Ocean Modelling 29 (2009) 15–26

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

Reconciling estimates of the free surface height in Lagrangian vertical coordinate ocean models with mode-split time stepping

Robert Hallberg *, Alistair Adcroft

NOAA Geophysical Fluid Dynamics Laboratory, Ocean Group, 201 Forrestal Rd., Princeton, NJ 08540, USA
Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ 08540, USA

ARTICLE INFO

Article history:
Received 27 August 2008
Received in revised form 17 February 2009
Accepted 19 February 2009
Available online 3 March 2009

Keywords:
Ocean model
Split time stepping
Lagrangian vertical coordinate
Free surface

ABSTRACT

In ocean models that use a mode splitting algorithm for time-stepping the internal- and external-gravity modes, the external and internal solutions each can be used to provide an estimate of the free surface height evolution. In models with time-invariant vertical coordinate spacing, it is standard to force the internal solutions for the free surface height to agree with the external solution by specifying the appropriate vertically averaged velocities; because this is a linear problem, it is relatively straightforward. However, in Lagrangian vertical coordinate ocean models with potentially vanishing layers, nonlinear discretizations of the continuity equations must be used for each interior layer. This paper discusses the options for enforcing agreement between the internal and external estimates of the free surface height, along with the consequences of each choice, and suggests an optimal, essentially exact, approach.
Published by Elsevier Ltd.

1. Introduction

Hydrostatic ocean models filter out sound waves, so the fastest motions in such models are external gravity waves, propagating horizontally at $\sqrt{gH}$ (where g is the gravitational acceleration and H is the total ocean depth) – of order 200 ms^{-1} in the deep ocean. Shallow-water external gravity waves have nearly vertically uniform horizontal velocities and are well characterized by two-dimensional equations. The next fastest motions are horizontal velocities and internal gravity waves, both with speeds of a few meters per second and rich three-dimensional structures. Ocean models are about two orders of magnitude less costly to integrate in time if they separate integration of the external mode from the internal evolution of the model.

In models with time-invariant vertical coordinates (sigma- or Z-coordinate models or their stretched equivalent with a free surface), gravity waves are typically handled with an external mode solver (using either a rigid lid or a free surface). In either case, the time-filtered evolution of the free surface height gives a boundary condition on the vertical velocity, which is determined diagnostically from the vertically structured continuity equation. The discretization of the continuity equation in such models is invariably linear in the velocities, and it is straightforward to use a finite volume formulation and obtain exact consistency between the time-averaged external mode solution and the internal model structure. Even when the free surface height does vary modestly with time (such that no levels ever vanish with a given definition of the vertical coordinate), the algorithm used is still essentially the same; most importantly a linear (in velocity) discretization of the continuity equation is still appropriate, and a finite volume reconciliation of the changes in the interior structure with the evolution of the free surface (Griffies et al., 2001; Campin et al., 2004). With this exact finite volume reconciliation, there are no issues with tracer or mass (volume if Boussinesq) conservation.

By contrast, Lagrangian vertical coordinate models[1] use the continuity equation prognostically to describe the evolution of the thickness, h_k, of each vertically discrete layer k:

$$\frac{\partial h_k}{\partial t} = -\nabla \cdot (\boldsymbol{u}_k h_k). \tag{1}$$

(For simplicity, vertical fluxes and precipitation minus evaporation are ignored here – they do not alter the discussion.) The layer thicknesses can be summed vertically to obtain an estimate of the free surface height

$$\eta_h \equiv \sum_{k=1}^{K} h_k - D, \tag{2}$$

* Corresponding author. Address: NOAA Geophysical Fluid Dynamics Laboratory, Ocean Group, 201 Forrestal Rd., Princeton, NJ 08540, USA. Tel.: +1 609 452 6508; fax: +1 609 987 5063.
E-mail addresses: Robert.Hallberg@noaa.gov (R. Hallberg), Alistair.Adcroft@noaa.gov (A. Adcroft).

[1] See Adcroft and Hallberg (2006) for a full discussion of the differences between the solution approaches in Lagrangian and Eulerian vertical coordinate ocean models.

1463-5003/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.ocemod.2009.02.008

where D is the time-invariant bottom depth.[2] The vertical sum of the layer continuity equations gives the barotropic continuity equation,

$$\frac{\partial \eta}{\partial t} = -\nabla \cdot \left(\sum_{k=1}^{K} \boldsymbol{u}_k h_k \right) = -\nabla \cdot (\boldsymbol{U} H), \tag{3}$$

which uses the definitions of the total thickness and the barotropic velocity,

$$H \equiv \sum_{k=1}^{K} h_k \quad \text{and} \quad \boldsymbol{U} \equiv \frac{1}{H} \left(\sum_{k=1}^{K} \boldsymbol{u}_k h_k \right). \tag{4}$$

In the horizontally and temporally continuous equations, the two estimates of the free surface height, η and η_h, are perfectly consistent. However, discretization in time or in the horizontal spatial directions can break the consistency between Eqs. (1) and (3). A widely used approach to solve Eqs. (1) through (3) for η and η_h is the barotropic–baroclinic split time stepping scheme, in which the two-dimensional shallow water equations [(3) and the vertically averaged momentum equations] are used to estimate the evolution of the free surface height, η, for the time interval Δt, over which the full three-dimensional equations are also advanced. The two-dimensional equations can be advanced either explicitly with many short time steps (e.g. Killworth et al., 1991; Bleck and Smith, 1990; deSzoeke and Higdon, 1997; Hallberg, 1997; Shchepetkin and McWilliams, 2005) or implicitly (e.g. Dukowicz and Smith, 1994; Campin et al., 2004). Both the explicit and implicit approaches can be represented schematically as

$$\frac{\eta^{n+1} - \eta^n}{\Delta t} = -\nabla \cdot \langle \boldsymbol{U} H \rangle = -\nabla \cdot \langle \boldsymbol{V}(\boldsymbol{U}, H) \rangle, \tag{5}$$

where the angle brackets are used to denote whatever time averaging of velocities and thicknesses are used to determine the volume fluxes that advance the free surface height over a time step, Δt. The function $\boldsymbol{V}$ represents the spatial discretization of the barotropic fluxes, and may be a nonlinear function of $\boldsymbol{U}$ or H. The superscripts n and $n+1$ refer to successive (baroclinic) time levels (there may be many shorter sub-cycled time levels averaged over by the angle brackets). The precise meaning of the angle brackets is determined by the choice of split time stepping scheme. For many split explicit schemes (e.g., Bleck and Smith, 1990; Killworth et al., 1991; Hallberg, 1997; Higdon, 2005) the angle brackets are approximately a simple time average, while for Shchepetkin and McWilliams (2005) it would be a weighted filter that extends past time level $n+1$. With an implicit scheme, the angle brackets are likely to be the values at time level $n+1$.

The layer continuity equations are integrated over the same time period as (5) with a single large time-step, Δt; this is represented schematically as

$$\frac{h_k^{n+1} - h_k^n}{\Delta t} = -\nabla \cdot (\boldsymbol{u}_k h_k) = -\nabla \cdot \boldsymbol{F}(\boldsymbol{u}_k, h_k). \tag{6}$$

The layer thicknesses, h_k, must be non-negative, which generally requires the use in (6) of a discretization (here represented as the function $\boldsymbol{F}$) of the horizontal volume fluxes ($\boldsymbol{u}_k h_k$) that depends nonlinearly on the velocities, inevitably reverting to upwind differencing for sufficiently strong flow out of a relatively thin cell. For the discrete (in time and space) equations to have a consistent (single) estimate of the free surface height, the time average barotropic fluxes, $\langle \boldsymbol{V} \rangle$, and vertically integrated baroclinic fluxes, $\sum \boldsymbol{F}$, must satisfy

$$\langle \boldsymbol{V}(\boldsymbol{U}, H) \rangle = \sum_{k=1}^{K} \boldsymbol{F}(\boldsymbol{u}_k, h_k). \tag{7}$$

Failure to satisfy this constraint implies the existence of two estimates of the free surface and a possible inconsistency in the model equations. Satisfying this constraint is non-trivial due to the non-local in time and space nature of the constraint, and is especially non-trivial when either of $\boldsymbol{V}$ or $\boldsymbol{F}$ are non-linear.

The accumulated horizontal volume fluxes used to update the free surface height, $\langle \boldsymbol{U}\, \mathrm{H} \rangle$, can be related to an effective time-mean barotropic velocity by

$$\langle \boldsymbol{U} \rangle = \frac{\langle \boldsymbol{U} H \rangle}{\langle H \rangle} = \frac{\langle \boldsymbol{V}(\boldsymbol{U}, H) \rangle}{\langle H \rangle}, \tag{8}$$

where $\langle H \rangle$ is an appropriate time-mean total thickness. $\langle \boldsymbol{U} \rangle$ is commonly used in strategies to reconcile Eqs. (5) and (6).

