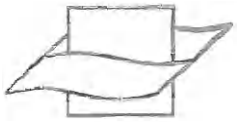

Another Reason Why Simple Discretizations of Rotated Diffusion Operators Cause Problems in Ocean Models: Comments on "Isoneutral Diffusion in a *z*-Coordinate Ocean Model"

J. M. BECKERS*

University of Liege, Liege, Belgium

H. BURCHARD

Joint Research Centre, Ispra, Italy

J. M. CAMPIN, E. DELEERSNIJDER,* AND P. P. MATHIEU

Université Catholique de Louvain, Louvain-La-Neuve, Belgium

28 July 1997 and 30 September 1997

1. Introduction

In recent papers (Griffies et al. 1998; Griffies 1998), the problem of isopycnal diffusion and Gent–McWilliams stirring in *z*-coordinate models was reviewed and a new discretization proposed. It was shown that classical discretization (Cox 1987; Gent and McWilliams 1990) needs rather heavy background diffusion along the grid lines in order to stabilize the scheme. Griffies et al. (1998) show the possible origin of the problem: one reason is the imperfect balancing of temperature and salinity diffusive fluxes along neutral surfaces and the other one the existence in the Cox discretization of a $2\Delta x$ mode invisible to the cross derivative in the isopycnal diffusion. The authors then present solutions, which they show to work properly in an experiment for long climate runs. Here, we will broaden the discussion of the problem by pinning up another basic numerical difficulty in discretizing rotated diffusion operators. We will also not limit ourselves to isopycnal diffusion in *z*-coordinate models, as the well-known GFDL MOM2 (Modular Ocean Model), but work in a more general context in which coastal ocean models such as SPEM (Semi-Spectral Primitive Equation Ocean Circulation Model) along terrain-following coordinates may use rotated diffusion operators to obtain isopycnal or geopotential diffusion. In the model descriptions and their applications (e.g., Cox 1987; Pacanowski 1995; Hedström 1994), this rotation of the diffusion operator is mentioned, but it is also always stated that some additional background diffusion along grid coordinates should be maintained, that rotations are limited to weak slopes, that intermittent filtering is added or that the rotated diffusion option does not work properly in all cases, a fact observed in several studies (e.g. Cox 1987; Gerdes et al. 1991). This has also been recognized by Griffies et al. (1998), and one could wonder why a diffusion term that, in principle, has a smoothing behavior needs such additional damping, potentially masking the desired diffusion along a specific coordinate surface. In addition to the problems established by Griffies et al. (1998), we will show that classical discretizations of rotated diffusion operators lead also to problems due to the violation of another physical property of diffusion: a well-known characteristic of pure diffusion is that diffusion operators are positive definite, which guarantees that, in a closed domain, the maximum and minimum of the variable being diffused do not increase and decrease, respectively, a property which will be referred to as the "min–max principle." In addition, the variance decreases if natural or Dirichlet boundary conditions are applied. The min–max principle should hopefully be recovered in the discretized version; otherwise, tracers would clearly show nonphysical behavior since positive definite quantities (e.g., turbulent kinetic energy or concentrations) could become negative and dynamical state variables as density would create unphysical pressure gradients. Even if this effect on

[*] Research associate, National Fund for Scientific Research, Belgium.

Corresponding author address: Dr. Jean-Marie Beckers, GHER, University of Liege, Sart-Tilman B5, B-4000 Liege, Belgium.
E-mail: JM.Beckers@ulg.ac.be

density and pressure may be limited in practice by some compensation of temperature and salinity errors when computing the corresponding density, this feedback generally has a destabilizing effect (e.g., Mathieu 1996).

