Ocean Modelling 26 (2009) 156–169

Contents lists available at ScienceDirect

Ocean Modelling

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

Spurious diapycnal mixing in terrain-following coordinate models: The problem and a solution

Patrick Marchesiello [a,*], Laurent Debreu [b], Xavier Couvelard [a]

[a] IRD, BP A5, 98848 Noumea, New Caledonia
[b] INRIA, Laboratoire Jean Kuntzmann, BP 53, 38041 Grenoble Cedex 9, France

ARTICLE INFO

Article history:
Received 10 June 2008
Received in revised form 5 September 2008
Accepted 18 September 2008
Available online 2 October 2008

Keywords:
Diapycnal mixing
Advection schemes
Terrain-following coordinates

ABSTRACT

In this paper, we identify a crucial numerical problem in sigma coordinate models, leading to unacceptable spurious diapycnal mixing. This error is a by-product of recent advances in numerical methods, namely the implementation of high-order diffusive advection schemes. In the case of ROMS, spurious mixing is produced by its third-order upwind advection scheme, but our analysis suggests that all diffusive advection schemes would behave similarly in all sigma models. We show that the common idea that spurious mixing decreases with resolution is generally false. In a coarse-resolution regime, spurious mixing increases as resolution is refined, and may reach its peak value when eddy-driven lateral mixing becomes explicitly resolved. At finer resolution, diffusivities are expected to decrease but with values that only become acceptable at resolutions finer than the kilometer. The solution to this problem requires a specifically designed advection scheme. We propose and validate the RSUP3 scheme, where diffusion is split from advection and is represented by a rotated biharmonic diffusion scheme with flow-dependent hyperdiffusivity satisfying the Peclet constraint. The rotated diffusion operator is designed for numerical stability, which includes improvements of linear stability limits and a clipping method adapted to the sigma-coordinate. Realistic model experiments in a southwest Pacific configuration show that RSUP3 is able to preserve low dispersion and diffusion capabilities of the original third-order upwind scheme, while preserving water mass characteristics. There are residual errors from the rotated diffusion operator, but they remain acceptable. The use of a constant diffusivity rather than the Peclet hyperdiffusivity tends to increase these residual errors which become unacceptable with Laplacian diffusion. Finally, we have left some options open concerning the use of time filters as an alternative to spatial diffusion. A temporal discretization approach to the present problem (including implicit discretization) will be reported in a following paper.

1. Introduction

Reducing the traditional errors in terrain-following coordinate ocean models (or sigma models) has been a focus of interest for the last two decades. The objective is to use this class of model in regional domains which include not only the continental shelf, but the slope and deep ocean as well. Two general types of error have been identified: (1) the pressure-gradient error and (2) spurious diapycnal diffusion associated with steepness of the vertical coordinate. At the time when state-of-the-art ocean models were using second-order centered advection schemes, large amounts of explicit lateral mixing was required to maintain stability. As a result, spurious diapycnal mixing was quickly identified as the main source of error in sigma models when applied to regions encompassing the continental slope (Barnier et al., 1998). Two different types of solution to this problem were then proposed. Firstly, Mellor et al. (1998) suggested applying diffusion only to a tracer perturbation with respect to a reference frame such as climatology. Barnier et al. (1998) argued for a more controllable approach: a rotated diffusion operator to align the diffusive fluxes along geopotential surfaces. This followed the earlier implementation of isopycnal diffusion in climate models (Cox, 1987), using the principle of coordinate rotation formalized by Redi (1982).

After having temporarily solved the diffusion problem, attention was then drawn to the pressure-gradient error problem, resulting in significant improvements (Beckmann and Haidvogel, 1993; McCalpin, 1994; and Shchepetkin and McWilliams, 2003). During the course of these improvements, numerical models were subjected to profound changes which were due to the demonstration that second-order schemes are not optimal methods for ocean (or atmospheric) modeling when accuracy is considered against computational costs (Sanderson, 1998). In the process, large efforts were made to reduce dispersion errors and lower excess diffusion in both temporal and spatial discretization. Tracer advection received particular attention and diffusive advection schemes were

* Corresponding author. Tel./fax: +687 26 07 24.
E-mail address: Patrick.Marchesiello@ird.fr (P. Marchesiello).

1463-5003/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ocemod.2008.09.004

devised (see Shchepetkin and McWilliams, 1998; Hecht et al., 2000, for a review). The success of these schemes was only shadowed quite recently by the discovery in climate models that implicit diffusion in diffusive advection schemes is large enough to produce excessive diapycnal mixing in geopotential coordinate models (Griffies et al., 2000). In this paper, we show that the issue is more dramatic in sigma-coordinate models, though as yet lacking acknowledgement. We then propose and validate a new scheme, which is a first step towards high quality advection schemes adapted to the sigma-coordinates.

2. Model description and problem identification

The model employed in this study is the Regional Oceanic Modeling System (ROMS). For a complete description, the reader is referred to Shchepetkin and McWilliams (2003, 2005). The model was recently used in a study of the southwest Pacific region (Couvelard et al., 2008) where the problem of spurious diapycnal mixing was identified. Here we briefly describe the model and the southwest Pacific configuration which is used to identify the diffusion problem and, in subsequent sections, to validate our solution to this problem.

2.1. Model overview

ROMS is a split–explicit, free-surface and terrain-following vertical coordinate oceanic model, where short time steps are used to advance the surface elevation and barotropic momentum equation and where a much larger time step is used for temperature, salinity, and baroclinic momentum. ROMS employs a two-way time-averaging procedure for the barotropic mode which satisfies the 3D continuity equation. A third-order, upstream-biased, dissipative advection scheme for momentum, implemented with a specially designed predictor–corrector time-step algorithm (Leapfrog–Adams–Moulton III, LF-AM3; Shchepetkin and McWilliams, 2005), allows the generation of steep gradients, enhancing the effective resolution of the solution for a given grid size (Shchepetkin and McWilliams, 1998). In the standard run, a third-order, upstream advection scheme is also used for tracers, which will reveal the diffusion problem. In this case, explicit lateral viscosity is null everywhere in the model, except in sponge layers near the open boundaries where it increases smoothly on several grid points. A K-profile parameterization (KPP) boundary layer scheme (Large et al., 1994) parameterizes the subgrid-scale vertical mixing processes. The algorithm developments on tracer advection presented in this study were implemented in the nested version of ROMS (ROMS_AGRIF) but are now being implemented also in the larger community version (myroms.org).

2.2. Configuration for the southwest Pacific

The presence of steep reef slopes and distinct intermediate water masses will make the southwest Pacific model configuration an interesting testbed for the present study. To encompass the steep topographic features most relevant to the circulation in this region, we have designed a rectangular grid extending from 30°S to 10°S in latitude and from 141°E to 191°E in longitude. The model grid, forcing initial and boundary conditions are built using the ROMSTOOLS package (Penven et al., 2007). The bottom topography is derived from a 2′ resolution database ETOPO2 from NOAA-NGDC (Fig. 1). Although a new pressure-gradient scheme associated to a modified equation of state limits computational errors of the pressure-gradient (Shchepetkin and McWilliams, 2003), the bathymetry h still needs to be smoothed, so that the "slope parameter" $r = \Delta h/2h$ (Beckmann and Haidvogel, 1993) is everywhere lower than a maximum value $r_0 = 0.2$. To preserve a sufficient resolution in the upper ocean, we use 35 vertical levels with stretched s-coordinates near the surface. The surface layer resolution ranges from 1.4 m to 15 m and a maximum grid size of 900 m is in the bottom layer.

The model external forcing is derived from observed and modeled climatology. At the surface, the model heat and fresh water fluxes are extracted from the COADS ocean surface monthly climatology at 1/2° resolution (Da Silva et al., 1994). The thermal feedback of the model ocean surface temperature to the surface heat flux is presented as a correction with respect to the COADS sea surface temperature (Barnier et al., 1995). A similar correction is applied to the surface salinity flux. For momentum fluxes, a monthly mean wind stress climatology is computed and interpolated on the model grid, using the QuikSCAT satellite scatterometer data for the period 1999–2006, provided by CERSAT (www.ifremer.fr/cersat/en/index.htm) on a half degree grid.

At the three lateral boundaries facing the open ocean, a mixed passive–active, implicit, radiation condition connects the model solution to the surroundings (Marchesiello et al., 2001). In the case of inflow conditions, the solution at the boundary is nudged toward monthly time-averaged outputs of the ORCA05 global ocean model, which had been run on a 1/2° resolution grid for the period

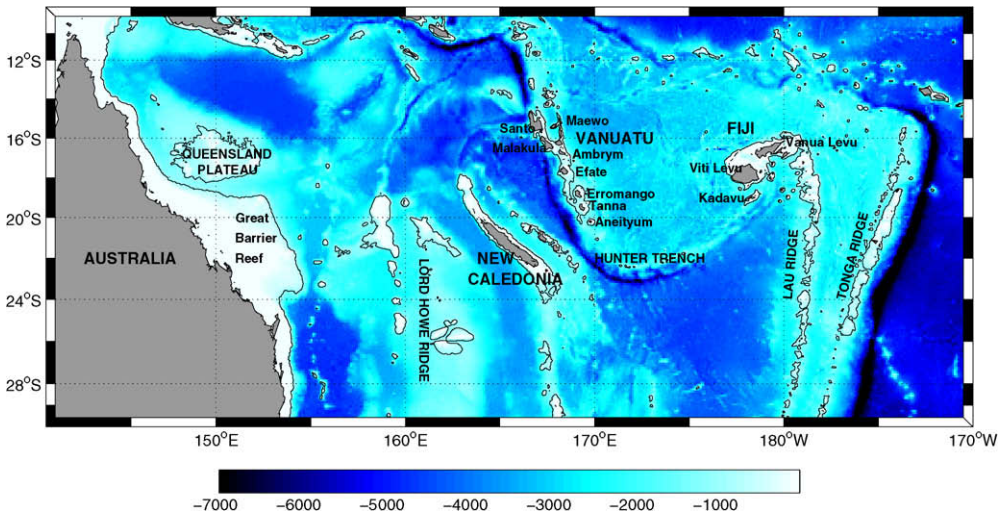

Fig. 1. Geography of the southwest Pacific region. The topography is essentially steep and rugged with meridional orientation.

1992–2002 (Kessler and Gourdeau, 2007). Forcing of inter-annual signals such as El-Niño is explicitly excluded from this study, either at the surface or lateral open boundaries. Nevertheless, intrinsic inter-annual variability is allowed to generate in the model as a result of non-linear instabilities of the large-scale circulation (Marchesiello et al., 2003). Finally, the model is initialized using January-mean values from ORCA05 climatology.

The model was run in parallel on a 10-node PC-cluster. The horizontal resolution used for the validation of our new advection scheme is 1/6°, but resolution sensitivity tests (at 1/2°, 1/4°, 1/6°, and 1/12° resolutions) serves as a measure of the convergence level of the model solution. In Couvelard et al. (2008) the statistical equilibrium solutions were analyzed, but in this study the first 3 years of integration only (spin-up phase) will be shown as it reveals the diffusion problem more clearly.