At this point the layer equations could be advanced with a velocity whose thickness weighted vertical mean has been replaced by the time-mean barotropic velocity, $\langle \boldsymbol{U} \rangle$. Replacing the instantaneous vertically averaged velocity,

$$\boldsymbol{U}_0 \equiv \frac{\sum_{k=1}^{K} \tilde{h}_k \boldsymbol{u}_k}{\sum_{k=1}^{K} \tilde{h}_k}, \tag{9}$$

by the time-mean barotropic velocity gives new layer velocities of

$$\widehat{\boldsymbol{u}}_k = \boldsymbol{u}_k + \langle \boldsymbol{U} \rangle - \frac{\sum_{j=1}^{K} \tilde{h}_j \boldsymbol{u}_j}{\sum_{j=1}^{K} \tilde{h}_j} = \boldsymbol{u}_k + \langle \boldsymbol{U} \rangle - \boldsymbol{U}_0, \tag{10}$$

whose vertical average matches the time-mean barotropic velocity. Here the $\tilde{h}_k$ are estimates of the thicknesses at the faces of the control volumes, but except in the linear limit, $\tilde{h}_k$ cannot be guaranteed to agree with the effective thicknesses from the continuity equation, defined by $\widehat{h}_k \equiv \hat{\boldsymbol{n}} \cdot \boldsymbol{F}\left(\widehat{\boldsymbol{u}}_k, h_k\right) / \left(\hat{\boldsymbol{n}} \cdot \widehat{\boldsymbol{u}}_k\right)$, where $\hat{\boldsymbol{n}}$ is the unit vector normal to the faces of the control volume. (In one-dimension, h_k is just the volume flux divided by the velocity.) Eq. (10) alters the vertical mean velocity, but deviations from this mean

$$\boldsymbol{u}'_k = \boldsymbol{u}_k - \boldsymbol{U}_0, \tag{11}$$

are unaffected by this adjustment. With the adjustment in (10), an estimate of the layer time-filtered thicknesses can be advanced by

$$h_k^{n+1} = h_k^n - \Delta t \nabla \cdot \boldsymbol{F}\left(\widehat{\boldsymbol{u}}_k, h_k\right) = h_k^n - \Delta t \nabla \cdot \left(\widehat{\boldsymbol{u}}_k \widehat{h}_k\right). \tag{12}$$

Like the definition of the averaging in the angle brackets, the timing of the layer velocities in (10) and of the thicknesses used for the fluxes in (12) are determined by the underlying baroclinic time-stepping scheme. If a predictor-corrector scheme is used (as in the examples in Section 3), mass conservation only requires that the final correction to the layer thicknesses be consistent with (12), although it is often useful for the overall stability of the scheme if similar constraints to be applied to the predictor steps as well (Higdon, 2008).

In the limit where the volume fluxes vary linearly with the velocities, it is possible to select the discretizations of the fluxes and the thickness weights such that

$$\tilde{h}_k = \widehat{h}_k = \overline{h_k}^x \quad \text{and} \quad \langle H \rangle = \sum_{k=1}^{K} \overline{h_k}^x, \tag{13}$$

where the overbar-x represents the arithmetic mean of adjacent thicknesses (or any other plausible interpolation of the thicknesses to the velocity points that is independent of the velocity). So in the linear limit, summing (12) over the layers and subtracting the bottom depth, D, combined with (10) and (13) and the identity

[2] The discussion presented here makes the Boussinesq approximation. Without it the thicknesses would be measured in units of Pascals instead of meters, the roles of the bottom depth and sea surface height are replaced by surface and bottom pressure, and volume conservation becomes mass conservation. The discussion presented here would be identical without the Boussinesq approximation if this change of variables were made.

$$\sum_{k=1}^{K}\left[\left(\boldsymbol{u_k}-\sum_{j=1}^{K}\tilde{h}_j\boldsymbol{u_j}\Big/\sum_{j=1}^{K}\tilde{h}_j\right)\tilde{h}_k\right]=\sum_{k=1}^{K}\boldsymbol{u_k}\tilde{h}_k-\left(\sum_{k=1}^{K}\tilde{h}_k\Big/\sum_{k=1}^{K}\tilde{h}_k\right)\sum_{k=1}^{K}\tilde{h}_k\boldsymbol{u}_k=0 \tag{14}$$

gives

$$\begin{aligned}\eta_h^{n+1}&=\sum_{k=1}^{K}\left\{h_k^n-\Delta t\nabla\cdot\left(\widehat{\boldsymbol{u}}_k\widehat{h}_k\right)\right\}-D=\eta_h^n-\Delta t\nabla\cdot\left[\langle\boldsymbol{U}\rangle\sum_{k=1}^{K}\overline{h_k}^{x}\right]\\&\quad-\Delta t\nabla\cdot\sum_{k=1}^{K}\left[\left(\boldsymbol{u_k}-\sum_{j=1}^{K}\tilde{h}_j\boldsymbol{u_j}\Big/\sum_{j=1}^{K}\tilde{h}_j\right)\tilde{h}_k\right]\\&=\eta_h^n-\eta^n+\eta^n-\Delta t\nabla\cdot(\langle\boldsymbol{U}\rangle\langle H\rangle)=\left(\eta_h^n-\eta^n\right)+\eta^{n+1}.\end{aligned} \tag{15}$$

So in the linear limit, the two estimates of the free surface – η^{n+1} from the advancing the barotropic equations, and η_h^{n+1} from the evolution of the sum of the layer thicknesses – are identical at one timestep if they were the same at the previous timestep.

Unfortunately, in the general case it is not possible to make all the equalities between the various thicknesses in (13) hold. As emphasized above, the layer continuity equations must be positive definite, so h_k typically depends nonlinearly on $\widehat{\boldsymbol{u}}_k$. But $\langle H\rangle$ must be known to calculate $\langle\boldsymbol{U}\rangle$ and $\tilde{h}_k$ is needed to determine $\widehat{\boldsymbol{u}}_k$, so it is impossible to guarantee either that $\tilde{h}_k=h_k$ or that $\langle H\rangle=\sum_{k=1}^{K}h_k$. In this more general case, simply advancing the layer equations with (12) leads to a mismatch between the barotropic estimate of the free surface height and the estimate from the sum of the layer thicknesses. Specifically

$$\begin{aligned}\eta_h^{n+1}-\eta^{n+1}&=\left(\eta_h^n-\eta^n\right)+\Delta t\nabla\cdot\left[\langle\boldsymbol{U}\rangle\left(\langle H\rangle-\sum_{k=1}^{K}\widehat{h}_k\right)\right]\\&\quad-\Delta t\nabla\cdot\sum_{k=1}^{k}\left[\left(\widehat{h}_k-\tilde{h}_k\frac{\sum_{j=1}^{K}\widehat{h}_j}{\sum_{j=1}^{K}\tilde{h}_j}\right)\boldsymbol{u}_k\right].\end{aligned} \tag{16}$$

With many upwind-biased advection schemes, the sums of the discrepancies between the thicknesses are often on the order of half the bottom depth change between adjacent grid points. With an advective baroclinic Courant number of order 0.1, it is not atypical to generate discrepancies between these two estimates of the free surface height of order a few percent of the depth of the ocean within a single time step (see Figs. 1 and 2). Without some strategy for reconciling these two estimates with one another, these discrepancies will grow with time – linearly at first, but more likely as a random walk later on. Without some corrective measure, the solutions quickly become absurd.

This paper describes some of the options for enforcing agreement between the external estimate of the free surface height and that due to the sum of the layer thicknesses in Lagrangian vertical coordinate ocean models. The disadvantages of several methods that have traditionally been used are described, and a new approach that avoids the primary difficulties is proposed.

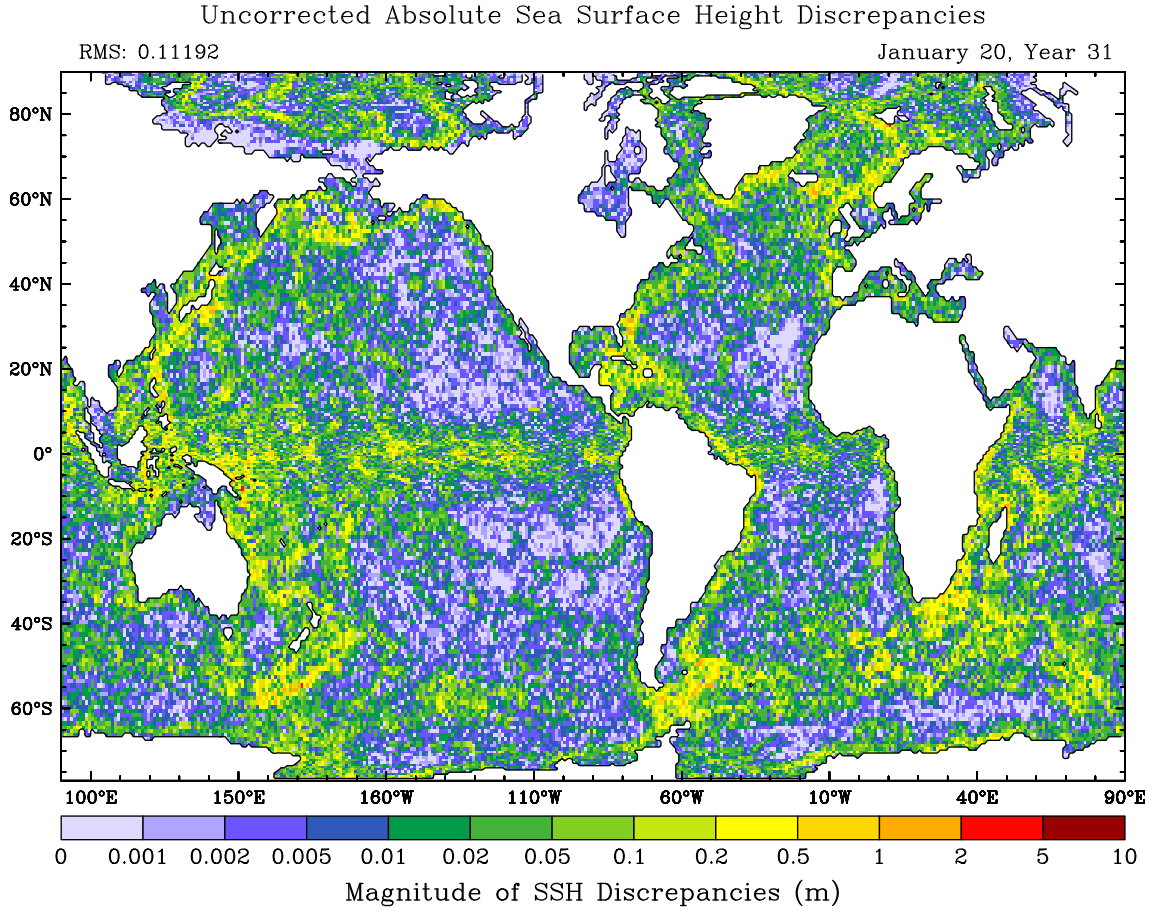

Fig. 1. The absolute value of the instantaneous discrepancy between the external and internal estimates of the free surface height after a single 3600 s baroclinic timestep, from a 49-layer global ocean-climate model with a 1° nominal resolution. This particular image is taken from January 20 of Year 31 of a CORE-forced run (see Griffies et al., 2009 for a full description of the CORE protocol), but it is representative of other times and configurations. Note that the greatest discrepancies occur where there are significant topographic variations or strong flows, such as near the equator or in western boundary currents. As both solutions conserve volume, the global average of the (signed) discrepancies is 0. Increasing the horizontal resolution will not tend to reduce the magnitude of these discrepancies, as the advective Courant number is typically held steady or even increases with increasing resolution, and finer resolution models will typically have comparable magnitude topographic and layer thickness variations at the smallest resolved scales.

