

A comparison of two formulations of barotropic-baroclinic splitting for layered models of ocean circulation

Robert L. Higdon

*Department of Mathematics
Oregon State University
Corvallis, Oregon 97331-4605, USA*

Abstract

In numerical models of ocean circulation, it is widespread practice to split the fast and slow motions into barotropic and baroclinic subsystems, respectively. In the case of the baroclinic equations, the dependent variables can either be (1) slowly-varying baroclinic quantities, obtained from splitting the original flow variables into barotropic and baroclinic components, or (2) the original unsplit variables, which can vary on both the fast and slow time scales. In the second case, the variables in each layer are adjusted after each (long) baroclinic time step to ensure compatibility with the results produced from the barotropic equations. The second approach can be applied to the layer thickness equation to ensure exact conservation of mass within each layer. In the case of the momentum equations, the second approach amounts to replacing unresolved fast portions of Coriolis and pressure forcing with time averages of well-resolved forcing from the barotropic system. In this study, both approaches for the momentum equations are evaluated, in several test problems, by comparing to analytical solutions or to solutions computed with an unsplit code that uses short time steps. The two methods give very similar results in some simple problems for which analytical solutions are known. However, in some eddying double-gyre simulations, the formulation with unsplit variables requires a significant reduction in the baroclinic time step in order to avoid numerical difficulties that include grid noise and inaccurate representation of the flow field. In contrast, the formulation with split variables does not display such difficulties, and in those same examples it can be used with zero explicit horizontal viscosity. All of these computations employ a two-level timestepping method that was previously developed by the author.

Key words: barotropic-baroclinic splitting, time-stepping

Email address: higdon@math.oregonstate.edu (Robert L. Higdon).

1 Introduction

The physical dynamics of the ocean include motions that vary on a wide range of time scales. In particular, the speeds of external gravity waves can be two orders of magnitude greater than the speeds of internal waves and advective motions. If an explicit time-stepping method were used to solve the governing equations numerically, then the presence of the external gravity waves would impose a severe restriction on the allowable time step. However, the fast external motions can be modeled accurately with a two-dimensional barotropic subsystem that resembles the shallow water equations for a homogeneous fluid. The remaining slow motions can be modeled by a three-dimensional baroclinic subsystem. The baroclinic equations can be solved explicitly with a long time step that is appropriate for resolving the slow motions, and the barotropic equations can be solved implicitly with the same time step or explicitly with many short substeps. The latter procedures are applied to a relatively simple two-dimensional subsystem instead of the full three-dimensional system, so the result is a major gain in efficiency relative to an algorithm that solves the three-dimensional system without splitting the dynamics into barotropic and baroclinic components. Barotropic-baroclinic splittings are therefore widely used in the numerical modeling of ocean circulation. In the present paper we consider such splittings in the context of isopycnic-coordinate ocean models, for which the vertical coordinate is a physical quantity related to density.

Of particular concern here is the choice of the dependent variables that are used in a barotropic-baroclinic splitting. The barotropic equations are obtained through a vertical averaging or summation of the three-dimensional governing equations, and for these equations the dependent variables are vertical averages or sums of the original flow variables. In the case of the baroclinic equations, two types of dependent variables have been widely used. One approach is to split the dependent variables (approximately) into barotropic and baroclinic components, and then use the baroclinic quantities as prognostic variables in the baroclinic equations. Another approach is to use the same dependent variables as in the original unsplit system, but at the end of each baroclinic time step adjust these variables to maintain consistency with the results computed with the barotropic equations. In the following, the two choices of dependent variables will be described as “split variables” and “unsplit (original) variables”, respectively. As noted in Section 3.4, there are strong reasons for using the second approach when solving the equation for conservation of mass. The purpose of the present paper is to explore which approach would be better when solving the equations for conservation of momentum.

In previous work by this author (Higdon, 2005), split variables were used to solve the baroclinic momentum equations, and the algorithm gave good results in some numerical experiments involving the model problems described in that paper. However, the formulation of the baroclinic momentum equations in such variables includes some terms that do not seem natural in a physical sense. This raises the possibility that a discretization based on this formulation could produce a distortion of the dynamics in the numerical solution, in some cases. On the other hand, a formulation in terms of unsplit (original) variables does not include the terms in question. It then seemed advisable to develop an algorithm with unsplit variables and to compare the two approaches.

Section 2 summarizes the governing equations, the choices of dependent variables, and the formulation of the baroclinic momentum equations in split variables. Section 3 develops a formulation of the baroclinic momentum equations in unsplit (original) variables. Section 4 describes some numerical computations involving those two formulations, and the discussion includes comparisons to analytical solutions or to solutions obtained with a relatively simple code that uses short time steps and no barotropic-baroclinic splitting. Section 4 concludes with a demonstration of the split-variable formulation in a long-time, high-resolution simulation of a double-gyre flow with active eddy fields. Section 5 discusses some issues raised by the differing behaviors of the two formulations of barotropic-baroclinic splitting, and Section 6 gives some conclusions.

2 Governing equations and summary of splitting

In an isopycnic-coordinate ocean model, the vertical coordinate is potential density or a related quantity, and a vertical discretization of such a model amounts to dividing the fluid into layers having distinct physical properties. An advantage of such a coordinate is that it facilitates the reduction or elimination of spurious diapycnal diffusion between distinct water masses. A systematic comparison of the advantages and disadvantages of isopycnic coordinates and other widely-used vertical coordinates (z and σ) is given, for example, by Griffies (2004).

For simplicity, it is assumed in the present discussion that the vertical coordinate is the specific volume $\alpha = 1/\rho$ (i.e., reciprocal of density) and that the horizontal coordinates are Cartesian coordinates (x, y) . Partition the fluid into R layers having specific volumes $\alpha_1, \dots, \alpha_R$, with the layer indices increasing downward. Assume that the vertical length scale is much less than the horizontal length scale, so that the hydrostatic assumption holds. Let $\Delta p_r > 0$ denote the vertical pressure difference across layer r ; because of the hydrostatic condition, Δp_r is equal to g times the mass per unit horizontal area in

layer r , where g is the acceleration due to gravity. The quantity Δp_r can also be regarded informally as the thickness of layer r . Let $\mathbf{u}_r = (u_r, v_r)$ denote the horizontal velocity in layer r , and let $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$. (In the vertically-continuous case, ∇ would act along surfaces of constant α .) If there is no transport of mass between layers, then the equation for conservation of mass is

$$\frac{\partial \Delta p_r}{\partial t} + \nabla \cdot (\mathbf{u}_r \Delta p_r) = 0, \quad (1)$$

and the equation for conservation of momentum is

$$\frac{\partial}{\partial t} (\mathbf{u}_r \Delta p_r) + \mathbf{A}_r + f \mathbf{u}_r^\perp \Delta p_r = -(\Delta p_r) \nabla M_r + \mathbf{D}_r \quad (2)$$

(e.g., Higdon (2006)). Here, $\mathbf{u}_r^\perp = (-v_r, u_r)$, f is the Coriolis parameter,

$$\mathbf{A}_r = \frac{\partial}{\partial x} \left(u_r (\mathbf{u}_r \Delta p_r) \right) + \frac{\partial}{\partial y} \left(v_r (\mathbf{u}_r \Delta p_r) \right) \quad (3)$$

represents the advection of momentum, $\mathbf{D}_r = \nabla \cdot (A_H \Delta p_r \nabla \mathbf{u}_r) + g \Delta \tau_r$ represents the effects of horizontal diffusion and the vertical differences of stresses at layer interfaces, and $M = \alpha p + gz$ is the Montgomery potential. Because of the hydrostatic condition $\partial p / \partial z = -\rho g$, M is independent of depth within a layer of constant density, and M_r refers to the value of this quantity in layer r . Equations (1) and (2) are supplemented with the jump condition $M_r - M_{r+1} = p_{r+1/2}(\alpha_r - \alpha_{r+1})$, where $p_{r+1/2}$ denotes the pressure at the interface between layers r and $r + 1$.

The quantity $\mathbf{u}_r \Delta p_r$ is equal to g times the horizontal momentum per unit horizontal area in layer r , so $\mathbf{u}_r \Delta p_r$ is a constant multiple of momentum density. Equations (1) and (2) could be combined to produce an equation for the velocity \mathbf{u}_r ; the momentum formulation (2) is used here because the nonlinear terms are expressed in the flux form (3), for which a numerical advection algorithm can be used.

In some work that led to the results reported by Higdon (2002), the velocity was initially used as the dependent variable, and the nonlinear and Coriolis terms were combined into a formulation involving vorticity and the gradient of kinetic energy. The gradient was approximated with centered spatial differences, and the vorticity and Coriolis terms were discretized with two different schemes of Sadourny (1975). However, these formulations allowed erratic and sometimes unstable behavior in some situations where layer thicknesses tend to zero. In contrast, the flux formulation described above, in conjunction with

a numerical advection scheme, produced stable numerical behavior in those same situations. The flux formulation is therefore the one that is considered by Higdon (2002) and in the present paper.

2.1 Barotropic-baroclinic splitting

A description of the process of barotropic-baroclinic splitting, along with a mathematical discussion of external and internal modes, is included in the review by Higdon (2006), and one particular formulation of a splitting is developed in Section 3 of the present paper. For the barotropic equations, the main idea is to employ a vertical summation and/or averaging of the three-dimensional mass and momentum equations. The barotropic mass variable can be taken to be

$$p_b(x, y, t) = \sum_{r=1}^R \Delta p_r, \quad (4)$$

which is g times the mass per unit horizontal area in a water column. Equivalently, one can use the perturbation $p'_b \eta$ in p_b , where p'_b is the value of p_b at a reference state, and η is the relative perturbation given by $\eta(x, y, t) = (p_b(x, y, t) - p'_b(x, y))/p'_b(x, y)$ (Bleck and Smith, 1990). For a dependent variable in the barotropic momentum equation, one can either use the mass-weighted vertically-averaged velocity

$$\bar{\mathbf{u}}(x, y, t) = \sum_{r=1}^R \mathbf{u}_r \frac{\Delta p_r}{p_b} \quad (5)$$

or the vertically-summed momentum $p_b \bar{\mathbf{u}} = \sum_{r=1}^R \mathbf{u}_r \Delta p_r$.

For the dependent variables in the baroclinic equations, one possibility is to split the mass and velocity fields into barotropic and baroclinic components and then use the baroclinic quantities as prognostic variables. In a splitting developed for isopycnic modeling, Bleck and Smith (1990) used the relations

$$\Delta p_r = (1 + \eta) \Delta p'_r \quad (6)$$

$$\mathbf{u}_r = \mathbf{u}'_r + \bar{\mathbf{u}} \quad (7)$$

to split the mass and velocity fields, respectively. The quantities η and $\bar{\mathbf{u}}$ are independent of depth and are intended to represent the fast external motions, whereas $\Delta p'_r$ and \mathbf{u}'_r are intended to represent the remaining (slow) motions. Equation (6) is based on the idea that an external signal causes all fluid layers

to thicken or thin by approximately the same proportion, at a given time and horizontal position.

The splittings of mass and velocity in (6) and (7) are not exact, in the sense that the barotropic variables η and $\bar{\mathbf{u}}$ can contain a small amount of energy from the internal modes, and the baroclinic variables $\Delta p'_r$ and \mathbf{u}'_r can contain a small amount of energy from the external mode. For example, in some numerical experiments involving external and internal Rossby waves in a two-layer fluid in a channel, Higdon (2005) found that the baroclinic velocity is nonzero but about three orders of magnitude smaller than the barotropic velocity in a pure external mode, in one particular configuration. The baroclinic equations are solved with a long time step suitable for resolving the slow internal motions, and this is consistent with the (essentially) slow time variation of the baroclinic variables $\Delta p'_r$ and \mathbf{u}'_r .

An alternative to using split variables in the baroclinic equations is to update the unsplit variables Δp_r and \mathbf{u}_r at each (long) baroclinic time step. However, this option poses two possible difficulties which are not encountered in the case of split variables.

(i) The quantities Δp_r and \mathbf{u}_r can include both external and internal motions, so in general these quantities can vary on the fast time scale. The long time step then raises the prospect of numerical instability.

(ii) Once Δp_r and \mathbf{u}_r are updated, the vertical sums $\sum_{r=1}^R \Delta p_r$ and $\sum_{r=1}^R \mathbf{u}_r \Delta p_r$ give values of p_b and $p_b \bar{\mathbf{u}}$, respectively. However, these values are not necessarily identical to the values of p_b and $p_b \bar{\mathbf{u}}$ that are computed with the barotropic equations, since different numerical methods are used for the barotropic and baroclinic systems. One can then ask whether such an inconsistency could cause any problems with the computed solution.

The second difficulty (ii) can be remedied by adjusting the values of Δp_r and \mathbf{u}_r slightly, at each time step, so that their vertical sums are equal to the values of p_b and $p_b \bar{\mathbf{u}}$ that are computed with the barotropic equations for that same time. Procedures for accomplishing this task are described in Section 3. However, as demonstrated in that section, this step can also be interpreted as follows. If a governing equation is solved with a (long) baroclinic time step, then each forcing term in that equation is the sum of a well resolved slowly-varying part and an unresolved rapidly-varying part. Enforcing barotropic consistency has the effect of discarding the unresolved rapidly-varying forcing and replacing it with the time-integrated effect of well resolved forcing that is computed when the barotropic equations are solved explicitly with short sub-steps. This substitution appears sufficient to maintain numerical stability. The enforcement of consistency thus also addresses the first difficulty (i) described above.