2.3. Problem identification

Fig. 2 shows the salinity field at 1000 m during the spin-up phase of the standard ROMS southwest Pacific simulation, using the third-order upstream advection scheme. The initialization is derived from ORCA05 which has been particularly well tuned to

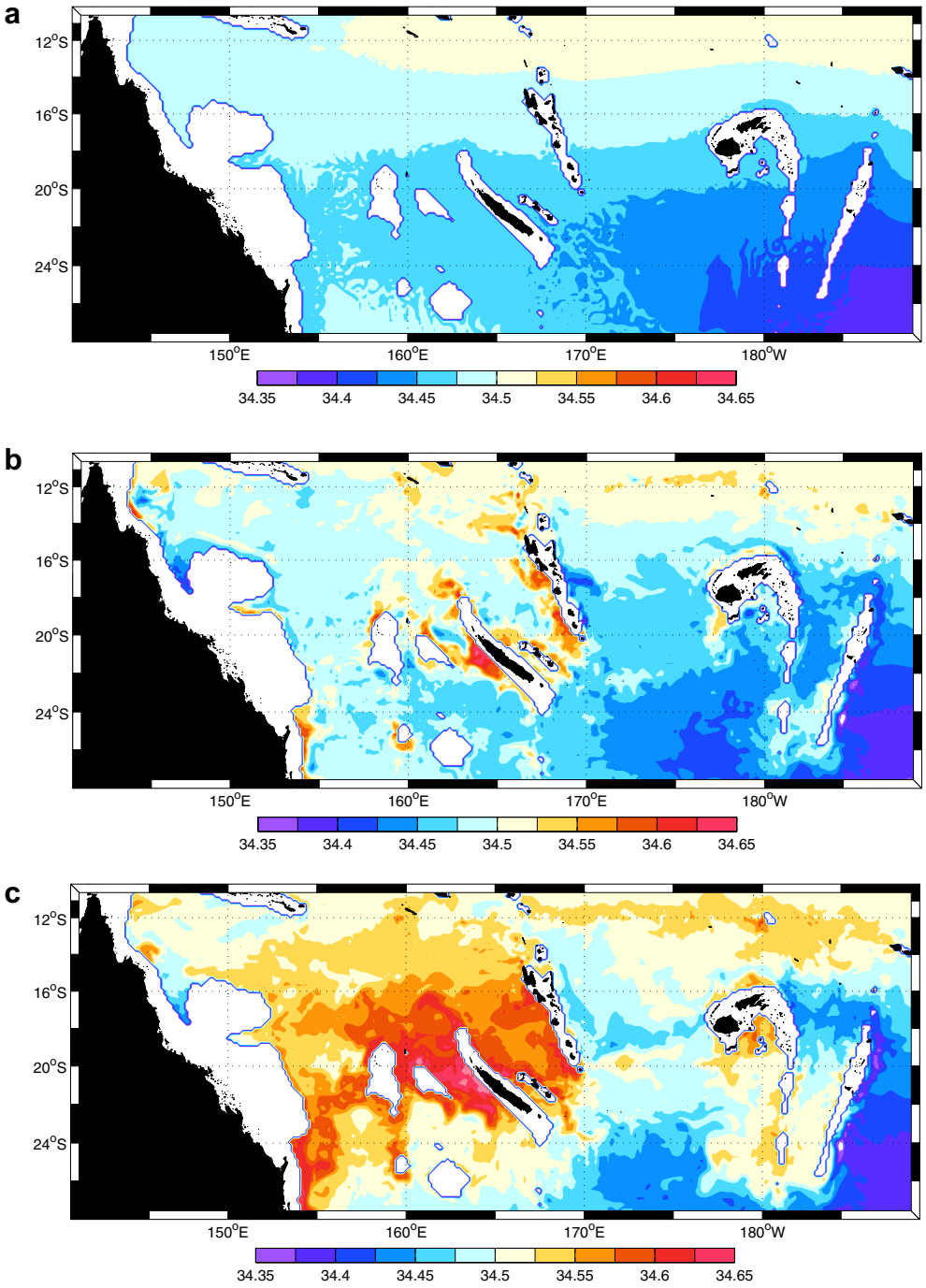

Fig. 2. Two-year evolution of salinity at 1000 m in the standard ROMS simulation (1/6°), using UP3 advection scheme. This phase corresponds to the model spinup, starting from ORCA05 mean state which is close to the Levitus climatology. Salinity at this depth is characterized by a minimum corresponding to the equatorward spreading of AAIW. (a) Initial state. (b) Solution at 6 months. (c) Solution at 2 years.

preserve the large-scale water masses. Therefore, temperature and salinity at depth are close to large-scale observations, as compiled for example in the CSIRO Atlas of Regional Seas (CARS: Dunn and Ridgway, 2002). Salinity at 1000 m is characterized by a vertical minimum corresponding to the equatorward spreading of Antarctic Intermediate Water (AAIW). On the time-scale of a few years and with seasonal forcing only, we expect this minimum to adjust dynamically to the regional solution, but with little large-scale variations. Clearly, this is not the case and the salinity minimum becomes quickly eroded, with the source of erosion being situated near the steepest island reefs, particularly the western side of New Caledonia. A salinity budget analysis (not shown), using built-in ROMS budget equations, confirms that the upstream advection term is responsible for the erosion of salinity, and that the maximum erosion occurs at the deepest level of the water column above the steepest slopes. By testing different advection and diffusion operators available in ROMS, it quickly became evident that the diffusive part of the advection term was responsible for the model error (the maximum of salinity advection over steep slopes disappears when non-diffusive centered schemes are used). In the following we will present a new scheme, test its performances, and verify that it solves our problem while retaining good numerical and physical properties.

3. A modified advection scheme to reduce diapycnal mixing

3.1. Diffusive advection schemes: short review

Let us write the one-dimensional tracer advection equation for a tracer T and advection velocity U:

$$\frac{\partial T}{\partial t} = -U\frac{\partial T}{\partial x} \tag{1}$$

This equation should be non-dispersive, meaning that the phase speed ω/k and group speed $\delta\omega/\delta k$ of any advected signal are equal (and equal to U). On the other hand, the discretized equation using second-order centered scheme is dispersive. To remedy the dispersion problem, we can introduce a numerical closure, following Bryan (1975) who solve an advection–diffusion equation:

$$\frac{\partial T}{\partial t} = -U\frac{\partial T}{\partial x} + \frac{\partial}{\partial x}\left[A\frac{\partial T}{\partial x}\right] \tag{2}$$

A is a harmonic diffusion coefficient. In this case, the condition of stability of the dispersive mode is dependent on the Peclet number (Bryan, 1975), which is the ratio of advection to diffusion:

$$\mathrm{Pe} = \frac{|U|\Delta x}{A} \tag{3}$$

Δx is the mesh size. The monotonicity condition, preventing computational modes and unphysical extrema, requires that Pe be lower than 2. If Pe = 2, then $A = 1/2|U|\Delta x$ and it can be shown that the discretized advection–diffusion equation is equivalent to a first-order upstream discretization of the advection equation. If monotonicity is a very valuable property, the first-order scheme is also known to be very diffusive, which results in unphysical degradation of the signal being transported.

A few solutions have been proposed to improve the advection problem. The earliest solution (Semtner and Mintz, 1977) was to use hyperdiffusion in the advection–diffusion equation to achieve a better scale selection:

$$\frac{\partial T}{\partial t} = -U\frac{\partial T}{\partial x} + \frac{\partial}{\partial x}\left[B\frac{\partial^3 T}{\partial x^3}\right] \tag{4}$$

However, the usual practice of choosing an arbitrary value for B is not considered satisfactory. The more recent approach has been to develop diffusive high-order advection schemes. Those can be classified into oscillatory and non-oscillatory schemes. QUICK (Quadratic Upstream Interpolation for Convective Kinematics) and QUICK-type third-order upstream schemes are the most common oscillatory schemes. They are not monotonic, unlike linear first-order advection schemes, according to the Godunov theorem (Godunov, 1959). The only way to circumvent this theorem is to devise a nonlinear discretization (in T). Monotonic, nonlinear schemes are built using anti-diffusion with non-linear limiters to prevent oscillations. Among the most widely used schemes of this type in ocean modeling, are FCT and TVD series, and more recently multidimensional monotone schemes such as MDPPM and the Prather family (Griffies, 2007).

In numerous idealized studies of advection schemes, it appears that monotonic schemes are often as diffusive as third-order upwind schemes. According to Griffies (2007; see also Griffies et al., 2000), only the most recent monotonic schemes are able to reduce diffusion sufficiently in geopotential models to bring spurious diapycnal mixing to reasonable levels. However, this is achieved at such cost that only intermediate solutions are currently affordable. We can easily foresee that the problem should be far worse in sigma models where the angle between isopycnal surfaces and sigma surfaces can be much greater. Nevertheless, it is generally believed that the problem of spurious mixing is only relevant to coarse-resolution models for climate studies and that the use of high resolution in regional models removes this complication. We will see in the following that this is not the case, at least not until a very high resolution is used. We will concentrate on the third-order upwind scheme for two reasons. First, it allows a scaling analysis to predict the evolution of spurious mixing with resolution. Second, it allows a straightforward correction to our problem.

3.2. Upstream third-order scheme UP3 and its split version SUP3

Let us write the spatial discretization of the one-dimensional advection equation in flux form:

$$\frac{\partial T_i}{\partial t} = -\frac{1}{\Delta x}\left[U_{i+\frac{1}{2}}T_{i+\frac{1}{2}} - U_{i-\frac{1}{2}}T_{i-\frac{1}{2}}\right] \tag{5}$$

ROMS tracer advection scheme (noted UP3 hereafter) is presented in Shchepetkin and McWilliams (2005). As in the QUICK scheme proposed by Leonard (1979), it is obtained by using a three-point, upstream-biased, quadratic (parabolic) interpolation of T to the interfaces of the tracer cells (mid-point):

$$T_{i+\frac{1}{2}} = \frac{1}{2}(T_{i+1} + T_i) - \begin{cases} \frac{1}{6}(T_{i+1} - 2T_i + T_{i-1}) \text{ if } U_{i+\frac{1}{2}} > 0 \\ \frac{1}{6}(T_{i+2} - 2T_{i+1} + T_i) \text{ if } U_{i+\frac{1}{2}} < 0 \end{cases} \tag{6}$$

In the UP3 scheme, the coefficient 1/6 was chosen instead of 1/8 in QUICK, as it can be shown (Webb et al., 1998) that only the former coefficient completely eliminates the error term of $O(\Delta x^2)$, which is the dispersive component associated with the third derivatives of tracer T (in the Taylor series expansion of the advection scheme).

