

HIM: The Hallberg Isopycnal Coordinate Primitive Equation Model

BY ROBERT HALLBERG¹

PRINCETON UNIVERSITY AND NOAA GEOPHYSICAL FLUID
DYNAMICS LABORATORY

LAST UPDATED APRIL 2, 1997

This technical note derives the continuous primitive equations on a sphere. It then describes the discretization used in the model, first into layers, then the horizontal discretization, and finally the discretization in time.

1. The Primitive Equations on a Sphere in Isopycnal

Coordinates

In spherical coordinates the inviscid Navier-Stokes equations can be written as:

$$\frac{Du}{Dt} - 2\Omega \sin\theta v + 2\Omega \cos\theta w - \frac{uv \tan\theta}{a} + \frac{uw}{a} = -\frac{1}{\rho a \cos\theta} \frac{\partial p}{\partial \phi}, \quad (1.1)$$

$$\frac{Dv}{Dt} + 2\Omega \sin\theta u + \frac{u^2 \tan\theta}{a} + \frac{vw}{a} = -\frac{1}{\rho a} \frac{\partial p}{\partial \theta}, \text{ and} \quad (1.2)$$

$$\frac{Dw}{Dt} - 2\Omega \cos\theta u - \frac{u^2 + v^2}{a} = -\frac{1}{\rho} \frac{\partial p}{\partial z} + g, \quad (1.3)$$

where θ is latitude, ϕ is longitude, z is height above the average radius of the sphere, a , (u, v, w) are the velocities in the (eastward, northward, upward) directions, g is the gravitational acceleration, ρ is the density of the fluid, p is pressure, and Ω is the sphere's rotation rate. In addition, these equations use the definition:

$$\frac{DA}{Dt} \equiv \left. \frac{\partial A}{\partial t} \right|_z + \frac{u}{a \cos\theta} \left. \frac{\partial A}{\partial \phi} \right|_z + \frac{v}{a} \left. \frac{\partial A}{\partial \theta} \right|_z + w \left. \frac{\partial A}{\partial z} \right|_z, \quad (1.4)$$

¹. Corresponding Author Address: Dr. Robert Hallberg, NOAA GFDL, Princeton University, Forrestal Campus, U. S. Route 1, P. O. Box 308, Princeton, NJ 08542; e-mail: rwh@gfdl.gov; phone (609)452-6508

where the subscripts z indicate partial derivatives taken on surfaces of constant z . The other equations that govern the motion are the incompressible continuity equation:

$$\left. \frac{1}{a \cos \theta} \frac{\partial u}{\partial \phi} \right|_z + \left. \frac{1}{a \cos \theta} \frac{\partial}{\partial \theta} (v \cos \theta) \right|_z + \frac{\partial w}{\partial z} = -\frac{1}{\rho} \frac{D\rho}{Dt}, \quad (1.5)$$

where the material derivative of density is filtered to remove high frequency signals, and the definition of some potential density, R , that is materially conserved except for an explicit diabatic forcing:

$$\frac{DR}{Dt} = \dot{R}. \quad (1.6)$$

It is assumed that density and potential density are related by some function $K(p, R)$:

$$R = \rho K(p, R). \quad (1.7)$$

The definition of potential density can be substituted into the continuity equation to give:

$$\begin{aligned} \left. \frac{1}{a \cos \theta} \frac{\partial u}{\partial \phi} \right|_z + \left. \frac{1}{a \cos \theta} \frac{\partial}{\partial \theta} (v \cos \theta) \right|_z + \frac{\partial w}{\partial z} &= -\left(\frac{1}{R} - \frac{1}{K} \frac{\partial K}{\partial R} \right) \dot{R} + \frac{1}{K} \frac{\partial K}{\partial p} \frac{Dp}{Dt}, \\ &\approx -\left(\frac{1}{R} - \frac{1}{K} \frac{\partial K}{\partial R} \right) \frac{\dot{R}}{R} + \frac{\rho g w}{K} \frac{\partial K}{\partial p} \approx -\frac{\dot{R}}{R} \end{aligned} \quad (1.8)$$

where the final approximation assumes that the vertical scale of the motion is small compared with the density scale height, and that thermobaric effects are negligible (a tenuous assumption for the ocean). The final approximation makes volume integrated potential density, rather than mass, conserved. The replacement of the material pressure derivative by the hydrostatic pressure change due to vertical motion filters out sound waves. If the Boussinesq approximation is made (which is of the same order as approximations used in (1.8)), volume is conserved instead of mass and (1.8) becomes:

$$\left. \frac{1}{a \cos \theta} \frac{\partial u}{\partial \phi} \right|_z + \left. \frac{1}{a \cos \theta} \frac{\partial}{\partial \theta} (v \cos \theta) \right|_z + \frac{\partial w}{\partial z} = 0. \quad (1.9)$$

The vertical momentum equation is replaced by the hydrostatic equation:

$$\frac{\partial p}{\partial z} = -\rho g. \quad (1.10)$$

For energy conservation, the horizontal momentum equations, (1.1) and (1.2), are modified by the traditional approximation and neglect of two of the metric terms:

$$\frac{Du}{Dt} - 2\Omega \sin \theta v - \frac{uv \tan \theta}{a} = -\left. \frac{1}{\rho a \cos \theta} \frac{\partial p}{\partial \phi} \right|_z, \text{ and} \quad (1.11)$$

$$\frac{Dv}{Dt} + 2\Omega \sin\theta u + \frac{u^2 \tan\theta}{a} = -\frac{1}{\rho a} \frac{\partial p}{\partial\theta} \Big|_z. \quad (1.12)$$

The primitive equations, (1.5), (1.6), (1.10), (1.11), and (1.12) can be rewritten in isopycnal coordinates. The relation between zonal derivatives at constant z and zonal derivatives at constant R , for example, is:

$$\frac{\partial p}{\partial\phi} \Big|_z = \frac{\partial p}{\partial\phi} \Big|_R - \frac{\partial p \partial z}{\partial z \partial\phi} \Big|_R = \frac{\partial p}{\partial\phi} \Big|_R - \frac{\partial p \partial R \partial z}{\partial R \partial z \partial\phi} \Big|_R. \quad (1.13)$$

The definition of w with an isopycnal vertical coordinate is:

$$w \equiv \frac{Dz}{Dt} = \frac{\partial z}{\partial t} \Big|_R + \frac{u}{a \cos\theta} \frac{\partial z}{\partial\phi} \Big|_R + \frac{v \partial z}{a \partial\theta} \Big|_R + \frac{DR \partial z}{Dt \partial R} = \frac{\partial z}{\partial t} \Big|_R + \mathbf{u}_h \cdot \nabla_{Rz} + \dot{R} \frac{\partial z}{\partial R}. \quad (1.14)$$

The volume integrated potential density conserving continuity equation, (1.8), becomes:

$$\begin{aligned} & \frac{1}{a \cos\theta} \left(\frac{\partial u}{\partial\phi} \Big|_R - \frac{\partial u \partial R \partial z}{\partial R \partial z \partial\phi} \Big|_R + \frac{\partial}{\partial\theta} (v \cos\theta) \Big|_R - \cos\theta \frac{\partial v \partial R \partial z}{\partial R \partial z \partial\theta} \Big|_R \right), \text{ or} \\ & + \frac{\partial R}{\partial z} \frac{\partial}{\partial R} \left(\frac{\partial z}{\partial t} \Big|_R + \frac{u}{a \cos\theta} \frac{\partial z}{\partial\phi} \Big|_R + \frac{v \partial z}{a \partial\theta} \Big|_R + \dot{R} \frac{\partial z}{\partial R} \right) = -\frac{\dot{R}}{R} \\ & \frac{\partial}{\partial t} \left(\frac{\partial z}{\partial R} \right) \Big|_R + \frac{1}{a \cos\theta} \left[\frac{\partial}{\partial\phi} \left(u \frac{\partial z}{\partial R} \right) \Big|_R + \frac{\partial}{\partial\theta} \left(v \frac{\partial z}{\partial R} \cos\theta \right) \Big|_R \right] + \frac{1}{R} \frac{\partial}{\partial R} \left(R \dot{R} \frac{\partial z}{\partial R} \right) = 0. \end{aligned} \quad (1.15)$$

The Boussinesq continuity equation, (1.9), becomes:

$$\begin{aligned} & \frac{1}{a \cos\theta} \left(\frac{\partial u}{\partial\phi} \Big|_R - \frac{\partial u \partial R \partial z}{\partial R \partial z \partial\phi} \Big|_R + \frac{\partial}{\partial\theta} (v \cos\theta) \Big|_R - \cos\theta \frac{\partial v \partial R \partial z}{\partial R \partial z \partial\theta} \Big|_R \right), \text{ or} \\ & + \frac{\partial R}{\partial z} \frac{\partial}{\partial R} \left(\frac{\partial z}{\partial t} \Big|_R + \frac{u}{a \cos\theta} \frac{\partial z}{\partial\phi} \Big|_R + \frac{v \partial z}{a \partial\theta} \Big|_R + \dot{R} \frac{\partial z}{\partial R} \right) = 0 \\ & \frac{\partial}{\partial t} \left(\frac{\partial z}{\partial R} \right) \Big|_R + \frac{1}{a \cos\theta} \left[\frac{\partial}{\partial\phi} \left(u \frac{\partial z}{\partial R} \right) \Big|_R + \frac{\partial}{\partial\theta} \left(v \frac{\partial z}{\partial R} \cos\theta \right) \Big|_R \right] + \frac{\partial}{\partial R} \left(\dot{R} \frac{\partial z}{\partial R} \right) = 0. \end{aligned} \quad (1.16)$$

The horizontal momentum equations become:

$$\begin{aligned} & \frac{\partial u}{\partial t} \Big|_R + \frac{u}{a \cos\theta} \frac{\partial u}{\partial\phi} \Big|_R + \frac{v \partial u}{a \partial\theta} \Big|_R + \dot{R} \frac{\partial u}{\partial R} - fv - \frac{uv \tan\theta}{a}, \\ & = -\frac{1}{\rho a \cos\theta} \left(\frac{\partial p}{\partial\phi} \Big|_R - \frac{\partial p \partial z}{\partial z \partial\phi} \Big|_R \right) \end{aligned} \quad (1.17)$$

$$\begin{aligned} \left. \frac{\partial v}{\partial t} \right|_R + \left. \frac{u}{a \cos \theta} \frac{\partial v}{\partial \phi} \right|_R + \left. \frac{v \partial v}{a \partial \theta} \right|_R + \dot{R} \frac{\partial v}{\partial R} + fu + \frac{u^2 \tan \theta}{a} , \\ = -\frac{1}{\rho a} \left(\left. \frac{\partial p}{\partial \theta} \right|_R - \left. \frac{\partial p \partial z}{\partial \theta} \right|_R \right) \end{aligned} \quad (1.18)$$

where these equations use the definition of the Coriolis parameter, $f = 2\Omega \sin \theta$. Using the hydrostatic equation and the definitions of the vertical component of relative vorticity:

$$\zeta = \frac{1}{a \cos \theta} \left(\left. \frac{\partial v}{\partial \phi} \right|_R - \left. \frac{\partial}{\partial \theta} (v \cos \theta) \right|_R \right) , \quad (1.19)$$

and the Bernoulli function, B, and Montgomery potential, M:

$$B = \tilde{K}(p, R)/R + gz + (u^2 + v^2)/2 \equiv M + (u^2 + v^2)/2 , \quad (1.20)$$

where $\tilde{K}(p, R)$ is defined by:

$$\left. \frac{\partial}{\partial p} [\tilde{K}(p, R)] \right|_R \equiv K(p, R) , \quad (1.21)$$

the horizontal momentum equations become:

$$\left. \frac{\partial u}{\partial t} \right|_R + \left. \dot{R} \frac{\partial u}{\partial R} - (f + \zeta)v \right|_R = -\left. \frac{1}{a \cos \theta} \frac{\partial B}{\partial \phi} \right|_R , \quad (1.22)$$

$$\left. \frac{\partial v}{\partial t} \right|_R + \left. \dot{R} \frac{\partial v}{\partial R} + (f + \zeta)u \right|_R = -\left. \frac{1}{a} \frac{\partial B}{\partial \theta} \right|_R . \quad (1.23)$$

The hydrostatic equation can be used to determine the Montgomery potential at some density:

$$\begin{aligned} \frac{\partial}{\partial z} (MR) &= \frac{\partial p}{\partial z} \frac{\partial}{\partial p} [\tilde{K}(p, R)] + \frac{\partial R}{\partial z} \frac{\partial}{\partial R} [\tilde{K}(p, R)] + Rg + gz \frac{\partial R}{\partial z} , \\ &= \left[\frac{\partial}{\partial R} [\tilde{K}(p, R)] + gz \right] \frac{\partial R}{\partial z} \end{aligned}$$

which can be integrated along a vertical path to give:

$$RM(R) = R_0 M(R_0) + \int_{\eta}^z \frac{d}{dz'} (MR) dz' = R_0 g \eta + \int_{R_0}^R \left[\frac{\partial}{\partial R} [\tilde{K}(p, R)] + gz \right] dR' ,$$

where η is the height of the upper surface (where the pressure is assumed to be zero), and R_0 is the potential density at the upper surface. The Montgomery potential at potential density R is:

$$M = \left(R_0 g \eta + \int_{R_0}^R \left\{ \frac{\partial}{\partial R} [\tilde{K}(p, R)] + gz \right\} dR' \right) / R . \quad (1.24)$$

If the compressibility is approximated as a function of pressure alone, (1.24) becomes:

$$M = (R_0 g \eta + \int_{R_0}^R g z dR') / R, \quad (1.25)$$

which is exactly the equation which would be obtained in density coordinates, rather than potential density coordinates, but with R replaced by ρ . Eq. (1.25) is an integral version of the hydrostatic equation.

In the model, momentum is subject to an along isopycnal biharmonic diffusion and a much weaker diapycnal Laplacian diffusion. With the definition of the along isopycnal gradient operator:

$$\nabla_R = \frac{\hat{\theta}}{a} \frac{\partial}{\partial \theta} \Big|_R + \frac{\hat{\phi}}{a \cos \theta} \frac{\partial}{\partial \phi} \Big|_R, \quad (1.26)$$

both horizontal momentum equations can be written with momentum conserving diffusion terms:

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} \Big|_R + \dot{R} \frac{\partial \mathbf{u}}{\partial R} + (f + \zeta) \hat{\mathbf{k}} \times \mathbf{u} \\ = -\nabla_R B + \frac{\partial R}{\partial z} \frac{\partial}{\partial R} \left(\kappa \frac{\partial R}{\partial z} \frac{\partial \mathbf{u}}{\partial R} \right) - \frac{\partial R}{\partial z} \nabla_R \cdot \left(A \frac{\partial z}{\partial R} \nabla_R (\nabla_R^2 \mathbf{u}) \right) \end{aligned} \quad (1.27)$$

where $\mathbf{u} = u \hat{\phi} + v \hat{\theta}$ is the horizontal velocity. Recall that in spherical coordinates:

$$\frac{\partial}{\partial \phi} \hat{\theta} = -\sin \theta \hat{\phi}, \text{ etc.}$$

which add a large number of metric terms to the horizontal biharmonic operator. In (1.27), both the diapycnal kinematic viscosity, κ , and the isopycnal biharmonic viscosity, A , can be either constant or variable. Wind stress is imposed as a stress boundary condition at the upper surface, and bottom drag is imposed as a no-slip bottom boundary condition. No diffusion is currently included in the continuity equation.