Unsplit variables have been used by Blumberg and Mellor (1987) and Shchepetkin and McWilliams (2005) in σ -coordinate models and by Hallberg (1997) in an isopycnic model. On the other hand, Bryan (1969) describes a treatment of split velocities (barotropic plus baroclinic) that is still used in the Bryan-Cox class of z -coordinate models (Griffies, 2004). As noted in Section 3.4, unsplit variables are well-suited for solving equation (1) for conservation of mass in a layered model.

For the case of split variables as defined by (6) and (7), the baroclinic mass and velocity satisfy the relations $\sum_{r=1}^R \Delta p'_r = p'_b$ and $\sum_{r=1}^R \mathbf{u}'_r \Delta p'_r = 0$. The second condition is equivalent to stating that the baroclinic velocity has mass-weighted vertical average equal to zero. In the formulation by Bleck and Smith (1990), these conditions are enforced by the implementation of certain forcing terms in the baroclinic mass and momentum equations, respectively. These processes are analogues of the enforcement of barotropic consistency of unsplit variables, which was described above. However, in the case of split variables the target values of $\sum_{r=1}^R \Delta p'_r$ and $\sum_{r=1}^R \mathbf{u}'_r \Delta p'_r$ are p'_b and 0, respectively, and these are time-independent quantities that are known a priori. Since these vertical sums do not involve any barotropic variables, they do not raise the possibility of inconsistency with the barotropic results. On the other hand, the unsplit variables Δp_r and \mathbf{u}_r contain the barotropic quantities η and $\bar{\mathbf{u}}$, and with that formulation it then becomes necessary to enforce the consistent computation of the quantities $p_b = \sum_{r=1}^R \Delta p_r$ and $p_b \bar{\mathbf{u}} = \sum_{r=1}^R \mathbf{u}_r \Delta p_r$. To put it another way, an update of Δp_r and \mathbf{u}_r makes an implicit statement about the barotropic variables η and $\bar{\mathbf{u}}$, but an update of $\Delta p'_r$ and \mathbf{u}'_r does not.

2.2 Baroclinic momentum equation with split variables

A prognostic equation for \mathbf{u}'_r can be used as a baroclinic momentum equation; such equations are given by Bleck and Smith (1990) and Higdon (2002). An alternative is to use the momentum-like quantity $\mathbf{u}'_r \Delta p'_r$, which varies mainly on the slow time scale and can be regarded as a baroclinic momentum density. Such an equation was used by Higdon (2005). The x -component of this equation has the form

$$\begin{aligned}
& \frac{\partial}{\partial t} (u'_r \Delta p'_r) + \frac{\partial}{\partial x} \left[u_r (u'_r \Delta p'_r) \right] + \frac{\partial}{\partial y} \left[v_r (u'_r \Delta p'_r) \right] \\
& = f v'_r \Delta p'_r - \Delta p'_r \left(\frac{\partial M_r}{\partial x} - (\nabla M)_x \right) + D^{u'} - G_x \Delta p'_r \\
& - u_r \Delta p'_r \frac{\partial \bar{u}}{\partial x} - v_r \Delta p'_r \frac{\partial \bar{u}}{\partial y} + \frac{u'_r \Delta p'_r}{p'_b} \nabla \cdot (p'_b \bar{\mathbf{u}})
\end{aligned} \tag{8}$$

and the y -component is analogous. Here, G_x is the term that is used to enforce the condition $\sum_{r=1}^R \mathbf{u}'_r \Delta p'_r = 0$, and $(\overline{\nabla M})_x$ is the x -component of the mass-weighted vertical average of ∇M . The left side of (8) includes terms involving the advection of the momentum density $u'_r \Delta p'_r$, and for these terms a numerical advection scheme can be used.

The formulation (8) yielded good results in the model problems used for the numerical computations reported by Higdon (2005). However, the last three terms in (8) do not seem natural in a physical sense. This raises the possibility that a discretization based on the form (8) might, in some way, lead to a distortion of the dynamics in the numerical solution in some cases. On the other hand, the equation (2) for the unsplit momentum $\mathbf{u}_r \Delta p_r$ contains no such terms. This suggests that it may be more appropriate, in physical terms, to use equation (2) and then enforce consistency with the results obtained with the barotropic equations. This idea is formulated in the next section. Section 4 then describes some numerical computations that compare these two approaches to barotropic-baroclinic splitting.

3 Momentum equations with unsplit variables

Here we develop a barotropic-baroclinic splitting with unsplit (original) variables and analyze some of its properties.

3.1 Development of a splitting

A vertical sum of the layer mass equation (1), coupled with the definitions (4) and (5) of p_b and $\bar{\mathbf{u}}$, yields the barotropic mass equation

$$\frac{\partial p_b}{\partial t} + \nabla \cdot (p_b \bar{\mathbf{u}}) = 0, \quad (9)$$

and a vertical sum of the layer momentum equation (2) yields the barotropic momentum equation

$$\frac{\partial}{\partial t} (p_b \bar{\mathbf{u}}) + f p_b \bar{\mathbf{u}}^\perp = -p_b \overline{\nabla M} + \sum_{r=1}^R (\mathbf{D}_r - \mathbf{A}_r). \quad (10)$$

Since p_b is equal to g times the mass per unit horizontal area over the depth of the fluid, the quantity $p_b \bar{\mathbf{u}}$ can be regarded as a barotropic mass flux in equation (9) and a barotropic momentum density in equation (10).

For a baroclinic momentum equation, we will use the layer momentum equation (2) minus $\Delta p_r/p_b$ times the barotropic momentum equation. This yields

$$\frac{\partial}{\partial t}(\mathbf{u}_r \Delta p_r) + \mathbf{A}_r + f \mathbf{u}_r^\perp \Delta p_r = -(\Delta p_r) \nabla M_r + \mathbf{D}_r + \frac{\Delta p_r}{p_b} \mathbf{B} \quad (11)$$

where

$$\mathbf{B}(x, y, t) = \frac{\partial}{\partial t}(p_b \bar{\mathbf{u}}) + f p_b \bar{\mathbf{u}}^\perp + p_b \overline{\nabla M} - \sum_{r=1}^R (\mathbf{D}_r - \mathbf{A}_r) . \quad (12)$$

According to the barotropic momentum equation (10), $\mathbf{B} = \mathbf{0}$, so it may initially appear that nothing is accomplished by moving from the layer equation (2) to equation (11). However, in the discretization described in Section 3.2, the term \mathbf{B} provides a mechanism for enforcing consistency with the results computed with the barotropic equations. In addition, the various terms in \mathbf{B} can be grouped with other terms in equation (11) to produce quantities that are baroclinic in nature.

In particular, assume that the flow is a small perturbation of a stationary state for which the layer interfaces are level. Also assume that the bottom of the fluid domain is level and that the viscosity and stresses are zero. Let $\Delta \tilde{p}_r$ denote the equilibrium pressure difference across layer r , and neglect products of small quantities. The result is the linearized equation

$$\frac{\partial \mathbf{u}_r}{\partial t} + f \mathbf{u}_r^\perp = -\nabla M_r + \left(\frac{\partial \bar{\mathbf{u}}}{\partial t} + f \bar{\mathbf{u}}^\perp + \overline{\nabla M} \right) .$$

The definition $\mathbf{u}'_r = \mathbf{u}_r - \bar{\mathbf{u}}$ of baroclinic velocity then yields

$$\frac{\partial \mathbf{u}'_r}{\partial t} + f (\mathbf{u}'_r)^\perp = -\left(\nabla M_r - \overline{\nabla M} \right) . \quad (13)$$

Equation (13) is the same as the linearization of the momentum equation (8) that was formulated in split variables. This paper is considering two different approaches to the momentum equation in a barotropic-baroclinic splitting, and at the linearized level these two approaches are the same.

In order to develop a numerical algorithm, it is also necessary to obtain a baroclinic mass equation for updating the layer thicknesses Δp_r . This point is discussed in Section 3.4.

3.2 Discretization of the baroclinic momentum equation

Assume that the baroclinic momentum equation (11) is discretized with a two-time-level method of the form

$$\begin{aligned}
(\mathbf{u}_r \Delta p_r)^{n+1} &= (\mathbf{u}_r \Delta p_r)^n \\
&+ \Delta t \left\{ -[\mathbf{A}_r] - f[\mathbf{u}_r^\perp \Delta p_r] - [(\Delta p_r) \nabla M_r] + [\mathbf{D}_r] \right\} \\
&+ \Delta t \left(\frac{\Delta p_r}{p_b} \right)^{n+1} \mathbf{B}^{n+1}.
\end{aligned} \tag{14}$$

Here, $(\mathbf{u}_r \Delta p_r)^n$ denotes the numerical approximation to $\mathbf{u}_r \Delta p_r$ at baroclinic time t_n , Δt is the (long) baroclinic time step, and the terms with square brackets denote discretizations with respect to space and/or time of the quantities enclosed in those brackets. For the sake of brevity in notation, horizontal spatial dependences are not indicated in equation (14).

The depth-independent quantity \mathbf{B}^{n+1} is computed after all other terms on the right side of (14) are accumulated, and it is defined by the conditions $\sum_{r=1}^R (\mathbf{u}_r \Delta p_r)^n = (p_b \bar{\mathbf{u}})^n$ and $\sum_{r=1}^R (\mathbf{u}_r \Delta p_r)^{n+1} = (p_b \bar{\mathbf{u}})^{n+1}$, where $(p_b \bar{\mathbf{u}})^n$ and $(p_b \bar{\mathbf{u}})^{n+1}$ denote results obtained with the barotropic equations at baroclinic times t_n and t_{n+1} , respectively. Vertical summation of equation (14) reveals that the quantity \mathbf{B}^{n+1} is a discretization of the quantity on the right side of equation (12), as expected, and thus $\mathbf{B}^{n+1} \rightarrow \mathbf{0}$ as $\Delta x \rightarrow 0$, $\Delta y \rightarrow 0$ and $\Delta t \rightarrow 0$.

Some further remarks on the implementation of the quantity \mathbf{B}^{n+1} are the following. Accumulating all of the other terms on the right side of (14) is equivalent to solving the momentum equation in layer r , for one time step, without regard to barotropic-baroclinic splitting. Some procedures regarding thin layers and Coriolis terms discussed by Higdon (2005) yield a velocity \mathbf{u}_r^* , from which a momentum density $\mathbf{u}_r^* \Delta p_r^{n+1}$ is obtained. The latter quantity is the result of implementing all terms on right side of (14) except for the one involving \mathbf{B}^{n+1} . Vertical summation of (14) then yields $\mathbf{B}^{n+1} \Delta t = (p_b \bar{\mathbf{u}})^{n+1} - \sum_{r=1}^R \mathbf{u}_r^* \Delta p_r^{n+1}$, and this result is inserted back into (14). The right side of equation (14) then becomes

$$\begin{aligned}
&\mathbf{u}_r^* \Delta p_r^{n+1} + \frac{\Delta p_r^{n+1}}{p_b^{n+1}} \left((p_b \bar{\mathbf{u}})^{n+1} - \sum_{k=1}^R \mathbf{u}_k^* \Delta p_k^{n+1} \right) \\
&= \Delta p_r^{n+1} \left(\mathbf{u}_r^* - \sum_{k=1}^R \mathbf{u}_k^* \frac{\Delta p_k^{n+1}}{p_b^{n+1}} + \bar{\mathbf{u}}^{n+1} \right).
\end{aligned}$$

Therefore, the effect of this procedure is to compute a tentative velocity \mathbf{u}_r^* , subtract its mass-weighted vertical average to obtain a baroclinic velocity having vertical average equal to zero, add the barotropic velocity to obtain the total velocity, and finally multiply by Δp_r^{n+1} to obtain the updated momentum density. This process resembles the treatments of baroclinic and barotropic velocities used by Bryan (1969) and by Bleck and Smith (1990).

3.3 Interpretation of the discrete baroclinic momentum equation

The action of the quantity \mathbf{B}^{n+1} can also be interpreted in terms of replacing unresolved forcing with well-resolved forcing, as follows. Assume that the barotropic equations are solved explicitly with time step $\Delta t/N$, where N denotes the number of barotropic substeps of each baroclinic step. Let $(p_b \bar{\mathbf{u}})^{n,m}$ denote the numerical approximation to $p_b \bar{\mathbf{u}}$ at the barotropic time step $t_n + m \left(\frac{\Delta t}{N}\right)$; thus $(p_b \bar{\mathbf{u}})^{n,0} = (p_b \bar{\mathbf{u}})^n$ and $(p_b \bar{\mathbf{u}})^{n,N} = (p_b \bar{\mathbf{u}})^{n+1}$. Assume that the discretization of the barotropic momentum equation (10) has the form

$$\begin{aligned} (p_b \bar{\mathbf{u}})^{n,m+1} &= (p_b \bar{\mathbf{u}})^{n,m} \\ &+ \frac{\Delta t}{N} \left\{ -f (p_b \bar{\mathbf{u}}^\perp)^{n,m} - (p_b \bar{\nabla} M)^{n,m} \right\} \\ &+ \frac{\Delta t}{N} \sum_{r=1}^R ([\mathbf{D}_r] - [\mathbf{A}_r]). \end{aligned} \quad (15)$$

The quantities $(p_b \bar{\mathbf{u}}^\perp)^{n,m}$ and $(p_b \bar{\nabla} M)^{n,m}$ are not necessarily evaluated at time $t_n + m(\Delta t/N)$, but instead these notations refer to whatever quantities are used to advance the solution from time $t_n + m(\Delta t/N)$ to time $t_n + (m+1)(\Delta t/N)$. The calculation of $(p_b \bar{\nabla} M)^{n,m}$ requires mass quantities, which are provided by a discretization of the barotropic mass equation (9). The term $\sum_{r=1}^R ([\mathbf{D}_r] - [\mathbf{A}_r])$ is assumed to be held constant in time on the baroclinic time interval $[t_n, t_{n+1}]$; in some experiments, a representation of the vertical sum $\sum_{r=1}^R [\mathbf{A}_r]$ of the advection terms was used to evaluate this quantity at each barotropic substep, but this had little or no effect on the solution and entailed extra computational cost.