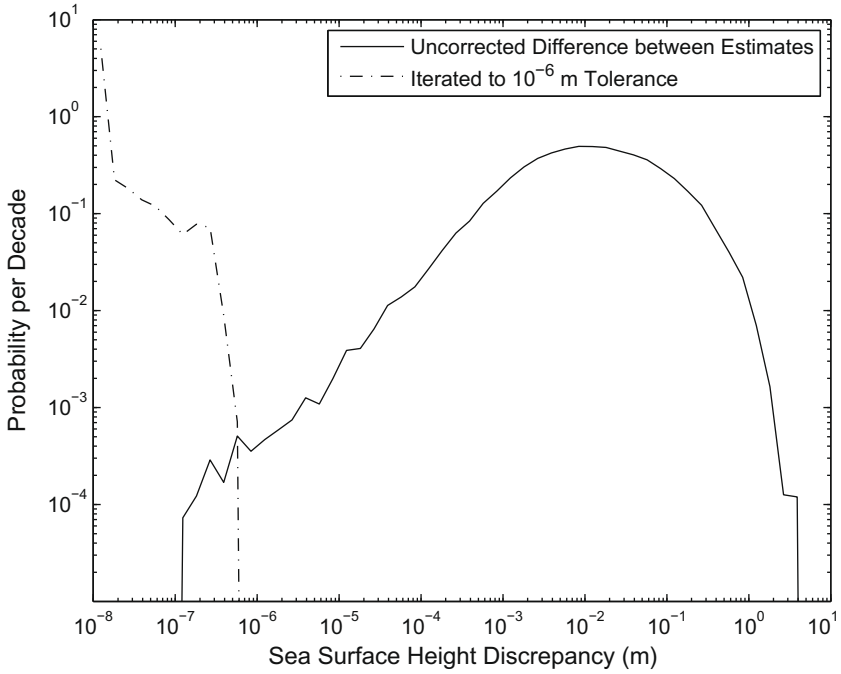

Fig. 2. The area-weighted histogram of the absolute value of the instantaneous discrepancy between the external and internal estimates of the free surface height after a single 3600 s baroclinic timestep, using a 49-layer global ocean-climate model with a 1° nominal resolution. The solid line shows the distribution from the same state as depicted in Fig. 1. The dashed-dotted line shows the distribution after iteration to within a SSH tolerance of less than 10^{-6} m for the barotropic velocities used in the layers' transports to be consistent with the barotropic transports, as described later in the paper.

2. Options for reconciling the estimates of the free surface height

2.1. Layer dilation

Bleck and Smith (1990) note that the barotropic estimate of the free surface height is a much more credible estimate of the sea surface height than the sum of the layer thicknesses. In essence, it has been subjected to the very strong flattening effects of divergent accelerations driven by the pressure gradients due to any slopes of the free surface, whereas the estimate from the sum of the layers has not. To force the sum of the layer thicknesses to conform to the barotropic free surface, one can uniformly dilate or compress all of the layers by whatever factor is necessary. The layer thicknesses then evolve as

$$h_k^{n+1} = S h_k^n - \Delta t S \nabla \cdot \boldsymbol{F}\left(\widehat{\boldsymbol{u}}_k, h_k\right) \tag{17}$$

where S is the horizontally varying but vertically uniform scaling factor given by

$$S = \frac{D + \eta^{n+1}}{\sum_{k=1}^{K}\left[h_k^n - \Delta t \nabla \cdot \boldsymbol{F}\left(\widehat{\boldsymbol{u}}_k, h_k\right)\right]}. \tag{18}$$

With this approach the two estimates of the free surface height agree exactly, and there is minimal additional computational expense and no additional lateral communication required.

The big problem with this approach is that the layer continuity equations are no longer cast in the flux-divergence form of (1) and (12), so the layers do not conserve their individual masses (Hallberg, 1997; Higdon, 2005). Globally, mass is still conserved because the barotropic solution conserves mass and the sum of the layer masses agree with the barotropic mass. But the nonconservation of mass by layers leads to spurious diapycnal mass fluxes and nonconservation of heat and salt. These fluxes are small enough that MICOM and subsequently HYCOM (Bleck, 2002) have been successfully applied to answering many questions about the ocean.[3] But these spurious diapycnal mass fluxes mean the abandonment of one of the most valuable properties of isopycnal coordinate ocean models – that they are uniquely able to robustly reproduce the exceedingly adiabatic nature of the interior ocean (Griffies et al., 2000). For long-term climate studies, nonconservation of heat and salt is a serious liability with this approach.

2.2. Flux-form reconciliation

The most obvious way to robustly conserve heat and salt and avoid spurious diapycnal mass fluxes is to keep the layer continuity equations in flux-conservative form. The problem then is how to deal with the differences in the two estimates of the free surface height.

The first step that can be taken with any approach is to guess values of $\langle \boldsymbol{U} \rangle$ that will minimize the discrepancy. With nonlinear layer continuity equations, this guess can never be perfect. But it will minimize the magnitude of any corrective fluxes or reduce the number of iterations required to obtain an acceptable reconciliation between the external and internal solutions for the free surface height.

2.2.1. Barotropic pseudo-mass source

For many applications, the instantaneous free surface height and barotropic flow are relatively unimportant. In this case, it may be acceptable to use the difference between the barotropic and layer estimates of the free surface height as a spurious mass source only in the barotropic equations.

[3] As detailed in Bleck and Smith (1990) and more explicitly in Appendix B of Bleck et al. (1992), MICOM and HYCOM use an upwind modification of their initial estimates of the layers' fluxes to partially compensate for some of the discrepancies between the baroclinic and barotropic mass fluxes, but they ultimately rely upon layer dilation to enforce consistency between the two estimates of the free surface height.

$$Src = \frac{1}{\Delta t}\left[\sum_{k=1}^{K} h_k^n - (D+\eta^n)\right] \tag{19}$$

The barotropic continuity equation becomes

$$\frac{\eta^{n+1}-\eta^n}{\Delta t} + \nabla\cdot\langle \boldsymbol{U}H\rangle = Src \tag{20}$$

This source term will drive a divergent barotropic flow field that will tend to drive the free surface height from the sum of the layer thicknesses toward the barotropic estimate. The mass of each layer is exactly conserved due to the unaltered layer continuity equations, as are the total amounts of salt, heat and other tracers.

The down-side of this approach is that it tends to create unphysical external gravity waves, but it was successfully used in Hallberg Isopycnal Model (HIM) simulations for about a decade, such as the eddy-permitting Southern Ocean simulations of Hallberg and Gnanadesikan (2006). The geostrophically adjusted barotropic residual from this source is quite small since the scale of the forcing is typically close to the grid scale and much smaller than the external deformation radius. However, with strongly variable forcing these external gravity waves can become quite intense, and nonlinearities in the equations may lead to significant nonphysical terms from the large external gravity waves. For some simulations, such as tidal studies, the external gravity waves are the signal of interest and this approach is unacceptable.

2.2.2. Upwind corrective layer fluxes

If the instantaneous free surface height and barotropic flow are important, the barotropic fluxes must be respected. If the layer continuity equations are to be kept in flux-form, the layer fluxes must be adjusted so that their sum agrees with the barotropic solution. One way to adjust the layer fluxes is by adding a corrective flux to the layer fluxes after they are calculated.

Higdon (2005) notes that a predictor-corrector approach can be combined with upwind corrective fluxes to force the layers' summed fluxes to agree with the barotropic fluxes.[4] An initial prediction of the updated layer thicknesses is

$$h_k^* = h_k^n - \Delta t\nabla\cdot\boldsymbol{F}\left(\widehat{\boldsymbol{u}}_k, h_k\right). \tag{21}$$

The difference between the barotropic fluxes and the sum of these initial-estimate fluxes is

$$\Delta(\boldsymbol{U}H) = \langle \boldsymbol{U}H\rangle - \sum_{k=1}^{K}\boldsymbol{F}\left(\widehat{\boldsymbol{u}}_k, h_k\right). \tag{22}$$

To be useful, whatever corrective fluxes are applied must respect the positive definiteness of the layer equations. Higdon (2005) shows that upwind fluxes based on these predictions of the layer thicknesses satisfy this requirement. That is, if $h_k^{*,Upwind}$ are the predicted layer thicknesses in the upwind direction based on the required direction of the corrective fluxes, the corrective velocity is

$$\boldsymbol{U}^{Cor} = \frac{\Delta(\boldsymbol{U}H)}{\sum_{k=1}^{N} h_k^{*,Upwind}}, \tag{23}$$

and the corrected layer thicknesses are

$$h_k^{n+1} = h_k^* - \Delta t\nabla\cdot\left(\boldsymbol{U}^{Cor}h_k^{*,Upwind}\right). \tag{24}$$