2. Vertical Discretization of the Isopycnal Primitive Equations

The continuous primitive equations in isopycnal coordinates are discretized by dividing the fluid into layers of constant density. The continuity equation, (1.15) or (1.16), is integrated in density across a layer to give a layer thickness conservation equation. Similarly the horizontal momentum equations are averaged through each layer; with constant layer densities, most of the terms in the momentum equations are constant within a layer.

The mass conservation equation for each layer is determined by integrating (1.15) or (1.16) between the potential densities $\rho_{n-1/2}$ and $\rho_{n+1/2}$ which bound layer n . (In this section, ρ is used for potential density rather than R as in the previous section.) With quantized density layers, the terms (z/ρ) act

as delta-functions in density space. The exact differential of a delta function on the left hand side of (1.15) and (1.16) is ill-posed with quantized density. The diapycnal mass flux term is obtained by redefining $F = \rho\dot{\rho}(\partial z/\partial\rho)$ to be the downward mass flux times the ratio of potential density to density across the interface between layers. With the definition of layer thickness:

$$h_n = -\int_{\rho_{n-1/2}}^{\rho_{n+1/2}} \frac{\partial z}{\partial \rho} d\rho, \quad (2.1)$$

a consistent definition of F for the potential density conserving continuity equation is:

$$F_{n+1/2} = \frac{[\max(\rho_n \dot{\rho}_n h_n, 0) + \min(\rho_{n+1} \dot{\rho}_{n+1} h_{n+1}, 0)]}{(\rho_{n+1} - \rho_n)}. \quad (2.2)$$

(n increases downward.) Multiplying (1.15) by potential density, integrating in potential density, and dividing by potential density gives the layer thickness conservation equation:

$$\frac{\partial h_n}{\partial t} + \nabla_\rho \cdot (h_n \mathbf{u}_n) = (F_{n-1/2} - F_{n+1/2})/\rho_n. \quad (2.3)$$

If the Boussinesq continuity equation, (1.16), is used instead, a consistent definition of F becomes:

$$F_{n+1/2} = \frac{\rho_0 [\max(\dot{\rho}_n h_n, 0) + \min(\dot{\rho}_{n+1} h_{n+1}, 0)]}{(\rho_{n+1} - \rho_n)}. \quad (2.4)$$

Integrating (1.16) in density gives the Boussinesq layer thickness conservation equation:

$$\frac{\partial h_n}{\partial t} + \nabla_\rho \cdot (h_n \mathbf{u}_n) = (F_{n-1/2} - F_{n+1/2})/\rho_0. \quad (2.5)$$

The horizontal momentum equation for layer n is found by multiplying (1.27) by (z/ρ) and integrating in potential density between $\rho_{n-1/2}$ and $\rho_{n+1/2}$. The diapycnal forcing term in the momentum equations requires some manipulation and the explicit assumption that velocity is constant within a layer:

$$\begin{aligned} \int_{\rho_{n-1/2}}^{\rho_{n+1/2}} \frac{\partial z}{\partial \rho} \dot{\rho} \frac{\partial \mathbf{u}}{\partial \rho} d\rho &= \int_{\rho_{n-1/2}}^{\rho_{n+1/2}} \frac{\partial}{\partial \rho} \left(\frac{\partial z}{\partial \rho} \dot{\rho} \mathbf{u} \right) d\rho - \int_{\rho_{n-1/2}}^{\rho_{n+1/2}} \mathbf{u} \frac{\partial}{\partial \rho} \left(\frac{\partial z}{\partial \rho} \dot{\rho} \right) d\rho, \\ &= \left[\frac{\partial z}{\partial \rho} \dot{\rho} \mathbf{u} \right]_{n-1/2}^{n+1/2} - \mathbf{u}_n \left[\frac{\partial z}{\partial \rho} \dot{\rho} \right]_{n-1/2}^{n+1/2} \\ &= (\mathbf{u}_n - \mathbf{u}_{n+1/2}) F_{n+1/2} / \rho_n - (\mathbf{u}_n - \mathbf{u}_{n-1/2}) F_{n-1/2} / \rho_n \end{aligned}$$

where $\mathbf{u}_{n+1/2}$ is the velocity right at the interface between layers. In the model, it is defined to be the ‘‘upwind’’ velocity based on the direction of the diapycnal fluxes. The integral of (z/ρ) times (1.27) in potential density gives:

$$\begin{aligned}
& h_n \frac{\partial \mathbf{u}_n}{\partial t} + [(\mathbf{u}_n - \mathbf{u}_{n+1/2})F_{n+1/2} - (\mathbf{u}_n - \mathbf{u}_{n-1/2})F_{n-1/2}] / \rho_n + q_n \hat{\mathbf{k}} \times (h_n^2 \mathbf{u}_n) \quad , \\
& = -h_n \nabla_\rho B_n + \left[\kappa \frac{\partial \rho}{\partial z} \frac{\partial \mathbf{u}}{\partial \rho} \right]_{n+1/2}^{n-1/2} - \nabla_\rho \cdot (A h_n \nabla_\rho (\nabla_\rho^2 \mathbf{u}_n))
\end{aligned}$$

which can be divided by h_n and rearranged to give

$$\begin{aligned}
\frac{\partial \mathbf{u}_n}{\partial t} + q_n \hat{\mathbf{k}} \times (h_n \mathbf{u}_n) &= -\nabla_\rho B_n - \frac{1}{h_n} \nabla_\rho \cdot [A h_n \nabla_\rho (\nabla_\rho^2 \mathbf{u}_n)] \quad . \\
& + \frac{[\tau_{n-1/2} - \tau_{n+1/2} - (\mathbf{u}_n - \mathbf{u}_{n-1/2})F_{n-1/2} + (\mathbf{u}_n - \mathbf{u}_{n+1/2})F_{n+1/2}]}{(h_n \rho_n)}
\end{aligned} \quad (2.6)$$

Eq. (2.6) uses the definition of the layer potential vorticity:

$$q_n = (f + \zeta_n) / h_n, \quad (2.7)$$

and of the stress between layers:

$$\tau_{n+1/2} = \frac{\rho_n + \rho_{n+1}}{2} \kappa \frac{\partial \rho}{\partial z} \frac{\partial \mathbf{u}}{\partial \rho} \Big|_{n+1/2}. \quad (2.8)$$

The layer Bernoulli function is given by $B_n = M_n + \frac{1}{2} \|\mathbf{u}_n\|^2$, while the layer Montgomery potential is given by the discretized version of (1.25):

$$M_n = \frac{\rho_0}{\rho_n} \sum_{j=1}^n (\eta_{j-1/2} g'_{j-1/2}). \quad (2.9)$$

Here $\eta_{n+1/2}$ is the height of the interface between layers n and $n+1$ (the index n increases downward), and $g'_{n+1/2} = g(\rho_{n+1} - \rho_n) / \rho_0$ is the reduced gravity at this interface. For N layers in a basin of depth D ,

$$\eta_{n-1/2} = -D + \sum_{j=n}^N h_j. \quad (2.10)$$

If the Boussinesq approximation is made, (2.9) is simply:

$$M_n = \sum_{j=1}^n (\eta_{j-1/2} g'_{j-1/2}). \quad (2.11)$$

Momentum is diffused by a biharmonic viscosity along isopycnal layers. In spherical coordinates the biharmonic operator includes many metric terms arising from the change with latitude of the grid spacing and unit vector directions. These metric terms are smaller than the leading order terms by something proportional to the ratio of the grid spacing to the radius of the earth, since the biharmonic

viscosity acts predominantly on scales of a few grid spacings. Since the domain does not include the poles, and the dissipation is only an arbitrary parameterization, these metric terms can be safely neglected. Apart from the diffusion parameterization, all the metric terms due to motion on a sphere are retained.

3. Spatial Discretization

Eq. (2.3) and (2.6) are discretized on a staggered Arakawa C-grid, with both meridional and zonal boundaries running through the PV grid points. B , M , h , and η are all evaluated at thickness grid points, which are displaced by 1/2 grid point in each direction from the PV grid points. The momentum equations are evaluated using a vorticity advection scheme of Arakawa and Hsu (1990). This scheme conserves both total energy and potential enstrophy in the limit of nondivergent mass fluxes. While this scheme is more complicated than the vorticity advection scheme of Sadourny (1975) and does not conserve potential enstrophy and total energy in general flow, as does a scheme from Arakawa and Lamb (1981), it tolerates arbitrarily thin layers. This feature is essential for simulating either outcropping isopycnals or large amplitude topography.

The Arakawa and Hsu (1990) vorticity advection scheme for a single layer is modified for spherical coordinates as described by Arakawa and Lamb (1981). The potential vorticity is:

$$q_{i,j} = (f_j + \zeta_{i,j})/h_{i,j}^{(q)}, \quad (3.1)$$

where the relative vorticity is:

$$\zeta_{i,j} = \frac{(v_{i+1/2,j} - v_{i-1/2,j})}{a \cos \theta_j \Delta \phi} - \frac{(\cos \theta_{j+1/2} u_{i,j+1/2} - \cos \theta_{j-1/2} u_{i,j-1/2})}{a \cos \theta_j \Delta \theta}, \quad (3.2)$$

the Coriolis parameter is:

$$f_j = 2\Omega \sin \theta_j, \quad (3.3)$$

and the layer thickness at PV grid points is:

$$h_{i,j}^{(q)} = \frac{\cos \theta_{j-1/2} (h_{i-1/2,j-1/2} + h_{i+1/2,j-1/2})}{4 \cos \theta_j} + \frac{\cos \theta_{j+1/2} (h_{i-1/2,j+1/2} + h_{i+1/2,j+1/2})}{4 \cos \theta_j}. \quad (3.4)$$

At the boundaries either free slip or no slip boundary conditions are applied. At the western boundary, for example, with a free slip boundary condition the relative vorticity becomes:

$$\zeta_{0,j} = 0, \quad (3.5)$$

while with a no slip boundary condition it becomes:

$$\zeta_{0,j} = (2v_{1/2,j})/(a \cos \theta_j \Delta \phi). \quad (3.6)$$

At the southern boundary the no slip boundary condition is analogous, but the free slip boundary condition becomes:

$$\zeta_{i,0} = (\tan \theta_0 u_{i,1/2})/a. \quad (3.7)$$

The layer thickness at the boundary PV grid points is determined by replacing undefined thickness points outside of the grid by the value of their neighbor inside the boundary.

In the horizontal momentum equations, the potential vorticity containing terms are represented as:

$$(qhv)_{i,j+1/2} = \alpha_{i,j+1/2}(hv)_{i+1/2,j+1} + \beta_{i,j+1/2}(hv)_{i-1/2,j+1} \quad \text{and} \quad (3.8)$$

$$+ \gamma_{i,j-1/2}(hv)_{i+1/2,j} + \delta_{i,j+1/2}(hv)_{i+1/2,j}$$

$$(qhu)_{i+1/2,j} = \alpha_{i,j-1/2}(hu)_{i,j-1/2} + \beta_{i+1,j-1/2}(hu)_{i+1,j-1/2} \quad , \quad (3.9)$$

$$+ \gamma_{i+1,j+1/2}(hu)_{i+1,j+1/2} + \delta_{i,j+1/2}(hu)_{i,j+1/2}$$

where the (hv) and (hu) terms are defined by the thickness advection scheme and the terms α , β , γ , and δ are combinations of the neighboring potential vorticities:

$$\alpha_{i,j+1/2} = (q_{i+1,j+1/2} + q_{i,j+1/2} + q_{i,j-1/2})/12,$$

$$\beta_{i,j+1/2} = (q_{i-1,j+1/2} + q_{i,j+1/2} + q_{i,j-1/2})/12,$$

$$\gamma_{i,j+1/2} = (q_{i-1,j-1/2} + q_{i,j+1/2} + q_{i,j-1/2})/12, \quad \text{and}$$

$$\delta_{i,j+1/2} = (q_{i+1,j-1/2} + q_{i,j+1/2} + q_{i,j-1/2})/12.$$

The energetically consistent form of the Bernoulli function is:

$$B_{i+1/2,j+1/2} = M_{i+1/2,j+1/2} + (u_{i,j+1/2}^2 + u_{i+1,j+1/2}^2)/4 \quad .$$

$$+ (\cos \theta_j v_{i+1/2,j}^2 + \cos \theta_{j+1} v_{i+1/2,j+1}^2)/(4 \cos \theta_{j+1/2})$$

4. The Thickness Advection Scheme

The layer thickness equation is stepped using a modified version of the positive definite scheme from Hsu and Arakawa (1990). This scheme is essentially the same as that of Takacs (1985), except when there is a large relative thickness difference between neighboring grid points. The Takacs scheme

is second order accurate in time and space (third order accurate for uniform flow), and has greatly reduced phase error compared with the standard Lax-Wendroff thickness advection scheme. When there are significant thickness variations between adjacent grid points, upwind differencing advection is used, but the overall scheme retains the higher order accuracy of the Takacs scheme. Whenever thickness falls below some prescribed thickness, ϵ (here 10^{-10} m), mass is added to return the thickness to this value.

The thickness conservation equation for a layer, (2.3), is horizontally discretized as:

$$\frac{\partial}{\partial t} h_{i+1/2, j+1/2} + \frac{(hu)_{i+1, j+1/2} - (hu)_{i, j+1/2}}{a \cos \theta_{j+1/2} \Delta \phi} + \frac{(hv \cos \theta)_{i+1/2, j+1} - (hv \cos \theta)_{i+1/2, j}}{a \cos \theta_{j+1/2} \Delta \theta} = \frac{\Delta F_{i+1/2, j+1/2}}{\rho} \quad (4.1)$$

where ΔF is the net diapycnal flux into a layer. At the boundaries, the no-normal flow boundary condition is applied easily, since the (hu) grid points are defined at the eastern and western boundaries and the (hv) grid points are defined at the northern and southern boundaries.

The diapycnal mass flux is applied before thickness is advected. Thickness is advected successively in each direction and a predictor-corrector is used in each direction. To illustrate the scheme, consider 1-dimensional advection in the x -direction in cartesian coordinates:

$$\frac{\partial}{\partial t} h_i + \frac{(hu)_{i+1/2} - (hu)_{i-1/2}}{\Delta x} = 0 \quad (4.2)$$

Thickness has integer indices, although this is not true elsewhere in this appendix. The predictor step for a positive velocity is:

$$h_i^* = h_i^m - (F_{i+1/2}^* - F_{i-1/2}^*) \quad (4.3)$$

$$F_{i+1/2}^* = \mu_{i+1/2} h_i^m \quad (4.4)$$

The corrector step is:

$$h_i^{m+1} = h_i^m - (F_{i+1/2} - F_{i-1/2}) \quad (4.5)$$

$$F_{i+1/2} = \mu_{i+1/2} h_i^m + \mu_{i+1/2} \frac{(2 - \mu_{i+1/2})}{6} (h_{i+1}^* - h_i^m) (1 - \gamma_{i+1/2}^+) + \hat{\mu}_{i+1/2} \frac{(1 + \mu_{i+1/2})}{6} (h_i^* - h_{i-1}^m) (1 - \gamma_{i+1/2}^+) \quad (4.6)$$

These equations use the definitions $\mu = u\Delta t/\Delta x$, and

$$\hat{\mu}_{i+1/2} = \mu_{i+1/2} \{ \max[(\mu_{i-1/2}/\mu_{i+1/2}), 0] \}^{1/2} .$$

If $\mu < 0$, (4.4) and (4.6) are replaced by the definitions:

$$F_{i+1/2}^* = \mu_{i+1/2} h_{i+1}^m . \quad (4.7)$$

$$F_{i+1/2} = \mu_{i+1/2} h_{i+1}^m + \mu_{i+1/2} \frac{(2 - \mu_{i+1/2})}{6} (h_i^* - h_{i+1}^m) (1 - \gamma_{i+1/2}^-) . \quad (4.8)$$

$$+ \hat{\mu}_{i+1/2} \frac{(1 + \mu_{i+1/2})}{6} (h_{i+1}^* - h_{i+2}^m) (1 - \gamma_{i+1/2}^-)$$

When $\gamma=0$, this scheme reduces to the Takacs (1985) thickness advection scheme. When $\gamma=1$, this scheme reduces to upwind differencing.