When equation (15) is applied over all barotropic substeps of the baroclinic interval $[t_n, t_{n+1}]$ (i.e., for $0 \leq m \leq N-1$), the cumulative effect is

$$\begin{aligned} (p_b \bar{\mathbf{u}})^{n+1} &= (p_b \bar{\mathbf{u}})^n \\ &+ \Delta t \left\{ -f \frac{1}{N} \sum_{m=0}^{N-1} (p_b \bar{\mathbf{u}}^\perp)^{n,m} - \frac{1}{N} \sum_{m=0}^{N-1} (p_b \bar{\nabla} M)^{n,m} \right\} \end{aligned} \quad (16)$$

$$+ \sum_{r=1}^R ([\mathbf{D}_r] - [\mathbf{A}_r]).$$

However, another way to compute barotropic momentum $p_b \bar{\mathbf{u}}$ is to sum the discrete baroclinic momentum equation (14) over all layers, and this summation yields

$$\begin{aligned} (p_b \bar{\mathbf{u}})^{n+1} &= (p_b \bar{\mathbf{u}})^n \\ &+ \Delta t \left\{ -f \sum_{r=1}^R [\mathbf{u}_r^\perp \Delta p_r] - \sum_{r=1}^R [(\Delta p_r) \nabla M_r] \right\} \\ &+ \Delta t \sum_{r=1}^R ([\mathbf{D}_r] - [\mathbf{A}_r]) + (\Delta t) \mathbf{B}^{n+1}. \end{aligned} \quad (17)$$

If equations (16) and (17) are to refer to the same values of $p_b \bar{\mathbf{u}}$, the quantity \mathbf{B}^{n+1} must equal the braced quantity in (16) minus the braced quantity in (17). The discretization (14) of the baroclinic momentum equation can then be written as

$$\begin{aligned} (\mathbf{u}_r \Delta p_r)^{n+1} &= (\mathbf{u}_r \Delta p_r)^n \\ &- (f \Delta t) \left\{ [\mathbf{u}_r^\perp \Delta p_r] + \left(\frac{\Delta p_r}{p_b} \right)^{n+1} \left(- \sum_{k=1}^R [\mathbf{u}_k^\perp \Delta p_k] + \frac{1}{N} \sum_{m=1}^N (p_b \bar{\mathbf{u}}^\perp)^{n,m} \right) \right\} \\ &- (\Delta t) \left\{ [(\Delta p_r) \nabla M_r] + \left(\frac{\Delta p_r}{p_b} \right)^{n+1} \left(- \sum_{k=1}^R [(\Delta p_k) \nabla M_k] + \frac{1}{N} \sum_{m=1}^N ((p_b \nabla M)^{n,m}) \right) \right\} \\ &+ (\Delta t) \left\{ -[\mathbf{A}_r] + [\mathbf{D}_r] \right\} \end{aligned} \quad (18)$$

The second line in equation (18), involving $f \Delta t$, is the Coriolis term in this discretization. The quantity $[\mathbf{u}_r^\perp \Delta p_r]$ is computed on the baroclinic time grid, and the barotropic part of this quantity is represented by $\sum_{k=1}^R [\mathbf{u}_k^\perp \Delta p_k]$; this sum can be regarded as an approximation to the quantity $p_b \bar{\mathbf{u}}^\perp$. In general, the time variation of this barotropic part is not resolved on the baroclinic time grid. On the other hand, the quantity $(p_b \bar{\mathbf{u}}^\perp)^{n,m}$ is computed at each barotropic substep when the barotropic equations are solved, and it is well-resolved on the finer barotropic time grid. The sum $\frac{1}{N} \sum_{m=1}^N (p_b \bar{\mathbf{u}}^\perp)^{n,m}$ represents the time-integrated effect of this forcing term over the entire baroclinic interval $[t_n, t_{n+1}]$. The effect of the second line in (18) is to subtract unresolved forcing and replace it with the time-integrated effect of well-resolved forcing, with the difference between the two being distributed proportionately over all of the

layers in the fluid. In addition, the vertical sum of the second line equals the time-integrated Coriolis term from the barotropic equations. The third line in (18) represents the lateral pressure forcing, and it can be interpreted similarly.

3.4 Analogy: mass equation

The preceding interpretation of the momentum equation has an analogue in the context of conservation of mass, and the purpose of this subsection is to describe this analogy.

When the mass is computed for the individual layers in the fluid, one option is to use a prognostic equation for the baroclinic pressure difference $\Delta p_r'$, which was derived by Bleck and Smith (1990). However, this equation is not in conservation form, and computational experience has shown that the total mass in individual layers can vary over time, even in situations where the governing equations do not allow mass transport between layers. This is an undesirable situation for long-term simulations, such as those arising in climate modeling.

An alternative is to use the layer mass equation (1), $\partial(\Delta p_r)/\partial t + \nabla \cdot (\mathbf{u}_r \Delta p_r) = 0$, and at the end of each baroclinic time step adjust the lateral mass fluxes in each layer so that the computed values of Δp_r are consistent with the results computed with the barotropic equations. Various procedures of this nature have been used by Hallberg (1997) and Higdon (2005) and by John Dukowicz, Mats Bentsen, and Paul Schopf (private communications). A summary is the following.

Assume that the mass equation (1) is discretized with an equation of the form

$$\begin{aligned}
 (\Delta p_r)_{ij}^{n+1} &= (\Delta p_r)_{ij}^n + \frac{\Delta t}{\Delta x} \left((F_r)_{i-\frac{1}{2},j}^n - (F_r)_{i+\frac{1}{2},j}^n \right) \\
 &\quad + \frac{\Delta t}{\Delta y} \left((G_r)_{i,j-\frac{1}{2}}^n - (G_r)_{i,j+\frac{1}{2}}^n \right), \tag{19}
 \end{aligned}$$

where $(\Delta p_r)_{ij}^n$ denotes an approximation to Δp_r in the mass cell centered at (x_i, y_j) at time t_n , $(F_r)_{i-\frac{1}{2},j}^n$ is a numerical approximation to the mass flux $u_r \Delta p_r$ on the time interval $[t_n, t_{n+1}]$ at the cell edge corresponding to minimal x , and $(G_r)_{i,j-\frac{1}{2}}^n$ is a numerical approximation to the flux $v_r \Delta p_r$ at the edge corresponding to minimal y . Equation (19) is in conservation form and thus yields exact conservation of mass in each layer, up to roundoff error. However, the formulation (19), as stated, has two flaws. One is that the quantity Δp_r can vary on the fast time scale associated with external gravity waves, and the long baroclinic time step Δt could then allow numerical instability. A second

flaw is that the vertical sum of (19) yields a statement about the water column mass (4), $p_b = \sum_{r=1}^R \Delta p_r$, which may or may not be consistent with the results produced by the barotropic mass equation (9).

These problems are remedied by enforcing consistency, as follows. Assume that the barotropic mass equation (9), $\partial p_b / \partial t + \nabla \cdot (p_b \bar{\mathbf{u}}) = 0$, is approximated using the barotropic time step $\Delta t / N$ and the equation

$$\begin{aligned} (p_b)_{ij}^{n,m+1} &= (p_b)_{ij}^{n,m} + \frac{1}{N} \frac{\Delta t}{\Delta x} \left(D_{i-\frac{1}{2},j}^{n,m} - D_{i+\frac{1}{2},j}^{n,m} \right) \\ &\quad + \frac{1}{N} \frac{\Delta t}{\Delta y} \left(E_{i,j-\frac{1}{2}}^{n,m} - E_{i,j+\frac{1}{2}}^{n,m} \right), \end{aligned} \quad (20)$$

where D and E denote approximations to the barotropic mass fluxes $p_b \bar{u}$ and $p_b \bar{v}$, respectively. Applying this relation over all barotropic substeps (i.e., for $0 \leq m \leq N-1$) leads to an expression for p_b^{n+1} in terms of p_b^n and time averages (i.e., time integrals) of the barotropic fluxes D and E , in analogy to equation (16) for barotropic momentum. However, a vertical sum of the discrete layer equation (19) produces an expression for p_b^{n+1} in terms of p_b^n and vertical sums of the lateral mass fluxes F and G . In order for these two relations to produce the same results, it suffices to have the vertical sum of the layer fluxes equal to the time averages of the barotropic fluxes. In general, this does not happen, due to different numerical methods being used for the layer equations and barotropic equations. However, consistency can be attained by adjusting the layer fluxes to obtain the numerical method

$$\begin{aligned} (\Delta p_r)_{ij}^{n+1} &= (\Delta p_r)_{ij}^n + \frac{\Delta t}{\Delta x} \left((\tilde{F}_r)_{i-\frac{1}{2},j}^n - (\tilde{F}_r)_{i+\frac{1}{2},j}^n \right) \\ &\quad + \frac{\Delta t}{\Delta y} \left((\tilde{G}_r)_{i,j-\frac{1}{2}}^n - (\tilde{G}_r)_{i,j+\frac{1}{2}}^n \right), \end{aligned} \quad (21)$$

where

$$\tilde{F}_r = F_r + \frac{\Delta p_r}{p_b} \left(- \sum_{k=1}^R F_k + \frac{1}{N} \sum_{m=0}^{N-1} D^{n,m} \right) \quad (22)$$

and a similar definition holds for \tilde{G}_r . Equation (22) implies

$$\sum_{r=1}^R \tilde{F}_r = \frac{1}{N} \sum_{m=0}^{N-1} D^{n,m},$$

as required.

In the formulation by Higdon (2005), the values of Δp_r in (22) are chosen in an upwind manner, based on the sign of the flux difference in the parentheses in (22). This is done in order to ensure that the flux adjustment does not produce negative layer thicknesses. In addition, the values of Δp_r are tentative values for time t_{n+1} obtained from equation (19), with the unadjusted fluxes F_r and G_r . The adjustments to F_r in (22), and the analogous adjustments to G_r , are then applied in a separate step. Finally, the upwind values of Δp_r are restricted to be values of “available mass” as defined by Higdon (2005), in order to avoid spurious mass transports near variable bottom topography, and the value of p_b in (22) is the vertical sum of those values of Δp_r .

The expression (22) for the adjusted mass flux \tilde{F}_r has a structure analogous to the expressions for the Coriolis and pressure terms in the discrete baroclinic momentum equation (18). In general, the mass flux F_r can vary on both the fast and slow time scales, as this flux arises from approximating the unsplit mass equation (1). The vertical sum $\sum_{k=1}^R F_k$ is the barotropic part of this flux; it approximates $\sum_{k=1}^R u_k \Delta p_k = p_b \bar{u}$, and it varies on the fast time scale but is not resolved on the baroclinic time grid. On the other hand, the barotropic flux $D^{n,m}$ approximates $(p_b \bar{u})^{n,m}$ and is well-resolved in the finer barotropic time grid. The flux adjustment stated in (22) amounts to replacing unresolved fast forcing with the time-integrated effect of well-resolved forcing, with the difference between the two distributed proportionately over all layers. This is an analogue of the interpretation of the discrete baroclinic momentum equation given in Section 3.3.

4 Numerical computations

This section describes some numerical computations that compare the two approaches to barotropic-baroclinic splitting described above.

4.1 Algorithms and codes used in the computations

These computations employ three different codes, as follows.

(I) *Split variables*. This code implements the methods described by Higdon (2005). In this case the dependent variable in the baroclinic momentum equation is $\mathbf{u}'_r \Delta p'_r$, as described in Section 2.2 of the present paper.

(II) *Unsplit (original) variables*. For reasons given in Section 2.2, the split-variable code was modified so that the dependent variable in the baroclinic momentum equation is $\mathbf{u}_r \Delta p_r$, and the barotropic-baroclinic splitting uses the

formulation described in Section 3.

(III) *No barotropic-baroclinic splitting.* Code (II) was modified by deleting everything related to barotropic-baroclinic splitting, and the resulting code solves the governing equations explicitly with a short time step determined by the speed of external gravity waves. Although it is much simpler, this code requires far longer computing times than codes (I) and (II). However, the lack of barotropic-baroclinic splitting means that there is no splitting error, so this code provides reference solutions that can be used to compare the two formulations of barotropic-baroclinic splitting represented in codes (I) and (II).