Combining Eqs. (21) and (24) gives

$$h_k^{n+1} = h_k^n - \Delta t\nabla\cdot\left(\boldsymbol{F}\left(\widehat{\boldsymbol{u}}_k, h_k\right) + \boldsymbol{U}^{Cor}h_k^{*,Upwind}\right), \tag{25}$$

which when summed vertically and combined with (22) and (23) gives

$$\begin{aligned}\sum_{k=1}^{K} h_k^{n+1} &= \sum_{k=1}^{K} h_k^n - \Delta t\nabla\cdot\left(\sum_{k=1}^{K}\boldsymbol{F}\left(\widehat{\boldsymbol{u}}_k, h_k\right) + \frac{\Delta(\boldsymbol{U}H)}{\sum_{k=1}^{K} h_k^{*,Upwind}}\sum_{k=1}^{K} h_k^{*,Upwind}\right)\\ &= \sum_{k=1}^{K} h_k^n - \Delta t\nabla\cdot\langle \boldsymbol{U}H\rangle.\end{aligned} \tag{26}$$

The problem with using barotropic corrective upwind fluxes is that the layers where the corrections are applied may be displaced from the layers that would be most nearly consistent with the layers' continuity equations. Near a sloping bottom, there are often layers that are massless on the slope but thick over the abyss. If the corrective fluxes fluctuate randomly, these deep layers will only partake of the corrective fluxes when they are upslope. This leads to an unphysical rectified numerical pumping of dense flow upslope, and an abyssal circulation around isolated topographic highs and around slopes, not dissimilar to the physical "Neptune" effect of Holloway (1992). In eddy-permitting North Atlantic simulations, barotropic upwind corrections can lead to unphysical abyssal gyres with transports of tens of Sverdrups (P. Schopf, personal communication).

One way to avoid these strong unphysical abyssal flows is to use only the layer thicknesses above the shallowest of the adjacent bottom depths, rather than the full upwind thicknesses (Higdon (2005) and M. Bentsen, personal communication). This expedient seems to give reasonable circulations in many instances. But with this approach, the corrective fluxes do not, in general, correspond to any purely barotropic flow added to the instantaneous layer velocities.

2.2.3. Iterative flux-form reconciliation

In some sense, the entire problem is that the barotropic transport, the time-mean barotropic velocity, and the layer transports determined by the layers' discretizations of the continuity equation using that time-mean barotropic velocity over-constrain the evolution of the free surface height. The approach of Bleck and Smith (1990) is to partially disregard the layers' transports, while the fictitious mass source for the barotropic equation partially disregards the barotropic transport. The upwind corrective fluxes only approximate an adjustment to the time-mean barotropic velocity; consequently this approach displaces the compensating fluxes from the layers in which they would occur in the limit of continuous time. The fourth option, described below, is to regard both the barotropic transport and the layers' discretizations as definitive and modify the time-averaged barotropic velocity (or acceleration) used to determine the layers' fluxes. In that this approach closely emulates what happens in the continuous limit, with the dominance of external gravity waves in determining the most rapid adjustment of the free surface height, this option might be preferable to the previous ones from consideration of first-principles.

Instead of replacing an estimate of the vertically averaged velocity with the time-mean barotropic velocity $\langle \boldsymbol{U}\rangle$ as in (12), the layer velocities can adjusted by a different, as-yet-unknown, barotropic velocity $\widetilde{\boldsymbol{U}}$, chosen such that the sum of the layer transports are

$$\sum_{k=1}^{K}\boldsymbol{F}\left[\left(\boldsymbol{u}_k + \widetilde{\boldsymbol{U}}\right), h_k\right] = \langle \boldsymbol{U}H\rangle. \tag{27}$$

In this case, the new estimate of the free surface height based on the layer thicknesses becomes

$$\begin{aligned}\eta_h^{n+1} &= \sum_{k=1}^{K} h_k^{n+1} - D = \sum_{k=1}^{K} h_k^n - D - \Delta t\sum_{k=1}^{K}\nabla\cdot\boldsymbol{F}\left[\left(\boldsymbol{u}_k + \widetilde{\boldsymbol{U}}\right), h_k\right]\\ &= \eta^n - \Delta t\nabla\cdot\langle \boldsymbol{U}H\rangle + \left(\eta_h^n - \eta^n\right) = \eta^{n+1} + \left(\eta_h^n - \eta^n\right).\end{aligned} \tag{28}$$

[4] This approach has also been independently developed multiple times in unpublished work by M. Bentsen, J. Dukowicz, and P. Schopf, and these ideas are partially present in Bleck and Smith (1990), but the clearest exposition in the literature can be found in Higdon (2005).

So the two estimates of the free surface height retain consistency.

The challenge for making use of (28) is how to determine $\widetilde{\boldsymbol{U}}$. The positive definite algorithms used for the fluxes in the continuity equation [schematically indicated by $\boldsymbol{F}(\boldsymbol{u}_k,h_k)$ in (28)] are typically continuous and monotonically increasing with $\boldsymbol{u}_k$, up to some fraction (perhaps all) of the mass in a cell within a timestep. So provided that the barotropic transport is reasonable (e.g. less than a quarter of the total mass of a water column is moved through each of the four faces within a timestep), an acceptable value for $\widetilde{\boldsymbol{U}}$ should exist. In the limit of horizontally uniform layer thicknesses, $\widetilde{\boldsymbol{U}}$ will be equivalent to the adjustments used previously in (10) to calculate $\widetilde{\boldsymbol{u}}_k$.

For some discretizations of the continuity equations, the transport through each face of a layer's tracer cell is independent of the velocities at the other faces of that tracer cell. In such cases, provided that the fluxes through a face vary continuously with the velocity component normal to the face, $\widetilde{V} \equiv \hat{\boldsymbol{n}} \cdot \widetilde{\boldsymbol{U}}$, and

$$\frac{\partial}{\partial \widetilde{V}} \left\{ \hat{\boldsymbol{n}} \cdot \boldsymbol{F} \left[\left(\boldsymbol{u}_k + \widetilde{V} \hat{\boldsymbol{n}} \right), h_k \right] \right\} > 0 \tag{29}$$

everywhere (even if the derivative itself is discontinuous), the solution for $\widetilde{\boldsymbol{U}} = \widetilde{V}\hat{\boldsymbol{n}}$ is unique (depending on the staggering of the discrete variables, it may be convenient to give a non-zero transverse component to $\widetilde{\boldsymbol{U}}$, but on an Arakawa C-grid this seems unnecessary). As (29) is essentially the thickness used for marginal changes in the fluxes, this is a very reasonable restriction, and one that is satisfied by a wide variety of positive definite continuity solvers. An iterative Newton's method approach should be able to find the appropriate values for the normal velocity components $\widetilde{V}$, iterating to find the solution to (27) using

$$\widetilde{V}^{m+1} = \frac{\hat{\boldsymbol{n}} \cdot \left\{ \langle \boldsymbol{U} H \rangle - \sum_{k=1}^{K} \boldsymbol{F} \left[\left(\boldsymbol{u}_k + \widetilde{V}^m \hat{\boldsymbol{n}} \right), h_k \right] \right\}}{\sum_{k=1}^{K} \frac{\partial}{\partial \widetilde{V}} \left\{ \hat{\boldsymbol{n}} \cdot \boldsymbol{F} \left[\left(\boldsymbol{u}_k + \widetilde{V}^m \hat{\boldsymbol{n}} \right), h_k \right] \right\}}. \tag{30}$$

There are many discretizations of the continuity equation that are not differentiable everywhere, and at these points a reasonable estimate of the effective thickness of the layer (which is essentially what the left side of (29) is) can be substituted. However, avoiding this minor complication does suggest that there may be a preference for selecting continuity solvers in which both the flux and the partial derivative of the flux with the velocity at the face are continuous with $\boldsymbol{u}$ (especially as $\boldsymbol{u} \cdot \hat{\boldsymbol{n}}$ changes sign), such as with a piecewise parabolic method (at least before it is limited to avoid overshoots). In any case, finding the correct velocities $\widetilde{\boldsymbol{U}}$ to agree with the barotropic transports is a one-dimensional search for the bounded root of a monotonic function, and there are many robust and efficient methods for solving this problem.

The facility of this approach is readily illustrated by considering in more detail how it would interact with the piecewise parabolic method (PPM) (e.g., Collela and Woodward, 1984). With PPM, the fluxes are constructed by first setting up a parabolic subgridscale distribution of thicknesses within each cell, $h_{i,k}(x)$. The integral of these subgridscale distributions always agrees with the discretized cell volume. The edge values are initially common interpolations, but the common edge values of neighboring cells may be adjusted separately to ensure monotonicity, or (by ensuring that the subgridscale distributions are positive throughout the cell) positive definiteness. The fluxes are then calculated by integrating these distributions over the area that is swept through the face within a timestep. That is, if the velocity $u_{i+1/2,k}$ at point $i + 1/2$ is positive, the flux is

$$F_{i+1/2,k} = \frac{1}{\Delta t} \int_{x_{i+1/2} - u_{i+1/2,k}\Delta t}^{x_{i+1/2}} h_{i,k}(x)\, dx, \tag{31}$$

so the marginal thickness becomes

$$\frac{\partial F_{i+1/2,k}}{\partial u_{i+1/2,k}} = h_{i,k}(x_{i+1/2} - u_{i+1/2,k}\Delta t), \tag{32}$$

Both (31) and (32) need to be calculated with each iteration, but the subgridscale distributions only need to be calculated once. With PPM, (31) is just the evaluation of a cubic expression, while (32) is a simple quadratic expression. In practice a single iteration is only about a third as costly as setting up the PPM distributions in the first place.