The definition of γ is crucial in assuring that the thickness remains positive definite. Hsu and Arakawa (1990) propose the definitions:

$$\gamma_{i+1/2}^+ = \left[\frac{(h_{i+1}^m - 2h_i^m + h_{i-1}^m)^2}{h_i^m h_{i+1}^m + (h_{i+1}^m - 2h_i^m + h_{i-1}^m)^2} \right]^2 .$$

$$\gamma_{i+1/2}^- = \left[\frac{(h_{i+2}^m - 2h_{i+1}^m + h_i^m)^2}{h_i^m h_{i+1}^m + (h_{i+2}^m - 2h_{i+1}^m + h_i^m)^2} \right]^2 . \quad (4.9)$$

The criterion imposed by Hsu and Arakawa (1990) in selecting these definitions are that $\gamma_{i+1/2}^+ \rightarrow 1$ and $\gamma_{i+1/2}^- \rightarrow 1$ as $h_i^m \rightarrow 0$ or $h_{i+1}^m \rightarrow 0$ and that $0 \leq \gamma \leq O(\Delta x^2)$. Unfortunately, the definitions of γ proposed by Hsu and Arakawa (1990), (4.9), fail to keep the thickness positive definite in all cases.

This can be seen by considering a case where all of the velocities are positive, $(h_{i+1}^m - 2h_i^m + h_{i-1}^m) = 0$, and $h_{i+1}^m \ll h_i^m$. Then $\gamma_{i+1/2}^+ = 0$, $h_{i+1}^* \approx \mu_{i+1/2} h_i^m$, $h_{i-1}^m \approx 2h_i^m$, $h_i^* \approx (1 - \mu_{i+1/2} + 2\mu_{i-1/2}) h_i^m$, and (4.6) becomes:

$$F_{i+1/2} \approx \frac{h_i^m}{6} [\mu_{i+1/2} (4 + 3\mu_{i+1/2} - 2\mu_{i+1/2}^2) . \quad (4.10)$$

$$- \hat{\mu}_{i+1/2} (1 + \mu_{i+1/2} - 2\mu_{i-1/2})]$$

$F_{i+1/2}$ can reverse direction, potentially depleting the downwind thickness grid point. For example, for $\mu_{i+1/2} = 0.000625$ and $\mu_{i-1/2} = 0.25$, $\hat{\mu}_{i+1/2} = 0.0125$ and $F_{i+1/2} \approx -0.00063 h_i^m$. This flux may easily be enough to make h_{i+1}^{m+1} negative.

The present model uses the definitions:

$$\begin{aligned} \gamma_{i+1/2}^+ &= \left[\frac{(h_{i+1}^* - h_i^m - h_i^* + h_{i-1}^m)^2 + (h_{i+1}^* - h_i^m)^2}{h_i^m h_{i+1}^m + (h_{i+1}^* - h_i^m - h_i^* + h_{i-1}^m)^2 + (h_{i+1}^* - h_i^m)^2} \right]^2 \\ \gamma_{i+1/2}^- &= \left[\frac{(h_{i+2}^m - h_{i+1}^m - h_{i+1}^* + h_i^*)^2 + (h_{i+1}^m - h_i^*)^2}{h_i^m h_{i+1}^m + (h_{i+2}^m - h_{i+1}^m - h_{i+1}^* + h_i^*)^2 + (h_{i+1}^m - h_i^*)^2} \right]^2, \end{aligned} \quad (4.11)$$

which assure that thickness remains positive as long as $|\mu| < 0.4517$. These definitions satisfy the criterion proposed by Hsu and Arakawa (1990) for the definition of γ . This value has been determined empirically when the velocities adjacent to a thickness point are both outward at the maximum velocity. For a given value of the maximum velocity, $\tilde{\mu}$, the minimum value of h_i^{m+1} relative to h_i^m is found for all of the values in the range $0 \leq \mu_{i+1/2} \leq \tilde{\mu}$, $-\tilde{\mu} \leq \mu_{i-1/2} \leq 0$, $0 < h_{i+1}^m \leq (h_{i+1}^* - \mu_{i+1/2} h_i^m) / (1 - \tilde{\mu})$, $0 < h_{i-1}^m \leq (h_{i-1}^* + \mu_{i-1/2} h_i^m) / (1 - \tilde{\mu})$, with the equality from (4.4) and (4.7) $h_i^* = (1 - \mu_{i+1/2} + \mu_{i-1/2}) h_i^m$. There is no upper limit placed on the size of h_{i+1}^* or h_{i-1}^* .

The definitions of γ also prevent thickness from becoming negative when both velocities are in the same direction. For this to be true, for positive values of $\mu_{i+1/2}$, there must be some value, α , such that:

$$-\alpha h_{i+1}^m < F_{i+1/2} < (1 - \alpha) h_i^m. \quad (4.12)$$

for any thickness and for velocities smaller than $\tilde{\mu}$. The maximum value of $F_{i+1/2}$ with $\hat{\mu}_{i+1/2} = 0$ is empirically less than $0.665 h_i^m$. Then (4.6) gives that:

$$\begin{aligned} F_{i+1/2} &< 0.665 h_i^m + \hat{\mu}_{i+1/2} \frac{(1 + \mu_{i+1/2})}{6} (h_i^* - h_{i-1}^m) (1 - \gamma_{i+1/2}^+) \\ &< 0.665 h_i^m + \tilde{\mu} \frac{(1 + \tilde{\mu})}{6} [(1 - \tilde{\mu})(h_i^m - h_{i-1}^m)] (1 - \gamma_{i+1/2}^+) \\ &< 0.665 h_i^m + \tilde{\mu} \frac{(1 - \tilde{\mu}^2)}{6} h_i^m < (0.665 + 1/16) h_i^m = 0.7275 h_i^m \end{aligned} \quad (4.13)$$

So the upper limit of the flux satisfies (4.12) for $\alpha = 0.25$. Using the fact that $0 \leq \mu_{i+1/2} \leq \tilde{\mu} < 0.5$ and the definition of $\hat{\mu}$, (4.6) shows that:

$$\begin{aligned} F_{i+1/2} &> \mu_{i+1/2} [4h_i^m + (h_{i+1}^* - h_i^m)(1 - \gamma_{i+1/2}^+)] / 4 \\ &+ \sqrt{\tilde{\mu} \mu_{i+1/2}} (h_i^* - h_{i-1}^m) (1 - \gamma_{i+1/2}^+) / 4 \end{aligned} \quad (4.14)$$

The only way that $F_{i+1/2}$ can be negative is if $h_{i-1}^m > h_i^*$, and the most negative value of the right hand side of (4.14) occurs when

$$\mu_{i+1/2} = \tilde{\mu} \frac{(1 - \gamma_{i+1/2}^+)^2 (h_i^* - h_{i-1}^m)^2}{4[4h_i^m + (h_{i+1}^* - h_i^m)(1 - \gamma_{i+1/2}^+)]},$$

and in this case (4.14) becomes:

$$\begin{aligned} F_{i+1/2} &> -\tilde{\mu} \frac{(1 - \gamma_{i+1/2}^+)^2 (h_i^* - h_{i-1}^m)^2}{16[4h_i^m + (h_{i+1}^* - h_i^m)(1 - \gamma_{i+1/2}^+)]} \\ &> -\tilde{\mu} \frac{(1 - \gamma_{i+1/2}^+)^2 (h_i^* - h_{i-1}^m)^2}{48h_i^m} \end{aligned} \quad (4.15)$$

The lower limit of the flux in (4.12) will be satisfied as long as:

$$\frac{48\alpha}{\tilde{\mu}} > \frac{(1 - \gamma_{i+1/2}^+)^2 (h_i^* - h_{i-1}^m)^2}{h_i^m h_{i+1}^m}. \quad (4.16)$$

The right hand side of (4.16) is found to have a maximum value of less than 1.24 with the definition (4.11) of $\gamma_{i+1/2}^+$, so (4.16) is clearly satisfied for $\alpha = 0.25$, and the modified thickness advection scheme is positive definite when both velocities adjacent to a thickness grid point are in the same direction. It has been shown previously to be positive definite when the two velocities are away from the grid point. By satisfying (4.14), the scheme also guarantees that the scheme is positive definite when both neighboring velocities are toward a thickness grid point. In short, the modified scheme is positive definite for any arrangement of velocities and positive thicknesses.

Returning to spherical coordinates, the sufficient conditions for the thickness advection scheme to be positive definite are that, for all i and j :

$$\begin{aligned} \frac{|u_{i,j+1/2}| \Delta t}{a \cos \theta_{j+1/2} \Delta \phi} &< 0.4517, \\ \frac{|v_{i+1/2,j}| \cos \theta_j \Delta t}{a \cos \theta_{j+1/2} \Delta \phi} &< 0.4517, \text{ and } \frac{|v_{i+1/2,j+1}| \cos \theta_{j+1} \Delta t}{a \cos \theta_{j+1/2} \Delta \phi} < 0.4517. \end{aligned}$$

These conditions are generally less restrictive than the time-step and velocity conditions required for stability of the linear gravity waves.

5. Viscous Terms

There are several types of dissipation terms in the momentum equations of this model. Momentum is diffused isopycnally with a biharmonic viscosity. A weak Laplacian diapycnal viscosity is also

imposed. Finally, the velocity components are truncated to a maximum value in extremely rare instances.

In the discretized formulations for the biharmonic diffusion operator, care was taken to insure that all layer thicknesses in the denominators also appear in the numerators in every term. This biharmonic formulation uses a flux conservative form, and is formally only first order accurate in the horizontal grid spacing. This should not be a difficulty, since the value of the biharmonic diffusivity is usually decreased as the cube of the grid spacing. At the boundaries, one of two sets of boundary conditions are imposed. One set of boundary conditions subjects the tangential velocity to a free slip boundary condition and a hyper slip (zero normal third derivative) boundary condition and the normal velocity is zero at the boundary and has zero normal second derivative. The other set of boundary conditions imposes no slip boundary condition and a super slip (zero normal second derivative) boundary condition on the tangential velocity and the normal velocity is zero at the boundary and has zero normal gradient. The first set of boundary conditions is used in most of the simulations described in this work. The exact formulation of the biharmonic operator discretization is far too complicated to warrant reproducing here.

In addition to the horizontal momentum diffusion within each layer, there is also a diapycnal (vertical) momentum viscosity. The vertical diffusion is fully implicit, as it must be to allow layers to be arbitrarily thin. At an interface the zonal stress is given by:

$$\tau_{n+1/2}^{(u)} = \frac{(\rho_{n+1} + \rho_n)v_{n+1/2}(u_{n+1} - u_n)}{h_n^{(u)} + h_{n+1}^{(u)}}, \quad (5.1)$$

The stress at the upper surface is an externally specified wind stress. Near the surface a viscosity profile which is inversely proportional to the depth is used with this definition to prevent the wind stress from creating excessively large velocities in the thin layers at the surface. At the bottom, a no slip bottom boundary condition makes the stress:

$$\tau_{N+1/2}^{(u)} = -(\rho_N v_{N+1/2} u_N) / H_{N+1/2}^-. \quad (5.2)$$

The definition of the layer thickness at velocity grid points is somewhat ambiguous. Two separate definitions are used in the model: a reduced thickness formulation:

$$h_{i,j+1/2,n}^{(u)} = h_{i,j+1/2,n}^{(\text{Reduced})} = \frac{2h_{i+1/2,j+1/2,n} h_{i-1/2,j+1/2,n}}{h_{i+1/2,j+1/2,n} + h_{i-1/2,j+1/2,n}}, \quad (5.3)$$

or an arithmetic mean formulation:

$${}^{(u)}_{i,j+1/2,n} = h_{i,j+1/2,n}^{(\text{Mean})} = \frac{h_{i+1/2,j+1/2,n} + h_{i-1/2,j+1/2,n}}{2}. \quad (5.4)$$

Both of these formulations are second order accurate. The meridional velocity point definitions are analogous. For greatly different magnitudes of neighboring thicknesses, the reduced thickness formulation gives 2 times the smaller thickness, while the arithmetic mean formulation gives 1/2 the larger thickness. Within 2 benthic boundary layer thicknesses (H_{BBL}) of the bottom (measured by the reduced thicknesses), the arithmetic mean formulation is used if the velocity is going from the thicker side to the thinner, and the reduced thickness formulation is used if the velocity is from the thinner side. Above this point the reduced thickness formulation is always used. This choice was made to compensate for artificially high Bernoulli functions where the thin meniscus of fluid extends over sloping topography. This formulation, along with the no-slip bottom boundary condition, allows a thin layer to inflate while keeping the artificial Bernoulli function gradients from generating spurious currents at the base of the slope. The kinematic viscosity is free to vary in the vertical. In the model, the kinematic viscosity is set to a high value within a specified mixed layer thickness (H_{ML}) of the surface, and this decreases quadratically in depth to the much smaller interior value at $2H_{ML}$ from the surface. Similarly, a high value of kinematic viscosity is used within H_{BBL} of the bottom, that decreases to the interior value by $2H_{BBL}$. If a point is within the influence of both the surface and the bottom, the largest viscosity is used. Typical values for the thickness are 25 m for H_{ML} and 10 m for H_{BBL} . Typical viscosities are $1 \text{ m}^2 \text{ s}^{-1}$ in the mixed layer, $10^{-4} \text{ m}^2 \text{ s}^{-1}$ in the benthic boundary layer, and $10^{-6} \text{ m}^2 \text{ s}^{-1}$ in the interior.

Finally, if the magnitude of any component of velocity exceeds some maximum upper bound, U_{\max} , that component of velocity is truncated to U_{\max} in the same direction. This is *not* meant to be a form of dissipation, but rather a means of keeping the model running when it would otherwise crash. The number of instances of this truncation occurring is reported, and is usually 0. This safety catch is quite effective in helping diagnose model failures by keeping the model running.

6. Time Discretization and Operator Splitting

There are two separate versions of the model that use different time stepping schemes. One scheme steps the entire equations using the same time steps. It is easy to ascertain the stability and conservation properties of this scheme, but it is relatively slow. The time step for this scheme is set by the speed of the barotropic gravity waves, which are much faster than any of the other signals in this system. The other time stepping scheme splits the equations into a simple, rapidly evolving subset which

contains the barotropic gravity waves and a remainder that evolves more slowly. The barotropic subset takes many small time steps for every time step taken by the full equations. This scheme allows the model to be integrated in time much more rapidly than the other scheme, but the stability and exact conservation properties are much more difficult to ascertain.