One of the most striking features of UP3 is its ability to be broken into its purely advective and diffusive parts. This feature was used almost simultaneously by Holland et al. (1998) and Webb et al. (1998), respectively for NCOM and OCCAM models. Their motivation was to solve a stability problem related to the Leapfrog time-stepping algorithm (the diffusive part of QUICK was to be evaluated at the old time step). In ROMS, the time-stepping algorithm is similar to the Leapfrog–Adams–Moulton III predictor–corrector (LF-AM3; Shchepetkin and McWilliams, 2005), which is stable for both advection and diffusion written at current time step. Therefore, the splitting of UP3 is also useful in our case but

for different reasons. The Split-UP3 scheme can be obtained by rewriting the interpolation of mid-point T:

$$T_{i+1/2} = \frac{1}{12}(-T_{i+2} + 7T_{i+1} + 7T_i - T_{i-1}) + \frac{1}{12}\frac{|U_{i+1/2}|}{U_{i+1/2}} \times (-T_{i+2} + 3T_{i+1} - 3T_i + T_{i-1}) \quad (7)$$

The first term is a local cubic polynomial interpolation of T, which results in a centered advection scheme of fourth-order spatial accuracy. Introducing the split interpolation into the advection scheme gives an equation of the form:

$$\frac{\partial T}{\partial t} = -\frac{\partial}{\partial x}[U\widehat{T}] - \frac{\partial}{\partial x}\left[B\frac{\partial^3 T}{\partial x^3}\right] \quad (8)$$

$\widehat{T}$ is the cubic polynomial interpolation of T to the respective interfaces of tracer cells. The first term is a fourth-order centered advection scheme; the second term is a third-order diffusion term which behaves like hyperdiffusion with a coefficient proportional to the absolute value of local velocity:

$$B = \frac{1}{12}|U|\Delta x^3 \quad (9)$$

In this form, B will be called Peclet hyperdiffusivity, since it is the coefficient required to satisfy an equivalent Peclet constraint for the third-order upstream scheme (i.e. eliminating the second-order dispersive error term). It is important to note that the exact equivalence of UP3 and SUP3 is only valid in the one-dimensional advection problem. A similar equivalence in the two-dimensional case is not available and we expect additional diffusion in the cross directions to occur in SUP3, as a result of the implementation of the biharmonic operator by applying the Laplacian operator twice. This point and the approximation made will be tested in Section 4.1.

3.3. Scaling of diapycnal mixing with grid spacing

One of the greatest advantages of splitting UP3 is to allow us to perform a scaling analysis. This is inappropriate with monotonic nonlinear schemes, where direct estimation of diapycnal mixing is therefore required (Griffies, 2007). Here, we extend the scaling analysis of Lee et al. (2002) to the case of sigma-coordinate models. In this case, the resolution dependence of the topographic slope is known since it is imposed by selective smoothing to safely limit pressure gradient errors. This particularity of sigma models results in different regimes of resolution dependence as will be shown in this section.

Diapycnal harmonic diffusivity A_D is generally estimated because it can be compared with observations (no observational estimates are available for the biharmonic diffusivity B_D which is less physically meaningful). Following Redi (1982), A_D can be derived from an equivalent harmonic diffusivity A:

$$A_D = A.S^2 \quad (10)$$

In sigma models, S is the slope between sigma surfaces (rather than geopotential) and isopycnal surfaces. However, S can be an order of magnitude larger than the slope between geopotential and isopycnal surfaces; therefore, in the following we will neglect the latter and consider vertical rather than diapycnal diffusion for simplicity. In sigma-coordinate, the maximum slope S^{max} is at the bottom level which follows the bottom topography. Then, mixing above the bottom varies with the coordinate position σ (which is -1 at the bottom and 0 at the surface):

$$A_D = A[\sigma S^{max}]^2 = \sigma^2 A[S^{max}]^2 = \sigma^2 A_D^{max} \quad (11)$$

The largest mixing value A_D^{max} within the water column is at the bottom and varies with the squared bottom slope S^{max}. In the following, we consider the largest bottom values and drop the "max" index for clarity. Writing S as a discretized form of the bottom slope and introducing r_0 the maximum r-value parameter allowed for acceptable levels of pressure-gradient error, we obtain two expressions of S:

$$S = \frac{\Delta h}{\Delta x} = r\frac{2h}{\Delta x} = \begin{cases} r_0 \frac{2h}{\Delta x} \ if\ r > r_0 \text{ coarse resolution} \\ s_{raw} \ if\ r < r_0 \text{ High resolution} \end{cases} \quad (12)$$

These relations express coarse and high resolution regimes, as topography smoothing is required only in the former. S_{raw} is the slope of "raw" bottom topography after interpolation to the model grid. The following equation uses the scaling proposed by Lee et al. (2002) to estimate the relation between harmonic and biharmonic diffusivity:

$$A = \frac{B}{L^2} = \frac{U\Delta x^3}{12L^2} \quad (13)$$

U is the velocity scale in the direction of the flow; L is a mixing length scale and can be taken as the Rossby radius L_0 if we consider that tracer variance is primarily associated with a stream of eddies in geostrophic turbulence (Lee et al., 2002). However, if the model grid cell has a greater size than L_0(turbulence is not fully resolved), we can use $L = \Delta x$ and consider that tracer variance occurs primarily at the grid cell level.

Taking the relations obtained for A and S, we can find a scale for diapycnal diffusivity:

$$A_D = \begin{cases} \frac{r^2 h^2 U}{3\Delta x} & \text{if } \Delta x > L_0 & \text{Coarse resolution} \\ \frac{r^2 h^2 U \Delta x}{3L_0^2} & \text{if } r > r_0 \text{ and } \Delta x < L_0 & \text{Medium resolution} \\ \frac{U\Delta x^3 S_r^2 aw}{12L_0^2} & \text{if } r < r_0 & \text{High resolution} \end{cases} \quad (14)$$

It appears that the spurious mixing scale varies differently in three resolution regimes, in relation with mesh size, bottom depth and slope, flow speed, and lateral mixing (definition of coarse-resolution: mesh size larger than the mixing scale; medium resolution: mesh size is smaller, but bottom smoothing is still required; and high resolution: bottom smoothing is not required as mesh size is smaller than 1–2 km).

To appreciate the relation of spurious mixing with spatial resolution, we have plotted in Fig. 3a, the values of A_D computed with a choice of averaged parameter values S_{raw} = 10%, r_0 = 0.2, U = 10 cm/s and L_0 = 20 km. Curves corresponding to different values of bottom depth h are represented. The most striking result is that the diffusion error increases with refined resolution in the coarse-resolution regime. At medium resolution the tendency reverses but the error only decreases significantly at high resolution (with a power of 3). Then, a resolution finer than 1 km is required to reduce spurious diffusivities to less than 1 cm^2/s, which may be considered acceptable in a regional model (as water properties are partially controlled by the open boundary conditions), but remains higher than measured values in the ocean interior (0.1 cm^2/s).

Fig. 3b shows an attempt at validating this scaling of resolution-dependent diffusivity by looking at the perturbation of the salinity minimum. If we assume the evolution of salinity minimum over a steep slope to be dominated by diapycnal diffusion, salinity anomalies are expected to evolve like A_D. Therefore, in the absence of a direct estimation of diapycnal mixing (as in Griffies et al., 2000), we use salinity anomalies at 1000 m as proxy for spurious mixing. Our model resolution sensitivity is based only on resolutions of 1/12, 1/6 1/4 and 1/2 degrees and miss out the high resolution regime, but the results presented in Fig. 3b confirm the increase of spurious diapycnal mixing with resolution at coarse-resolution. It also suggests a change of regime at medium resolution. A finer resolution sensitivity experiment (with direct estimation of diapycnal

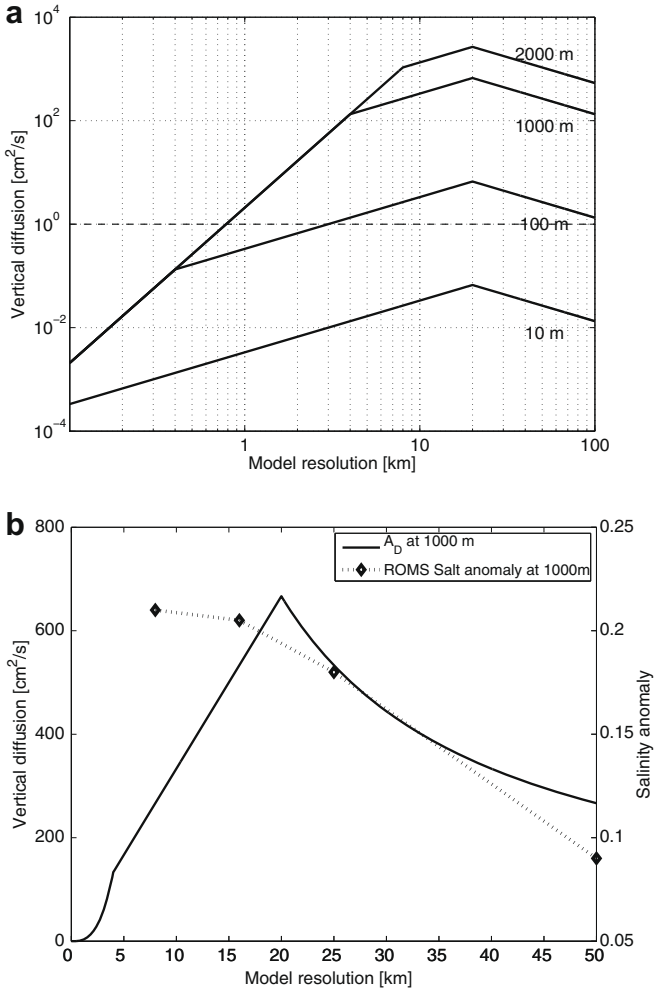

Fig. 3. Vertical diffusion coefficient A_D [cm^2/s] plotted as a function of horizontal resolution (see text). (a) Scaling of A_D for various depth values, where three regimes of evolution are predicted (coarse, medium and fine). (b) Comparison of resolution dependence of A_D scaling at 1000 m with ROMS salinity anomaly (using UP3).

mixing) is still needed to complete the validation task, but we can nevertheless safely contest the general admitted idea that spurious mixing can be reduced by refining the model resolution, unless resolution is greatly refined.

It also appears that A_D increases with depth at coarse and medium resolution (it is depth independent at high resolution). The increase of spurious mixing with depth is consistent with our model solution showing larger errors at depth where the AAIW salinity minimum lies. However, the error is present at all depths and very clearly affects even the barotropic circulation of the southwest Pacific (see Section 4).

3.4. Rotated diffusion operator: RSUP3

The second advantage provided by the splitting of UP3 is that it gives us access to lateral mixing and therefore an opportunity to correct the diapycnal part of the mixing. Here, we face the original mixing problem, discussed by Mellor et al. (1998) and Barnier et al. (1998) who have provided two different solutions. Following Mellor et al. (1998), we can apply diffusion only to a tracer perturbation with respect to a reference frame (climatology). This tends to reduce mixing that takes place across reference surfaces, crudely approximating isopycnal mixing behavior. It may also keep one advantage of isosigma diffusion (Mellor and Blumberg, 1985), by preventing lateral mixing from interfering with subgrid-scale vertical mixing in the bottom boundary layer (diffusivity should approach zero where the logarithmic law of the wall prevails). We have implemented this scheme and will show in Section 4 that it considerably reduces spurious mixing, but it also appears to produce unstable behavior of uncertain origin, showing vertical oscillations of the tracer profile and grid-scale noise at the bottom in the vertical velocity field. For this reason, our preference goes to the solution proposed by Barnier et al. (1998), which in essence uses the same technique as Cox (1987) who numerically implemented Redi's (1982) coordinate rotation. The principle is a straightforward rotation of the diffusion tensor to align the diffusive fluxes along the geopotential surfaces. Note that geopotential diffusion is justified in regional applications for two reasons: eddy-driven mixing is generally resolved explicitly in these applications (although Roberts and Marshall, 1998; argue that adiabatic subgrid diffusion may be required even in eddy-resolving models); and water characteristics are largely controlled by the open boundary conditions on decadal time-scales. However, the use of larger domains, coarser resolution and longer time integration will require addressing the issue of isopycnal diffusion.