This iterative approach works exactly as in the one-dimensional discussion of (31) and (32) with directionally split discretizations for the continuity equations, in which the fluxes through the faces in one direction are calculated and used to partially update the thickness before the fluxes in the other direction are calculated. Positive definiteness is easily assured with a directionally split discretization. By alternating the order in which the directions are calculated, a first-order in time discretization error can be avoided (Strang, 1968; Durran, 1999). For the tracer equations the remaining directional-splitting errors can cause violation of monotonicity, unless particular care is taken (Easter, 1993). The continuity equations, though, are not inherently monotonic, and small-scale imbalances will tend to mostly radiate away as gravity waves. There do not appear to be particularly significant undesirable consequences from using a directionally-split discretization of the continuity equation.

For other discretizations of the continuity equation, such as the (directionally split) scheme of Hsu and Arakawa (1990) traditionally used in HIM, or the (unsplit) MPDATA scheme (Smolarkiewicz, 1984) used in HYCOM, the fluxes are the result of a multi-stage calculation, and therefore the fluxes through a face depend (in some cases nonlinearly) on the neighboring velocities as well as the velocity at that face. If the flux algorithms are applied robustly, then the appropriate values of $\widetilde{\boldsymbol{U}}$ are not independent, but involve nonlocal contributions. For such discretizations, it may be impractical to calculate the partial derivatives of the flux to find the exact set of velocities such that $\sum_{k=1}^{K} \boldsymbol{F} \left[\left(\boldsymbol{u}_k + \widetilde{\boldsymbol{U}} \right), h_k \right] = \langle \boldsymbol{U} H \rangle$. The nonlocal terms (which are typically relatively small, especially with a small Courant number) could probably be treated as though they were fixed with the neighboring first guess velocities, but this could destroy the positive-definiteness of the schemes. A flux-corrected-transport style limiter could be applied to retain positive definiteness, while keeping the flux calculations local to each water column. Alternately an iterative approach could be used to determine the complete field of $\widetilde{\boldsymbol{U}}$ including the nonlocal terms and without altering the discretization, but this would introduce potentially costly communications with each iteration. If the iterations introduce small updates to the fluxes, the full algorithm need not even be used for these updates. For instance, a final upwind iteration (upwind in the sense of the final corrective flux, not the full fluxes) along the lines described in Section 2.2.2 might be convenient as it can be used to enforce exact consistency between the barotropic flux and the sum of the layer fluxes in a single step. On balance, though, flux discretizations (such as PPM) that depend only locally on the velocity will be much more convenient to work with.

There is one other subtlety to the reconciliation that could be considered: a vertically uniform force does not lead to a vertically uniform acceleration if vertical viscosity and bottom drag are taken into account. Starting with Hallberg and Rhines (1996), vertical viscosity and bottom drag using upwind biased estimates of thickness have proven quite adept at representing the flow near the point where layers intersect the sloping bottom by suppressing apparent down-slope pressure-driven accelerations in nearly massless layers while permitting on-slope flows to inflate such layers. Vertical viscosity must be treated implicitly (after linearization, if necessary), and may greatly reduce the near-bottom velocity that results from a barotropic force. Schematically, if the change in the velocity

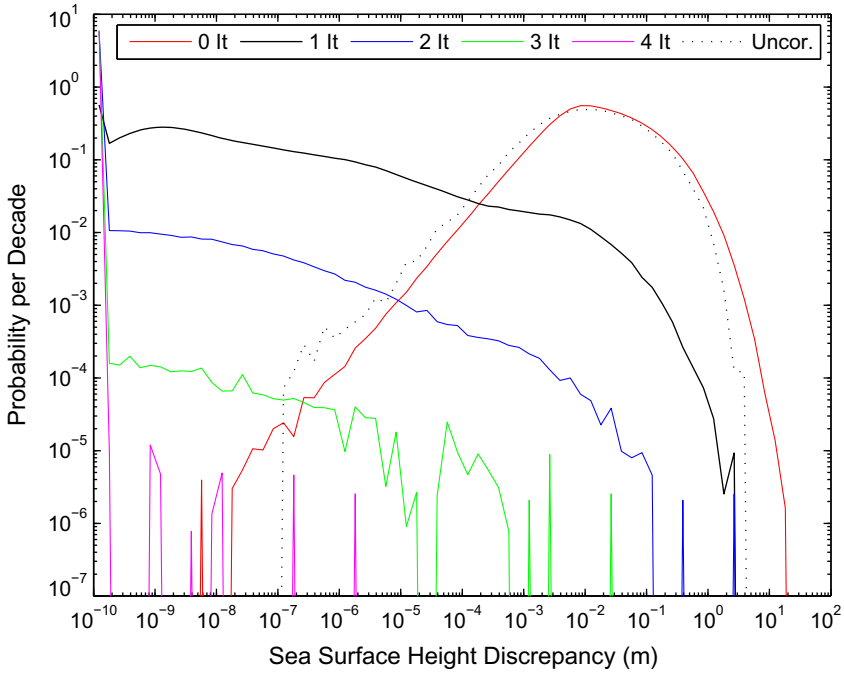

Fig. 3. The area-weighted histogram of the absolute discrepancies between the two estimates of the free surface height, from the same model as shown in Fig. 1, as a function of the number of corrective iterations used, averaged over 73 realizations 5 days apart. With 6 bins per decade, a single instance contributes about 10^{-6} to this curve. The curves move progressively to the left with each iteration, and converge rapidly once the velocities do not change sign in any layer between iterations. Also shown here, with a dotted line, is the histogram of the absolute discrepancy at one time from Fig. 2 using $\widetilde{\boldsymbol{U}} = \langle \boldsymbol{U} \rangle - \boldsymbol{U}_0$ to approximately adjust the velocities to agree with the time-mean barotropic velocity without iterative corrections. A PPM discretization is used for the continuity solver, as described in (31). The first iteration simply uses $\widetilde{\boldsymbol{U}}^0 = 0$ in (30), which is clearly worse than using $\widetilde{\boldsymbol{U}}^0 = \langle \boldsymbol{U} \rangle - \boldsymbol{U}_0$, but this simplification does not appreciably slow the iteration toward convergence.

after a timestep, δu_k, due to an incremental barotropic acceleration δA is only a fraction r_k of the integrated acceleration, so that

$$\delta u_k = r_k(\Delta t \delta A), \tag{33}$$

the corrective velocity in each layer should not be the vertically uniform $\widetilde{U}$, but proportional to r_k. The effects of vertical viscosity can be written as

$$h_k u_k^{n+1} = h_k(u_k^n + A_k \Delta t) + \frac{\nu_{k+1/2}\Delta t}{h_{k+1/2}}(u_{k+1}^{n+1} - u_k^{n+1}) - \frac{\nu_{k-1/2}\Delta t}{h_{k-1/2}}(u_k^{n+1} - u_{k-1}^{n+1}) \tag{34}$$

where h_k is the thickness of a layer, while $h_{k+1/2}$ is the effective thickness over which the stress between layers is calculated and $\nu_{k+1/2}$ is the viscosity between layers, with boundary conditions that the velocity at the bottom is zero, and the viscosity at a free-slip free surface is zero. Eq. (34) is linear in the velocities and accelerations, so the r_k can be determined by solving the tridiagonal equation

$$h_k r_k = h_k(1) + \frac{\nu_{k+1/2}\Delta t}{h_{k+1/2}}(r_{k+1} - r_k) - \frac{\nu_{k-1/2}\Delta t}{h_{k-1/2}}(r_k - r_{k-1}) \tag{35}$$

with the boundary conditions at the free surface that $r_0 = r_1$, and at the no-slip bottom that $r_{N+1} = -r_N$. The r_k are between 0 and 1, asymptoting to 1 far from the no-slip bottom. This tridiagonal equation is very similar to that for the velocities themselves (34) and can be solved efficiently at the same time.

Taking the vertical viscosity into account, it is not a barotropic velocity that is needed, as in (27), but rather a barotropic acceleration averaged over a timestep, so that (27) is replaced by the search for the acceleration A such that

$$\sum_{k=1}^{N} F[(u_k + r_k A \Delta t), h_k] = \langle UH \rangle \tag{36}$$

The layer transports increase monotonically with A, making this again a one-dimensional search for the root of a monotonically increasing function.

This iterative, flux-form reconciliation between the two estimates of the free surface height has been used in a variety of different applications, including a global 1° climate model similar to that shown in Fig. 1, and high-resolution tidal simulations following on from the work described by Simmons et al. (2004). As shown in Fig. 3, the two estimates typically agree everywhere to within a tolerance of 10^{-6} m after three or four iterations. The cost of this approach is not a substantially larger fraction of the ocean models CPU time than other continuity solvers, and may even be dramatically less in parallel models since it requires no communication between processors. As a desirable side-effect, determining the barotropic velocity correction that yields consistent barotropic and layer volume fluxes also eliminates the potential for temporal instability due to aliasing of rapidly varying external gravity waves into the layer thickness fluxes, as described by Higdon (2008). This approach seems to be a highly attractive method for forcing the two estimates of the free surface height to agree, without any obvious significant liabilities.

3. Two illustrations of the liabilities with traditional approaches

Two examples can be used to illustrate the complications with some approaches to reconciling the two estimates of the free surface height.

3.1. Spurious diapycnal mixing in DOME

The Dynamics of Overflows and Mixing Experiment (DOME) established a set of standard idealized test cases that have been used to study the representation entraining gravity currents in numerical ocean models (Legg et al., 2006), with physical parameters that are broadly characteristic of real oceanic overflows. Here, one of these test cases is used to illustrate the magnitude of the problems with non-conservation that arise when layer dilation is used to reconcile the two estimates of the free surface height.

The DOME test-case 1 was spun up for 50 days using the Generalized Ocean Layered Dynamics (GOLD)[5] model in a purely isopycnal coordinate configuration with an unsplit time stepping scheme. This *f*-plane case has an initial linear density stratification with a density range of $2\,\mathrm{kg\,m^{-3}}$, and a 1% bottom slope extending down 3600 m, and is forced by a geostrophically balanced inflow atop the slope at the density at the bottom. A vigorously entraining dense plume develops and extends along the slope. Sponges at the ends of the domain allow it to attain a statistically steady state. In the present configuration, the diapycnal mixing (and plume entrainment) is due to the shear-mixing parameterization of Jackson et al. (2008) and a weak background diffusivity. See Legg et al. (2006) for full details of the DOME configuration, or Legg et al. (2008) for a discussion of this particular model configuration.