With the non-split explicit scheme, the momentum equations are time stepped using a third order accurate scheme proposed by Matsuno (1966). Three evaluations of the time derivative are required with this scheme, but it is stable for a CFL number up to $\sqrt{3}$, which is more efficient than either a simulated backward Euler or leapfrog time stepping scheme. As an added benefit, this scheme weakly dissipates high frequency signals. All of the high frequency motions are relatively short gravity waves, which are not usually of interest. The thickness equation is stepped alternately with the integer and half integer time step velocities in a leapfrog pattern. There is no possibility of any splitting instability, and the overall scheme is second order accurate. If the layer momentum equations, (2.6), are schematically written as:

$$\frac{\partial \mathbf{u}}{\partial t} = G(\mathbf{u}, h, (\mathbf{u}h)) \quad (6.1)$$

and the volume integrated potential density conserving layer thickness equation, (2.3), is schematically written as:

$$\frac{\partial h}{\partial t} = J(\mathbf{u}, h) , \quad (6.2)$$

the non-split explicit scheme can be written as:

$$\begin{aligned} h^{m+1/4} &= h^{m-1/4} + (\Delta t/2)J(\mathbf{u}^m, h^{m-1/4}) , \\ \mathbf{u}^{m+1/3} &= \mathbf{u}^m + (\Delta t/3)G(\mathbf{u}^m, (h^{m-1/4} + h^{m+1/4})/2, (\mathbf{u}h)^m) , \\ \mathbf{u}^{m+1/2} &= \mathbf{u}^m \\ &\quad + (\Delta t/2)G(\mathbf{u}^{m+1/3}, (7h^{m+1/4} - h^{m-1/4})/6, [5(\mathbf{u}h)^m - 2(\mathbf{u}h)^{m-1/2}]/3) , \\ h^{m+3/4} &= h^{m+1/4} + (\Delta t/2)J(\mathbf{u}^{m+1/2}, h^{m+1/4}) , \\ \mathbf{u}^{m+1} &= \mathbf{u}^m + (\Delta t)G(\mathbf{u}^{m+1/2}, (h^{m+3/4} + h^{m+1/4})/2, (\mathbf{u}h)^{m+1/2}) , \end{aligned}$$

where the $(\mathbf{u}h)$ are determined as a part of the thickness time steps. The appropriate rate of diapycnal advection of momentum and diapycnal viscosity are applied immediately after each momentum time step.

Barotropic gravity waves are resolved by this model. With the non-split-explicit scheme, the time step is typically extended to a reasonable value by reducing the gravitational restoring force of the free surface to only 10 times the sum of the internal interfaces' reduced gravities. This retards barotropic gravity wave speeds to about 12 times the baroclinic wave speeds. Tests comparing simulations with the actual restoring force on the free surface with simulations with the weaker restoring force show no significant effect of the smaller free surface reduced gravity for many simulations. This approach is the opposite of applying a rigid lid, but is in much the same spirit.

The following description of the split-explicit time stepping scheme is essentially an exact reproduction of Hallberg (1997). The notation, unfortunately, does not conform with that of the rest of this manuscript. Instead, Bleck and Smith's (1990) notation of pressure thicknesses (Δp) replaces my previous notation for the layer thicknesses (h). The description which follows does not make the Boussinesq approximation - there are slight differences in the splitting if the Boussinesq approximation is made.

6.1 Introduction

The primitive equations used in numerical simulations of the ocean circulation are frequently split into a rapidly evolving, simple set of equations describing surface gravity waves and the more slowly evolving remainder. This approach can allow a great increase in the efficiency of numerical simulations of the large scale ocean circulation. Unfortunately, this splitting cannot in general be done exactly. Higdon and Bennett (1996) have recently shown that the split time stepping scheme of Bleck and Smith (1990) is unstable at all wavenumbers due to coupling between external and internal gravity wave modes. This note derives a family of split explicit time stepping schemes that is stable at all wavelengths, subject to a CFL condition based on the internal gravity wave speed and the Coriolis parameter. With these schemes, internal gravity waves and inertial oscillations are subject to frequency dependent damping that is controlled by the values of two free parameters.

The ocean has many dynamically important timescales, from millennial variations of the thermohaline circulation to the rapid timescales of surface gravity waves. Large-scale ocean circulation models typically eliminate the fastest oscillations through use of the primitive equations. Sound waves are removed by assuming that flow is incompressible, while the hydrostatic approximation to the vertical momentum equation eliminates the need to solve a three-dimensional elliptic equation for pressure or resolve the timescales associated with vertically propagating gravity waves. The fastest

remaining timescales are associated with horizontal propagation of external gravity waves, with a speed of about \sqrt{gD} where g is the gravitational acceleration and D is the depth of the ocean. In the deep ocean, this speed is typically on the order of 225 m/s. The next fastest timescales are associated with internal gravity wave propagation or horizontal advection, both with speeds of order a few m/s. There is a strong incentive to use a time stepping scheme with time steps that are determined by the slower internal timescales, rather than by the fast timescales of external gravity waves.

There is a long history of removing the external gravity wave modes in ocean models by replacing the free surface with a rigid lid. This effectively makes the external gravity waves infinitely fast, and the effect of the external gravity waves is reproduced by solving a two-dimensional elliptic equation at every time step. The velocities associated with the external mode in this case are exactly vertically uniform, and it is easy to exactly split the internal and external gravity wave modes. While this approach allows a numerical simulation to take long time steps based on the internal timescales, with irregular bathymetry or coastlines or with large numbers of islands relaxation methods can converge slowly and a large fraction of the computer time is spent solving the two-dimensional elliptic equation. Also, a rigid lid distorts the properties of large-scale barotropic Rossby waves and complicates inclusion of a fresh water flux surface boundary condition and assimilation of sea surface height data.

Another approach is to eliminate the gravity waves altogether by using the simplified quasigeostrophic equations. However, the assumptions leading to the quasigeostrophic equations do not hold for many interesting cases, such as flow over large amplitude topography or when isopycnals outcrop into the surface mixed layer.

A popular alternate approach, used by Bleck and Smith (1990) with an isopycnal coordinate ocean model, is to split the governing equations into a simplified two-dimensional set of equations describing the evolution of the external gravity wave field and a remainder that evolves more slowly. The simple external equations can rapidly be integrated over many time-steps, while the much more complicated three dimensional remainder can safely take long time steps dictated by the internal dynamics. A similar split time stepping scheme has been used for a depth-coordinate ocean model by Killworth et al. (1991) and Dukowicz and Smith (1994) and for a terrain-following coordinate model by Blumberg and Mellor (1987). Unfortunately, without a rigid lid the velocities associated with the external modes are not quite vertically uniform, and it is not possible to split the equations exactly into external and internal modes. Higdon and Bennett (1996) demonstrate that the Bleck and Smith (1990) time stepping scheme for a two-layer system is linearly unstable at all wavelengths for any size time

step. While Higdon and Bennett (1996) demonstrate that this instability is due to an inexact splitting between the internal and external modes, they do not offer a stable split time-stepping scheme. De Szoeke and Higdon (1997) propose an alternative split time stepping scheme that is stable at most wavelengths, but this scheme is also unstable at some wavelengths.

The present paper offers a family of split time stepping schemes that are stable at all wavelengths for time steps that resolve the inertial frequency and satisfy a CFL criterion based on the internal gravity wave speed. Nonrotating internal gravity waves are neutrally stable for one member of this family and subject to frequency dependent dissipation for the other members of the family. Inertial oscillations are weakly damped with these schemes, subject to the control of a free parameter. The next section describes the proposed separation of the rapidly evolving barotropic equations from the more slowly evolving vertically varying equations. In section 3, the stability of a large number of candidate schemes is assessed for linear, nonrotating two-layer flow, sufficient conditions for the stability of a scheme are derived, and a stable scheme is described. A stable treatment of the Coriolis terms for the proposed scheme is described in section 4. Section 5 presents a nonlinear time stepping scheme for the primitive equations that is consistent with the stable linear scheme described here.

6.2 *A new proposed split time stepping scheme*

Higdon and Bennett (1996) identify an inexact splitting into barotropic and baroclinic modes and interaction between the split modes as the cause of the instability in the Bleck and Smith (1990) split time stepping scheme. The barotropic mode in the Bleck and Smith scheme does not resolve the rapid variations of the vertical average of several of the linear terms in the momentum equations. Here we propose a different approach to splitting the time integration that efficiently includes all of the rapid variations in the vertical average of the linear momentum terms in the two-dimensional barotropic equations.

An appropriate starting point is the inviscid, unforced primitive equations in isopycnal layers:

$$\frac{\partial \mathbf{u}_n}{\partial t} + \mathbf{u}_n \cdot \nabla \mathbf{u}_n + f \hat{\mathbf{k}} \times \mathbf{u}_n = -\nabla M_n, \quad (6.3)$$

$$\frac{M_{n+1} - M_n}{\alpha_{n+1} - \alpha_n} = p_{n+1/2}, \text{ and} \quad (6.4)$$

$$\frac{\partial}{\partial t}(\Delta p_n) + \nabla \cdot (\mathbf{u}_n \Delta p_n) = 0 \quad (6.5)$$

where \mathbf{u} is the horizontal velocity, f is the Coriolis parameter, $\hat{\mathbf{k}}$ is a vertical unit vector, p is pressure, α is the specific volume (the inverse of density), and $M \equiv \alpha p + gz$ is the Montgomery potential. The subscripts indicate the layer with an index that increases downward. A half-integer subscript indicates the value at the interface between layers. $\Delta p_n = p_{n+1/2} - p_{n-1/2}$ is the (positive) pressure thickness of a layer. The specific volume, α , is a constant within each layer, but \mathbf{u} , M , and p vary in the horizontal. See Higdon and Bennett (1996) or Bleck and Smith (1990) for a more detailed discussion of these equations.

The Montgomery potential can be expressed directly for each layer (and is vertically constant within each layer, from the hydrostatic equation), but is most easily found by integrating from a known value. At the sea surface an atmospheric pressure is assumed, but the sea surface height must also be calculated. At the sea floor, the height relative to mean sea level, $-D$, is known and the pressure is easily calculated by summing the pressure thicknesses over all the layers. The Montgomery potential for each layer is then given by

$$M_n = \alpha_N p_B - gD + \sum_{i=n}^{N-1} (\alpha_i - \alpha_{i+1}) p_{i+1/2}, \quad (6.6)$$

where N is the total number of layers and

$$p_B = \sum_{n=1}^N \Delta p_n$$

is the bottom pressure.

Following Higdon and Bennett (1996), define the barotropic velocities as the mass weighted vertical average velocity

$$\bar{\mathbf{u}} = \frac{1}{p_B} \sum_{n=1}^N \mathbf{u}_n \Delta p_n. \quad (6.7)$$

The sum of the layer continuity equations then gives an evolution equation for the bottom pressure,

$$\frac{\partial}{\partial t} p_B + \nabla \cdot (\bar{\mathbf{u}} p_B) = 0. \quad (6.8)$$

The mass weighted vertical average of the horizontal momentum equations gives

$$\frac{\partial \bar{\mathbf{u}}}{\partial t} + f \hat{\mathbf{k}} \times \bar{\mathbf{u}} + \overline{\nabla M} = -\frac{1}{p_B} \sum_{n=1}^N \Delta p_n (\mathbf{u}_n \cdot \nabla \mathbf{u}_n) + \sum_{n=1}^N \mathbf{u}_n \frac{\partial}{\partial t} \left(\frac{\Delta p_n}{p_B} \right). \quad (6.9)$$

The nonlinear momentum advection term on the right hand side of this equation, perhaps along with forcing or viscous terms, is treated as a constant over the integration of the barotropic equations. The evolution of the fast external gravity waves is described by (6.8) and (6.9); all of the linear terms that vary with the rapid gravity wave time scale appear on the left hand side of these equations.

The barotropic Montgomery potential gradients are given by

$$\begin{aligned}
 \overline{\nabla M} &= \nabla(\alpha_N p_B - gD) + \frac{1}{p_B} \sum_{n=1}^N \left\{ \Delta p_n \nabla \left[\sum_{i=n}^{N-1} (\alpha_i - \alpha_{i+1}) p_{i+1/2} \right] \right\} \quad . \quad (6.10) \\
 &= \nabla(\alpha_N p_B - gD) + \frac{1}{p_B} \sum_{n=1}^{N-1} [(\alpha_n - \alpha_{n+1}) p_{n+1/2} \nabla p_{n+1/2}] \\
 &= \nabla(\alpha_N p_B - gD) + \sum_{n=1}^{N-1} \left\{ (\alpha_n - \alpha_{n+1}) \left(\frac{p_{n+1/2}}{p_B} \right) \nabla \left[p_B \left(\frac{p_{n+1/2}}{p_B} \right) \right] \right\}
 \end{aligned}$$

Since the velocities associated with long external gravity waves are nearly uniform with depth, to good approximation the ratios of the interface pressures to the bottom pressure do not vary rapidly with external gravity wave time scales. In the final expression in (6.10), only the p_b in the numerators are assumed to vary rapidly with time, so (6.10) is a linear function of p_b . This is one of the important differences from the scheme of Bleck and Smith (1990), which treats the entire summation in (6.10) as a constant over the barotropic integration. The addition or removal of massless layers at the bottom does not affect the value of $\overline{\nabla M}$, even though it changes α_N . De Szoeke and Higdon (1997) suggest that inclusion of all of the rapid variations of the Montgomery potential gradient in the barotropic integration is essential for limiting destabilizing interactions between the barotropic and baroclinic modes. The present work supports this suggestion.

The layer momentum equation can now be written as

$$\frac{\partial \mathbf{u}_n}{\partial t} = \frac{\partial \bar{\mathbf{u}}}{\partial t} - f \hat{\mathbf{k}} \times (\mathbf{u}_n - \bar{\mathbf{u}}) - (\nabla M_n)' - \mathbf{u}_n \cdot \nabla \mathbf{u}_n + \frac{1}{p_B} \sum_{n=1}^N \Delta p_n (\mathbf{u}_n \cdot \nabla \mathbf{u}_n) , \quad (6.11)$$

where the perturbation Montgomery potential gradients are given by

$$\begin{aligned}
(\nabla M_n)' &= \nabla M_n - \overline{\nabla M} \\
&= \nabla \left[\sum_{i=n}^{N-1} (\alpha_i - \alpha_{i+1}) \left(\frac{P_{i+1/2}}{P_B} \right) p_B \right] \\
&\quad - \sum_{n=1}^N \left\{ (\alpha_n - \alpha_{n+1}) \left(\frac{P_{n+1/2}}{P_B} \right) \nabla \left[\left(\frac{P_{n+1/2}}{P_B} \right) p_B \right] \right\}
\end{aligned} \tag{6.12}$$

In (6.12) the ratios of an interface's pressure to the bottom pressure will not vary much with the rapid external gravity wave time scales, while the bottom pressures in the numerators are taken from the integration of the barotropic equations, (6.8) and (6.9). The layer continuity equations are modified to insure that the time average barotropic velocity equals the vertical average of the velocities that are used to step the layer thicknesses:

$$\frac{\partial}{\partial t} (\Delta p_n) + \nabla \cdot (\mathbf{u}_n \Delta p_n) + \nabla \cdot \left\{ \left[\widehat{\mathbf{u}} - \sum_{i=1}^N \mathbf{u}_i \left(\frac{\Delta p_i}{P_B} \right) \right] \Delta p_n \right\} = 0, \tag{6.13}$$

where $\widehat{\mathbf{u}}$ is the time average of the barotropic velocities. This serves to filter from the layer continuity equations those external gravity waves with unresolvably high frequencies. The mass in each layer is easily conserved with this time splitting; with the Bleck and Smith (1990) splitting only the total mass in all the layers is conserved.

