^2 + (1+\lambda)^2\left(\frac{c_{GW}\Delta t}{\Delta x}\right)^2 \sin^2\left(\frac{k\Delta x}{2}\right) = 0 .$$

(61)

Using (41), we can show that this is equivalent to

$$\lambda^2 - 2\lambda\left(\frac{16-b}{16+b}\right) + 1 = 0 .$$

(62)

The solutions are

$$\lambda = \left(\frac{16-b}{16+b}\right) \pm i\sqrt{1-\left(\frac{16-b}{16+b}\right)^2} \ . \tag{63}$$

It follows that $|\lambda|^2 = 1$ for all modes, i.e., the trapezoidal scheme is unconditionally neutral.

The disadvantage of such implicit schemes is that they give rise to matrix problems, i.e., the various unknowns must be solved for simultaneously at all grid points.

A simple alternative, which is conditionally stable but allows a longer time step, is the *"forward-backward" scheme*, given by

$$\frac{u^{n+1}_{j+\frac{1}{2}} - u^{n}_{j+\frac{1}{2}}}{\Delta t} + g\left(\frac{h^{n+1}_{j+1} - h^{n+1}_{j}}{\Delta x}\right) = 0 \ , \tag{64}$$

$$\frac{h^{n+1} - h^{n}_{j}}{\Delta t} + H\left(\frac{u^{n}_{j+\frac{1}{2}} - u^{n}_{j-\frac{1}{2}}}{\Delta x}\right) = 0 \ . \tag{65}$$

This scheme can be called "partially implicit," because the end-of-time-step mass field predicted using (64) is used to compute the pressure-gradient force in (63). The continuity equation uses a forward time step. There is no need to solve a matrix problem.

We know that the forward scheme for both equations is unconditionally unstable, and that the backward scheme for both equations is unconditionally stable and damping. When we "combine" the two approaches, in the forward-backward scheme, the result turns out to be conditionally stable with a fairly long allowed time step, and neutral when stable. From (64) and (65), we get

$$(\lambda - 1)\hat{u}^n + \lambda \frac{g\Delta t}{\Delta x} 2i \sin\left(\frac{k\Delta x}{2}\right)\hat{h}^n = 0 \ , \tag{66}$$

$$\frac{H\Delta t}{\Delta x} 2i \sin\left(\frac{k\Delta x}{2}\right)\hat{u}^n + (\lambda - 1)\hat{h}^n = 0 \ . \tag{67}$$

This leads to

$$(\lambda - 1)^2 + 4\lambda\left(\frac{c_{GW}\Delta t}{\Delta x}\right)^2 \sin^2\left(\frac{k\Delta x}{2}\right) = 0 \ , \tag{68}$$

which is equivalent to

$$\lambda^2 + \left(\frac{b}{4} - 2 \right) \lambda + 1 = 0 \, .$$

(69)

The solutions are

$$\lambda = \frac{\left(2 - \frac{b}{4} \right) \pm \sqrt{\left(2 - \frac{b}{4} \right)^2 - 4}}{2} = \left(1 - \frac{b}{8} \right) \pm i \sqrt{\frac{b}{4} - \left(\frac{b}{8} \right)^2} \, .$$

(70)

The discriminant is non-negative for

$$b \le 16 \, ,$$

(71)

which corresponds to

$$\left(\frac{c_{GW} \Delta t}{\Delta x} \right)^2 \sin^2 \left(\frac{k \Delta x}{2} \right) \le 1 \, ,$$

(72)

It follows that

$$\frac{c_{GW} \Delta t}{\Delta x} \le 1 \text{ is required for stability.}$$

(73)

The time step can thus be twice as large as with the leapfrog scheme. When (71) is satisfied, we have $|\lambda|^2 = 1$ for all modes, i.e., the scheme is neutral when stable (like the leapfrog scheme). The forward-backward scheme is thus very attractive: It allows a long time step, it is neutral when stable, it is non-iterative, and it has no computational modes.