For spatial discretization, all three codes use the C-grid. Mass quantities are defined at the centers of grid cells, and normal components of velocity are defined at the centers of edges of grid cells.

For solving the equation for conservation of mass, codes (I) and (II) both employ the method described in Section 3.4; unsplit (original) variables are used, and the lateral mass fluxes in the layers are adjusted to ensure compatibility with the barotropic mass flux. Code (III) solves the mass equation (1) directly, without any reference to splitting. All three codes thus conserve the discrete mass in each layer exactly, up to roundoff error.

For a time-stepping method, codes (I) and (II) both use a two-level time-stepping method developed by Higdon (2005) for barotropic-baroclinic splitting. After an initial prediction from baroclinic time t_n to time t_{n+1} , the correction steps involve centered time differencing and averaging between those two time levels. In a linearized stability analysis involving a fluid with two layers (Higdon, 2002), the time-stepping method is stable, subject to the usual Courant-Friedrichs-Lewy condition. In addition, the method is very nearly nondissipative; for spatial Fourier modes with a time dependence of the form λ^n , the two-layer stability analysis yields $1 - \epsilon < |\lambda| \leq 1$, with ϵ typically near 10^{-4} . Code (III) uses, in each layer, the natural reduction of the two-level time-stepping method to the case of no barotropic-baroclinic splitting.

For the computations reported here, the execution times for code (II) are typically about ten to fifteen percent greater than the times for code (I), when the codes use the same baroclinic and barotropic time steps. This discrepancy is due, at least in part, to extensive computation of time averages of various quantities over the barotropic substeps of a baroclinic interval. Some examples of such averaging are described in Section 4.2(c). Another example is related to the concepts of available mass and available momentum that are used to prevent nonphysical transports of mass and momentum near steep bottom topography (Higdon, 2005). This process requires that the elevations of layer interfaces be compared to the elevation of the bottom topography. In code (II),

time averages are used when computing the elevations of the layer interfaces both at mass points and at momentum points.

4.2 Sources of damping

A significant issue in the subsequent discussions is the presence or absence of grid noise and whether viscosity must be introduced explicitly in order to dampen such noise. For the sake of that discussion, we list the sources of damping that are already present in the algorithms used here.

(a) *Implicit viscosity due to the advection scheme.* The advection terms in the mass and momentum equations are implemented with the basic form of the multidimensional positive definite advection transport algorithm (MP-DATA) (Smolarkiewicz and Margolin, 1998), and like any advection scheme this method causes some diffusion in the computed solution. As implemented in the present computations, the method consists of an upwind step followed by two antidiffusive iterations. This method was implemented in the same manner in all three codes.

(b) *Time averaging in the barotropic solver.* The forcing in the barotropic momentum equation includes quantities that are held constant in time on each baroclinic time interval but change slightly from one baroclinic interval to the next. The resulting discontinuity in forcing can induce temporal oscillations in the barotropic solution (Higdon, 2005). This problem was diagnosed in a test computation by plotting the barotropic variables as functions of time at a fixed position in space; during the first baroclinic time interval after startup, these quantities varied smoothly over the barotropic substeps, but at the beginning of the second baroclinic interval the barotropic variables suddenly displayed sharp oscillations. As described by Higdon (2005), these oscillations are suppressed by the following procedure. At the beginning of each baroclinic interval, solve the barotropic equations for one (short) barotropic step and then compute a two-point time average of each of the dependent variables. This suppresses grid-scale oscillations, but it also gives a solution at a half-step. Repeat the procedure once more, in order to give a solution at an integer step, and then proceed for the rest of the baroclinic time interval without any more time averaging. This procedure is used in codes (I) and (II), but it does not apply to code (III), which does not use barotropic-baroclinic splitting. The barotropic equations in codes (I) and (II) are solved with a nondissipative forward-backward algorithm.

(c) *Time averaging of barotropic forcing in the baroclinic equations.* During the correction steps for the baroclinic equations, time averages are used in various terms in order to represent the time-integrated effects of those terms. For

example, in codes (I) and (II) the advective velocity for mass and momentum consists of the average of baroclinic velocity between times t_n and t_{n+1} plus the average of barotropic velocity over all of the barotropic substeps of the interval $[t_n, t_{n+1}]$. In the case of code (II), the momentum density $\mathbf{u}_r \Delta p_r$ in the advection term (3) is composed of baroclinic quantities \mathbf{u}'_r and $\Delta p'_r$ at time t_n and weighted time averages of the barotropic variables $\bar{\mathbf{u}}$ and η (see (6) and (7)) about several barotropic steps centered at t_n . For the sake of computing time-averaged quantities in code (II), the baroclinic quantities \mathbf{u}'_r and $\Delta p'_r$ are extracted from the computed values of \mathbf{u}_r , Δp_r , and $p_b \bar{\mathbf{u}}$ by using the relations (6) and (7); the advective velocity and the momentum density are then reconstructed from the extracted baroclinic quantities and time averages of barotropic variables. Essentially, the advective velocity and the momentum density are taken apart and then re-built by using averaged barotropic variables. During the prediction step for codes (I) and (II) the advective velocity and momentum density are evaluated at time t_n , due to a lack of further information at that stage of the computation.

(d) *Bottom friction and interface friction.* In most of the tests described here, the model is subjected to steady wind forcing. In such cases, the model employs friction at the bottom of the fluid and friction between fluid layers in order to prevent the fluid velocities from increasing without bound.

A time-averaging procedure that was tested but not employed in these computations is the following. In the case of code (II), which uses original (unsplit) variables, it is necessary to enforce the conditions that the vertical sum of the layer momenta $\mathbf{u}_r \Delta p_r$ equals the barotropic momentum $p_b \bar{\mathbf{u}}$ and the vertical sum of the layer masses Δp_r equals the barotropic mass p_b . For the barotropic momentum and mass, one could either use instantaneous values at baroclinic time t_{n+1} or weighted time averages over several barotropic steps centered about t_{n+1} . The latter procedure requires that the barotropic equations be integrated beyond time t_{n+1} , which imposes an extra computational cost. Weighted averaging was used successfully by Shchepetkin and McWilliams (2005) in their setting. However, in some experiments, this procedure did not seem to make any difference in the computed solutions for the context considered in the present paper, so it was not used further. Instead, the vertical sums are matched to instantaneous values of barotropic momentum and mass at time t_{n+1} .

Unless otherwise stated, the explicit horizontal viscosity is zero, i.e., $A_H = 0$ in the term \mathbf{D}_r in the momentum equation (2).

4.3 Some comparisons involving analytical solutions

This subsection summarizes the results of some computations involving model problems for which analytical solutions are known. These problems are formulated in Higdon (2005), and the computational results obtained with code (I) are described in detail in that paper. For these problems, code (II) yields solutions which are very similar to those obtained with code (I), so the results are only summarized here. More substantial differences appear in the computations reported in Section 4.4, and those are described in more detail.

(a) *Upwelling and downwelling in a straight channel.* In this case the fluid consists of two layers occupying an infinite straight channel with sloping sides. The system starts at rest, and a constant wind stress is applied along the length of the channel. Due to the Coriolis effect, the fluid in the upper layer shifts laterally. Along one side of the channel, the layer interface moves downward along the sloping bottom topography, and along the other side the interface moves upward so that the upper layer has negligible thickness over a portion of the fluid domain. All three codes produce solutions that converge to an analytical steady state solution as $t \rightarrow \infty$.

(b) *Upwelling and downwelling in a circular channel.* Here, the channel in (a) is bent into a circle, and the wind stress is directed along the length of the channel. Again, the fluid in the upper layer shifts laterally. Each code produces a solution that is close to an analytical steady state, for sufficiently large t . However, the steady state appears to be physically unstable, as all solutions display wavy patterns in contour plots of the free-surface elevation. The solutions produced by the three codes are very similar.

(c) *External and internal Rossby waves in a channel.* In this problem the fluid consists of two layers in an infinite straight channel with vertical sides and level bottom. The wind stress, bottom stress, and friction between layers are all zero, and the flow is assumed to be a small perturbation of a stationary state. The Coriolis parameter varies linearly across the channel (y -direction). To construct analytical solutions, discretize in space on a C-grid and leave time continuous, and apply Fourier transforms in the along-channel direction (x) and time. For fixed wavenumber in x , the result is a matrix eigenvalue problem for which the eigenvalues yield time frequencies and the eigenvectors yield dependences in y for modal solutions. These solutions are exact solutions of the semi-discrete problem, up to the numerical accuracy in computing eigenvalues and eigenvectors of the associated matrix. The modal solutions include Rossby, Poincaré, and Kelvin waves, both external and internal. The values of such solutions at a fixed time provide initial data for testing time-stepping methods; external modes primarily test the barotropic solver, and internal modes primarily test the baroclinic routines. Rossby waves are considered in

the present computations. For the first internal Rossby wave and two different wavenumbers in x , codes (I) and (II) give results that are nearly identical. For the first external Rossby mode and two different wavenumbers in x , the results from code (II) are slightly less accurate. One source of error could be some spatial averaging that is required to implement the coefficient p_b of $\overline{\nabla M}$ in the barotropic momentum equation (10) on the C-grid.

4.4 *Double-gyre circulations*

The computations described in the present subsection involve double-gyre circulations for which eddies and meanders are prominent. Analytical solutions are not available in this case. In Sections 4.4.1 and 4.4.2 the results obtained with codes (I) and (II) are compared to the solutions obtained with code (III), which does not use a barotropic-baroclinic splitting. The computations described in Section 4.4.4 give a demonstration of the split-variable code (I) on a larger domain over a longer time interval.

In these computations the spatial domain is square and has the form $0 < x < L$, $0 < y < L$, with x and y eastward and northward, respectively. The bottom of the fluid domain is level. The Coriolis term is given by the β -plane approximation $f = f_0 + \beta(y - L/2)$, where f_0 and $\beta = \partial f / \partial y$ are the values associated with latitude 45° N on the earth. The system begins at rest and is forced by a steady wind stress $\tau = (\tau_x, \tau_y)$, where $\tau_y = 0$ and $\tau_x = \tau_{max} \cos\left(\frac{2\pi}{L}(y - L/2)\right)$, with $\tau_{max} = 0.1$ N/m². The wind stress thus points eastward in the middle of the domain and westward at the northern and southern boundaries. This pattern of wind stress generates a double-gyre circulation, with a counter-clockwise circulation in the northern half of the domain, a clockwise circulation in the southern half, and intense western boundary currents.

4.4.1 *Test 1. Thin upper layer*

The computations described in this subsection employ a 200×200 array of mass cells, which includes a border of massless cells to generate a solid boundary. The grid cells have dimensions $\Delta x = \Delta y = 10$ km, so the size of the fluid domain is 1980 km by 1980 km. This relatively small domain is used here because of the computational cost of computing reference solutions with code (III). The fluid consists of two layers with specific volumes 0.976×10^{-3} m³/kg and 0.972×10^{-3} m³/kg. The friction between layers and the stress at the bottom of the fluid are parameterized in the manner described by Higdon (2005). The upper and lower layers are initially 100 meters thick and 900 meters thick, respectively. Due to lateral mass transports in the upper layer, the lower layer

outcrops to the surface in a portion of the northern half of the domain, i.e., the upper layer has negligible thickness on that portion.

With upper and lower layer thicknesses equal to 100 and 900 meters, respectively, and with the given densities in the layers, the speed of external gravity waves is approximately 99 m/sec and the speed of internal gravity waves is $c_1 \approx 1.9$ m/sec. (Formulas are given, for example, by Higdon (2002).) The internal Rossby radius c_1/f is approximately 19 km, or about two grid intervals.

In the case of code (I), the time step used for the baroclinic equations is $\Delta t = 2400$ seconds (36 steps per day), and the nominal barotropic step is 70 seconds. Here, “nominal” means that the number N of barotropic sub-steps is the smallest integer N such that $\Delta t/N$ is less than or equal to the nominal barotropic step. In this case, the actual barotropic time step is approximately 68.6 seconds. In the case of code (III), which does not use a barotropic-baroclinic splitting, the time step is 70 seconds. Code (I) was run for a total of 4000 model days, and code (III) was run for 2000 model days. For both codes, the explicit horizontal viscosity was set to zero, i.e., $A_H = 0$.

Code (II) was initially run with $A_H = 0$ and with the same time steps as code (I). However, between model days 604 and 605 the computation produced a negative value of layer thickness at some location in the upper layer, and the computation was terminated automatically. Code (II) uses the same procedures for handling thin layers as codes (I) and (III), and these are described by Higdon (2005). An inspection of the solution at day 604 suggested that the problem could be related to large fluid velocities in the western boundary layer, but in addition some grid-scale noise in the velocity field was evident in an otherwise quiescent region near the southwestern corner of the fluid domain, well away from the location of the largest velocities in the boundary layer. In contrast, no grid noise was found in the solutions obtained with codes (I) and (III). If the baroclinic time step for code (II) is reduced from 2400 seconds to 2160 seconds (40 steps per day) but the nominal barotropic step of 70 seconds and the zero viscosity $A_H = 0$ are retained, then the code runs for long times without grid noise. Another option for code (II) is to retain the baroclinic time step of 2400 seconds but use nonzero horizontal viscosity; the value $A_H = 100$ m²/sec is sufficient to eliminate the noise and crash, but the value $A_H = 10$ m²/sec is not. The Reynolds number corresponding to $A_H = 100$ m²/sec, and velocity scale $U = 1$ m/sec and length scale $\Delta x = \Delta y = 10^4$ m, is $Re = U\Delta x/A_H = 100$. This grid Reynolds number does not include the effect of the implicit viscosity induced by the advection terms. Although the behavior of code (II) is of concern in the present test, the grid noise is more pervasive in the computations described in Section 4.4.2, and a discussion of grid noise is deferred to that section.