After the model has spun up, the inflow of dense water and the sponges at the edges were then turned off, and the run carried on using each of the methods of reconciling the flows described above and the split time-stepping algorithm of Hallberg (1997), or with an unsplit time stepping scheme as the reference solution, for two test cases. In the first test case, the parameterized diapycnal mixing was deliberately disabled. As this is a purely isopycnal-coordinate configuration, it should be able to preserve the histogram of watermasses without change. Fig. 4a shows that for both the non-split run and for a run using a split time-stepping scheme with the iterative flux-form reconciliation proposed in this paper, the volume of water below each isopycnal is unchanged. This is also true of the upwind flux form solution. By contrast, when layer dilation is used to reconcile the two estimates (as in Bleck and Smith (1990)), there are significant changes in the volume of water below isopycnals, equivalent to domain-averaged fluxes of order $0.1\,\mathrm{m\,day^{-1}}$ averaged over 5 days in this example. These changes are the basin-averaged residuals of even larger localized non-conservation of layer volume of both signs, but they are not completely random. As they arise from differences between the baroclinic and barotropic estimates of the mass fluxes, they do lead to systematic and accumulating biases over timescales of days or longer, as illustrated in Fig. 5.

The significance of these changes can be assessed with a second set of DOME runs, in which the diapycnal mixing was left on (although the inflow and sponges are still disabled). There is significant shear-driven diapycnal mixing, with the Jackson et al. (2008) scheme yielding 5-day average diffusivities averaged over the dense water plume of order $10^{-2}\,\mathrm{m^2\,s^{-1}}$, and peak values of time mean diffusivities as high as $10^{-1}\,\mathrm{m^2\,s^{-1}}$. This vigorous mixing leads to downward movements of the isopycnals near the bottom and upward isopycnal movements at the top of the plume as it entrains, as seen in Fig. 4b. The iterative flux-form solution agrees reasonably well with the unsplit solution (as does the upwind flux-form solution), while there are clear discrepancies between the mean interface height changes in the unsplit solution and the solution that uses layer dilation. The diapycnal mixing in this case is strongly nonlinear, so it is impossible to attribute all of these discrepancies in layer

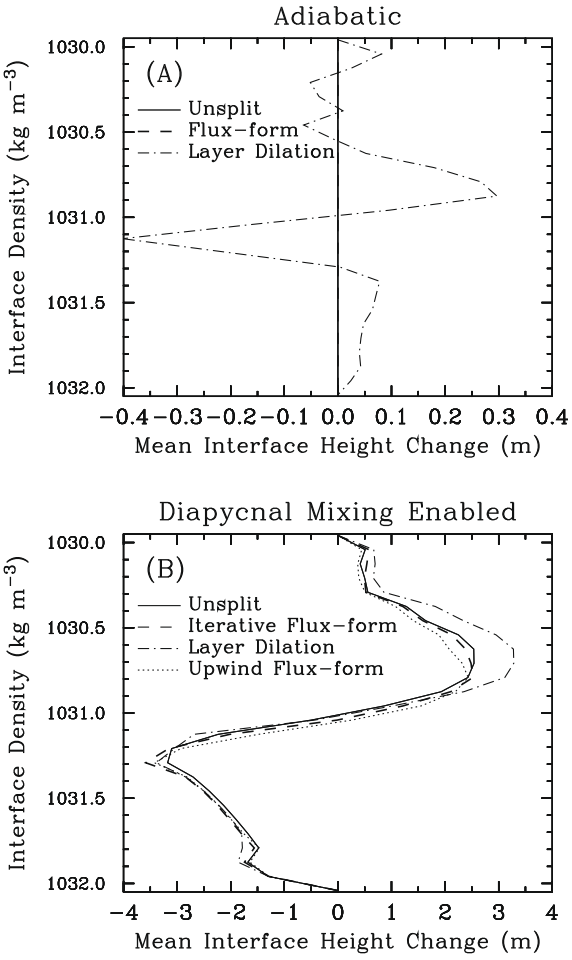

Fig. 4. The change after 5 days in the depth that each isopycnal would have if the model were adiabatically adjusted to the minimum potential energy state. That is, the change in the depth below which the volume of ocean equals the volume of water below each interface. This is a measure of how well the census of watermasses is preserved. In the top panel (A), the model was run with all parameterized cross-isopycnal mixing disabled; the correct solution is uniformly 0 change in this case. The bottom panel (B) uses the Jackson et al. (2008) shear mixing parameterization and a background diapycnal diffusivity of $10^{-4}\,\mathrm{m^2\,s^{-1}}$. An unsplit integration (solid) is here as a reference, and is run with a very short time step to resolve the barotropic waves. The iterative flux-form scheme (dashed) and upwind flux-form scheme (dotted in B) agree very well with the reference solution. By contrast when layer dilation is used to reconcile the two estimates (as proposed by Bleck and Smith (1990) and used in MICOM and HYCOM), there is significant nonconservation of integrated layer volume in (A) and significant departures from the reference solution in (B).

volume directly to the dilation, but it is highly suggestive that the magnitude of the discrepancy is comparable to that seen in the adiabatic case in Fig 4a. These volume conservation imbalances due to the dilation are of haphazard sign, and cannot be directly translated into a diffusivity, but it may be worth noting that in this particular test case, the spurious regionally-averaged watermass volume modifications due to the reconciliation approach of Bleck and Smith (1990) are comparable in magnitude (if not everywhere of corresponding sign) to that given by a diapycnal diffusivity of order $10^{-3}\,\mathrm{m^2\,s^{-1}}$, which is two orders of magnitude larger than the observationally inferred background diffusivity of the interior ocean (Ledwell et al., 1993; Kunze and Sanford, 1996).

There is also a significant non-conservation of integrated density in this test case with the dilation approach, but not with either the flux-form or the unsplit solutions. To the extent that this comparison and this test case are meaningful, they strongly suggest that the use of this dilation will be highly problematic for the

[5] GOLD is a generalized ocean model being developed at NOAA GFDL as the successor to both HIM and the Modular Ocean Model (MOM). As configured here, the solutions are essentially the same as would have been obtained with HIM.

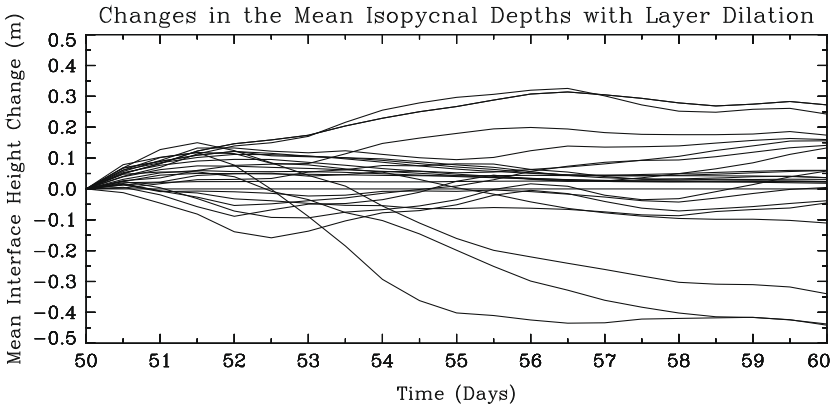

Fig. 5. The change over time in the same mean interface anomalies shown in Fig. 4a for the adiabatic case using the layer dilation proposed by Bleck and Smith (1990). The correct solution is uniformly 0 change in this case. This figure illustrates that these spurious watermass changes accumulate over timescales of days, but are likely to be akin to a random walk on longer timescales.

use of such a model for long-term climate studies. There is circumstantial evidence to support this conjecture, for example in the inter-model comparisons of Griffies et al. (2009).

3.2. Surface gravity waves in a climate model's Black Sea

The importance of an essentially exact approach for reconciling the two estimates of the free surface height can be illustrated by considering the Mediterranean and Black Seas taken from a 1° resolution global model. This region is particularly illustrative for several reasons. The density ranges of the Black and Mediterranean Seas are quite distinct from the main body of the ocean; the Black Sea is much fresher and lighter, while the Mediterranean is atypically salty and dense. This poses a particular challenge for an isopycnal coordinate model that is configured for global applications, as there will be relatively few layers of the right density classes; in the example shown in Fig. 6, there are just 10 interior layers active in the Mediterranean and just 1 interior layer in the Black Sea, in addition to the 4 variable density mixed- and buffer-layers that are available everywhere. With so few layers, the assumption in (13), that the various estimates of the layer thicknesses at velocity points are equivalent, is rather poor near topography. In addition, the large-scale boundary currents and the large-scale external gravity wave fields in these isolated seas are relatively weak compared to the open ocean basins, so any signal radiated from numerical imbalances will be easily detected.

The differences in the quality of the solutions using different methods of reconciling the free surface height estimates are particularly evident in Fig. 6, which shows the instantaneous deviations of the sea surface height from its 2-day mean in a model of the Mediterranean and Black Seas, extracted from the global model shown in Fig 1. The large values with the pseudo-mass-source reconciliation, of order tens of cm up to a few m, are indicative of external gravity waves bouncing around these two seas. There are physical sources for large-scale external gravity waves – namely wind-driven seiches, but these are much smaller than the numerical error induced signals in the bottom panel of Fig. 6. By contrast, the iterative flux form (top panel of Fig. 6) gives quite plausible magnitudes for the high frequency variability in the sea-surface height.