The geopotential biharmonic operator is computed by applying the rotated Laplacian operator twice. In the three-dimensional Cartesian coordinates, we obtain a mathematical expression for the lateral diffusion term of the tracer equation:

$$D = -\left[\frac{\partial}{\partial x}(B_x F_x) + \frac{\partial}{\partial y}(B_y F_y) + \frac{\partial}{\partial \sigma}(B_\sigma F_\sigma)\right]$$

$$\text{with} \begin{cases} F_x = \frac{\partial Y}{\partial x}\big|_\sigma - \frac{S_x}{\Delta z}\frac{\partial Y}{\partial \sigma} \\ F_x = \frac{\partial Y}{\partial y}\big|_\sigma - \frac{S_y}{\Delta z}\frac{\partial Y}{\partial \sigma} \\ F_\sigma = -\frac{S_x}{\Delta z}F_x - \frac{S_y}{\Delta z}F_y \end{cases}$$

$$\text{and } Y = \frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y} + \frac{\partial f_\sigma}{\partial \sigma}$$

$$\text{with} \begin{cases} f_x = \frac{\partial T}{\partial x}\big|_\sigma - \frac{S_x}{\Delta z}\frac{\partial T}{\partial \sigma} & S_x = \frac{\partial z}{\partial x}\big|_\sigma \\ f_y = \frac{\partial T}{\partial y}\big|_\sigma - \frac{S_y}{\Delta z}\frac{\partial T}{\partial \sigma} & S_y = \frac{\partial z}{\partial y}\big|_\sigma \\ f_\sigma = -\frac{S_x}{\Delta z}f_x - \frac{S_y}{\Delta z}f_y \end{cases} \tag{15}$$

(B_x, B_y, B_σ) are the Peclet hyperdiffusion coefficients, and (S_x, S_y) are the isosigma slopes. The numerical implementation of this diffusion operator on the C grid requires some averaging of the hyperdiffusion coefficients and isosigma slopes:

$$\begin{aligned} D &= -[B_x|_{i+\frac{1}{2},j,k}F_x|_{i+\frac{1}{2},j,k} - B_x|_{i-\frac{1}{2},j,k}F_x|_{i-\frac{1}{2},j,k}] \\ &= -[B_y|_{i,j+\frac{1}{2},k}F_y|_{i,j+\frac{1}{2},k} - B_y|_{i,j-\frac{1}{2},k}F_y|_{i,j-\frac{1}{2},k}] \\ &\quad - \left\{ \begin{array}{l} \overline{\overline{B}}^{x,\sigma}|_{i,j,k+\frac{1}{2}}\frac{\overline{\overline{S}}^{x,\sigma}|_{i,j,k+\frac{1}{2}}}{\Delta z}\overline{\overline{F}}^{x,\sigma}|_{i,j,k+\frac{1}{2}} - \overline{\overline{B}}^{x,\sigma}|_{i,j,k-\frac{1}{2}}\frac{\overline{\overline{S}}^{x,\sigma}|_{i,j,k-\frac{1}{2}}}{\Delta z}\overline{\overline{F}}^{x,\sigma}|_{i,j,k-\frac{1}{2}} + \\ \overline{\overline{B}}^{y,\sigma}|_{i,j,k+\frac{1}{2}}\frac{\overline{\overline{S}}^{y,\sigma}|_{i,j,k+\frac{1}{2}}}{\Delta z}\overline{\overline{F}}^{y,\sigma}|_{i,j,k+\frac{1}{2}} - \overline{\overline{B}}^{y,\sigma}|_{i,j,k-\frac{1}{2}}\frac{\overline{\overline{S}}^{y,\sigma}|_{i,j,k-\frac{1}{2}}}{\Delta z}\overline{\overline{F}}^{y,\sigma}|_{i,j,k-\frac{1}{2}} \end{array} \right\} \end{aligned} \tag{16}$$

with:

$$\begin{cases} B_x|_{i+\frac{1}{2},j,k} = \frac{1}{12}|U|_{i+\frac{1}{2},j,k}\left(\overline{\Delta}^x|_{i+\frac{1}{2},j,k}\right)^3 \\ B_y|_{i,j+\frac{1}{2},k} = \frac{1}{12}|V|_{i,j+\frac{1}{2},k}\left(\overline{\Delta}^x|_{i,j+\frac{1}{2},k}\right)^3 \end{cases} \tag{17}$$

With this rotated diffusion operator, we name the split advection scheme RSUP3. We have noted in Section 3.2 that the exact equivalence between UP3 and SUP3 is lost in the two-dimensional problem. Here, with the rotated diffusion operator, the sources of numerical discrepancy between the two schemes are more numerous, due to additional averaging and differentiation. It is useful here to recall other known problems associated with rotated diffusion operators.

There are various kinds of discretization errors, which have been discussed within the isopycnal diffusion framework (Griffies et al., 1998; Roberts and Marshall, 1998; and Beckers et al., 2000). The first kind is associated with computational modes that are unaltered (or even amplified) by the rotated diffusion if simultaneous averaging and differentiation operators are applied in the same spatial direction (Griffies et al., 1998). In ROMS, first- and second-order diffusion operators are designed to avoid this error. In particular, in both f_x and F_x the six-point stencil used in Cox (1987) to compute the vertical derivatives (source of the computational modes) is reduced to a four-point stencil designed to avoid combinations of averaging and differentiation. However, there is a second kind of discretization errors argued by Beckers et al. (2000). It arises from the fact that no linear discretization of the rotated diffusion operator (first- or second-order) is monotonic. This means that new local extrema may be created by the numerical version of the diffusion operator. This problem may be accentuated by large truncation errors. As for the pressure gradient problem, the rotated diffusion term is the sum of two relatively large terms (corresponding to horizontal and vertical fluxes) which compensate each other; a small error in either could result in a relatively large error in the total diffusion. For this reason, vertical resolution should be sufficiently fine when using rotated diffusion operators in sigma models. In our case, a minimum of 30–40 sigma levels (with surface refinement) is needed to maintain numerical integrity.

In spite of potential errors with rotated operators, the validations presented in Section 4 will show that the essential qualities of low dispersion and diffusion of the UP3 scheme are retained. This is due in part to the variable hyperdiffusion coefficient in RSUP3, which seems to reduce discretization errors to acceptable levels.

3.5. Treatment of stability limits

Our main limitation in proposing a rotated isopycnal (rather than geopotential) diffusion operator comes from the fact that the linear stability of the diffusion equation is strongly constrained by large vertical fluxes along steep isopycnal slopes (especially near the surface where vertical resolution is increased). With Laplacian diffusion, the largest constraint applies to the diagonal–vertical piece of the tensor (Griffies et al., 1998) which can be solved implicitly, together with the vertical diffusion term. However with second-order biharmonic diffusion operator, implicit treatment of vertical fluxes is not straightforward. An semi-implicit solution to this problem will be sought in a following paper, where the present hyperdiffusion scheme, extended to isopycnal diffusion, may also prove useful to the climate modeling community.

If RSUP3 is to be solved explicitly, stability criteria need to be considered depending on the slope of surfaces along which the tensor is rotated. Isopycnal slopes are too limiting near the surface, where the vertical mesh size is reduced to properly resolve the surface mixing layer, and cannot be used without serious slope clipping. We have seen that geopotential diffusion may be acceptable in regional models. In this case, the problem is more manageable, and isosigma slopes impose a lower restriction to the size of the time-step. The restriction is no longer at the surface, where the slope approaches zero, but near the bottom in deep water, where the slope is maximum. If the diffusion term is not solved implicitly, we are left with two propositions: improving the temporal scheme to obtain less restrictive criteria; and tapering diffusivity by using a clipping method, similar to that of Cox (1987) but adapted to the sigma-coordinates.

3.5.1. Extension of linear stability

In ROMS, the baroclinic time-stepping algorithm LF-AM3 (predictor–corrector) allows tracer diffusion to be applied either at the current time of the corrector step $n + 1/2$ (standard) or at the current time of the predictor step n. We have performed a Fourier stability analysis (see for example Durran, 1999) which indicates that a 50% gain in the stability range is obtained by applying the biharmonic operator at time $n + 1/2$ rather than n, but another 30% is given by interpolating between these two time levels, i.e. the tracer value used in the diffusion term at time n is:

$$\overline{T}_n = \frac{1}{4}T_n + \frac{3}{4}T_{n+\frac{1}{2}} \tag{18}$$

To obtain this result, we assume that the main stability restriction is set by hyperdiffusion in the vertical direction (of coefficient μ), and then focus on extending the stability limit of an explicit implementation for a pure biharmonic diffusion term. For an explicit one time level scheme:

$$\frac{T_{n+1} - T_n}{\Delta t} = -\mu\Delta^2 T_n, \tag{19}$$

the stability condition obtained by a Fourier analysis is given by:

$$\mu\frac{\Delta t}{\Delta z^4} \leqslant \frac{1}{8} \tag{20}$$

Δz is the variable vertical mesh size. In the two time level approach of ROMS, diffusion is applied during the corrector step. In this case, the scheme:

$$\frac{T_{n+1} - T_n}{\Delta t} = -\mu\Delta^2\left[\frac{3}{4}T_n + \frac{1}{4}T_{n-1}\right] \tag{21}$$

leads to the less restrictive stability condition:

$$\mu\frac{\Delta t}{\Delta z^4} \leqslant \frac{1}{4} \tag{22}$$

With this modification in the biharmonic operator, the permissible time step is increased, although still limiting.

3.5.2. Clipping method

Since we do not want the model time step to be limited by the diffusion term, we apply a clipping procedure to the diffusivities. To be practical, this procedure should not be left as a tunable parameter; instead we can obtain a relation for the maximum hyperdiffusivity which is authorized at a given time step. Starting from the numerical stability condition (22) given for an equivalent vertical hyperdiffusivity $B_D = BS^4$, and writing the isosigma slope as:

$$S = \frac{\Delta z|_\sigma}{\Delta x} \tag{23}$$

where $\Delta z|_\sigma$ is the variation of z along a tracer cell following a sigma surface, we obtain a limit for B:

$$B \leqslant B_{\max} = \frac{\Delta x^4}{4\Delta t}\left[\frac{\Delta z}{\Delta z|_\sigma}\right]^4 \tag{24}$$

If we assume that clipping is most needed at the CFL limit, we can use $U = \Delta x/\Delta t$, and write:

$$B_{\max} = C_{\text{clip}}\left[\frac{\Delta z}{\Delta z|_\sigma}\right]^4 * B \tag{25}$$

B is here the unrestricted coefficient given by RSUP3. In this scaling analysis C_{clip} has a value of 3; however, with regards to the various simplifying hypotheses that we used, this coefficient should be reduced for safety: we propose a value of 1. Our clipping formulation says that the maximum hyperdiffusion coefficient allowed is a fraction of the unrestricted coefficient, which depends on vertical mesh-size relative to the vertical extent of the horizontal mesh. The restriction is close to zero at the surface and maximum above steep slopes in deep water, but somewhat compensated by coarser resolution at depth. In our configuration, the minimum fraction of B

(maximum restriction) is above the western barrier reef of New Caledonia. The spatial localization of the limiting factor is coincidental with the location of maximum spurious diapycnal mixing. Therefore, our clipping method may help to further reduce the amount of spurious mixing in deep water, without major penalty in terms of numerical dispersion. More importantly, clipping is avoided at the surface where the strongest currents are present.