For the model configuration considered here, the circulation pattern is ini-

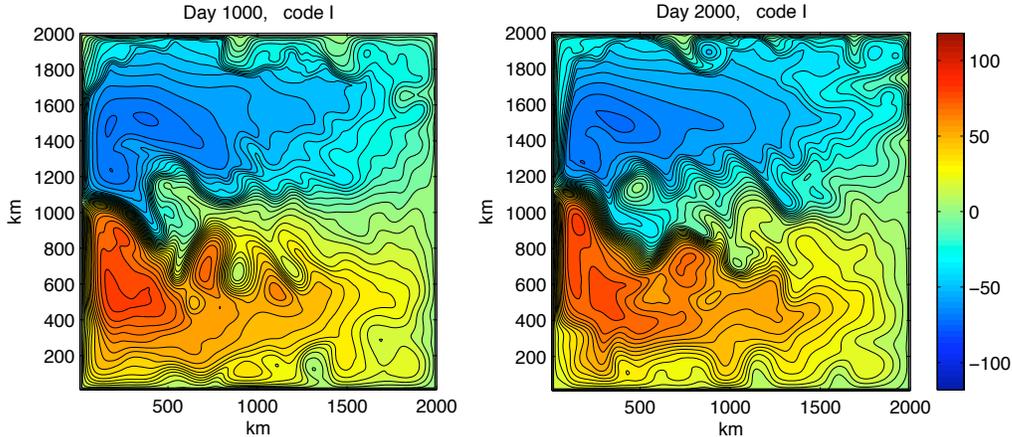


Fig. 1. Test 1 (thin upper layer). The graphs are contour plots of the free-surface elevations at model days 1000 and 2000, as computed with the code that uses split variables. The contour interval is 5 centimeters, with lower elevations in the northern region and higher elevations in the southern region. The (maximum, minimum) elevations at days 1000 and 2000 are approximately $(83, -71)$ and $(83, -72)$ centimeters, respectively, where elevation zero corresponds to the rest state.

tially smooth, but over time the flow develops meanders and eddies. Figures 1 through 4 show contour plots of the elevation of the free surface at the top of the fluid at days 1000 and 2000, for each computation. Figure 1 shows the results obtained with code (I), Figures 2 and 3 show the results obtained with the two configurations of code (II) described above, and Figure 4 was produced by code (III). These plots are instantaneous snapshots, without any time-averaging of data. Due to the approximate geostrophic balance in the flow, the curves of constant elevation are approximate streamlines for the flow at the top of the fluid. By day 1000 the lower layer has outcropped to the surface over a portion of the northern half of the domain. In each plot the outcrop region corresponds approximately with the region where the flow lines are relatively smooth; on that region the governing equations reduce to the barotropic equations, which are nearly linear.

The solutions obtained with these codes are not point-wise identical, due to the sensitivity of the flow to perturbations. However, the general flow patterns are the same, and the levels of eddy activity are similar. This relation between the solutions holds throughout the 2000-day comparison period.

Another comparison of the solutions is given by Figure 5, which shows plots of the absolute values of the discrete Fourier transforms of the kinetic energy in each solution. More precisely, in each graph the quantity plotted is $\log_{10}(|\hat{E}(k, \ell, t)| / \max |\hat{E}|)$ where $E(x, y, t)$ is the kinetic energy per unit horizontal area (summed vertically over both layers), $\hat{E}(k, \ell, t)$ is the discrete Fourier transform with respect to (x, y) for fixed t , and $\max |\hat{E}|$ denotes the maximum over all wavenumbers for that t . The maximum value plotted is

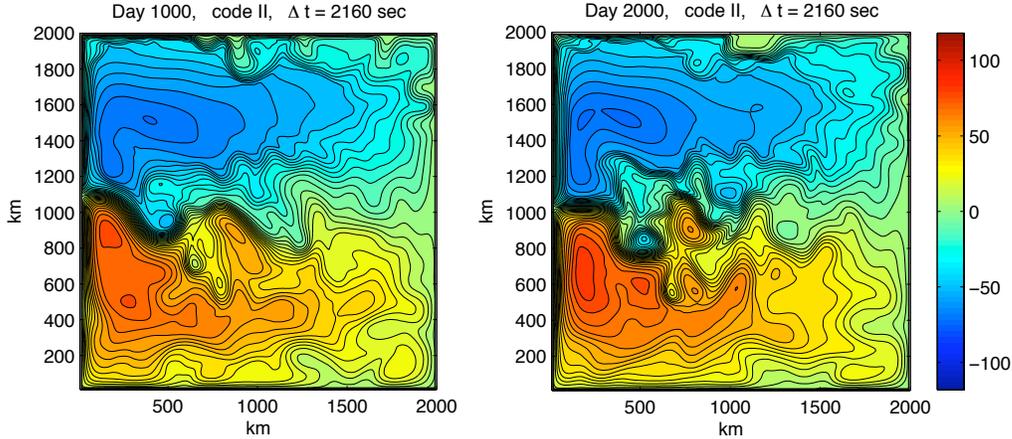


Fig. 2. Test 1. Free-surface elevations at days 1000 and 2000, as computed with the version that uses unsplit (original) variables with $\Delta t = 2160$ seconds and viscosity $A_H = 0$. The (maximum, minimum) elevations at days 1000 and 2000 are approximately $(78, -70)$ and $(83, -71)$ centimeters, respectively.

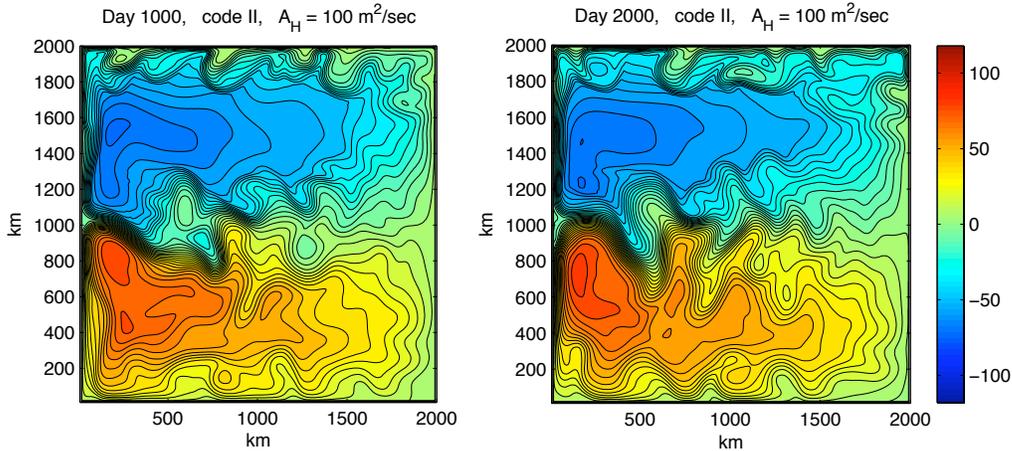


Fig. 3. Test 1. Free-surface elevations at days 1000 and 2000, as computed with the version that uses unsplit (original) variables with $\Delta t = 2400$ seconds and viscosity $A_H = 100 \text{ m}^2/\text{sec}$. The (maximum, minimum) elevations at days 1000 and 2000 are approximately $(80, -72)$ and $(84, -70)$ centimeters, respectively.

thus 0, and the values shown in these plots range from 0 down to -2 . Since the Fourier transform is normalized relative to the maximum, Figure 5 does not account for differences in the total kinetic energy between different solutions and different times; differences in total kinetic energy are discussed below. The discrete Fourier transforms are defined for wavenumbers (k, ℓ) for which $|k\Delta x| \leq \pi$ and $|\ell\Delta y| \leq \pi$, but only the positive wavenumbers need to be shown, due to symmetries in $|\hat{E}|$. The extreme values $|k\Delta x| = \pi$ and $|\ell\Delta y| = \pi$ correspond to sawtooth modes with wavelengths $2\Delta x$ and $2\Delta y$ in x and y , respectively. The present plots are restricted to the range $0 \leq k\Delta x \leq \pi/2$, $0 \leq \ell\Delta y \leq \pi/2$ in order to display better the lower wavenumbers, where

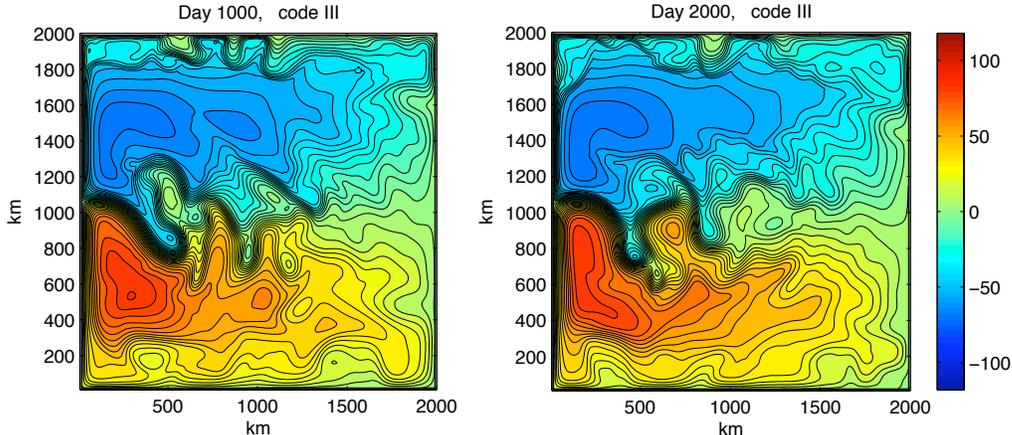


Fig. 4. Test 1. Free-surface elevations at days 1000 and 2000, as computed with the code that uses no barotropic-baroclinic splitting. The (maximum, minimum) elevations at days 1000 and 2000 are approximately (85, -69) and (82, -70) centimeters, respectively.

most of the kinetic energy resides. At day 20 the energy is confined mainly to regions where either $k \approx 0$ or $\ell \approx 0$, which correspond to modes that are nearly constant in x or constant in y , respectively. The flow at that time (not shown here) consists mainly of a smooth eastward flow in the middle of the domain, smooth westward flows near the northern and southern boundaries, and strong northward and southward flows near and parallel to the western boundary. However, as t increases, the flow develops meanders and eddies, and in Figure 5 this development is represented by a migration of energy through wavenumber space. The figure indicates that the relative distributions of kinetic energy in wavenumber space are similar for the three codes.

Of the two codes that employ barotropic-baroclinic splitting, code (I) behaves more stably than code (II), in the present test. These codes use split variables and unsplit (original) variables, respectively. This comparison raises the question of whether the split-variable formulation (8) of the momentum equation embodies some kind of implicit damping that is not present in the formulation with unsplit variables. Accordingly, Figure 6 shows plots of the total kinetic energy in the system as a function of time. The top graph shows that the kinetic energy with code (I) is very similar to that of code (III), over the first 2000 days. The middle graph shows that codes (I) and (II) yield kinetic energies that are very similar, when the latter code is used with the reduced time step and $A_H = 0$. However, the bottom graph shows that the kinetic energy with code (II) is noticeably less than the kinetic energy with code (I) when the viscosity $A_H = 100 \text{ m}^2/\text{sec}$ is used in code (II). These results do not suggest any extra damping in the split-variable formulation, but they do show the damping effect of nonzero viscosity if it is used to maintain stability in code (II).

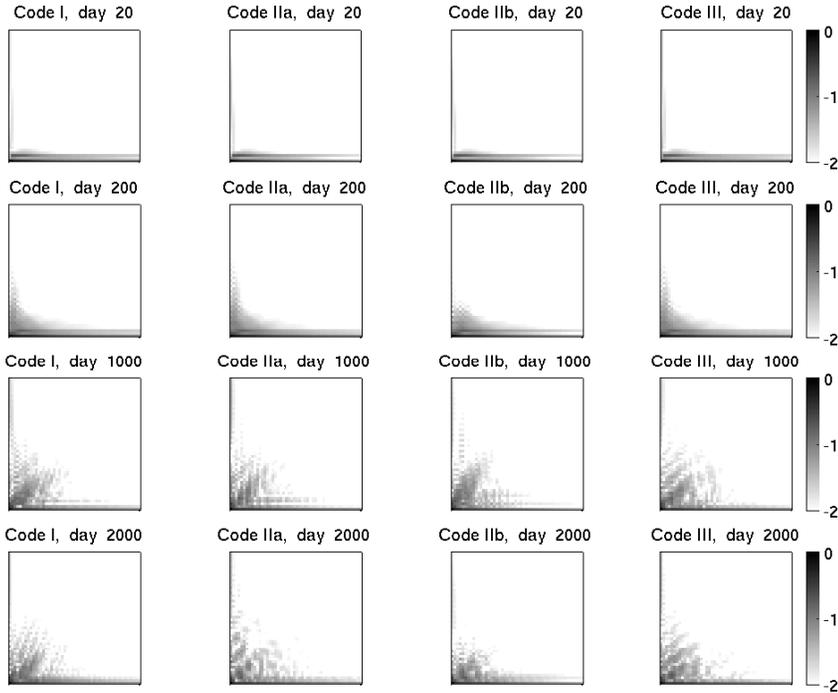


Fig. 5. Test 1. Spectra of kinetic energy. Each plot shows the absolute value of the Fourier transform of the kinetic energy per unit horizontal area, regarded as a function of (x, y) for fixed t . The horizontal axes represent values of k for which $0 \leq k\Delta x \leq \pi/2$, and the vertical axes represent values of ℓ for which $0 \leq \ell\Delta y \leq \pi/2$. Here, k and ℓ are wavenumbers with respect to x and y , respectively. The quantities plotted here are values of $\log_{10}(|\hat{E}|/\max|\hat{E}|)$ over the range from 0 down to -2 . In the second column of plots, the notation “Code IIa” refers to code (II) with the reduced time step $\Delta t = 2160$ seconds and viscosity $A_H = 0$. The third column was produced by code (II) with time step $\Delta t = 2400$ seconds and viscosity $A_H = 100 \text{ m}^2/\text{sec}$. The first and fourth columns correspond to codes (I) and (III), respectively.