This difference in the quality of the solutions can also be easily seen in the RMS deviations of the Black Sea surface height from its instantaneous spatial mean (Fig. 7). The Black Sea is small enough that the sea surface height deviations associated with the geostrophic flows are relatively modest (e.g., Staneva et al., 2001), and this makes the deviations from the spatial mean a good measure of the amplitude of the surface gravity wave field. With the pseudo-mass-source reconciliation the numerically induced large-scale external gravity wave field is roughly 20 times stronger than in a model run with a very short time step and no mode splitting. The unsplit model has very similar amplitude gravity waves with the iterative flux form reconciliation. Although there are many applications of ocean models for which the surface gravity wave field may be unimportant, there are undesirable consequences from having such large amplitude waves as found with the pseudo-mass-source version. These waves steepen near the boundaries (they are dynamically equivalent to tsunamis), and can lead to such large velocities that the model's CFL limits are violated and the model solution diverges.[6] But even beyond numerical stability, the ocean is in fact a nonlinear system, and such large gravity waves can compromise the physical relevance of the system, for example by driving excessive shear-driven mixing or enhancing the quadratic bottom drag. From these simulations, it would seem advisable not to use the pseudo-mass-source reconciliation of the two estimates of the free surface height in realistic ocean models, while the iterative flux form reconciliation works quite well.

In these simulations, the upwind flux form approach gives a similar external gravity wave field to the iterative flux form approach and to an unsplit solution – which is to be expected since by construction the two estimates of the sea surface height agree. The same is not true everywhere for the long-term mean circulation. The 1-year time mean sea surface height relative to the time- and spatial-mean sea surface height over a marginal sea, $\langle \eta' \rangle$, is roughly equivalent to a time-mean geostrophic streamfunction. For the Black Sea in these simulations, the mean-squared difference between $\langle \eta' \rangle$ for an iterative flux-form solution and an unsplit solution is 9 mm^2, while for an upwind flux form solution this difference is 80 mm^2, compared with a mean-squared value for $\langle \eta' \rangle$ itself of 268 mm^2. The Black Sea circulations are qualitatively similar overall, but the iterative flux form circulation is clearly much closer to the unsplit solution. The Black Sea is particularly challenging because the free modes can be prominent. In the Mediterranean, both iterative and upwind flux form solutions are quite similar to the unsplit solution. Both have mean-squared differences in $\langle \eta' \rangle$ from the unsplit solution of 1 cm^2, which is small compared with a mean-squared value for $\langle \eta' \rangle$ itself of 37 cm^2. The Mediterranean has better vertical resolution in these simulations than does the

[6] The Black Sea occasionally generated such large velocities in an early version of the global model shown here (which used a time-filtered variant of pseudo-mass-source reconciliation) that an unduly short timestep was required for stability. These problems were largely the motivation for pursuing this work in the first place.

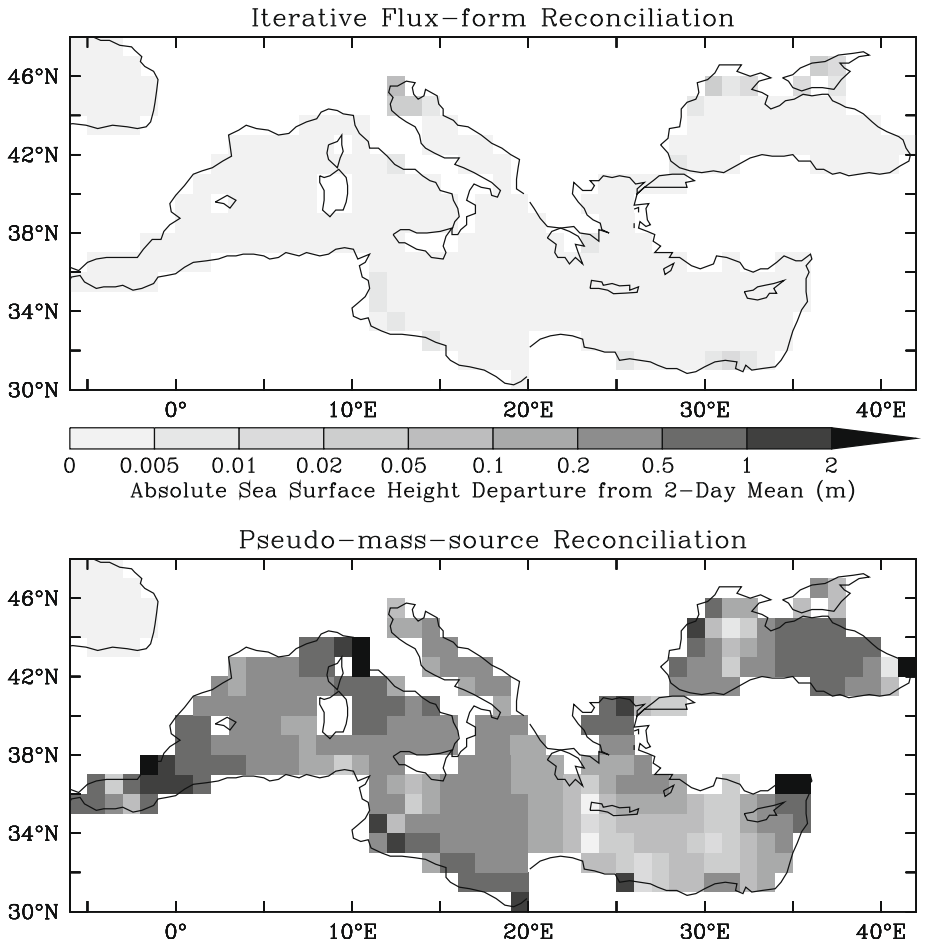

Fig. 6. The magnitude of the deviation of the free surface height from its 2-day mean value in a 1° CORE-forced simulation of the Black and Mediterranean seas extracted from the global model shown in Fig. 1. The top panel uses the iterative flux-form reconciliation, while the bottom panel uses the pseudo-mass source reconciliation. A model run with a short time step and no mode splitting (not shown) is similar in magnitude to the top panel. Both models have been spun up for 165 days, and this time is typical of the magnitudes of the sea-surface height anomalies throughout the spring.

Black Sea, and the overall circulation is better governed by large-scale dynamics, so it is perhaps not surprising that the simulated Mediterranean circulation is less sensitive to numerical choices than is the Black Sea. Based on these runs, both the upwind and iterative flux form approaches are stable and may give acceptable results, but the iterative approach appears to be the more robust and accurate of the two.

4. Conclusion and discussion

Time-split ocean models generate two distinct estimates for the evolution of the free surface, one from the barotropic solver and the other from the vertical integration of the interior continuity equation. In models with a linear continuity solver (as is true of essentially all fixed-grid models), it is straightforward to cause the vertical integral of the interior continuity equation to exactly agree with the external solution by using the appropriate barotropic velocity. Lagrangian vertical coordinate models, however, must use nonlinear continuity solvers to ensure positive definiteness of the vertical coordinate. This has led to several different approaches for how to reconcile these two estimates of the free surface height in practical Lagrangian vertical coordinate ocean models.

The approach of dilating the water column to agree with the barotropic estimate of the free surface height (Bleck and Smith, 1990) has been used extensively in a wide variety of simulations, particularly with HyCOM and MICOM. Any spurious generation of external gravity waves is limited, but at a serious price. Although total mass is conserved, neither heat, total salt, nor any other tracer can be exactly conserved with this approach. Also, there can be significant numerical transfer of water between density classes. In the overflow test case examined here, the spurious transfer of water was quite large – equivalent in magnitude to that arising from a diapycnal diffusivity of order 10^{-3} m^2 s^{-1}. This particular test case is oceanographically relevant, but is surely an over-estimate of the magnitude of the spurious watermass modification that should be expected for the ocean as a whole. But as minimizing spurious watermass modifications is ostensibly one of the primary motivations for using an isopycnal (or hybrid-isopycnal) coordinate model for long-term oceanographic studies and climate simulations, abandoning this desirable property seems very ill-advised. For climate studies, non-conservation of heat or salt will seriously complicate the analysis of an ocean model. The dilation approach is probably only advisable for relatively short-term ocean simulations, when careful watermass conservation is not an especially important property.

Flux-form methods for reconciling the two estimates conserve tracers by construction. The challenge for these techniques is to ensure that the resulting circulations are reasonable. Adding a fake mass source to the barotropic solver is one way to drive the two estimates of the free surface height back toward one another. Tracers and watermasses are exactly conserved because the layer equations retain their original flux-convergence form. Although it

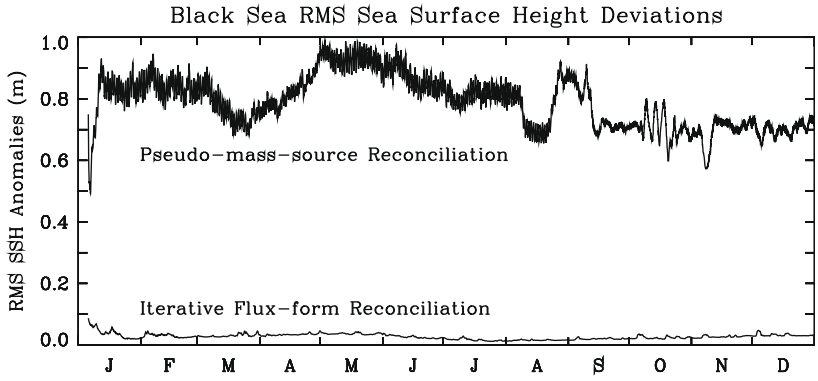

Fig. 7. The RMS deviation of the free surface height in the Black Sea from the instantaneous spatial average sea surface height in the Black Sea, from the same simulations as depicted in Fig. 6, using the pseudo-mass-source and iterative flux-form methods for reconciling the sea surface height. The equivalent curves for an unsplit simulation with short timesteps, or for upwind corrective layer fluxes are almost indistinguishable form the iterative flux-form curve; the temporary increases in the RMS height deviations in those cases are dominated by the passage of storms in the wind stress datasets.

has been used for about a decade in HIM (e.g., Hallberg and Gnanadesikan, 2006), this approach can lead to the generation of large-amplitude external gravity waves that will be highly problematic in any simulation where the exact sea-surface height is of interest. These waves are particularly prominent in cases with high-frequency (i.e., realistic) forcing variability. Given that this is not very much more computationally efficient than other flux-form approaches to reconciling the two estimates of the free surface height, this approach has clearly been superseded.