6.3 Linear Stability Analysis

To facilitate a Von Neumann stability analysis, we consider only the two-layer, flat-bottom case and introduce several definitions to make the linearization clear. Let $\bar{\mathbf{u}} = (\Delta p_1 \mathbf{u}_1 + \Delta p_2 \mathbf{u}_2) / (\Delta p_1 + \Delta p_2)$, $\mathbf{u}'_1 = \mathbf{u}_1 - \bar{\mathbf{u}}$, $p_B = \tilde{p}_B (1 + \eta) = (\Delta \tilde{p}_1 + \Delta \tilde{p}_2) (1 + \eta)$, $\Delta p'_1 = -\Delta \tilde{p}_1 + \Delta p_1 / (1 + \eta)$, and $\Delta p'_2 = -\Delta \tilde{p}_2 + \Delta p_2 / (1 + \eta)$, where the variables with tildes are the constant pressure thicknesses that the layers would have at rest. The linearized versions of (6.8), (6.9), (6.11), and (6.13) can be written as

$$\frac{\partial \eta}{\partial t} + \nabla \cdot \bar{\mathbf{u}} = 0, \tag{6.14}$$

$$\frac{\partial \bar{\mathbf{u}}}{\partial t} + f \hat{\mathbf{k}} \times \bar{\mathbf{u}} + c_0^2 \nabla \eta = -\Gamma c_0^2 \nabla (\Delta p'_1 / \Delta \tilde{p}_1), \tag{6.15}$$

$$\frac{\partial \mathbf{u}'_1}{\partial t} = -f \hat{\mathbf{k}} \times \mathbf{u}'_1 - c_1^2 \nabla (\eta + \Delta p'_1 / \Delta \tilde{p}_1), \text{ and} \tag{6.16}$$

$$\frac{\partial}{\partial t}(\Delta p'_1/\Delta \tilde{p}_1) = -\nabla \cdot \mathbf{u}'_1, \quad (6.17)$$

where $c_0^2 \equiv \alpha_2 \tilde{p}_B + \Delta \alpha (\Delta \tilde{p}_1)^2 / \tilde{p}_B$, $c_1^2 \equiv \Delta \alpha \Delta \tilde{p}_1 \Delta \tilde{p}_2 / \tilde{p}_B$, and $\Gamma \equiv (\Delta \tilde{p}_1 / \Delta \tilde{p}_2)(c_1^2 / c_0^2)$.

The Coriolis parameter complicates the analysis of the schemes, so in the interest of brevity assume that $f=0$, which effectively eliminates one of the horizontal dimensions and the velocity in that direction from the problem. A stable treatment of the Coriolis parameter is described in the next section.

It can now be assumed that the wavevector is in the x -direction without loss of generality. Assuming periodic boundary conditions, the wavenumbers that can be represented on a discrete C-grid are quantized by

$$k = [4\sin^2(\pi m/M)/\Delta x^2 + 4\sin^2(\pi n/N)/\Delta y^2]^{1/2}, \quad (6.18)$$

where m and n are integers from $(-M/2$ to $M/2)$ and $(-N/2$ to $N/2)$ and M and N are the number of grid-points in the x - and y -directions. (On a C-grid the zonal velocities are displaced half a grid point to the east of the thickness grid points, while the meridional velocities are displaced half a grid point to the north of the thickness points.) The largest wavenumber that can be resolved is

$$k_{\max} \equiv 2(1/\Delta y^2 + 1/\Delta x^2)^{1/2}, \quad (6.19)$$

If (6.14) through (6.17) are Fourier transformed in space, the finite difference representation of the gradients can be replaced by their Fourier transform, $(ik\hat{x})$.

A large family of time stepping schemes for (6.14) through (6.17) based on a generalization of the forward-backwards scheme can now be explored (see Haltiner and Williams (1980) for a description of the forwards-backwards scheme). Following Higdon and Bennett (1996), the left hand side of the barotropic equations, (6.14) and (6.15), are simultaneously integrated analytically. If the superscript represents the time level of a variable and an asterisk denotes the result of a predictor time step, the discrete Fourier transformed equations become

$$i\eta^* = i\eta^n \cos T + (\bar{u}^n/c_0) \sin T - i\Gamma(1 - \cos T)(\Delta p'_1/\Delta \tilde{p}_1)^n, \quad (6.20)$$

$$(\bar{u}/c_0)^* = (\bar{u}/c_0)^n \cos T - i\eta^n \sin T - i\Gamma \sin T (\Delta p'_1/\Delta \tilde{p}_1)^n, \quad (6.21)$$

$$(u'_1/c_1)^* = (u'_1/c_1)^n - i\tau[(1 - \alpha)\eta^n + \alpha\eta^*] - i\tau(\Delta p'_1/\Delta \tilde{p}_1)^n, \quad (6.22)$$

$$i(\Delta p'_1/\Delta \tilde{p}_1)^* = i(\Delta p'_1/\Delta \tilde{p}_1)^n + \tau[(1 - \beta)(u'_1/c_1)^n + \beta(u'_1/c_1)^*], \quad (6.23)$$

$$i\eta^{n+1} = i\eta^n \cos T + (\bar{u}^n/c_0) \sin T, \quad (6.24)$$

$$-(i\Gamma)(1 - \cos T)[(1 - \gamma)(\Delta p'_1/\Delta \tilde{p}_1)^n + \gamma(\Delta p'_1/\Delta \tilde{p}_1)^*]$$

$$(\bar{u}/c_0)^{n+1} = (\bar{u}/c_0)^n \cos T - i\eta^n \sin T, \quad (6.25)$$

$$-i\Gamma \sin T[(1 - \gamma)(\Delta p'_1/\Delta \tilde{p}_1)^n + \gamma(\Delta p'_1/\Delta \tilde{p}_1)^*]$$

$$(u'_1/c_1)^{n+1} = (u'_1/c_1)^n - i\tau[(1 - \delta)\eta^n + \delta\eta^{n+1}], \quad \text{and} \quad (6.26)$$

$$-i\tau[(1 - \zeta)(\Delta p'_1/\Delta \tilde{p}_1)^n + \zeta(\Delta p'_1/\Delta \tilde{p}_1)^*]$$

$$i(\Delta p'_1/\Delta \tilde{p}_1)^{n+1} = i(\Delta p'_1/\Delta \tilde{p}_1)^n + \tau[(1 - \mu)(u'_1/c_1)^n + \mu(u'_1/c_1)^{n+1}]. \quad (6.27)$$

These equations use the definitions $T \equiv c_0 k \Delta t$ and $\tau \equiv c_1 k \Delta t$. Any interpolation between the old and latest values of the variables can be used through the choice of the constants α , β , γ , δ , ζ and μ . Inclusion of six free parameters significantly complicates these equations, but a stability analysis with these free parameters reveals the stable two-step predictor-corrector time stepping schemes that evaluate equations (6.14) through (6.17) in the order given here. The full range of these parameters have been explored, but in the interest of simpler expressions, the results presented here set $\gamma = 0$ and $\alpha = \delta$. The stable scheme suggested later is found by setting $\alpha = \delta = 1/2$, $\gamma = 0$, $\mu = 1$, with β and ζ left as free parameters.

Matrix notation can be used to rewrite (6.20) through (6.27) with $\gamma = 0$ and $\alpha = \delta$ as

$$\begin{bmatrix} 1 & -\mu\tau & 0 & 0 \\ 0 & 1 & \delta\tau(1 - \beta\zeta\tau^2) & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} i(\Delta p'_1/\Delta \tilde{p}_1) \\ u'_1/c_1 \\ i\eta \\ \bar{u}/c_0 \end{bmatrix}^{n+1} \quad (6.28)$$

$$= \begin{bmatrix} 1 & (1-\mu)\tau & 0 & 0 \\ -\tau(1 - \beta\zeta\tau^2) & 1 - \zeta\tau^2 & -\tau(1 - \delta)(1 - \beta\zeta\tau^2) & 0 \\ -\Gamma(1 - \cos T) & 0 & \cos T & \sin T \\ -\Gamma \sin T & 0 & -\sin T & \cos T \end{bmatrix} \begin{bmatrix} i(\Delta p'_1/\Delta \tilde{p}_1) \\ u'_1/c_1 \\ i\eta \\ \bar{u}/c_0 \end{bmatrix}^n$$

The eigenvalues of the product of the inverse of the matrix on the left hand side of (6.28) and the matrix on the right hand side are given by the quartic equation for λ

$$[\lambda^2 - \lambda(2\cos T) + 1] \times \{\lambda^2 - \lambda[2 - (\zeta + \mu)\tau^2 + \beta\zeta\mu\tau^4] + [1 + (1 - \zeta - \mu)\tau^2 + \beta\zeta(1 - \mu)\tau^4]\} - \Gamma\tau^2(1 - \beta\zeta\tau^2)(1 - \cos T) \times [\lambda^3\delta\mu + \lambda^2(\delta + \mu - \delta\mu) + \lambda(1 - \delta\mu) + (1 - \delta - \mu + \delta\mu)] = 0 \quad (6.29)$$

The time stepping scheme is linearly stable if all of the roots of the quartic equation in (6.29) have magnitude less than or equal to 1.

To illustrate the desired properties of (6.29), consider the product of two quadratic equations with complex conjugate pairs of roots. If the roots have magnitudes $\|a\|$ and $\|b\|$ and real parts a_R and b_R , the product of the quadratic equations is:

$$\begin{aligned} & (\lambda^2 - 2a_R\lambda + \|a\|^2)(\lambda^2 - 2b_R\lambda + \|b\|^2) \\ &= \lambda^4 - 2(a_R + b_R)\lambda^3 + (\|a\|^2 + \|b\|^2 + 4a_R b_R)\lambda^2 \\ & \quad - 2(a_R\|b\|^2 + \|a\|^2 b_R)\lambda + \|a\|^2\|b\|^2 = 0 \end{aligned} \quad (6.30)$$

Now suppose that a quartic equation cannot be factored exactly, but can be written as

$$\begin{aligned} & \lambda^4 - C_3\lambda^3 + C_2\lambda^2 - C_1\lambda + C_0 \\ &= \lambda^4 - 2(\tilde{a}_R + \tilde{b}_R)\lambda^3 + (\|\tilde{a}\|^2 + \|\tilde{b}\|^2 + 4\tilde{a}_R\tilde{b}_R + \varepsilon)\lambda^2 \\ & \quad - 2(\tilde{a}_R\|\tilde{b}\|^2 + \|\tilde{a}\|^2\tilde{b}_R)\lambda + \|\tilde{a}\|^2\|\tilde{b}\|^2 \\ &= (\lambda^2 - 2\tilde{a}_R\lambda + \|\tilde{a}\|^2)(\lambda^2 - 2\tilde{b}_R\lambda + \|\tilde{b}\|^2) + \varepsilon\lambda^2 = 0 \end{aligned} \quad (6.31)$$

where ε is a small perturbation. The perturbed roots are found by starting from the unperturbed roots, and following the roots as the perturbation is added to the coefficient of λ^2 , but the coefficients of λ^3 , λ , and λ^0 (C_3 , C_1 , and C_0) are unchanged.

First consider the case where $C_0=1$, $C_3=C_1$ and all four roots of the unperturbed case have unit magnitude. Assuming that the perturbed roots still have unit magnitude, the real part of the perturbed roots must satisfy

$$a_R + b_R = \tilde{a}_R + \tilde{b}_R \quad (6.32)$$

and

$$a_R b_R = \tilde{a}_R \tilde{b}_R + \frac{\varepsilon}{4}, \quad (6.33)$$

where the tildes mark the unperturbed roots, while the perturbed roots have no tildes. The solutions are

$$a_R, b_R = \frac{1}{2} \left\{ \tilde{a}_R + \tilde{b}_R \pm [(\tilde{a}_R - \tilde{b}_R)^2 - \varepsilon]^{1/2} \right\}. \quad (6.34)$$

The solutions to (6.34) are inconsistent with our assumption that the roots have unit magnitude either if they are complex or if they have magnitude greater than 1. A positive perturbation causes the (assumed) real part of the perturbed solutions to become complex when the real parts of the unperturbed roots coincide. When the solutions to (6.34) are complex, one pair of the perturbed roots will

have magnitude greater than 1, and the time stepping scheme will be unstable. The solutions to (6.34) are always real with a negative perturbation, and a negative perturbation does not cause the roots to have a magnitude greater than 1 as long as

$$|\varepsilon| < 4(1 - \tilde{a}_R)(\tilde{a}_R - \tilde{b}_R) \text{ as } \tilde{a}_R \rightarrow 1 \text{ and } |\varepsilon| < 4(1 + \tilde{a}_R)|\tilde{a}_R - \tilde{b}_R| \text{ as } \tilde{a}_R \rightarrow -1 \quad (6.35)$$

with similar conditions as \tilde{b}_R approaches 1 or -1. (By assumption $|\tilde{a}_R| \leq 1$ and $|\tilde{b}_R| \leq 1$.)