4.4.2 Test 2. Layers of equal thickness

Next assume that both layers are initially 500 meters thick, with all other aspects of the problem configuration remaining the same. In this case, the lower layer does not outcrop to the surface, so there is no possibility of any numerical difficulties arising from the presence of a thin fluid layer. For the present configuration, the speed of internal gravity waves is approximately 3.2 m/sec.

In the case of code (I), the time step used for the baroclinic equations is 1800 seconds (48 steps per day), and the nominal barotropic step is 70 seconds. With code (III), which does not use a barotropic-baroclinic splitting, the time step is 70 seconds. Both codes were run with horizontal viscosity $A_H = 0$. The solutions obtained with codes (I) and (III) are illustrated in Figures 7 and 8.

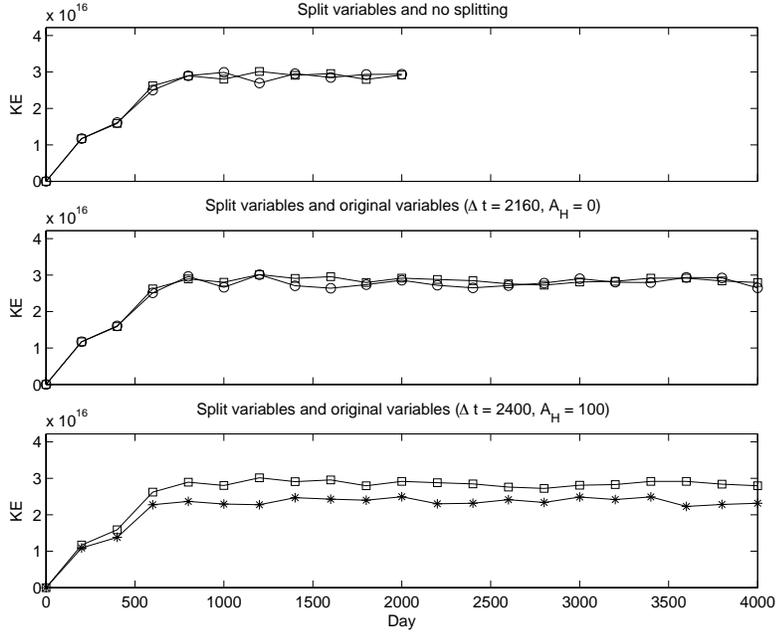


Fig. 6. Test 1. Total kinetic energy in the system, as a function of time. In all of the plots, the squares indicate values of kinetic energy obtained with code (I), which uses split variables. In the top graph, the circles indicate values obtained with code (III), which uses no barotropic-baroclinic splitting. In the middle graph, the circles indicate values obtained with code (II), which uses unsplit (original) variables, when that code is used with a reduced baroclinic time step and zero horizontal viscosity. In the bottom graph, the asterisks indicate values obtained with code (II) when it uses viscosity $A_H = 100 \text{ m}^2/\text{sec}$ and the same time steps as code (I).

The figures show the solutions at model days 400 and 1000; the computations were actually run to day 2000, but the results between days 1000 and 2000 are qualitatively similar to the solutions at day 1000. The solutions produced by the two codes are not point-wise identical, but the general flow patterns are the same, and the levels of eddy activity are similar. Plots of spectra of kinetic energy, not shown here, are similar.

Code (II), which uses unsplit (original) variables, was run to model day 2000 with the same time steps as code (I) and with $A_H = 0$. However, grid-scale numerical noise in the solution becomes evident by day 100, and the noise is distributed widely throughout the fluid domain. Figure 9 shows the solution at day 400, which is the first day for which solutions are shown in the preceding two plots. By that time, the noise has become so strong that a contour plot of the free-surface elevation is nearly unintelligible. For the sake of producing a readable contour plot, the elevation at that time was filtered with simple four-point spatial averages. When such a filter is applied to a checkerboard +1/-1 pattern the result is zero, so the effect of the filter in the present case is to suppress the grid-scale noise but leave larger features intact. (The filtering was performed off-line and did not affect the computed solution at later times.)

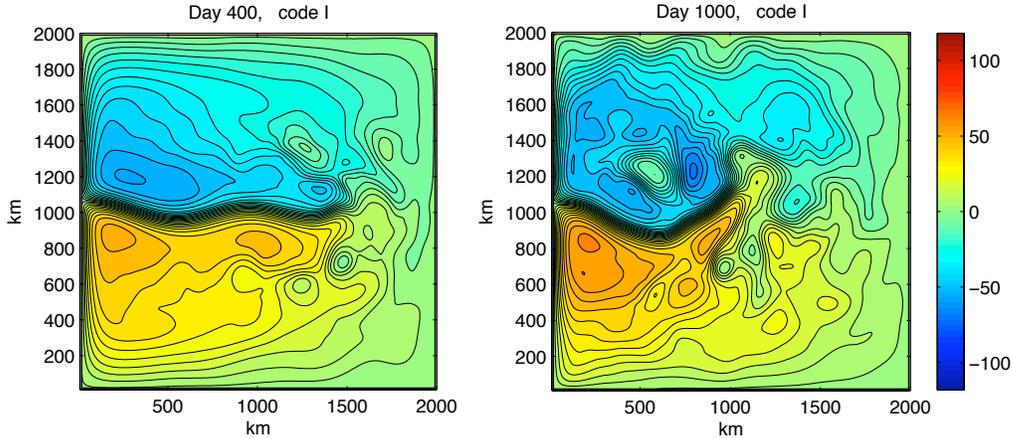


Fig. 7. Test 2 (layers initially of equal thickness). The graphs are contour plots of the free-surface elevations at model days 400 and 1000, as computed with the code that uses split variables. The plotting format is the same as in Figures 1–4.

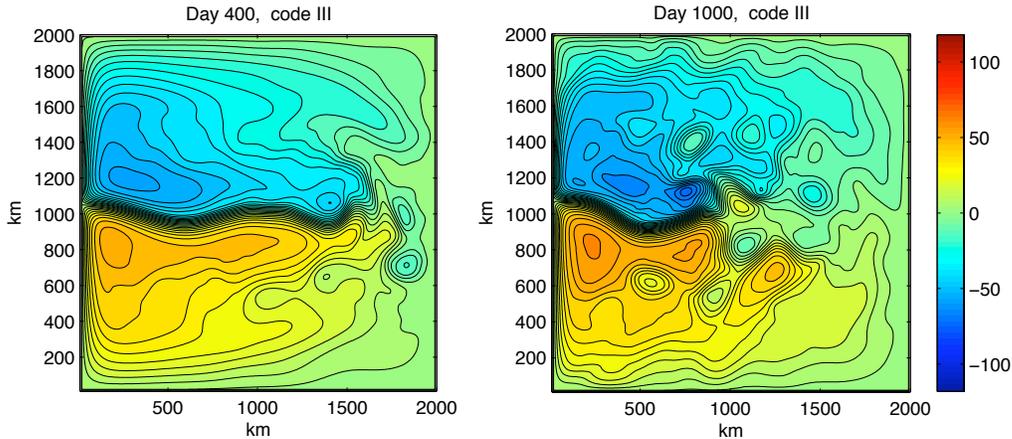


Fig. 8. Test 2. Free-surface elevations at days 400 and 1000, as computed with the code that uses no barotropic-baroclinic splitting.

A comparison of the left frame of Figure 9 with the left frames of Figures 7 and 8 reveals that the present solution disagrees radically with the solutions computed with codes (I) and (III).

The right frame of Figure 9 shows a close-up view of the velocity field near the southwestern corner of the fluid domain. Roughly speaking, the velocity vectors are shown at each grid point. To be more precise, the present computations use the staggered C-grid, and with this grid the components of velocity are defined at different locations. For the sake of producing a plot, at each mass cell the value of u at the western edge of the cell and the value of v at the southern edge of the cell are combined to produce a vector defined at the center of the cell. Spatial averaging would produce values of u and v defined at cell centers, but this would conceal the grid noise. The grid-scale noise seen in Figure 9 is typical of what is present throughout the spatial domain.

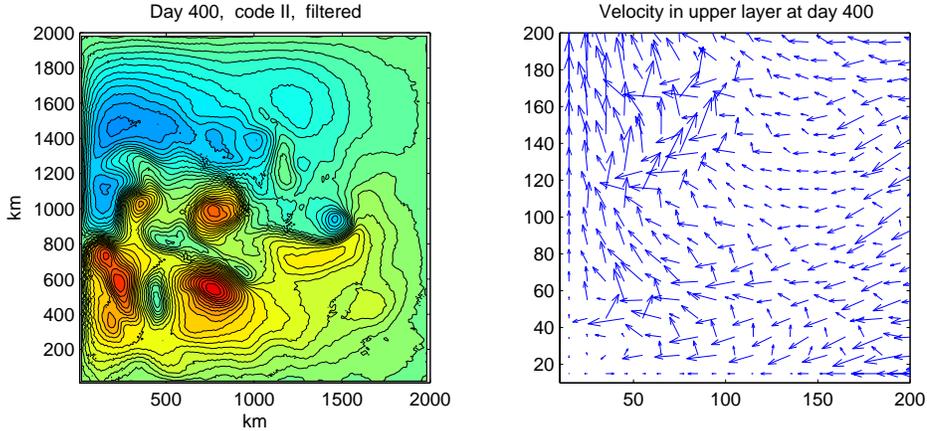


Fig. 9. Test 2. Solution at day 400, as computed with the code that uses unsplit (original) variables. In this case the free-surface elevation contains too much numerical noise to yield an intelligible contour plot. Instead, for purposes of plotting, the elevation was filtered with a four-point spatial average which suppresses grid-scale noise but leaves larger features intact. The result is displayed in the left frame. The right frame shows a close-up view of the velocity field near the southwestern corner, with velocity vectors shown at each grid point and without any filtering. This amount of noise is typical of what is seen throughout the fluid domain.

As with the computations described in Section 4.4.1, some options for code (II) are to reduce the baroclinic time step or introduce nonzero viscosity. Figure 10 shows the solution that is obtained when the baroclinic time step is reduced from 1800 seconds to 1440 seconds (60 steps per day). As before, the nominal barotropic step is 70 seconds and the horizontal viscosity is $A_H = 0$. In this case, the solution is free of grid noise. In another experiment with baroclinic time step 1600 seconds (54 steps per day), some grid noise is present, although it is not as strong as with time step 1800 seconds. The solution shown in Figure 10 does not agree point-wise with the solutions obtained with codes (I) and (III) which are shown in Figures 7 and 8, but the flow patterns and eddy characteristics are similar. It then appears that a sufficient reduction of the baroclinic time step can lead to a credible computed solution.

However, the situation is different if nonzero horizontal viscosity is used to suppress the grid noise while retaining the baroclinic time step of 1800 seconds. Experiments reveal that the value $A_H = 400 \text{ m}^2/\text{sec}$ is not enough to remove the noise, but the value $A_H = 1000 \text{ m}^2/\text{sec}$ is sufficient. With these values of viscosity, the solution is smoothed greatly and bears little resemblance to the solutions shown in Figures 7, 8, and 10. The results obtained with this code are not shown here, but instead Figure 11 shows the solution obtained with code (III) with viscosity $A_H = 400 \text{ m}^2/\text{sec}$ and baroclinic time step 1800 seconds. The only difference between the computation shown in Figure 11 and the one shown in Figure 8 is that the earlier computation uses $A_H = 0$. Code (III) is used for this illustration because it uses no barotropic-baroclinic

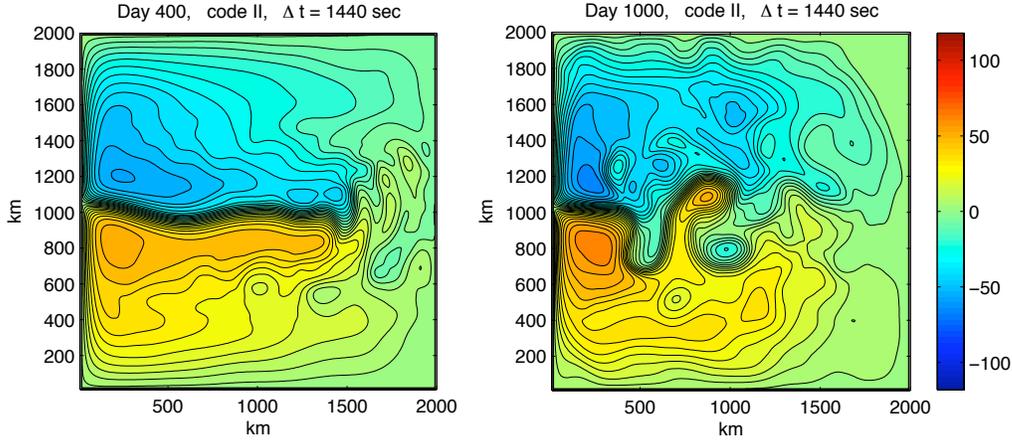


Fig. 10. Test 2. Free-surface elevations at days 400 and 1000, as computed with the code that uses unsplit (original) variables but with the baroclinic time step reduced from 1800 seconds to 1440 seconds.