An upwind velocity correction can be used to force the sum of the layers' fluxes to equal the time-mean barotropic fluxes (e.g., Higdon, 2005). Tracers and watermasses are conserved by construction, and there is not spurious generation of large external gravity waves. This approach also has the advantage that the corrections can be determined in a single step, and it is essentially independent of the choice of discretization for the continuity equation. The challenge with this approach is that a full-depth (barotropic) correction can lead to (spurious) rectified flow of the densest water up slopes, while limiting the correction to the water column that is entirely above the bathymetry introduces baroclinic shears to counteract an essentially barotropic imbalance and can lead to a small rectified down-slope drift of dense bottom boundary currents relative to the true solution (as occurs in the DOME test case).[7] Because different thicknesses are used to determine the original and upwind-corrective fluxes, it is even possible for the net mass fluxes in a layer to be in the opposite direction to the sum of the original and corrective velocities. The success of this scheme is sensitive to the estimate of the barotropic velocity, $\langle U \rangle$, used in the uncorrected internal continuity equations; as an extreme test, when a lagged barotropic velocity is used in place of the time average, the solutions with the upwind velocity correction diverge for the Black and Mediterranean seas test case in Section 3, even though the two estimates of the free surface height agree perfectly. Despite these caveats about the care with which this scheme must be implemented, it can lead to acceptable simulations and avoids the greatest problems with the dilation or pseudo-mass-source approaches.

Finally, a new approach described here is to determine the additional barotropic acceleration (or velocity increment) that will cause the sum of the layer fluxes to agree with the barotropic fluxes. Like the upwind flux-form correction, this approach exactly conserves tracers and watermasses and avoids spurious generation of large external gravity waves. It also has the advantage that the vertical distribution of the fluxes are entirely consistent with the model's dynamics and the chosen discretization for the internal continuity equations; in this sense it can be expected to closely reproduce what would have been obtained with an unsplit time-stepping scheme with a very short timestep – something that in fact happens in all the tests presented here. This approach is insensitive to the time-averaged estimate of the barotropic velocity, $\langle \boldsymbol{U} \rangle$, except perhaps as a starting point for the iterations to determine the internally consistent barotropic velocity $\widetilde{\boldsymbol{U}}$. The difficulty with this approach is that it is necessarily iterative with a nonlinear continuity discretization, and its implementation will vary strongly with the choice of continuity discretization. There are viable continuity solvers where the volume fluxes depend non-locally on the velocities (e.g., Hsu and Arakawa, 1990), for which there may not even be a unique solution. If an appropriate continuity discretization (e.g., PPM) is used, though, the iterations converge to a unique solution very rapidly and this approach is not significantly more time-consuming than other approaches. Based on the examples shown here and many others that were not shown, this new iterative flux-form reconciliation appears to be a very attractive approach to advancing the free surface of a time-split Lagrangian vertical coordinate ocean model, without any compromise of conservation and without generation of spurious waves or flows.

[7] This inconsistency is avoided if the same restrictions of the transport to the water above the shallowest neighboring topography are applied everywhere in the continuity equations, in the upwind corrections, and in the barotropic solver, although this will lead to only a first-order accurate representation of the topographic influence on the ocean flow.

Acknowledgements

The authors thank a number of people, including Mitsuhiro Kawase, who pointed out this particular challenge and started RWH thinking about it a very long time ago. Stephen Griffies, Matt Harrison, and Laurent White gave us thorough internal reviews that substantially improved this paper, as did comments from Bob Higdon and an anonymous reviewer. We also thank Paul Schopf, and Matts Bentsen for useful conversations about these issues.

References

Adcroft, A., Hallberg, R., 2006. On methods for solving the oceanic equations of motion in generalized vertical coordinates. Ocean Modell. 11, 224–233.
Bleck, R., 2002. An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Modell. 4, 55–88.
Bleck, R., Rooth, C., Hu, D., Smith, L.T., 1992. Salinity-driven thermocline transients in a wind- and thermohaline-forced isopycnic coordinate model of the North Atlantic. J. Phys. Oceanogr. 22, 1486–1505.
Bleck, R., Smith, L., 1990. A wind-driven isopycnic coordinate model of the North and Equatorial Atlantic Ocean. 1. Model development and supporting experiments. J. Geophys. Res. 95, 3273–3285.

Campin, J.-M., Adcroft, A., Hill, C., Marshall, J., 2004. Conservation of properties in a free-surface model. Ocean Modell. 6, 221–244.
Collela, P., Woodward, P., 1984. The piecewise-parabolic method (PPM) for gas-dynamical simulations. J. Comp. Phys. 87, 174–201.
deSzoeke, R., Higdon, R., 1997. Barotropic–baroclinic time splitting for ocean circulation modeling. J. Comp. Phys. 135, 30–53.
Dukowicz, J.K., Smith, R.D., 1994. Implicit free-surface method for the Bryan–Cox–Semtner ocean model. J. Geophys. Res. 99, 7991–8014.
Durran, D.R., 1999. Numerical Methods for Wave Equations in Geophysical Fluid Dynamics. Springer. 465 p.
Easter, R.C., 1993. Two modified versions of Bott's positive-definite numerical advection scheme. Mon. Wea. Rev. 121, 297–304.
Griffies, S.M., Pacanowski, R.C., Hallberg, R., 2000. Spurious diapycnal mixing associated with advection in a *z*-coordinate ocean model. Mon. Wea. Rev. 128, 538–564.
Griffies, S.M., Pacanowski, R.C., Schmidt, M., Balaji, V., 2001. Tracer conservation with an explicit free surface method for *z*-coordinate ocean models. Mon. Wea. Rev. 129, 1081–1098.
Griffies, S.M., Biastoch, A., Böning, C., Bryan, F., Chassignet, E., England, M., Gerdes, R., Haak, H., Hallberg, R., Hazelager, W., Jungclaus, J., Large, W.G., Madec, G., Samuels, B.L., Scheinert, M., Severijns, C.A., Simmons, H.L., Treguier, A.M., Winton, M., Yeager, S., Yin, J., 2009. A proposal for Coordinated Ocean-ice Reference Experiments (COREs). Ocean Modell 26, 1–46.
Hallberg, R., 1997. Stable split time stepping schemes for large-scale ocean modeling. J. Comp. Phys. 135, 54–65.
Hallberg, R., Rhines, P., 1996. Buoyancy-driven circulation in an ocean basin with isopycnals intersecting the sloping bottom. J. Phys. Oceanogr. 26, 913–940.
Hallberg, R., Gnanadesikan, A., 2006. The role of eddies in determining the structure and response of the wind-driven Southern Hemisphere overturning: Results from the Modeling Eddies in the Southern Ocean (MESO) project. J. Phys. Oceanogr. 36, 2232–2252.
Higdon, R., 2005. A two-level time-stepping method for layered ocean circulation models: further development and testing. J. Comp. Phys. 206, 463–504.
Higdon, R., 2008. A comparison of two formulations of barotropic–baroclinic splitting for layered models of ocean circulation. Ocean Modell. 24, 29–45.
Holloway, G., 1992. Representing topographic stress for large-scale ocean models. J. Phys. Oceanogr. 22, 1033–1046.
Hsu, Y.-J.G., Arakawa, A., 1990. Numerical modeling of the atmosphere with an isentropic vertical coordinate. Mon. Wea. Rev. 118, 1933–1959.
Jackson, L., Hallberg, R., Legg, S., 2008. A Parameterization of shear-driven turbulence for ocean climate models. J. Phys. Oceanogr. 38, 1033–1053.
Killworth, P.D., Webb, D.J., Stainforth, D., Paterson, S.M., 1991. The development of a free-surface Bryan–Cox–Semtner ocean model. J. Phys. Oceanogr. 21, 1333–1348.
Kunze, E., Sanford, T.B., 1996. Abyssal mixing: where it is not. J. Phys. Oceanogr. 26, 2286–2296.
Ledwell, J.R., Watson, A.J., Law, C.S., 1993. Evidence for slow mixing across the pycnocline from an open-ocean tracer-release experiment. Nature 364, 701–703.
Legg, S., Hallberg, R., Girton, J., 2006. Comparison of entrainment in overflows simulated by *z*-coordinate, isopycnal and non-hydrostatic models. Ocean Modell. 11, 69–97.
Legg, S., Jackson, L., Hallberg, R., 2008. Eddy-resolving modeling of overflows. In: M. Hecht, M. Hasumi (Eds.), Eddy-resolving Ocean Modeling. AGU, pp. 63–81.
Shchepetkin, A., McWilliams, J., 2005. The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. Ocean Modell. 9, 347–404.
Simmons, H., Hallberg, R.W., Arbic, B.K., 2004. Internal wave generation in a global baroclinic tide model. Deep Sea Res. II 51, 3043–3068.
Smolarkiewicz, P.K., 1984. A fully multidimensional advection transport algorithm with small implicit diffusion. J. Comp. Phys. 67, 325–362.
Staneva, J.V., Dietrich, D.E., Stanev, E.V., Bowman, M.J., 2001. Rim current and coastal eddy mechanisms in an eddy-resolving Black Sea general circulation model. J. Mar. Sys. 31, 137–157.
Strang, G., 1968. On the construction and comparison of difference schemes. SIAM J. Numer. Anal. 5, 506–517.