Now consider the case when $C_0 < 1$ in (6.31) (that is, the product of the magnitude of the roots is less than 1). The effect of the perturbation in (6.31) can be determined by looking at $dC_2/d(\|b\|^2)$ when the coefficients of λ^3 , λ , and λ^0 are the constants C_3 , C_1 , and C_0 . These three constraints require that

$$\|a\|^2 = \frac{C_0}{\|b\|^2}, \quad a_R = \frac{C_3 C_0 - C_1 \|b\|^2}{2C_0 - 2\|b\|^4}, \quad \text{and } b_R = \frac{C_1 \|b\|^2 - C_3 \|b\|^4}{2C_0 - 2\|b\|^4}.$$

These expressions can be substituted into the expression for C_2 to give

$$C_2 = \|a\|^2 + \|b\|^2 + 4a_R b_R = \frac{C_0}{\|b\|^2} + \|b\|^2 + \frac{C_3 C_1 C_0 \|b\|^2 - (C_3^2 C_0 + C_1^2) \|b\|^4 + C_3 C_1 \|b\|^6}{(C_0 - \|b\|^4)^2}. \quad (6.36)$$

Taking the derivative of (6.36) with $(\|b\|^2)$ gives

$$\frac{dC_2}{d(\|b\|^2)} = -\frac{C_0}{\|b\|^4} + 1 + \frac{[C_3 C_1 (C_0^2 + 6C_0 \|b\|^4 + \|b\|^8) - 2(C_1^2 + C_3^2 C_0)(C_0 \|b\|^2 + \|b\|^6)]}{(C_0 - \|b\|^4)^3}. \quad (6.37)$$

Introducing the definition $\xi = -1 + \|b\|^2/C_0^{1/2}$ [or $\|b\|^2 = C_0^{1/2}(1 + \xi)$], (6.37) becomes

$$\frac{dC_2}{d(\|b\|^2)} = \frac{\xi(2 + \xi)}{(1 + \xi)^2} - \frac{1}{C_0^3 \xi^3 (2 + \xi)^3} \times \quad (6.38)$$

$$[C_3 C_1 C_0^2 (8 + 16\xi + 12\xi^2 + 4\xi^3 + \xi^4) - C_0^{3/2} (C_1^2 + C_3^2 C_0) (4 + 8\xi + 6\xi^2 + 2\xi^3)]$$

$$= \frac{\xi(2 + \xi)}{(1 + \xi)^2} + \frac{(C_1 - C_0^{1/2} C_3)^2 (4 + 8\xi + 6\xi^2 + 2\xi^3)}{C_0^{3/2} \xi^3 (2 + \xi)^3} - \frac{C_3 C_1 \xi}{C_0 (2 + \xi)^3}$$

$$= \xi \left\{ \frac{(16 - C_3 C_1 / C_0)(1 + \xi)^2 + 8\xi^2 + 8\xi^3 + \xi^4}{(1 + \xi)^2 (2 + \xi)^3} + \frac{(C_1 - C_0^{1/2} C_3)^2 [4(1 + \xi)^2 + 2\xi^2 (1 + \xi)]}{C_0^{3/2} \xi^4 (2 + \xi)^3} \right\}$$

Some manipulation is required to show that the numerator of the first term inside of the braces in the final expression of (6.38) is always positive. Note that

$$\xi = -1 + \|b\|^2/C_0^{1/2} = -1 + \|b\|^2/(\|a\|\|b\|) = \|b\|/\|a\| - 1. \quad (6.39)$$

From the definitions of C_3 , C_1 , and C_0 ,

$$\begin{aligned} C_3 C_1 / C_0 &= \frac{4(a_R + b_R)(\|b\|^2 a_R + \|a\|^2 b_R)}{\|a\|^2 \|b\|^2} = 4 \left(\frac{\|a\| a_R}{\|b\| \|a\|} + \frac{b_R}{\|b\|} \right) \left(\frac{\|b\| a_R}{\|a\| \|a\|} + \frac{b_R}{\|b\|} \right), \quad (6.40) \\ &= 4 \left[\frac{a_R^2}{\|a\|^2} + \frac{b_R^2}{\|b\|^2} + \frac{a_R b_R}{\|a\| \|b\|} \left(\frac{\|a\|}{\|b\|} + \frac{\|b\|}{\|a\|} \right) \right] \leq 8 + 4 \left(\frac{1}{1 + \xi} + 1 + \xi \right) = 8 + 4 \frac{\xi^2 + 2\xi + 2}{1 + \xi} \end{aligned}$$

where it is assumed that $|a_R| \leq \|a\|$ and $|b_R| \leq \|b\|$ to be consistent with the assumption that complex conjugate pairs of roots are being perturbed. The numerator of the first term in the final expression of (6.38) can now be written as

$$\begin{aligned} (16 - C_3 C_1 / C_0)(1 + \xi)^2 + 8\xi^2 + 8\xi^3 + \xi^4 \\ \geq 8(1 + \xi)^2 - 4(\xi^2 + 2\xi + 2)(1 + \xi) + 8\xi^2 + 8\xi^3 + \xi^4 = 4\xi^2(1 + \xi) + \xi^4 \end{aligned} \quad (6.41)$$

Since by definition $\xi > -1$, both terms inside of the braces in the final expression of (6.38) are always positive. The derivative in (6.38) is positive if $\|b\| > \|a\|$ ($\xi > 0$) and negative if $\|b\| < \|a\|$ ($\xi < 0$). This means that a positive perturbation in the original quartic equation will increase the magnitude of the larger pair of roots (and decrease the magnitude of the smaller pair of roots) and could make the time stepping scheme unstable if the unperturbed roots are not sufficiently damped. A negative perturbation brings the magnitude of the roots together and makes all of the roots damping. The time stepping scheme is stable if the perturbation is always negative, the unperturbed roots are stable, and the perturbation goes to 0 when the imaginary part of either set of roots goes to 0.

No choice of parameters will make (6.29) exactly factorable independent of wavelength (T or τ). However, if there is no coupling term (involving Γ) in the λ^0 term of (6.29) and the coupling terms in the λ^3 and λ terms are equal, (6.29) can be approximately factored with an additional λ^2 term, as described in (6.31). [There is actually a less restrictive criterion for (6.29) to be approximately factored, and this leads to other stable schemes, but this restrictive criterion leads to a versatile scheme for which incorporation of the Coriolis terms is straightforward.] These restrictions require that $\delta = 1$ and $\mu = 1/2$ or that $\mu = 1$ and $\delta = 1/2$. With the second of these solutions ($\mu = 1$ and $\delta = 1/2$), (6.29) can be written as

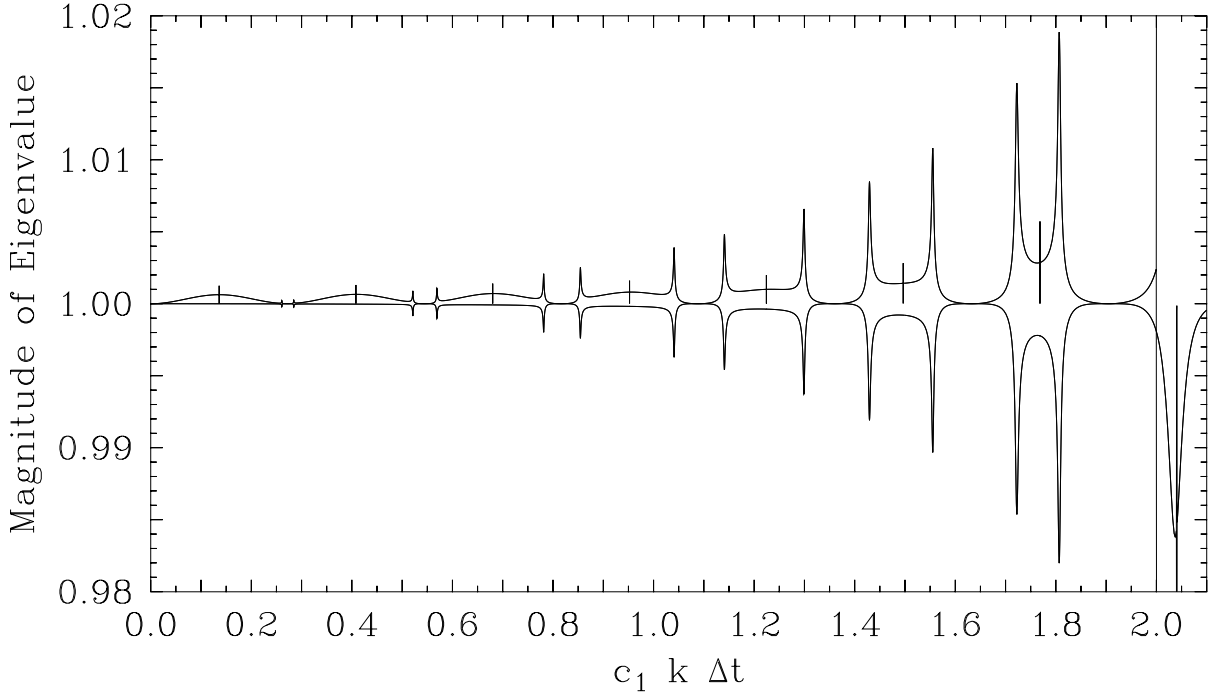


Figure 1. Magnitude of the four eigenvalues as a function of wavenumber for the Bleck and Smith (1990) time stepping scheme for two layers with $\Delta\alpha/\alpha_2 = 0.01$, $\Delta\tilde{p}_1/\tilde{p}_{Bot} = 0.25$, and $\Delta\tilde{p}_2/\tilde{p}_{Bot} = 0.75$. Only two lines are visible at most wavelengths because the roots are in complex conjugate pairs. Only positive wavenumbers are shown because the curves are symmetric about 0. Instability results when the magnitude of an eigenvalue is greater than 1.

$$\begin{aligned}
 & [\lambda^2 - \lambda(2\cos T) + 1] \\
 & \times \{ \lambda^2 - \lambda[2 - (\zeta + 1)\tau^2 + \beta\zeta\tau^4 + (\Gamma\tau^2/2)(1 - \beta\zeta\tau^2)(1 - \cos T)] + [1 - \zeta\tau^2] \} \\
 & - \lambda^2\Gamma\tau^2(1 - \beta\zeta\tau^2)(1 - \cos^2 T) = 0
 \end{aligned} \tag{6.42}$$

With this scheme, the perturbation to the exact factoring is negative as long as $\tau^2 \leq 1/(\beta\zeta)$, and goes to zero when the real part of the unperturbed barotropic roots goes to 1 or -1, meeting the requirements for all of the roots to be stable. (The baroclinic roots are always complex in the range of interest.) This scheme is stable at all wavenumbers up to a limit determined by a baroclinic gravity wave CFL condition.

The stability of the proposed family of schemes is most clearly demonstrated by plotting the magnitude of the four eigenvalues as a function of wavenumber. The four roots of (6.29) are calculated from the analytic expression for the roots of a quartic equation and have been checked for accuracy as described by Abramowitz (1965). Very high resolution in wavenumber is used when two roots are near each other, insuring that all spikes have been found. A stability plot of the Bleck and Smith (1990)

scheme is shown in Fig. 1. There are broad regions where this scheme is unstable, as evidenced by eigenvalues with magnitude greater than one, as well as spikes of greater instability where the baroclinic and barotropic roots or the two barotropic roots nearly coincide. The vertical line at $c_1 k \Delta t = 2$ occurs where the unperturbed baroclinic roots become stable and unstable real roots.

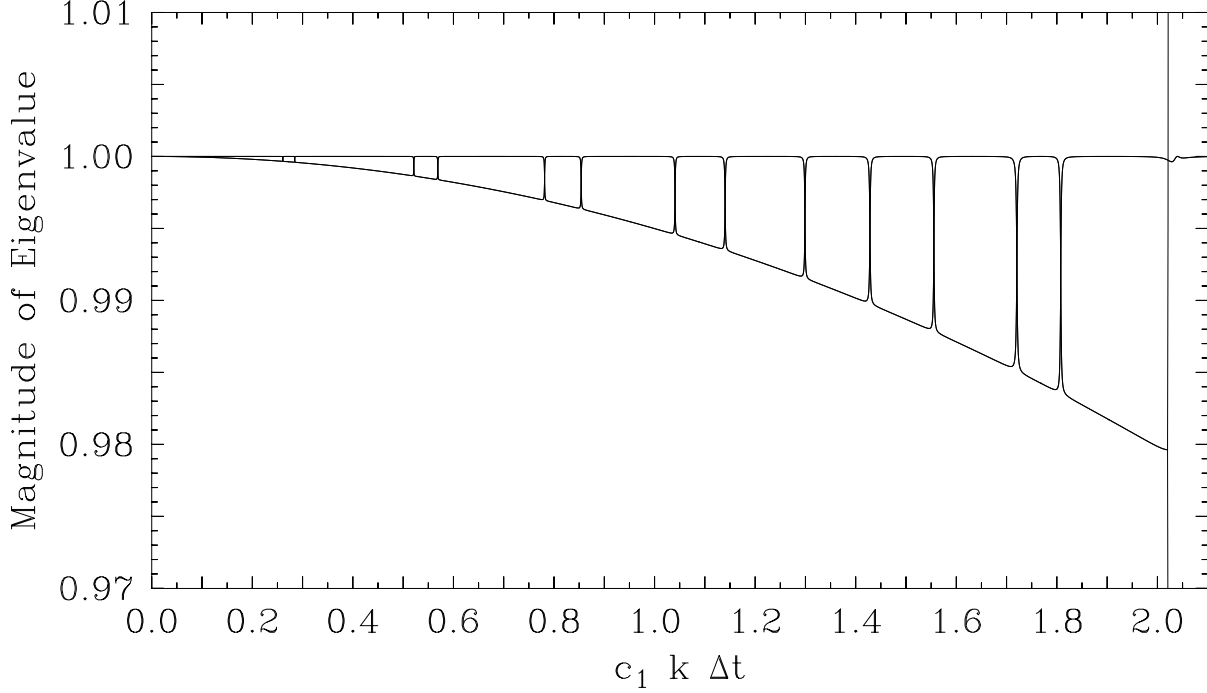


Figure 2. Magnitude of the four eigenvalues for the scheme described in (6.20) through (6.27) with $\delta = 0.5$, $\zeta = 0.01$, $\mu = 1$, and $\beta = 1$. As in Fig. 1, there are two layers with $\Delta\alpha/\alpha_2 = 0.01$, $\Delta\tilde{p}_1/\tilde{p}_{Bot} = 0.25$, and $\Delta\tilde{p}_2/\tilde{p}_{Bot} = 0.75$. With these parameters this scheme is linearly stable for all time steps up to $2.02/c_1 k_{max}$.

With a weakly dissipative member of the proposed family of time stepping schemes, all wavenumbers are stable up to a critical wavenumber, as seen in Fig. 2. Without the coupling terms (i.e. if $\Gamma=0$), (6.29) becomes:

$$[\lambda^2 - \lambda(2 \cos T) + 1] \times \{\lambda^2 - \lambda[2 - (\zeta + 1)\tau^2 + \beta\zeta\tau^4] + [1 - \zeta\tau^2]\} = 0. \quad (6.43)$$

All of the roots of (6.43) are complex and have magnitude less than or equal to 1 as long as $\zeta \geq 0$ and Δt is small enough that

$$-4 + [4\beta\zeta + (1 + \zeta)^2]\tau^2 - 2\beta\zeta(1 + \zeta)\tau^4 + \beta^2\zeta^2\tau^6 \leq 0 \quad (6.44)$$

for all resolved wavenumbers. A simpler, more restrictive form of this last constraint is that $\tau_{max}^2 \leq 4/[(1 + \zeta)^2 + 4\beta\zeta]$. When ζ is small, a more accurate form of this constraint is

$\tau_{max}^2 < 4 + 8\zeta(2\beta - 1) + O(\zeta^2)$. Damping may either restrict or extend the range of stable time steps. On a C-grid, these schemes are stable provided that

$$\Delta t \leq \frac{1}{[(1 + \zeta)^2 + 4\beta\zeta]^{1/2} c_1 (1/\Delta x^2 + 1/\Delta y^2)^{1/2}}, \quad (6.45)$$

or, more accurately in the limit of small ζ , provided that

$$\Delta t \leq \frac{1 + 2\zeta(2\beta - 1) + O(\zeta^2)}{c_1 (1/\Delta x^2 + 1/\Delta y^2)^{1/2}}. \quad (6.46)$$

If the neutrally stable scheme ($\zeta = 0$) had been shown, Fig. 2 would have been a straight line at a magnitude of 1 out to an abscissa of 2. The vertical lines in Fig. 2 occur when the two pairs of eigenvalues nearly coincide and there is strong interaction between barotropic and baroclinic gravity waves.