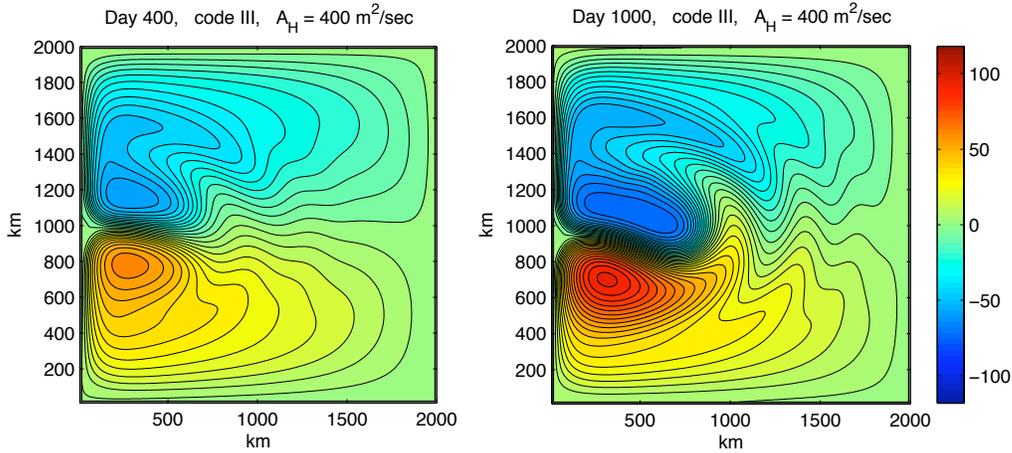


Fig. 11. Test 2. Effects of viscosity. These results were obtained with the version that does not use barotropic-baroclinic splitting, with horizontal viscosity $A_H = 400 \text{ m}^2/\text{sec}$. The only difference between this computation and the one shown in Figure 8 is that the latter uses $A_H = 0$.

splitting, and therefore the results shown in Figure 11 cannot be influenced by any errors associated with such a splitting. The amount of viscosity used for Figure 11 is not enough to remove the noise produced by code (II) without a reduced time step, but this amount of viscosity is sufficient to distort the solution severely. The solution produced by code (II) with $A_H = 1000 \text{ m}^2/\text{sec}$ has the same general character as the solution shown in Figure 11. This result indicates that viscosity is not a viable method for removing the grid noise in this example.

4.4.3 Further experiments with code (II) in Test 2.

Some additional experiments with code (II) were performed in the context of Test 2 in an attempt to identify the cause of the problems cited above. For the sake of brevity, this subsection only summarizes the results and does not include graphs.

In one set of experiments, the advection terms in the momentum equations were deactivated. All other aspects of the problem configuration were the same as described in Section 4.4.2, including $A_H = 0$. With this modification the code produces smooth, noise-free gyres, with western boundary currents that are narrower than those in the solutions described above. The deletion of the advection terms means that an advection scheme is not used in the momentum equations; this removes a source of numerical viscosity, and this change may explain the narrower boundary layer. The narrowness of the boundary currents produces fluid velocities that are greater than the ones seen in the solutions described in Section 4.4.2. With a baroclinic time step equal to 1800 seconds, an instability therefore originates in the western boundary layer, but elsewhere the solution is free of grid noise. If the baroclinic time step is reduced according to the velocity in the boundary layer, then the computation proceeds for long times without grid noise. This experience suggests that the problems with code (II) may be due to the momentum advection terms.

There then arises the question of whether the particular choice of numerical advection scheme may be a factor. As noted in Section 4.2, codes (I), (II), and (III) all use the standard form of the multidimensional positive definite advection transport algorithm (MPDATA) (Smolarkiewicz and Margolin, 1998) to implement the advection terms in the mass and momentum equations. As implemented in these three codes, the method consists of an upwind step, which is diffusive, followed by two antidiffusive corrections. In codes (I) and (III), this choice causes no apparent difficulty.

In another experiment with code (II), the antidiffusive iterations were turned off in the momentum equation, so that the momentum advection was simulated with the upwind method. As before, the baroclinic time step is 1800 seconds. Substantial grid noise remains, and again the general flow pattern is very different from the circulations computed with codes (I) and (III). This experience suggests that the problem may be more fundamental than the choice of momentum advection scheme. For this experiment the antidiffusive iterations were retained in the mass equation; however, these iterations are unlikely to cause the grid noise, as they were also used in the experiment in which the momentum advection terms were deactivated entirely, and that case the solution does not exhibit grid noise. In other words, with no momentum advection the solution is free of noise, but if momentum advection via the upwind method is introduced, then noise appears. In a further experiment, the

antidiffusive iterations were also turned off in the mass equation, so that the momentum and mass advection were both simulated with the upwind method. The resulting solution displays no grid noise, but it is smoothed severely.

In summary, these additional experiments suggest that the problems with code (II) arise from the momentum advection terms but are not due to the particular choice of the advection scheme that is used to implement those terms. A further discussion is given in Section 5.

4.4.4 Test 3. Larger domain, split-variable formulation

In the preceding double-gyre simulations, the split-variable code (I) yields better results than code (II). The present subsection provides a further demonstration of code (I).

As before, the spatial domain is square and has a level bottom, the Coriolis parameter is the β -plane approximation centered at 45° N, and the wind stress is a cosine pattern that generates a double-gyre circulation. In the preceding computations, the spatial domain is much smaller than real ocean basins, and there are only two layers. In the present case the computation uses a 500×500 array of mass cells, with a border of massless cells to generate a solid boundary. The cells have dimensions $\Delta x = \Delta y = 10$ km; the size of the fluid domain is then 4980 km by 4980 km, which is closer to a typical basin size. Three fluid layers are used, with specific volumes 0.976×10^{-3} , 0.974×10^{-3} , and 0.972×10^{-3} m³/kg and initial thicknesses 150, 300, and 550 meters, ordered from top to bottom. During the simulation, the second layer outcrops to the surface over a portion of the northern half of the domain. The model was run for a total of 6000 model days, with $A_H = 0$.

Figure 12 shows the elevation of the free surface at day 3000. The flow shows substantial meandering and eddy activity, especially near the western boundary, where the intense boundary currents separate from the side wall. Figure 13 shows close-up views of the region $0 < x < 2000$, $1500 < y < 3500$ at days 3000, 4000, 5000, and 6000. This region is centered on the western boundary, at $y = 2500$. The wind stress pattern is also centered at $y = 2500$, and the curl of the wind stress is zero along that line. Figure 14 shows the velocity field in the top layer at day 6000 on the smaller region $250 < x < 700$, $1800 < y < 2200$. The velocity field is shown at each grid point, or more precisely, for each mass cell the value of u at the western edge and the value of v at the southern edge are used to create a vector located at the center of the cell. Spatial averaging was not used, as this could conceal grid noise. The results shown in the figures indicate that the model supports vigorous eddy activity and strong shears, and it behaves stably over long times.

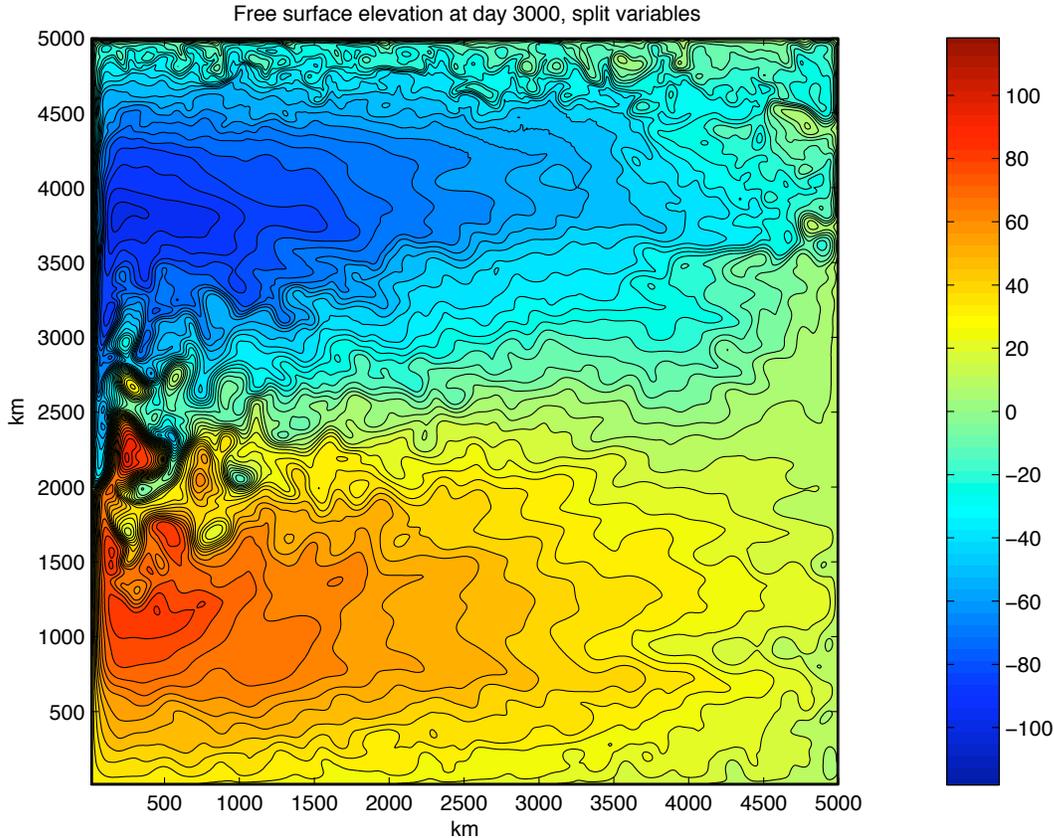


Fig. 12. Test 3 (larger domain). The graph is a contour plot of the free-surface elevation at model day 3000, as computed with the code that uses split variables. The maximum and minimum elevations are approximately 96 and -97 centimeters, respectively.

5 Discussion

In the double-gyre computations described in the preceding section, both formulations of the baroclinic momentum equation produce results that agree well with the reference solutions computed with code (III), which does not use a barotropic-baroclinic splitting. However, the formulation with unsplit (original) variables, as implemented in code (II), requires a reduction in the baroclinic time step in order to produce credible results. Regularizing the solution with horizontal viscosity A_H is not a viable option, for reasons stated in Sections 4.4.1 and 4.4.2. As noted in Section 4.4.3, the difficulties with code (II) are related to the momentum advection terms. A complete explanation of the problem is not known to this author at this time, but the following remarks may be pertinent.