4. Validation of the new tracer advection scheme

In practice, the implementation of RSUP3 in ROMS is relatively straightforward because the main pieces are already in place: the fourth-order centered scheme (the purely advective part of RSUP3) is implemented with the proper weighting coefficients (Shchepetkin and McWilliams, 2005); a geopotential biharmonic operator is also implemented (and applied in Haidvogel et al., 2000; for example) in the standard version, but with constant hyperdiffusion coefficients. To obtain RSUP3, we combined these two pieces and computed the hyperdiffusion coefficients with the averaging procedure described in the above equation. We then implemented modifications to the temporal scheme for extended stability criteria and our specially designed clipping method (as described in Section 3.5).

We show in this section that the new tracer advection scheme RSUP3 provides us with an acceptable solution to our problem. Fig. 4 (bottom panel) presents a proper representation of the salinity field at 1000 m depth, to be compared with the UP3 solution (Fig. 1). The correction of water masses has greatly affected the dynamical solution in the whole water column (not shown), which confirms that the problem identified in our study is relevant to all regional modeling studies. In the following, we will proceed to demonstrate the performance of the RSUP3 scheme relevant to regional modeling, after first evaluating the splitting of UP3 in a multidimensional context.

4.1. Equivalence of UP3 and SUP3

The first step in validating our new advection scheme is to evaluate the quality of the splitting technique alone. In the one-dimensional problem, SUP3 is an exact equivalent of UP3 (this can easily be verified mathematically and numerically). In the two-dimensional problem, an exact mathematical expression of SUP3 is not available, so that an extension of the one-dimensional scheme is used (as in previous work by Holland et al., 1998; Webb et al., 1998). We present here a comparison of UP3 and SUP3 solutions on the southwest Pacific domain after two years of integration of ROMS.

Fig. 5 presents salinity fields at 1000 m depth for a sub-region around New Caledonia, where the maximum diffusion error is found (associated with steep reef slopes). The top panels show UP3 and SUP3 solutions compared with ORCA05 climatology, which is used to initialize and force the regional model. Both schemes produce large erroneous salinity anomalies which develop from the steepest slope areas, i.e. essentially the western side of New Caledonia. The pattern of spuriously diffused salinity is simi-

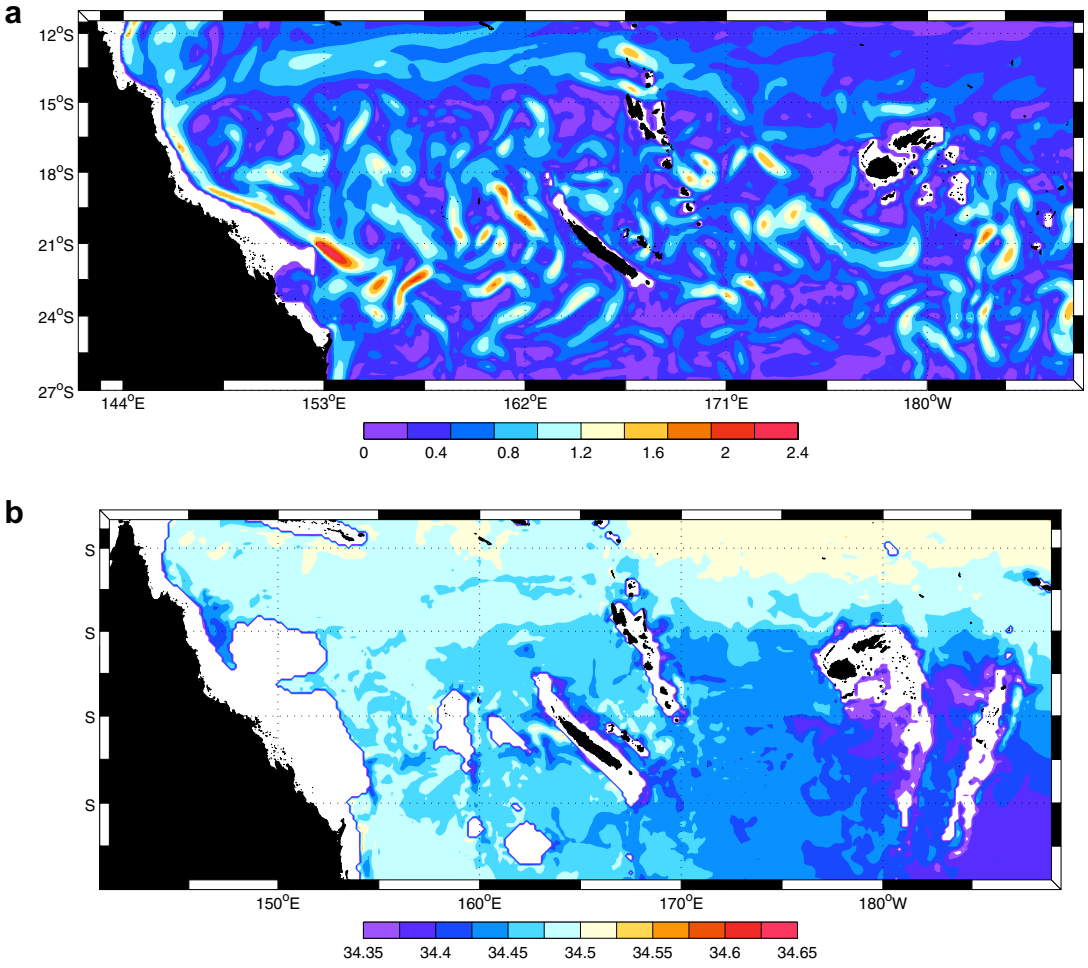

Fig. 4. ROMS simulation using the new advection scheme RSUP3. (a) Peclet hyperdiffusivity at the surface [10^{11} m^4/s]. (b) Salinity at 1000 m.

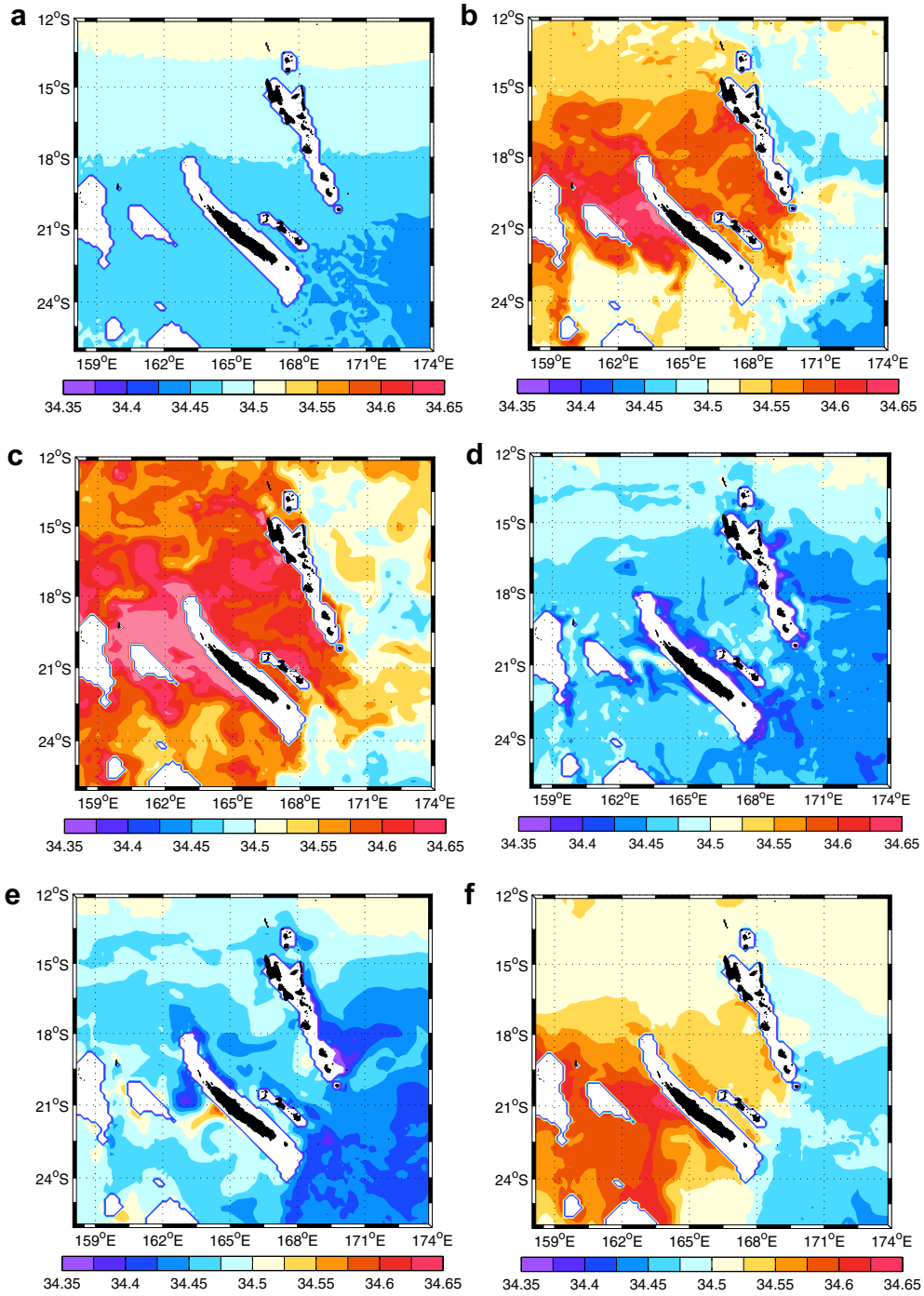

Fig. 5. ROMS Salinity at 1000 m for various choices of advection–diffusion schemes zoomed in around New Caledonia after 2 years of integration. (a) Initial state from ORCA05 climatology. (b) UP3 scheme. (c) SUP3 scheme. (d) RSUP3 scheme. (e) RSUP3 modified with constant hyperdiffusivity 6×10^{10} m^4/s. (f) RSUP3 modified with Laplacian diffusion and constant diffusivity 400 m^2/s. UP3 is ROMS standard Upstream third-order scheme (Shchepetkin and McWilliams, 2005); SUP3 is split UP3 between advection and diffusion terms along isosigma surfaces; RSUP3 is split UP3 with diffusion operator rotated along geopotential surfaces.

lar in both cases, but the anomalies are larger in the SUP3 case. Thus, the SUP3 scheme appears to overestimate the diffusion compared with the implicit amount provided by the upstream-biased advection scheme. The excess vertical diffusion appears also in comparisons of averaged vertical salinity profiles (Fig. 7), as well as averaged salinity anomalies at 1000 m (Fig. 8). The latter calculation reveals anomalies of about 15% higher in SUP3 compared with UP3. This excess diffusion is probably related to additional diffusion cross terms in SUP3, due to the biharmonic operator implementation as twice applied Laplacian operator.

Another way of comparing the amounts of diffusion provided consists in observing its effect on the dynamical spectrum. Fig. 9a shows the energy spectrum for the entire domain and second year of simulation. It shows a very comparable energy spectrum for UP3 and SUP3 versions. Therefore the difference in diffusion is not large enough to significantly affect the dynamics at any scale. SUP3 presents the same ability as UP3 to satisfy the Peclet constraint without excessive diffusion. This suggests that the difference between UP3 and SUP3 salinity profiles is probably associated with small differences in lateral diffusion fluxes, but

they do translate into significant spurious diapycnal diffusion. The latter will now be corrected.