Going to two dimensions and adding rotation does not change much. The Coriolis terms can easily be made implicit if desired, since they are linear in the dependent variables and do not involve spatial derivatives.

8.6 Boundary conditions for wave propagation

At a real, physical wall, the normal component of the velocity has to be zero. If we use the C-grid, for example, we should position and orient the walls so that they are located at wind points, and oriented perpendicular to the locally defined velocity component. This means that the walls correspond to the edges of mass boxes on the C-grid. See Fig. 8.5 for a one-dimensional example.

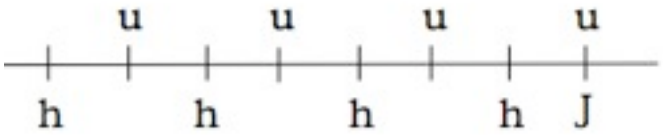

Figure 8.5: A one-dimensional staggered grid for solution of the shallow water equations, near a wall where $j = J$.

The case of a fictitious wall is more difficult We want waves to propagate out of the domain, instead of bouncing off a wall that isn't really there.

xxxx more to be added here.

8.7 Other meshes

In order to define the grids A-E for a square mesh, we have had to specify both the locations and the orientations of the velocity components that are used to represent the horizontal wind. We now broaden the discussion to include the triangular and hexagonal meshes, applying the definitions consistently in all cases.

The A-grid has the both velocity components co-located with the mass, and the directions of the velocity components are perpendicular to the cell walls.

The A-Grid and Z-grid do not involve any staggering, so they can be unambiguously defined on triangular or hexagonal meshes, or for that matter meshes of any other shape.

The B-grid can be generalized by defining it to have the horizontal velocity vector at the corners of mass cells. *The vector is represented using components that point along the walls that intersect at the corners.* On a triangular mesh, there are 6 intersecting walls at each corner, on a quadrilateral mesh there are two, and on a hexagonal mesh there are three. From this point of view, the B grid is really only compatible with quadrilateral meshes.

The C-grid can be generalized by defining it to have the normal component of the velocity on the edges of all mass-cells.

The generalized D-grid has the tangential velocity component on the edges of mass-cells.

As with the B-grid, the generalized E-grid has wind vectors on the corners of the mass-cells. In contrast to the B-grid, however, *the E-grid's wind components point diagonally across the cells*, as shown for the grey cells in the illustration of the E-grid in Fig. 8.2. There would be six such components on a triangular mesh, two on a quadrilateral mesh, and three on a hexagonal mesh. With this definition, the E grid cannot be defined for the triangular mesh or hexagonal meshes.

Alternatively, we can define the E grid as the superposition of multiple hexagonal C grids, such that two-dimensional velocity vectors, represented by the tangential and normal components, are defined on each cell wall, as with the orange grid cells shown for the E grid in Fig. 8.2. It is not possible to create triangular or hexagonal E grids in this way. So again with this definition, the E grid cannot be defined for the triangular mesh or hexagonal meshes.

From this point of view, the E grid is really only compatible with quadrilateral meshes. It is possible, however, to create an E grid by combining a hexagonal C grid with a triangular C grid. The resulting grid suffers from computational modes.

The Z-grid represents the velocity in terms of the vorticity and divergence, so no velocity components are defined.

Table 8.2 lists the numbers of corners and edges per face, on the triangular, square, and

	Triangles	Squares	Hexagons
Corners per face	1/2	1	2
Edges per face	3/2	2	3

Table 8.1: The numbers of corners and edges per face, on the triangular, square, and hexagonal meshes.

hexagonal meshes. Table 8.1 lists the number of prognostic degrees of freedom in the wind field per mass point, for the generalized A-E and Z grids, on triangular, square, and hexagonal meshes. From a physical point of view, *there should be two prognostic degrees of freedom in the wind*

Grid	Triangles	Squares	Hexagons
A	2	2	2
B	1	2	4
C	3/2	2	3
D	3/2	2	3
E	Does not exist	2	Does not exist
Z	2	2	2

Table 8.2: The number of prognostic degrees of freedom in the horizontal wind field, per mass point, on grids A-E and Z, and for triangular, square, and hexagonal meshes. For the Z-grid, the vorticity and divergence carry the information about the wind field.

field per mass point. The A-grid and Z-grid achieve this ideal on all three meshes. All of the

other combinations fall short.