De Szoeke and Higdon (1997) have found a neutrally stable scheme that is close to the scheme described by (6.20) through (6.27) with $\beta = \mu = 0$, $\delta = \zeta = 1$ and α irrelevant. In de Szoeke and Higdon's scheme the barotropic equations [(6.15) and (6.14)] are integrated analytically assuming that the baroclinic pressure term varies linearly between the value based on $(\Delta p'_1 / \Delta \tilde{p}_1)^n$ and the value based on $(\Delta p'_1 / \Delta \tilde{p}_1)^*$. A scheme with a constant baroclinic pressure term with the same mean value is obtained from (6.20) through (6.27) by setting $\gamma = 1/2$. However, this difference causes de Szoeke and Higdon's scheme to be unstable at some wavelengths. With their scheme, the analogous equation to (6.43) is

$$\begin{aligned} & [\lambda^2 - (2 - \tau^2 + \Gamma\tau^2)\lambda + 1] \quad , \quad (6.47) \\ & \times \left[\lambda^2 - \left(2\cos T - \frac{\sin T}{T} \Gamma\tau^2 \right) \lambda + 1 \right] - \Gamma\tau^4 (1 - \Gamma) \frac{\sin T}{T} \lambda^2 = 0 \end{aligned}$$

while if they had used $\gamma = 1/2$, their characteristic equation would have been

$$\begin{aligned} & [\lambda^2 - (2\cos T)\lambda + 1] \quad , \quad (6.48) \\ & \times \{ \lambda^2 - [2 - \tau^2 + (\Gamma\tau^2/2)(1 - \cos T)]\lambda + 1 \} - \Gamma\tau^2 (1 - \cos^2 T) \lambda^2 = 0 \end{aligned}$$

and the scheme would have been stable. The representation of the barotropic and baroclinic gravity waves with the de Szoeke and Higdon's scheme is more accurate than for the schemes with characteristic equations (6.43) or (6.48), but the perturbations are not always negative. The perturbations are positive when $(2N - 1)\pi < T < 2N\pi$, and there are unstable roots when the real part of all the roots are nearly the same.

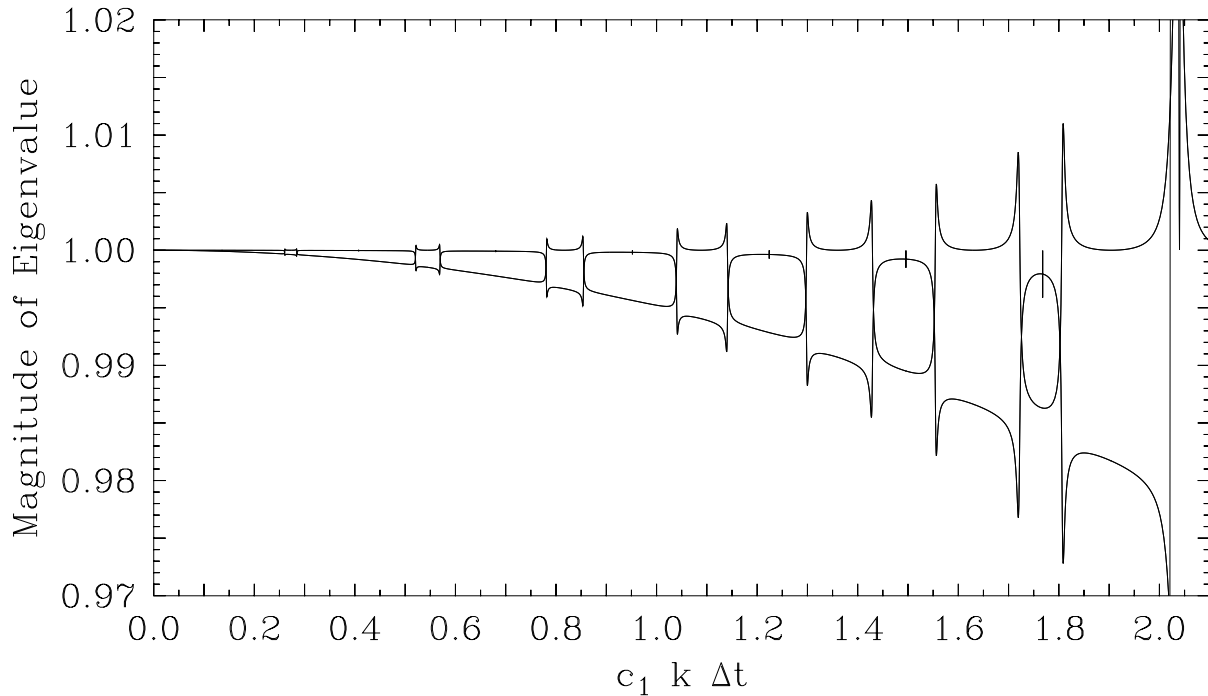


Figure 3. Magnitude of the four eigenvalues for the scheme described in (6.20) through (6.27) with $\delta = 1$, $\zeta = 0.01$, $\mu = 1$, and $\beta = 1$. Again, there are two layers with $\Delta\alpha/\alpha_2 = 0.01$, $\Delta\tilde{p}_1/\tilde{p}_{Bot} = 0.25$, and $\Delta\tilde{p}_2/\tilde{p}_{Bot} = 0.75$. This scheme is linearly unstable at wavelengths determined by the external gravity wave speed. Only the coupling terms in this scheme differ from the scheme in Fig. 2.

The importance of the exact choice of a splitting scheme can be seen in the eigenvalues of a scheme that is not a member of a stable family. The scheme presented in Fig. 3 differs from the scheme in Fig. 2 only in its choice of the time levels of the bottom pressure used to force the baroclinic velocity. Although the product of the magnitudes of the eigenvalues is exactly the same as in the stable scheme in Fig. 2, this scheme is unstable at many wavenumbers determined by the barotropic wave speed. Without a careful stability analysis, it is unlikely that a time splitting scheme will just happen to be stable at all wavenumbers. Even the use of a dissipative barotropic or baroclinic time stepping scheme or both is not sufficient to insure stability of the overall split scheme.

6.4 The Coriolis Terms

The Coriolis terms may be included in the baroclinic momentum equations using the original velocity values in the predictor momentum equation, (6.22), and some interpolation between the orig-

inal and predicted velocity values in the corrector momentum equation, (6.26). This is exactly the same approach suggested by de Szoeke and Higdon (1997). That is, (6.22) becomes

$$(u'_1/c_1)^* = (u'_1/c_1)^n + \phi(v'_1/c_1)^n - i\tau[(1 - \alpha)\eta^n + \alpha\eta^*] - i\tau(\Delta p'_1/\Delta \tilde{p}_1)^n, \quad (6.49)$$

where $\phi = f\Delta t$, and there is a meridional velocity equation

$$(v'_1/c_1)^* = (v'_1/c_1)^n - \phi(u'_1/c_1)^n, \quad (6.50)$$

while (6.26) becomes

$$(u'_1/c_1)^{n+1} = (u'_1/c_1)^n + \phi[(1 - \psi)(v'_1/c_1)^n + \psi(v'_1/c_1)^*] - i\tau[(1 - \delta)\eta^n + \delta\eta^{n+1}] - i\tau[(1 - \zeta)(\Delta p'_1/\Delta \tilde{p}_1)^n + \zeta(\Delta p'_1/\Delta \tilde{p}_1)^*], \quad (6.51)$$

where ψ is a free parameter between 0 and 1. The equation for v'_1^{n+1} is

$$(v'_1/c_1)^{n+1} = (v'_1/c_1)^n - \phi[(1 - \psi)(u'_1/c_1)^n + \psi(u'_1/c_1)^*]. \quad (6.52)$$

(There are no pressure terms in the equations for v'_1 since only gradients in the \hat{x} -direction are being considered). The barotropic equations also include Coriolis terms, as seen in (6.9).

The stability of the scheme including the Coriolis terms is evaluated exactly as was the scheme without these terms. The algebra is more complicated, since the determinant gives a sixth order polynomial, but with a constant Coriolis parameter two of the roots are always 1, corresponding to steady geostrophically balanced flow. In the interest of brevity the intermediate steps are omitted here. The equivalent of (6.42) (the characteristic equation for the scheme with $\delta = 1/2$ and $\mu = 1$) including the Coriolis terms is

$$\begin{aligned} & (\lambda - 1)^2([\lambda^2 - \lambda(2 \cos T) + 1] \\ & \times \{\lambda^2 - \lambda[2 - (1 + \zeta)\tau^2 + \beta\zeta\tau^4 - 2\psi\phi^2 - (\Gamma\tau^2/2)(1 - \beta\zeta\tau^2)(1 - \cos T)] \\ & + [1 - \zeta\tau^2 + (1 - 2\psi)\phi^2 + \psi^2\phi^4 + \zeta(\psi - \beta)\phi^2\tau^2] \\ & - \lambda^2\Gamma\tau^2(1 - \beta\zeta\tau^2)(1 - \cos^2 T)) = 0 \end{aligned}, \quad (6.53)$$

where T has been redefined as $T \equiv \Delta t(c_0^2 k^2 + f^2)^{1/2}$. The coupling (Γ) terms are the same as they were when the Coriolis parameter was neglected (although this is not true for some other schemes). Again, the perturbation to the exact factoring is negative, and does not destabilize the scheme, as long as $\tau^2 \leq 1/(\beta\zeta)$.

Neglecting the coupling terms, the roots of (6.53) are stable as long as $\psi \leq 1/2$, $\Delta t \leq (2\psi - 1)^{1/2}/(\psi f)$, and as long as Δt is small enough that

$$\begin{aligned} -4(\phi^2 + \tau^2) + [4\beta\zeta + (1 + \zeta)^2]\tau^4 + 4(\psi + \zeta\beta)\tau^2\phi^2 \\ - 2\beta\zeta(1 + \zeta)\tau^6 - 4\beta\zeta\psi\tau^4\phi^2 + \beta^2\zeta^2\tau^8 \leq 0 \end{aligned} \quad (6.54)$$

for all resolved wavenumbers. If we introduce the variable $\sigma = c_1 k_{max}$, a more restrictive version of this last constraint is that

$$\Delta t \leq \left\{ \frac{4(f^2 + \sigma^2)}{\sigma^2 [4\beta\zeta(f^2 + \sigma^2) + 4\psi f^2 + (1 + \zeta)^2 \sigma^2]} \right\}^{1/2}. \quad (6.55)$$

In the limit of small ζ , a more accurate version of the last constraint is

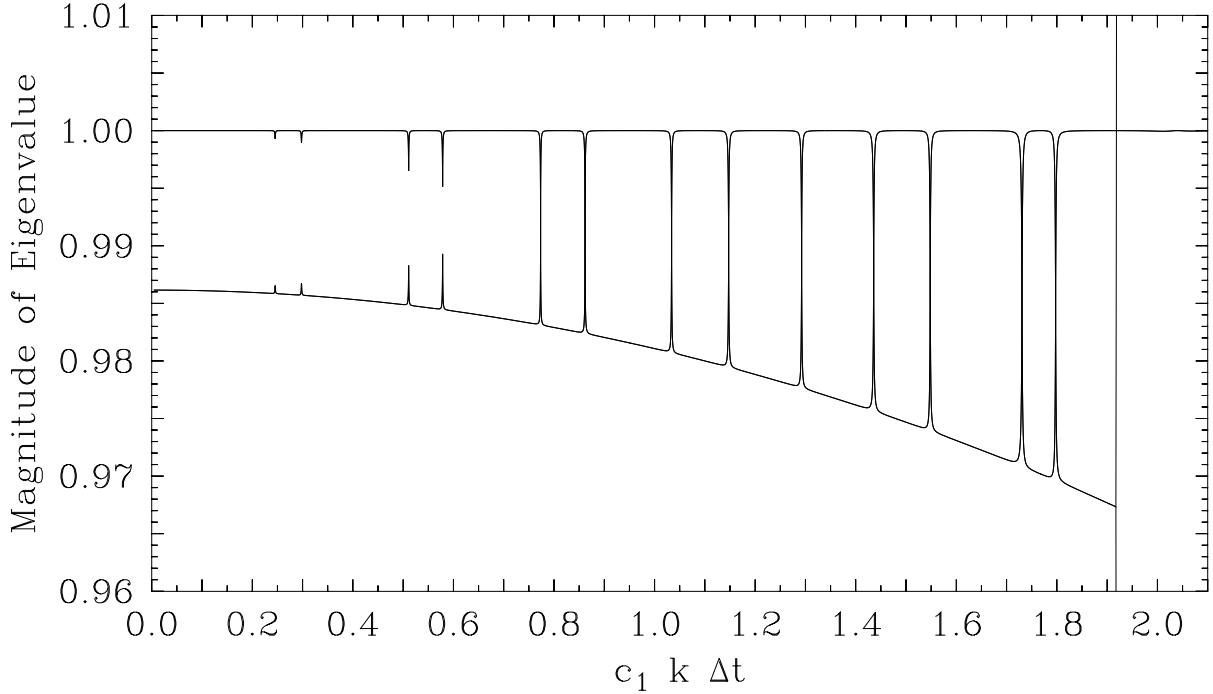


Figure 4. Magnitude of the four eigenvalues for the scheme including the Coriolis terms with $f\Delta t = 0.5$, $\delta = 0.5$, $\xi = 0.01$, $\mu = 1$, and $\beta = \psi = 0.6$. As in Fig. 1, there are two layers with $\Delta\alpha/\alpha_2 = 0.01$, $\Delta\tilde{p}_1/\tilde{p}_{Bot} = 0.25$, and $\Delta\tilde{p}_2/\tilde{p}_{Bot} = 0.75$. With these parameters the scheme is linearly stable for all time steps up to $1.92/c_1 k_{max}$.

$$\Delta t \leq \left\{ \frac{4(f^2 + \sigma^2)}{\sigma^2 (4\psi f^2 + \sigma^2)} \left[1 + \frac{2\zeta\sigma^2}{4\psi f^2 + \sigma^2} \left(\frac{2\beta(f^2 + \sigma^2)}{4\psi f^2 + \sigma^2} - 1 \right) \right] \right\}^{1/2}. \quad (6.56)$$

The stability of this scheme for a particular value of the Coriolis parameter is depicted in Fig. 4. Baroclinic inertial oscillations with even the longest wavelengths are subject to damping.

6.5 Stable Time Stepping Scheme

Only the linear stability of the proposed time stepping scheme has been evaluated here, but a time stepping scheme including all of the nonlinear terms which is consistent with the proposed linear

scheme is included here for clarity. When nonlinear terms are included, the stability becomes much more difficult to assess, and the result depends strongly on the horizontal discretization. Any claim of nonlinear stability for this scheme would go well beyond the scope of the present note. Explicit damping of some sort will probably be necessary to make the following scheme stable for flows with non-negligible nonlinearities.

In the following scheme, four of the free parameters in (6.20) through (6.27) are fixed at $\alpha = \delta = 1/2$, $\gamma = 0$, $\mu = 1$, while β and ζ left as free parameters. In addition, this scheme assumes that $\psi = \beta$ and only a partial time step $\beta\Delta t$ is taken for the predicted velocity, rather than interpolating between the original and predicted velocities in the final step. The predicted velocities are used in the nonlinear momentum advection terms in (6.67) and (6.68) to avoid an obvious instability, subject to the time step limit $\Delta t \leq (2\beta - 1)^{1/2} / (\beta k_{max} U)$, where U is the maximum velocity realized. The predicted pressure thicknesses would have been used in the thickness advection terms of (6.68), had it not been assumed that a positive definite thickness advection scheme will be used for the continuity equations.