4.2. Performance of RSUP3

Three performance criteria are proposed to validate the desirable abilities of our new scheme: (1) maintaining numerical integrity associated with the Peclet constraint (2) preserving the large-scale water masses entering through the lateral boundaries; (3) preserving the energy spectrum given by the original UP3 scheme.

4.2.1. Maintaining numerical integrity

The first criterion for validating the new scheme resides in its ability to prevent computational modes from generating unphysical extrema when the Peclet constraint is violated. Fig. 4 presents the surface Peclet hyperdiffusion coefficient B in the model solution with RSUP3. The hyperdiffusivity shows large spatial variations which span two orders of magnitude (between 1×10^9 and 2×10^{11} m^4/s). Highest values occur where the flow is most intense in the western boundary current region but strong eddies are also shed all over the domain from unstable zonal jets (Couvelard et al., 2008). In each region of the domain, the diffusivity is adequate to maintain numerical integrity, satisfying the Peclet constraint, since the solution is smooth at the grid scale.

Fig. 5e,f present additional experiments to estimate the solution sensitivity to the coefficient magnitude and order of the diffusion operator. In the first experiment, a constant hyperdiffusion coefficient is used instead of the Peclet relation. Numerical stability requires a constant coefficient capable of controlling the largest velocities at any place and time (Griffies and Hallberg, 2000). We found the value of $B = 6 \times 10^{10}$ m^4 s^{-1} to be satisfactory. Lower values produced unphysical extrema at locations where the currents are strongest (East Australian Current and its eddy field) as predicted by the Peclet constraint. On the other hand, higher values are not permitted by surface stability limits of the temporal scheme.

4.2.2. Preserving water masses

This test requires that large-scale water masses entering through the lateral boundaries are not modified unrealistically within a timescale that is shorter than the flushing timescale of the regional domain (a few years). It appears sufficient to concentrate on the AAIW around 1000 m depth where water characteristics are the most affected. It should be noted that a complete preservation of the large-scale salinity profile may not be desirable, as dynamical adjustments of the mass field also occur in the spin-up process. But these changes should remain small, within acceptable values compared with observations (salinity between 34.35 and 34.55). Fig. 5c and d show a comparison of salinity at 1000 m for SUP3 and RSUP3 solutions respectively. Clearly, the rotated operator has almost completely corrected the previous diffusion error, as salinity remains in the range of ORCA05 climatology (Fig. 5a). Temperature–salinity diagrams (Fig. 6) and mean vertical salinity profiles (Fig. 7b) show that the correction is observable over the whole southwest Pacific region and water column, while some residual salinity anomaly persists. Fig. 8 provides a quantification of the 1000-m residual anomaly, which is about 20% of the error made with SUP3. As said before this average anomaly is not necessarily erroneous but possibly created by dynamical adjustments, as the salinity values remain in the normal range.

If the large-scale salinity signal appears satisfactory, there is persistence of local salinity extrema near the Caledonian slope. They are too localized to impact the large-scale water-masses and may also be associated with small-scale dynamical adjustments, but our sensitivity study on coefficient and order of diffusion reveals the presence of discretization errors from the rotated diffusion operator. In Fig. 5e, it appears that a constant hyperdiffusion coefficient produces a generally smoothed tracer field and preserves the large-scale AAIW signal (its area average anomaly is only slightly higher than in RSUP3; Fig. 8). However, the salinity extrema located near the Caledonian western reef is more clearly defined compared with the RSUP3 solution. Fig. 7d presents an averaged vertical profile of the Peclet hyperdiffusivity B in the New Caledonia region, which shows that the decrease of B with depth is exponential. When choosing a constant diffusivity value for B, numerical integrity must be satisfied at the surface in energetic regions. Therefore, this constant value tends to largely overestimate diffusion at depth with two results: one is that the tracer field is smoother than necessary; the other is that discretization errors are emphasized. However, this issue becomes truly relevant with first-order diffusion operators. Using rotated Laplacian diffusion with constant coefficient $A = 400$ m^2 s^{-1} (Fig. 5f), the salinity error becomes such that the large-scale AAIW signal is affected (the average anomaly is 0.13 in Fig. 8). We conclude that discretization errors are not amplified by the use of second-order versus first-order operators despite complexity of the former. On the contrary, acceptable residual errors are achieved only when using biharmonic operators, especially with dynamically-dependent coefficients.

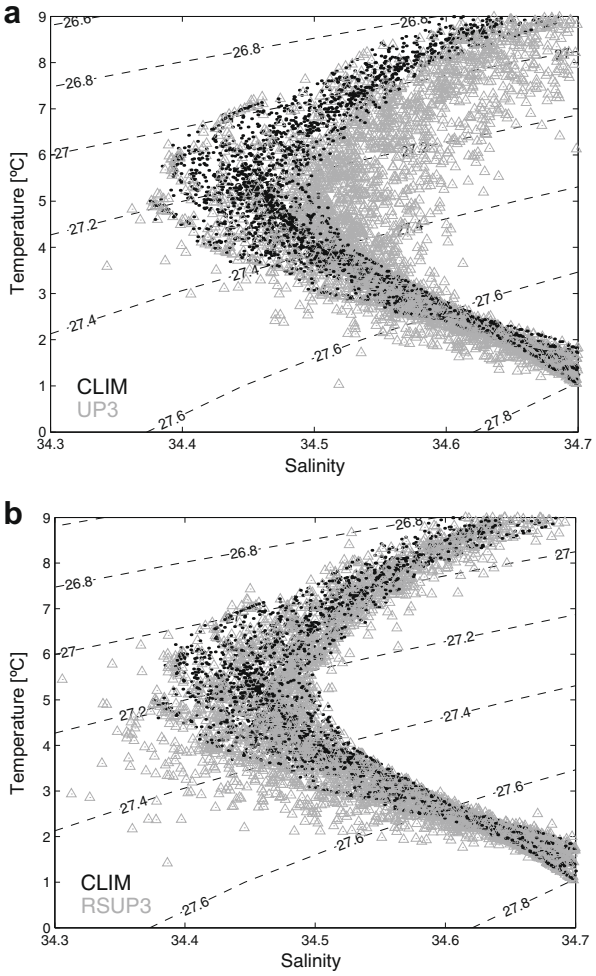

Fig. 6. Temperature–salinity diagrams for the southwest Pacific region centered around AAIW characteristics. (a) Comparison of ROMS solutions using UP3 with ORCA05 January climatology. (b) Same but with ROMS solution using RSUP3.

4.2.3. Preserving the energy spectrum

The last criterion for a suitable UP3 alternative is to allow realistic dynamics at all scales. This is achieved if diffusion is small enough that isopycnal slopes are preserved, i.e. available potential

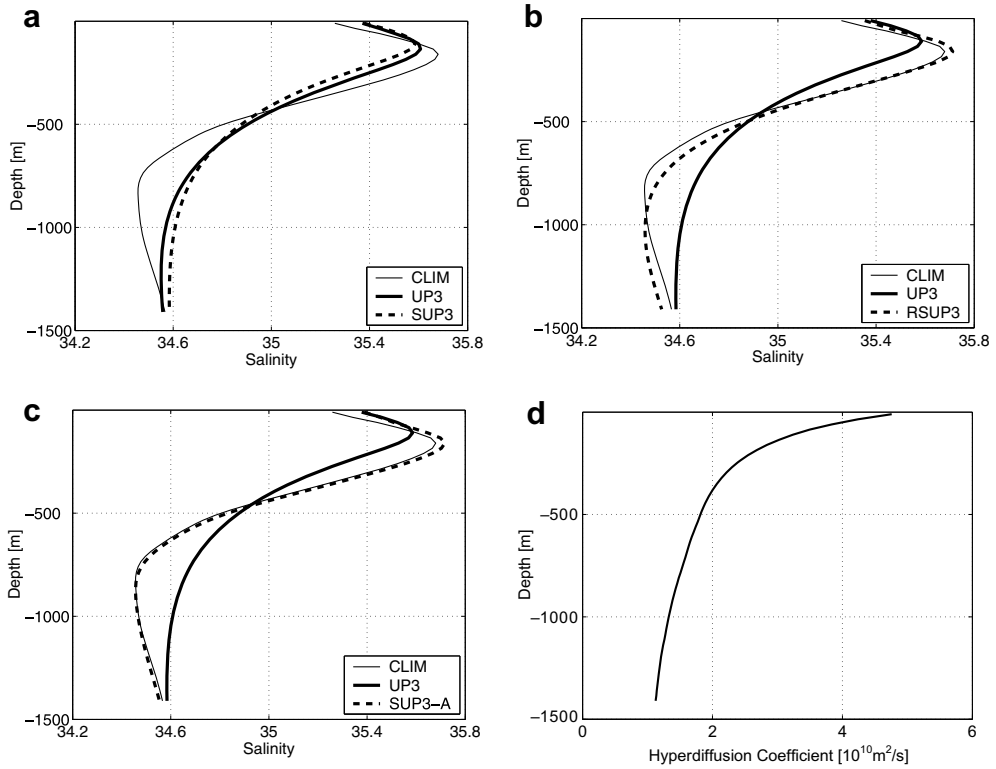

Fig. 7. Vertical profiles of salinity and hyperdiffusivity averaged over the Caledonia region (spanning longitudes158–174°E and latitudes 26–12°S, as in the maps of Fig. 5) after 2 years of integration. (a) Comparison of ROMS solutions using UP3 and SUP3, referred to ORCA05 January climatology. (b) Comparison of ROMS solutions using UP3 and RSUP3 (c) Comparison of ROMS solutions using UP3, and SUP3-A; SUP3-A is the non-rotated version of SUP3 in which diffusion is applied to salinity anomalies. In this case the profile remains very close to the climatology. (d) Peclet hyperdiffusivity vertical profile.

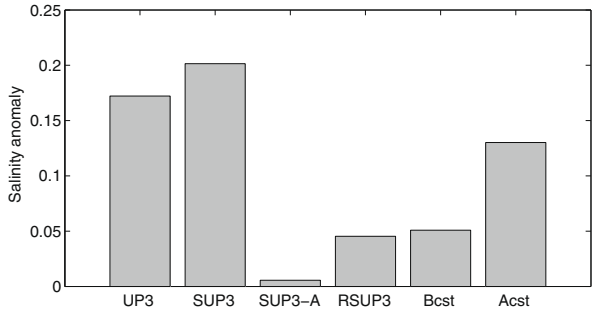

Fig. 8. ROMS salinity anomalies at 1000 m depth, averaged over the Caledonia region (spanning longitudes158–174°E and latitudes 26–12°S, as in Fig. 7) after 2 years of integration. The solutions are referred to ORCA05 January climatology. Bcst and Acst are modified RSUP3 versions with constant diffusion coefficients and lower order of diffusion (Acst); SUP3-A is the non-rotated version of SUP3 in which diffusion is applied to salinity anomalies.

energy is not excessively dissipated (our interest here is mostly concerned with the thermocline area where available potential energy is largest). Fig. 9b shows the surface energy spectrum for various choices of diffusion: RSUP3, constant biharmonic and harmonic diffusion coefficients. We have already shown in Section 4.1 that SUP3 is able to maintain the high levels of energy given by UP3 (Fig. 9a), featuring active mesoscale activity even at such medium resolution. When using constant hyperdiffusivity, the energy spectrum is not dramatically changed, but the smaller scales in the mesoscale regime (under $\sim$ 100 km) present lower energy levels consistent with our previous observation that the tracer fields are smoother in this case. If excess diffusion occurs it may affect the effective resolution of the model (estimated around $7\Delta x$, i.e. the wavelength where the model's spectrum begins to decay relative to the k^{-3} mesoscale regime), and compromise frontal and other sub-mesoscale formations. The difference with RSUP3 then decreases towards larger scales, although it does not vanish completely.