In the case of the split-variable code (I), the momentum fluxes have the form $u_r(u'_r \Delta p'_r)$. The momentum density $u'_r \Delta p'_r$ varies mainly on the slower time

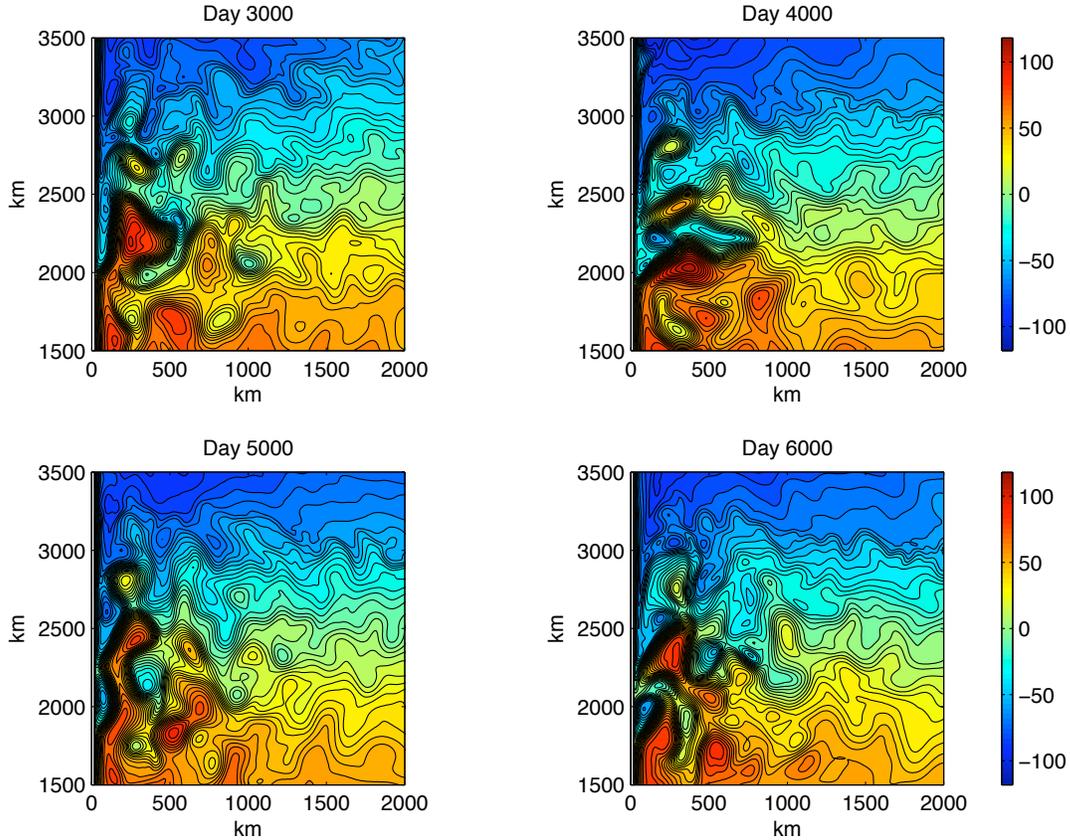


Fig. 13. Test 3. Close-up views of the free-surface elevations at days 3000, 4000, 5000, and 6000. The graphs show a $2000 \text{ km} \times 2000 \text{ km}$ subregion centered on the western boundary.

scales, and it is advected by a velocity u_r that can vary on the fast time scale. On the other hand, in the code (II) with original variables, the momentum fluxes have the form $u_r(u_r \Delta p_r)$, and in general both the momentum density $u_r \Delta p_r$ and the advective velocity u_r can vary on the fast time scale.

Suppose that a two-dimensional density ψ is advected by a velocity $\mathbf{u} = (u, v)$, where \mathbf{u} can vary on the fast time scale, and ψ may be either slowly-varying or rapidly-varying. The corresponding flux in the x -direction, i.e., rate of flow per unit time per unit cross-sectional length, is the product $u\psi$. For the sake of simplicity in the following discussion, assume that ψ and u are known at the edge of a grid cell. (In practice, ψ is associated with a cell center and u is defined at a cell edge.) Under this assumption, the net transport across an edge between baroclinic times t_n and $t_{n+1} = t_n + \Delta t$ is given by $\int_{t_n}^{t_{n+1}} u\psi dt$; spatial integration along the cell edge is deleted here for notational simplicity. For the sake of numerical discretization, the transport can be approximated

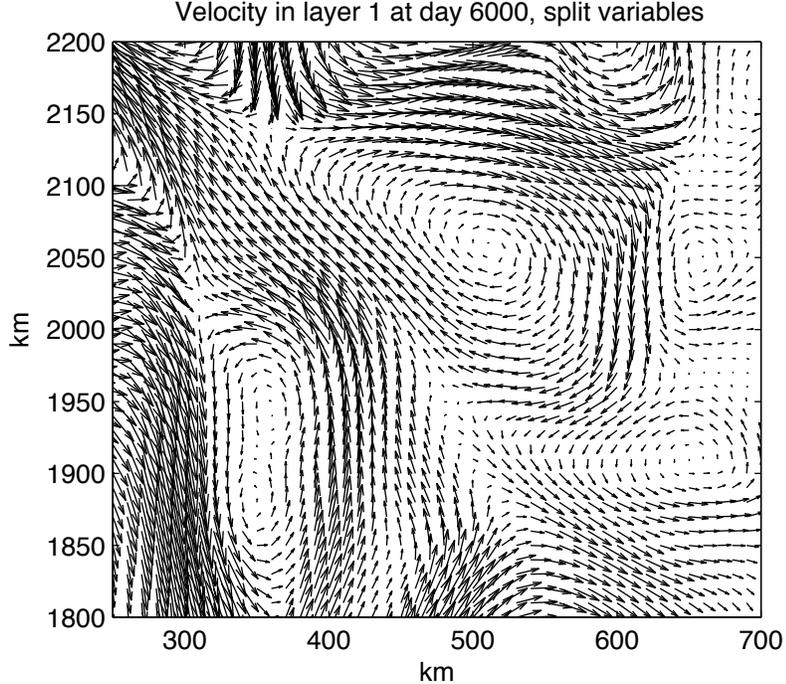


Fig. 14. Test 3. Close-up view of the horizontal velocity field in the uppermost layer at day 6000.

by

$$\int_{t_n}^{t_{n+1}} u\psi dt \approx \Delta t \left(\frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} u dt \right) \psi^n \equiv (\Delta t) u_{ave} \psi^n, \quad (23)$$

where ψ^n denotes the value of ψ at time t_n . The time-averaged velocity u_{ave} includes averaging of fast external signals over the baroclinic time interval $[t_n, t_{n+1}]$.

If the density ψ varies on a time scale that is resolved well by the time step Δt , then the approximation (23) has some validity. In the case where both u and ψ are well-resolved, Smolarkiewicz and Margolin (1998) show that if the flux in a two-time-level advection scheme is discretized in time with the quantity $u^{n+1/2}\psi^n$, then the method is second-order accurate in time; here, $u^{n+1/2}$ denotes the velocity at the midpoint of the interval $[t_n, t_{n+1}]$. In the present case, the time-averaged velocity u_{ave} takes the place of the velocity $u^{n+1/2}$. The split-variable formulation of the momentum equation uses the (essentially) slowly-varying momentum density $u'_r \Delta p'_r$, with the approximation in (23), and during the correction step the advective velocity is obtained from time averages over all barotropic substeps, as noted in Section 4.2. In the numerical experiments reported here, this approach yields good results.

On the other hand, if the density ψ in (23) varies on the fast time scale, then the validity of the approximation in (23) is questionable. This consideration applies to the formulation of the momentum equation involving unsplit (original) variables, since the momentum density $u_r \Delta p_r$ can vary on the fast time scale. In code (II), some time-averaging is used to represent this density, as noted in Section 4.2, and this may serve to reduce numerical irregularities. However, there remains the fact that the average of a product $u\psi$ is not equal to the product of the individual averages of u and ψ . Whether the advected momentum density $u_r \Delta p_r$ is an instantaneous value or a time average, the resulting inaccuracy in the computed momentum transport may play a role in the irregular and inaccurate results that are obtained with code (II) if the time step is not reduced sufficiently. For example, the solution shown in Figure 9 is not only noisy, but the large-scale features are incorrect.

The preceding remarks could also be applied to the layer mass equation (1), $\partial(\Delta p_r)/\partial t + \nabla \cdot (\mathbf{u}_r \Delta p_r) = 0$. In the mass flux $\mathbf{u}_r \Delta p_r$, both the advective velocity \mathbf{u}_r and the density Δp_r can vary on the fast time scale. However, with the procedure described in Section 3.4, the lateral mass fluxes are adjusted at each baroclinic time step by removing the unresolved fast portion of the flux and substituting the time-integrated barotropic mass flux. This flux adjustment appears to prevent difficulties of the kind described above.

This then raises the question of whether the same operation could be employed for the momentum equation. An examination of the analysis in Section 3.3 indicates that this operation would be performed if the vertical average of the momentum advection terms were incorporated explicitly into the barotropic momentum equation and thus computed at each barotropic substep. Such a procedure was implemented and tested in code (II), but it did not produce an improvement in the computed results. The reason for this is unclear to this author, but a possible factor could be some feedback between the momentum density $u_r \Delta p_r$ and the velocity u_r which transports it. This procedure also increases the complexity of the barotropic solver and consequently increases the execution time significantly.

Another possible way to approximate the time integral in (23) is to compute the product $u\psi$ at each barotropic substep and then compute the average over all such substeps. With the two-level time-stepping method used here, this quantity could be computed during the correction step by using predicted values of baroclinic and barotropic variables. In the present context, ψ denotes a component of momentum density; this quantity is defined at the center of a momentum cell, whereas the advective velocity u is defined at an edge of such a cell. In order to compute the integral $\int_{t_n}^{t_{n+1}} u\psi dt$ at the edge of a grid cell, it would then be necessary to obtain values of ψ at such an edge. In fact, in many cases, one of the major steps in the construction of an advection scheme is to obtain a suitable reconstruction or interpolation of the advected density

field (LeVeque, 2002). The explicit computation of a time average of $u\psi$ will not be attempted in the present paper, for reasons of length and potential complexity. In addition, the apparent success of the split-variable formulation reduces the need for improving the formulation involving unsplit (original) variables.

The discussion in this section has been based on the formulations (2) and (8) of the momentum equation, in which the dependent variable is momentum density and the advective terms are written in flux form. An alternative would be to try velocity as the dependent variable, and combine the nonlinear and Coriolis terms into a formulation involving vorticity and the gradient of kinetic energy. However, as noted in Section 2, previous computational experience suggests that the momentum flux form produces numerical algorithms that are more robust, especially in locations where layer thicknesses tend to zero. The momentum flux form has therefore been the focus of the present investigation.

6 Conclusions

The main goal of this paper is to compare two formulations of dependent variables in barotropic-baroclinic splittings for layered ocean circulation models. One approach is to split the dependent variables into barotropic and baroclinic components, where the baroclinic variables are mostly devoid of motions varying on the fast time scale associated with external gravity waves. Another approach is to use unsplit (original) variables in each layer and, at the end of each baroclinic step, adjust their values to ensure consistency with the barotropic variables. This adjustment has the effect of removing unresolved fast forcing and replacing it with the time-integrated effect of well-resolved fast forcing that is computed when the barotropic equations are solved explicitly with short substeps.

Using unsplit (original) variables is particularly valuable when solving the equation for conservation of mass, as it is possible to maintain conservation form and thus guarantee conservation of mass in each layer. This is especially important in long-term simulations.

The main focus here is on the momentum equations, in the particular setting where the equations are written in flux form so that a numerical advection scheme can be used for the flux terms. The two formulations, split variables and unsplit (original) variables, are identical at the level of a linearized analysis, and they give very similar results in some simple test problems where analytical solutions are known. In addition, in some eddying double-gyre computations, both formulations are able to produce results that agree well with reference solutions computed with a code that uses short time steps and no

barotropic-baroclinic splitting.

However, in the double-gyre computations, there are significant differences between the two formulations. With the formulation involving unsplit (original) variables, it is necessary to reduce the baroclinic time step significantly, relative to the time step that can be used with split variables. Otherwise, in one test problem the formulation with unsplit variables produces some grid noise and a negative layer thickness, and in another problem this formulation produces pervasive grid noise and a highly inaccurate large-scale flow. The problem can be traced to the momentum advection terms, and it may be due to advecting a rapidly-varying momentum density with a rapidly-varying velocity. The grid noise can be removed by introducing a sufficient amount of horizontal viscosity, but in one example the required viscosity causes a noticeable reduction in the total kinetic energy in the computed solution, and in another case the required viscosity causes a major distortion of the flow.

In contrast, the formulation with split variables does not experience these difficulties. One effect of the computations described here is to provide a further validation of the split-variable formulation for the momentum equation, beyond the testing previously reported by Higdon (2005). The analysis and computations described in this paper suggest that, in the context considered here, a good approach is to use unsplit (original) variables in the equation for conservation of mass and to use split variables in the equations for conservation of momentum.

7 Acknowledgments

This work was supported by National Science Foundation grant DMS-0511782. Most of the computations were performed using computational facilities provided by the ocean and climate modeling group at Los Alamos National Laboratory.

References

- Bleck, R., Smith, L. T., 1990. A wind-driven isopycnic coordinate model of the north and equatorial atlantic ocean 1. model development and supporting experiments. *J. Geophysical Research* 95C, 3273–3285.
- Blumberg, A. F., Mellor, G. L., 1987. A description of a three-dimensional coastal ocean circulation model. In: Heaps, N. (Ed.), *Three-dimensional Coastal Ocean Models*. American Geophysical Union, Washington, D.C., pp. 1–16.

- Bryan, K., 1969. A numerical method for the study of the circulation of the world ocean. *J. Computational Physics* 4, 347–376.
- Griffies, S. M., 2004. *Fundamentals of Ocean Climate Models*. Princeton University Press, Princeton, N.J.
- Hallberg, R., 1997. Stable split time stepping schemes for large-scale ocean modeling. *J. Computational Physics* 135, 54–65.
- Higdon, R. L., 2002. A two-level time-stepping method for layered ocean circulation models. *J. Computational Physics* 177, 59–94.
- Higdon, R. L., 2005. A two-level time-stepping method for layered ocean circulation models: further development and testing. *J. Computational Physics* 206, 463–504.
- Higdon, R. L., 2006. Numerical modelling of ocean circulation. *Acta Numerica* (2006) 15, 385–470.
- LeVeque, R. J., 2002. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2002.
- Sadourny, R., 1975. The dynamics of finite-difference models of the shallow-water equations. *J. Atmospheric Sciences* 32, 680–689.
- Shchepetkin, A. F., McWilliams, J. C., 2005. The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling* 9, 347–404.
- Smolarkiewicz, P. K., Margolin, L. G., 1998. MPDATA: A finite-difference solver for geophysical flows. *J. Computational Physics* 140, 459–480.