By definition, at the start of a time step

$$p_B^m = \sum_{n=1}^N \Delta p_n^m \text{ and } \bar{\mathbf{u}}^m = \sum_{n=1}^N \mathbf{u}_n^m \left(\frac{\Delta p_n}{p_B} \right)^m. \quad (6.57)$$

The barotropic equations are integrated with a series of short time steps

$$p_B^* = p_B^m - \int_{t^m}^{t^m + \Delta t} \nabla \cdot [\bar{\mathbf{u}}(t) p_B(t)] dt \text{ and} \quad (6.58)$$

$$\begin{aligned} \bar{\mathbf{u}}^* = \bar{\mathbf{u}}^m - \int_{t^m}^{t^m + \Delta t} \left\{ f \hat{\mathbf{k}} \times \bar{\mathbf{u}}(t) + \nabla [\alpha_N p_B(t) - gD] \right. \\ \left. + \sum_{n=1}^{N-1} (\alpha_n - \alpha_{n+1}) \left(\frac{p_{k+1/2}}{p_B} \right)^m \nabla \left[p_B(t) \left(\frac{p_{k+1/2}}{p_B} \right)^m \right] + \sum_{n=1}^N \left(\frac{\Delta p_n}{p_B} \right)^m \mathbf{u}_n^m \cdot \nabla \mathbf{u}_n^m \right\} dt \end{aligned} \quad (6.59)$$

where the integrals symbolically represent a number of partial time steps. The time average mass fluxes from the barotropic integrations are also calculated for later use:

$$\widehat{\bar{\mathbf{u}} p_B} = \frac{1}{\Delta t} \int_{t^m}^{t^m + \Delta t} [\bar{\mathbf{u}}(t) p_B(t)] dt. \quad (6.60)$$

The predicted total (barotropic plus baroclinic) layer velocity is

$$\begin{aligned}
\mathbf{u}_n^* &= \mathbf{u}_n^m + \beta(\bar{\mathbf{u}}^* - \bar{\mathbf{u}}^m) - \beta\Delta t f \hat{\mathbf{k}} \times (\mathbf{u}_n^m - \bar{\mathbf{u}}^m) \\
&\quad - \beta\Delta t \left(\nabla \left[\sum_{i=n}^{N-1} (\alpha_i - \alpha_{i+1}) \left(\frac{p_{i+1/2}}{p_B} \right)^m \frac{(p_B^m + p_B^*)}{2} \right] \right. \\
&\quad \quad \left. - \sum_{n=1}^N \left\{ (\alpha_i - \alpha_{i+1}) \left(\frac{p_{i+1/2}}{p_B} \right)^m \nabla \left[\left(\frac{p_{i+1/2}}{p_B} \right)^m \frac{(p_B^m + p_B^*)}{2} \right] \right\} \right) \\
&\quad - \beta\Delta t \left[\mathbf{u}_n^m \cdot \nabla \mathbf{u}_n^m - \sum_{i=1}^N \left(\frac{\Delta p_i}{p_B} \right)^m (\mathbf{u}_i^m \cdot \nabla \mathbf{u}_i^m) \right]
\end{aligned} \tag{6.61}$$

The pressure thickness is stepped using a velocity constructed from the predicted layer velocities and the time mean of the barotropic velocities,

$$\tilde{\mathbf{u}}_n^* = \mathbf{u}_n^* + \frac{1}{\Delta t} \int_{t^m}^{t^m + \Delta t} \bar{\mathbf{u}}(t) dt - \sum_{i=1}^N \mathbf{u}_i^* \left(\frac{\Delta p_i}{p_B} \right)^m, \tag{6.62}$$

which will remove much of the mismatch between the barotropic divergence as calculated by (6.58) and the sum of the layer divergences. The equation for the predicted thickness then is

$$\Delta p_n^* = \Delta p_n^m - \Delta t \nabla \cdot (\tilde{\mathbf{u}}_n^* \Delta p_n^m) - \Delta t \nabla \cdot \left[\left(\bar{\mathbf{u}} \widehat{p}_B - \sum_{i=1}^N \tilde{\mathbf{u}}_i^* \Delta p_i^m \right) \left(\frac{\Delta p_n}{p_B} \right)^m \right]. \tag{6.63}$$

In practice, a positive definite thickness advection scheme must be used, and the final term in (6.63) will be applied as a final correction to insure that the sum of the layer pressures is the bottom pressure calculated by (6.59). The time levels of the pressure thicknesses inside of the divergences in (6.63) will effectively be determined by the positive definite thickness advection scheme. Note that \mathbf{u}_n^* is effectively defined for time $t^m + \beta\Delta t$, while Δp_n^* , p_B^* , and $\bar{\mathbf{u}}^*$ are all defined for time $t^{m+1} \equiv t^m + \Delta t$.

The barotropic equations are now integrated again with slightly different nonlinear forcing terms

$$p_B^{m+1} = p_B^m - \int_{t^m}^{t^m + \Delta t} \nabla \cdot [\bar{\mathbf{u}}(t) p_B(t)] dt \text{ and} \tag{6.64}$$

$$\begin{aligned}
\bar{\mathbf{u}}^{m+1} &= \bar{\mathbf{u}}^m - \int_{t^m}^{t^m + \Delta t} \left\{ f \hat{\mathbf{k}} \times \bar{\mathbf{u}}(t) + \nabla(\alpha_N p_B(t) - gD) \right. \\
&\quad \left. + \sum_{n=1}^{N-1} (\alpha_n - \alpha_{n+1}) \left(\frac{p_{k+1/2}}{p_B} \right)^m \nabla \left[p_B(t) \left(\frac{p_{k+1/2}}{p_B} \right)^m \right] + \sum_{n=1}^N \left(\frac{\Delta p_n}{p_B} \right)^m \tilde{\mathbf{u}}_n^* \cdot \nabla \tilde{\mathbf{u}}_n^* \right\} dt
\end{aligned} \tag{6.65}$$

The time mean thickness fluxes from this second barotropic integration are also calculated:

$$\overline{\tilde{\mathbf{u}} p_B} = \frac{1}{\Delta t} \int_{t^m}^{t^{m+\Delta t}} [\bar{\mathbf{u}}(t) p_B(t)] dt. \quad (6.66)$$

The next time step's layer velocities are

$$\begin{aligned} \mathbf{u}_n^{m+1} = & \mathbf{u}_n^m + (\bar{\mathbf{u}}^{m+1} - \bar{\mathbf{u}}^m) - \Delta t f \hat{\mathbf{k}} \times \left[\mathbf{u}_n^* - \sum_{i=1}^N \left(\frac{\Delta p_i}{p_B} \right)^m \mathbf{u}_i^* \right], \quad (6.67) \\ & - \Delta t \left[\nabla \cdot \left\{ \sum_{i=n}^{N-1} (\alpha_i - \alpha_{i+1}) \left[(1 - \zeta) \left(\frac{p_{i+1/2}}{p_B} \right)^m + \zeta \left(\frac{p_{i+1/2}}{p_B} \right)^* \right] \frac{(p_B^m + p_B^{m+1})}{2} \right\} \right. \\ & \quad \left. - \sum_{i=1}^N \left((\alpha_i - \alpha_{i+1}) \left(\frac{p_{i+1/2}}{p_B} \right)^m \right. \right. \\ & \quad \left. \left. \nabla \cdot \left[\left((1 - \zeta) \left(\frac{p_{i+1/2}}{p_B} \right)^m + \zeta \left(\frac{p_{i+1/2}}{p_B} \right)^* \right) \frac{(p_B^{m+1} + p_B^m)}{2} \right] \right\} \right] \\ & - \Delta t \left[\tilde{\mathbf{u}}_n^* \cdot \nabla \tilde{\mathbf{u}}_n^* - \sum_{i=1}^N \left(\frac{\Delta p_i}{p_B} \right)^m (\tilde{\mathbf{u}}_i^* \cdot \nabla \tilde{\mathbf{u}}_i^*) \right] \end{aligned}$$

Defining $\tilde{\mathbf{u}}_n^{m+1}$ as

$$\tilde{\mathbf{u}}_n^{m+1} = \mathbf{u}_n^{m+1} + \frac{1}{\Delta t} \int_0^{\Delta t} \bar{\mathbf{u}}(t') dt' - \sum_{i=1}^N \mathbf{u}_i^{m+1} \left(\frac{\Delta p_i}{p_B} \right)^m, \quad (6.68)$$

the next time step's pressure thicknesses are

$$\Delta p_n^{m+1} = \Delta p_n^m - \Delta t \nabla \cdot (\tilde{\mathbf{u}}_n^{m+1} \Delta p_n^m) - \Delta t \nabla \cdot \left[\left(\overline{\tilde{\mathbf{u}} p_B} - \sum_{i=1}^N \tilde{\mathbf{u}}_i^{m+1} \Delta p_i^m \right) \left(\frac{\Delta p_n}{p_B} \right)^m \right]. \quad (6.69)$$

Again, a positive definite thickness advection scheme will be used in (6.69), with the correction to insure that the bottom pressure agrees with the barotropic calculation in (6.64).

Finally the barotropic velocity must agree with the vertical average of the layer velocities, while the sum of the layer pressure thicknesses must agree with the bottom pressure. The pressures should already be consistent because of the final correction term in (6.69). However, the barotropic velocities will not be consistent with the average of the layer velocities, in part because of the nonlinear term in (6.9) which has been neglected so far:

$$\frac{\partial \bar{\mathbf{u}}}{\partial t} + \dots = \dots + \sum_{n=1}^N \mathbf{u}_n \frac{\partial}{\partial t} \left(\frac{\Delta p_n}{p_B} \right). \quad (6.70)$$

This is accomplished quite simply by replacing the previously calculated barotropic velocity with

$$\bar{\mathbf{u}}^{m+1} = \sum_{n=1}^N \mathbf{u}_n^{m+1} \left(\frac{\Delta p_n}{p_B} \right)^{m+1}. \quad (6.71)$$

A similar step might have been necessary even if the nonlinear velocity-thickness change correlation term had been explicitly included in the barotropic integration, depending on how viscous or adiabatic effects had been included.

6.6 Conclusions of the split time stepping scheme

Splitting the time stepping operator into barotropic and baroclinic parts allows great efficiency improvements in the integration of free surface primitive equation numerical ocean models. Unfortunately, such splitting often creates linear instability due to interactions between external and internal gravity wave modes. Higdon and Bennett (1996) recently demonstrated this type of linear instability in the time splitting scheme proposed by Bleck and Smith (1990) for an isopycnal coordinate ocean model. The present work offers a family of time stepping schemes that is stable for time steps of the baroclinic equations up to a limit determined by the internal gravity wave speed and the Coriolis parameter, rather than being limited by the external gravity wave speed. One member of this family is neutrally stable for nonrotating gravity waves, although inertial oscillations are always damped.

The simplest of these time stepping schemes, with $\zeta = 0$, does not dissipate nonrotating internal gravity waves. A frequency dependent dissipation might be useful for some simulations because it removes marginally resolved gravity waves without affecting geostrophically balanced flow with the same horizontal scale. Such a scheme might permit stable non-linear flow with a smaller explicit horizontal diffusion than would otherwise be necessary. Otherwise, the scheme with $\zeta = 0$ is probably the most useful.

The stability of these schemes is only demonstrated here for linear, two-layer, flat bottom flows of infinite extent, but other experiments demonstrate a much wider validity. A linear geostrophic adjustment initial value problem in a flat-bottom beta-plane channel without any explicit dissipation shows no amplification of any gravity, Kelvin, or Rossby waves over 100000 baroclinic time steps for the full scheme presented in Section 5 with $\zeta = 0$, $\beta = 0.55$, and Δt set to 94% of the stability limit predicted by (6.56). Also, this time stepping scheme has successfully been used with eddy-rich simulations with multiple layers and large amplitude topography and only modest amounts of explicit dissipation.

Various ocean models have used time splitting with time filtering or heavily dissipative barotropic integrations to eliminate weak instabilities [Bleck and Smith (1990), Killworth et al. (1991), Dukowicz and Smith (1994), Blumberg and Mellor (1987)]. In many of these instances the time filtering is already present to suppress the splitting instability of the leapfrog time integration scheme. Tsumi (1983) demonstrates that time filtering can effectively damp instabilities that would otherwise plague a model with split integration. In the case of the Bleck and Smith (1990) scheme, the fact that the instability occurs even at the longest wavelengths may require extremely heavy time filtering.

The split time stepping scheme presented here offers an efficient, stable method for integrating primitive equation ocean models. This scheme requires explicit dissipation only to control nonlinear instabilities and insure that important boundary currents are well resolved.

References

- Abramowitz, M., 1965: Elementary Analytical Methods. *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun, (Dover, New York, 1965), p.9.
- Arakawa, A., and Y.-J. G. Hsu, 1990: Energy conserving and potential-enzrophy dissipating schemes for the shallow water equations. *Mon. Wea. Rev.*, **118**, 1960-1969.
- _____, and V. R. Lamb, 1981: A potential enstrophy and energy conserving scheme for the shallow water equations. *Mon. Wea. Rev.*, **109**, 18-36.
- Bleck, R., and L. T. Smith, 1990: A wind-driven isopycnic coordinate model of the North and Equatorial Atlantic Ocean; 1. Model development and supporting experiments. *J. Geophys. Res.*, **95**, 3273-3285.
- Blumberg, A. F., and G. L. Mellor, 1987: A description of a three-dimensional coastal ocean circulation model. *Three-Dimensional Coastal Ocean Models*, edited by N. Heaps, (Am. Geophys. Union, Washington, DC, 1987), p. 1.
- R. A. de Szoeke and R. L. Higdon, 1997: Barotropic-baroclinic time splitting for ocean circulation modeling. Submitted to *J. Comput. Phys.*
- Dukowicz, J. K. and R. D. Smith, 1994: Implicit free-surface method for the Bryan-Cox-Semtner ocean model, *J. Geophys. Res. C*, **99**, 7991-8014.

- G. J. Haltiner and R. T. Williams, 1980: *Numerical Prediction and Dynamic Meteorology*, John Wiley & Sons, New York, p. 143.
- Hallberg, R., 1997: Stable split time stepping for large-scale ocean modeling. Submitted to *J. Comp. Phys.*
- Higdon, R. L., and A. F. Bennett, 1996: Stability analysis of operator splitting for large-scale ocean modeling. *J. Comput. Phys.*, **123**, 311-329.
- Hsu, Y.-J. G., and Arakawa, A., 1990: Numerical modeling of the atmosphere with an isentropic vertical coordinate. *Mon. Wea. Rev.*, **118**, 1933-1959.
- P. D. Killworth, D. Stainforth, D. J. Webb, and S. M. Paterson, 1991: The development of a free-surface Bryan-Cox-Semtner ocean model. *J. Phys. Oceanogr.*, **21**, 1333-1348.
- Matsuno, T., 1966: A finite difference scheme for time integrations of oscillatory equations with second order accuracy and sharp cut-off for high frequencies. *J. Met. Soc. Japan*, **44**, 85-88.
- Takacs, L. L., 1985: A two-step scheme for the advection equation with minimized dissipation and dispersion errors. *Mon. Wea. Rev.*, **113**, 1050-1065.
- Y. Tatsumi, 1983: An economical explicit time integration scheme for a primitive model. *J. Met. Soc. Japan* **61**, 269-287.