However, more dramatic differences appear for the constant first-order diffusion case. The energy levels in this case are many times lower. The large-scales and mesoscale are particularly affected, as also illustrated in Fig. 10, where eddy activity is clearly lacking in the northern part of the domain (compare with UP3 and RSUP3 solutions, which are both very energetic at the mesoscale). Note that rotated Laplacian operators are commonly used in climate models for physical closure, even in eddy-resolving models, instead of more complex rotated biharmonic schemes (Roberts and Marshall, 1998). In the case of regional sigma-coordinate models, we conclude that the RSUP3 with variable hyperdiffusivity provides the best dynamical solution, while avoiding excessive diapycnal diffusion.

5. Further considerations

5.1. Diffusion of tracer anomalies: SUP3-A

SUP3-A is an alternative scheme designed for diffusion of tracer anomalies as in Mellor et al. (1998) but using SUP3; some results are plotted in Figs. 7c and 8. In the averaged salinity profile (Fig. 7c), the solution is almost confounded with climatology, suggesting that the water masses seem perfectly conserved. However, the vertical velocity field (not shown) can be very noisy, with a maximum numerical check-board mode at the bottom. This may also result in local oscillations of tracer profiles. A general explanation for these numerical instabilities may be that the artificial constraint to preserve climatology is inconsistent with

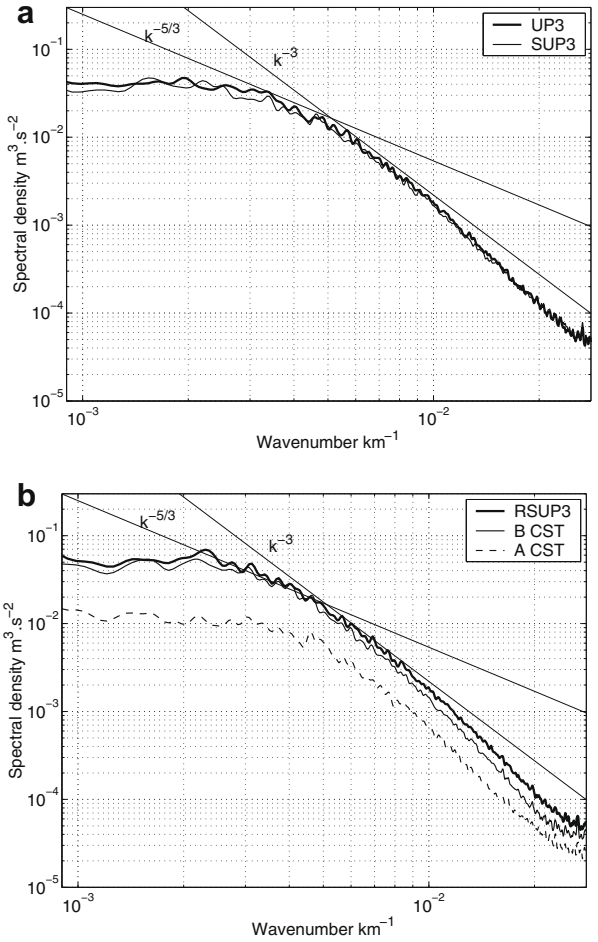

Fig. 9. Energy spectrum of ROMS solutions. (a) Comparison of UP3 and SUP3 schemes. (b) Comparison of RSUP3, and modified versions to test coefficients and order of diffusion.

dynamical adjustments of the mass field. Mellor et al. (1998) proposed (we can assume for this general reason as no explanation was given) to let the reference field evolve from the climatology. We have tried this option but so far without further success. Nevertheless, this solution has the real advantage of avoiding rotated operators and could be useful even in climate models if the numerical instability problem can be solved. We therefore leave this option open.

5.2. Smagorinsky diffusivity

The only widely used formulation of flow-dependent diffusivity is proposed by Smagorinsky (1963). This is a physical rather than numerical closure, developed for the momentum equations; it is sometimes applied to tracers (Blumberg and Mellor, 1987; in POM for example), but with unclear physical justification. The Smagorinsky viscosity for one-dimensional Laplacian operators is given by:

$$A_{\text{smag}} = C_A|D|\Delta x^2|D| = \sqrt{D_T^2 + D_S^2} = \sqrt{\left[\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right]^2 + \left[\frac{\partial u}{\partial y} - \frac{\partial v}{\partial x}\right]^2} \tag{26}$$

D is the deformation rate (D_T is horizontal tension and D_S is horizontal shearing strain), C_A is a dimensionless scaling parameter. If the deformation rate scales as $|D| = U/\Delta x$ (Griffies and Hallberg, 2000), then a scale for the Smagorinsky coefficient is:

$$A_{\text{smag}} \approx C_A U \Delta x \tag{27}$$

From physical arguments, the value of C_A is generally given around 0.1. However to prevent numerical instability, the Peclet constraint requires a value of 0.5, which is much larger. In practice then, the Smagorinsky formulation with usual coefficients should under-dissipate or under-diffuse the solution. However, numerical instabilities may be suppressed in some models by excessive temporal filtering. In particular, Durran (1991) has shown that the Robert–Asselin filter applied in many ocean models to the Leapfrog time-stepping is able to suppress numerical instabilities arising from centered advection operators (in this case the temporal scheme becomes first-order accurate and over-diffusive). This explains the observation by Mellor et al. (1998) concerning the good stability of POM tracer fields, even in absence of any explicit diffusion. In ROMS, our third-order time-stepping allows fine-scale physical structures to develop, as well as numerical instabilities. In our case then, the Smagorinsky formulation with $C_A = 0.1$ leads to a very noisy solution (not shown). On the other hand, imposing higher values may lead to excessive diffusion as well as truncation errors similar to the constant harmonic diffusion case. The relevant outcome of the present discussion is that time filters may present a valuable alternative to spatial diffusion for numerical closure in sigma models. This aspect of the problem, which is related to the implicit discretization issue (Section 3.5), will be addressed in a following study devoted to temporal schemes for advection.

5.3. Suggestions for momentum advection

If the first-order Smagorinsky formulation is numerically inappropriate, let's explore, as Griffies and Hallberg (2000) did, the possibility of extending the formulation to second-order diffusion, but using our own scaling analysis. The Smagorinsky hyperdiffusivity should scale as:

$$B_{\text{smag}} \approx \alpha A_{\text{smag}} \Delta x^2 \approx \alpha C_A U \Delta x^3 \tag{28}$$

The Peclet constraint for our fourth-order centered advection scheme imposes a value of 1/12 for αC_A. With this constant value, the biharmonic Smagorinsky formulation can be used in a stable way. For tracer advection, there is no real physical justification to replace the diffusivity formulation given by the split advection scheme, i.e. replacing U by. However, the parameterization of frictional boundary layers in the momentum equation requires a formulation of viscosity which depends on shearing strain. It has been shown (Chassignet and Garraffo, 2001) that the absence of such dependence leads to western boundary currents (WBC) that are too energetic, with generation of eddies and separation mechanism produced upstream of the observed locations. In ROMS also, the use of UP3 in the momentum equation leads to physically unstable WBC, i.e. the Aghulas Current (Penven, personal communication) and the Gulf Stream (see Haidvogel et al., 2000). A natural way of combining numerical and frictional boundary layer constraints for the momentum equation would be to adopt a biharmonic Smagorisnky formulation adapted to the advection scheme. In our case, a combination of physical and numerical closure would give:

$$B_{\text{smag}} = \frac{1}{12}\max\left(\frac{|u|}{\Delta x}, |D|\right)\Delta x^4 \tag{29}$$

This suggested formulation is to be validated in further studies.

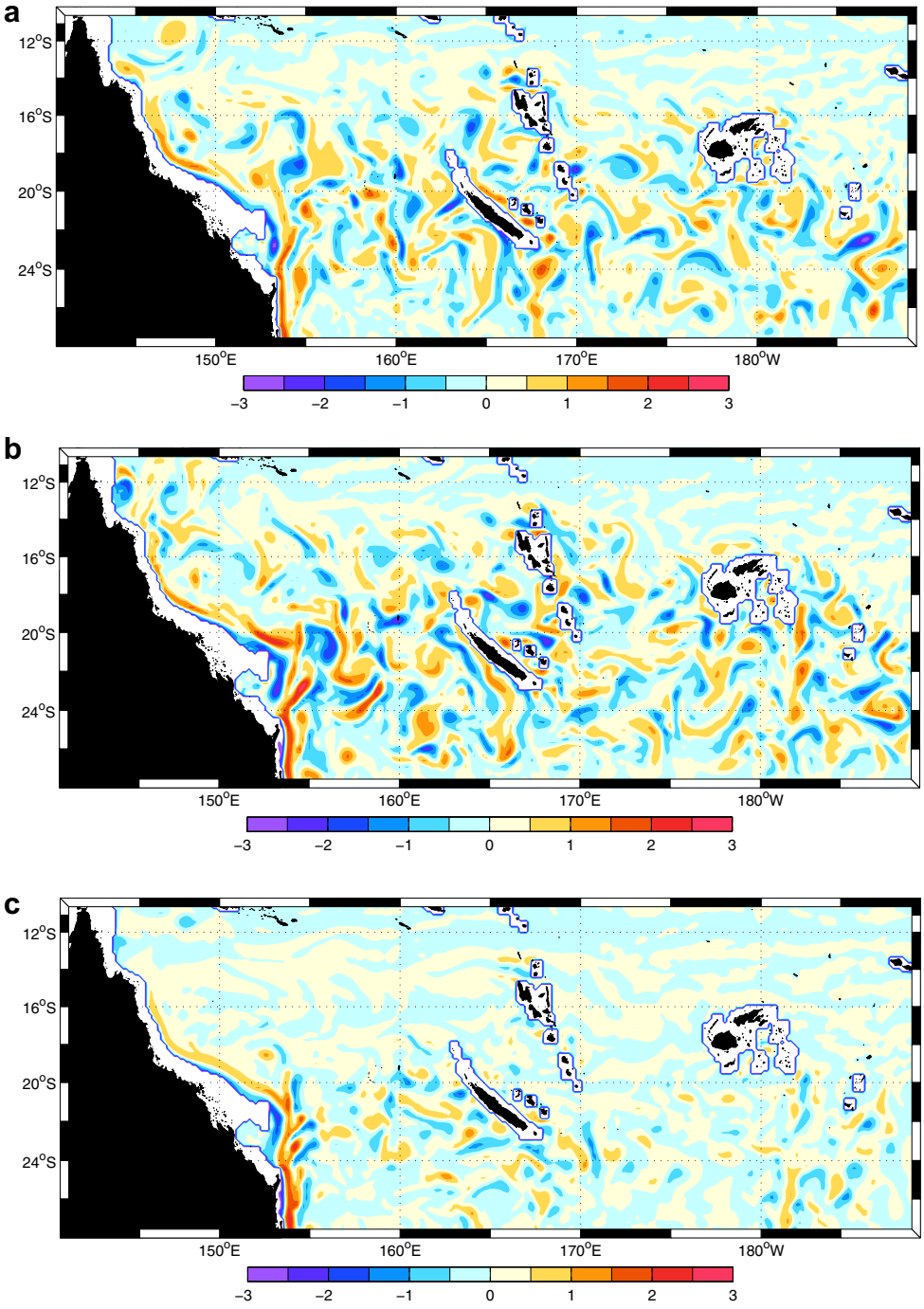

Fig. 10. ROMS surface relative vorticity after 2 years of integration. (a) UP3. (b) RSUP3. (c) RSUP3 with first-order diffusion and constant diffusivity 400 m^2/s.