8.8 Summary and conclusions

Horizontally staggered grids are important because they make it possible to avoid or minimize computational modes in space, and to realistically simulate geostrophic adjustment. The Z-grid gives the best overall simulation of geostrophic adjustment, for a range of grid sizes relative to the radius of deformation. In order to use the Z-grid, it is necessary to solve a pair of Poisson equations on each time step.

The rapid phase speeds of external gravity waves limit the time step that can be used with explicit schemes. Implicit schemes can be unconditionally stable, but in order to use them it is necessary to solve the equations simultaneously for all grid points.

Problems

1. Derive the dispersion equation for the C-grid, as given in Fig. 8.2.

2. Consider the linearized (about a resting basic state) shallow-water equations without rotation on the one-dimensional versions of the A grid and the C grid. Let the distance between neighboring mass points be d on both grids. Use leapfrog time differencing and centered space differencing. Derive the stability criteria for both cases, and compare the two results.

3. Write down differential-difference equations for the linearized (about a resting basic state) one-dimensional shallow water equations without rotation on an *unstaggered* grid (the A grid), using fourth-order accuracy for the spatial derivatives. (Just use the fourth-order scheme discussed in Chapter 2; you are not required to prove the order of accuracy in this problem.) Perform an analysis to determine whether or not the scheme has computational modes. Compare with the corresponding second-order scheme.

4. Program the two-dimensional linearized shallow water equations for the square A-grid and the square C-grid, using a mesh of 101 x 101 mass points, with periodic boundary conditions in both directions. Use leapfrog time differencing. Set $g = 0.1$ m s^{-1}, $H = 10^3$ m, and $d = 10^5$ m. In the square region

$$\begin{aligned} 45 \le i \le 55, \\ 45 \le j \le 55, \end{aligned} \tag{74}$$

apply a forcing in the continuity equation, of the form

$$\left(\frac{\partial h}{\partial t} \right)_{\text{noise}} = (-1)^{i+j} N \sin(\omega_N t), \tag{75}$$

and set $\left(\frac{\partial h}{\partial t} \right)_{\text{noise}} = 0$ at all other grid points. Adopt the values $\omega_N = 2\pi \times 10^{-3}$ s^{-1}; and $N = 10^{-4}$ m s^{-1}. In addition, apply a forcing to the entire domain of the form

$$\left(\frac{\partial h}{\partial t} \right)_{\text{smooth}} = S \sin\left(\frac{2\pi x}{L} \right) \sin\left(\frac{2\pi y}{L} \right) \sin(\omega_S t) \tag{76}$$

with $\omega_S = \frac{2\pi\sqrt{gH}}{L}$ s^{-1} and $S = 10^{-4}$ m s^{-1}. Here L is $101 \times d$, the width of the domain. Finally, include damping in the momentum equations, of the form

$$\left(\frac{\partial u}{\partial t}\right)_{\text{fric}} = -Ku,$$
$$\left(\frac{\partial v}{\partial t}\right)_{\text{fric}} = -Kv,$$

(77)

where $K = 2\times10^{-5}$ s^{-1}. Because the model has both forcing and damping, it is possible to obtain a statistically steady solution.

a) Analyze the stability of the two models without the forcing or damping terms. Using your results, choose a suitable time step for each model. Note: The forcing and damping terms are not expected to limit the time step, so simply omit them in your analysis of the stability criterion.

b) As initial conditions, put $u = 0$, $v = 0$, and $h = 0$. Run both versions of the model for at least 10^5 simulated seconds, and analyze the results. Your analysis should compare various aspects of the solutions, in light of the discussion given in this chapter.

c) Repeat your runs using $f = 2\times10^{-4}$ s^{-1}. Discuss the changes in your results.