6. Conclusion

In this paper, we identify a crucial numerical problem in sigma coordinate models, leading to unacceptable spurious diapycnal mixing. The error is a byproduct of recent advances in numerical methods, namely the implementation of high-order diffusive advection schemes. Spurious mixing in the case of ROMS is produced by its third-order upwind advection scheme, but our analysis suggests that all diffusive advection schemes, including the most recent monotonic schemes with complex nonlinear limiters, should give equivalent problems in all sigma models. Most regional applications are focused on surface fields where mixing errors are least detectable. This may explain why the problem has been so far unnoticed, although the mixing problem also affects the surface layers. Another prevailing yet false idea is that spurious mixing should decrease with resolution. We show that in a coarse-resolution regime, spurious mixing increases as resolution is refined, and may reach its peak value when eddy-driven lateral mixing becomes explicitly resolved. At finer resolution, diapycnal diffusivities should decrease but only become physically acceptable at resolutions finer than the kilometer.

The solutions to this problem are few because it requires a specifically designed advection scheme with either very small implicit diffusion along coordinate surfaces or rotated explicit

diffusion. We have proposed and validated the latter, which at this time appears as the only valid alternative. The RSUP3 scheme is a modified version of the scheme proposed in Holland et al. (1998) and Webb et al. (1998), where diffusion is split from advection and is represented by a rotated biharmonic diffusion scheme with flow-dependent hyperdiffusivity satisfying the Peclet constraint (named here Peclet hyperdiffusivity). Our modified version is composed of a rotated diffusion operator modified for numerical stability; it comes with improvements of stability limits by adjustment of the temporal scheme, and a clipping method adapted to the sigma-coordinate. RSUP3 is able to preserve low dispersion and diffusion capabilities of the original third-order upwind scheme while preserving water mass characteristics. There are likely to be residual errors from the rotated diffusion operator but they remain acceptable. In this respect, we have shown that a rotated biharmonic diffusion operator performs better than a Laplacian operator, despite increased numerical complexity of the second-order rotation. The use of a constant diffusivity rather than the Peclet hyperdiffusivity also tends to increase residual errors which become unacceptable with Laplacian diffusion. Finally, some options have been left open concerning the application of non-rotated SUP3 to tracer anomalies, and perhaps more importantly that of time filters as an alternative to spatial diffusion. A temporal discretization approach to the present problem (including semi-implicit discretization) is under way and will be reported in a following paper. It should provide in particular an extension of our techniques to high-order isopycnal diffusion for non-regional applications of sigma-coordinate models (and geopotential coordinate models as well). Time filters may also help damping numerical extrema from oscillatory schemes, but the accurate application of strictly monotonic schemes to the sigma-coordinates remains an unsolved and largely unacknowledged issue.

Acknowledgement

We appreciate financial support from the IRD. In particular, the PC-cluster used for the ROMS simulations was financed through IRD's scientific computing effort project SPIRALES.

References

Barnier, B., Marchesiello, P., de Miranda, A.P., Molines, J.M., Coulibaly, M., 1998. A sigma-coordinate primitive equation model for studying the circulation in the South Atlantic. Part I: model configuration with error estimates. Deep-Sea Res. I 45, 543–572.

Barnier, B., Siefried, L., Marchesiello, P., 1995. Thermal forcing for a global ocean circulation model using a three-year climatology of ECMWF analyses. J. Mar. Sys. 6, 363–380.

Beckers, J.-M., Burchard, H., Deleersnijder, E., Mathieu, P.P., 2000. Numerical discretization of rotated diffusion operators in ocean models. Mon. Weather Rev. 128 (8), 2711–2733.

Beckmann, A., Haidvogel, D.B., 1993. Numerical simulation of flow around a tall isolated seamount. Part I: problem formulation and model accuracy. J. Phys. Oceanogr. 23, 1737–1753.

Blumberg, A., Mellor, G.L., 1987. A description of a three-dimensional coastal ocean circulation model. In: Heaps, N.S. (Ed.), Three-Dimensional Coastal Ocean Models, vols. 1–16. American Geophysical Union, Washington, DC.

Bryan, K., 1975. Three-dimensional numerical models of the ocean circulation. In: Numerical Models of Ocean Circulation. Natl. Acad. Sci., pp. 94–105.

Chassignet, E.P., Garraffo, Z.D. 2001. Viscosity parameterization and the Gulf Stream separation. In: Muller, P., Henderson, D. (Eds.), From Stirring to Mixing in a Stratified Ocean. Proceedings'Aha Huliko'a Hawaiian Winter Workshop. U. of Hawaii. January 15–19, 2001. pp. 37–41.

Couvelard, X., Marchesiello, P., Gourdeau, L., Lefevre, J. 2008. Barotropic zonal jets induced by islands in the southwest Pacific. J. Phys. Oceanogr., Doi: doi:10.1175/2008JPO3903.1 (online release).

Cox, M.D., 1987. Diffusion in a z-coordinate ocean model. Ocean Modell. 74, 1–5. unpublished manuscript.

Da Silva, A.M., Young, C.C., Levitus, S. 1994. Atlas of surface marine data 1994. vol. 1, algorithms and procedures, technical report, Natl. Oceanogr. Data. Cent., Sylver, spring, Md.

Dunn, J.R., Ridgway, K.R., 2002. Mapping ocean properties in regions of complex topography. Deep-Sea Res. 49, 591–604.

Durran, D.R., 1991. The third-order Adams–Bashforth method: an attractive alternative to leapfrog time differencing. Mon. Weather Rev. 119, 702–720.

Durran, D.R., 1999. Numerical Methods for Wave Equations in Geophysical Fluid Dynamics. Springer Verlag. 465pp.

Godunov, S.K., 1959. A difference scheme for numerical solution of discontinuous solution of hydrodynamic equations. Math. Sbornik, 47, 271–306. Translated US Joint Publ. Res. Service, JPRS 7226, 1969.

Griffies, S, 2007. Spurious diapycnal mixing in ocean models, Workshop on Numerical Methods in Ocean Models, Bergen, August 2007.

Griffies, S.M., Hallberg, R.W., 2000. Biharmonic friction with a Smagorinsky-like viscosity for use in large-scale eddy-permitting ocean models. Mon. Weather Rev. Part 2 128 (8), 2935–2946.

Griffies, Stephen M., Pacanowski, Ronald C., Hallberg, Robert W., 2000. Spurious diapycnal mixing associated with advection in a z-coordinate ocean model. Mon. Weather Rev. 128 (3), 538–564.

Griffies, S.M., Gnanadesikan, A., Pacanowski, R.C., Larichev, V.D., Dukowicz, J.K., Smith, R.D., 1998. Isoneutral diffusion in a *z*-coordinate ocean model. J. Phys. Oceanogr. 28 (5), 805–830.

Haidvogel, D.B., Arango, H.G., Hedstrom, K., Beckmann, A., Malanotte-Rizzoli, P., Shchepetkin, A.F., 2000. Model evaluation experiments in the north atlantic basin: simulations in nonlinear terrain-following coordinates. Dyn. Atmos. Oceans 32, 239–281.

Hecht, M.W., Wingate, B.A., Kassis, P., 2000. A better, more discriminating test problem for ocean tracer transport. Ocean Modell. 2, 1–15.

Holland, W.R., Chow, J., Bryan, F., 1998. Application of a third-order upwind scheme in the NCAR ocean model. J. Clim. 11, 1487–1493.

Kessler, W.S., Gourdeau, L., 2007. The annual cycle of circulation of the southwest subtropical Pacific analysed in an ocean GCM. J. Phys. Oceanogr. 37, 1610–1627.

Large, W.G., McWilliams, J.C. Doney, 1994. Oceanic vertical mixing: a review and a model with a nonlocal boundary layer paramétrisation. Rev. Geophys. 32, 363–403.

Lee, M-M., Coward, A.C., Nurser, A.J.G., 2002. Spurious diapycnal mixing of the deep waters in an eddy-permitting global ocean model. J. Phys. Oceanogr. 32 (5), 1522–1535.

Leonard, B.P., 1979. A stable and accurate convective modelling procedure based on quadratic upstream interpolation. Comput. Methods Appl. Mech. Eng. 19, 59–98.

Marchesiello, P., McWilliams, J.C., Shchepetkin, A., 2001. Open boundary conditions for long-term integration of regional oceanic models. Ocean Modell. 3, 1–20.

Marchesiello, P., McWilliams, J.C., Shchepetkin, A., 2003. Equilibrium structure and dynamics of the California current system. J. Phys. Oceanogr. 33, 753–783.

McCalpin, J.D., 1994. A comparison of second-order and fourth-order pressure gradient algorithms in a sigma coordinate ocean model. Int. J. Num. Methods Fluids 18, 361–383.

Mellor, G.L., Oey, L.-Y., Ezer, T., 1998. Sigma coordinate pressure gradient errors and the seamount problem. J. Atmos. and Ocean. Technol. 15 (5), 1122–1131.

Mellor, G.L., Blumberg, A.F., 1985. Modeling vertical and horizontal diffusivities with the sigma coordinate system. Mon. Weather Rev. 113 (8), 1379–1383.

Penven, P., Marchesiello, P., Debreu, L., Lefevre, J., 2007. Software tools for pre- and post-processing of oceanic regional simulations. Environ. Model. Softw. 23, 660–662.

Redi, M.H., 1982. Oceanic isopycnal mixing by coordinate rotation. J. Phys. Oceanogr. 12, 1154–1158.

Roberts, M., Marshall, D., 1998. Do we require adiabatic dissipation schemes in eddy-resolving ocean models. J. Phys. Oceanogr. 28, 2050–2063.

Sanderson, B.G., 1998. Order and resolution for computational ocean dynamics. J. Phys. Oceanogr. 28, 1271–1286.

Semtner, A.J., Mintz, Y., 1977. Numerical simulation of the Gulf Stream and mid-ocean eddies. J. Phys. Oceanogr. 7, 208–230.

Shchepetkin, A.F., McWilliams, J.C., 1998. Quasi-monotone advection schemes based on explicit locally adaptive dissipation. Mon. Weather Rev. 126, 1541–1580.

Shchepetkin, A.F., McWilliams, J.C., 2003. A method for computing horizontal pressure-gradient force in an ocean model with a non-aligned vertical coordinate. J. Geophys. Res. 108 (C3), 3090. doi:10.1029/2001JC001047.

Shchepetkin, A.F., McWilliams, J.C., 2005. The regional oceanic modeling system (roms): a split–explicit, free-surface, topography-following-coordinate oceanic model. Ocean Modell. 9, 147–404.

Smagorinsky, J., 1963. General circulation experiments with the primitive equations. Mon. Weather Rev. 91, 99–164.

Webb, D.J., Cuevas, B.D., Richmond, C., 1998. Improved advection schemes for ocean models. J. Atmos. Ocean. Technol. 15 (1998), 1171–1187.