In the large-scale Cox model, the major problems are those described in Griffies et al. (1998), namely the noncancellation of temperature and salinity flux contributions to density fluxes along isoneutral surfaces and the inapproriate computation of products of coordinate and tracer gradients. But for other types of models or passive tracers, the violation of a min–max principle may lead to grid noise that amplifies to a level where heavy filtering or diffusion along the numerical grid is needed to get rid of the numerical noise. Then, obviously, the potential advantages of rotated diffusion operators are lost since filtering is carried out along grid lines. Violation of the min–max principle or the decreasing variance principle is thus a potential problem. Beside the MOM2, for which Griffies et al. (1998) bring some important improvements, there is a range of models using different coordinate systems and a series of assumptions on the preferred direction of diffusion, which generally does not coincide with the model coordinates, unless the diffusion is purely a numerical artifact to damp noise of the numerical scheme. A few well-known implementations of the preceding concepts are the following:

1) diffusion on potential density surfaces, be it isopycnal (Redi 1982) or isoneutral (McDougall and Church 1986), in the Cox GFDL model (Cox 1987)
2) isopycnal diffusion in terrain-following coordinates, as for example the implemenation of Hedström (1994) in SPEM (Haidvogel et al. 1991). Here the authors clearly state in the user manual of SPEM version 5.1: *rotated mixing tensors do* not *work yet* (Hedström 1996)
3) z diffusion in σ-coordinate models (Stelling and Van Kester 1994).

We will thus concentrate on a very general formulation of the subgrid-scale parameterization formulated in a coordinate system not coinciding with the numerical grid.

Here we analyze the pure diffusion term associated with the subgrid-scale mixing processes, and no additional eddy-induced advection term (Gent and McWilliams 1990) is retained here. The latter is indeed generally dealt with by the advection scheme of the numerical model (e.g., Gerdes et al. 1991) or as the antisymmetric part of a generalized diffusion tensor (Visbeck et al. 1997; Griffies 1998).

Here we have to mention that dealing with the advection part in the antisymmetric part of a diffusion tensor is generally not a way to ensure a positive defined method since, when it is a linear scheme, only the upwind scheme obeys this property in the abscence of diffusion. In particular, for a Gent–McWilliams advection velocity that is uniform and horizontal, the scheme of Griffies (1998) reduces to a classical horizontal centered advection scheme when no diffusion is added. In this case, we know the behavior. Of course when combined with the isopycnal diffusion the behavior of the scheme improves, but unless the isopycnal diffusion coefficient $\mathcal{A}$ is larger or similar to the thickness diffusion coefficient κ, the centered nature of this skew flux computation may be grid-noise producing.

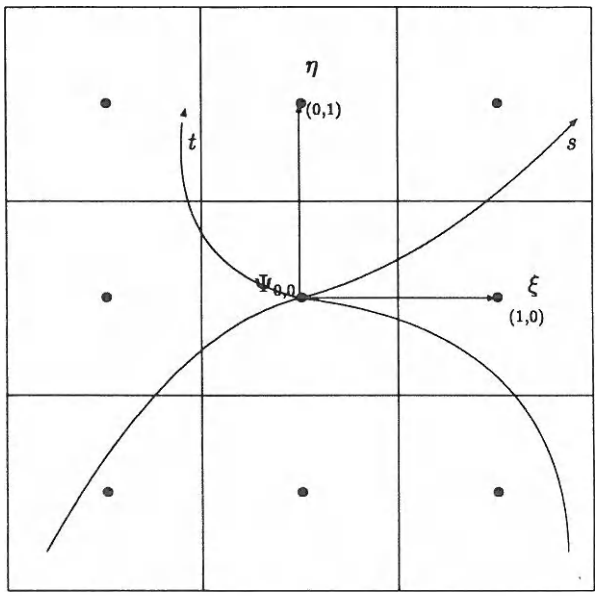

FIG. 1. Grid and naming convention.

Limiting our investigation to the diffusion part implies that a symmetric positive definite diffusion tensor is assumed.

2. Generic formulation

In order to analyze the problem of rotated diffusion, we will restrict the dimensions to a vertical section. This limitation to 2D is not penalizing since at least any method should work in 2D. If problems appear in two dimensions, then they will surely be present in general 3D situations, at least for the long waves in the additional direction. We even expect more severe problems in 3D because of the appearance of cross-derivative terms coupling the two horizontal directions when rotating diffusion tensors. As our purpose is to show that the rotation of diffusion operators leads to a fundamental numerical problem, we will remain in the framework of a 2D case. Because we often do not only wish to express the diffusion tensor to be diagonal in a specific coordinate system, but also often introduce a numerical coordinate change, which is used to operate in a convenient discrete space, we shall analyze the configuration of Fig. 1, which contains all possibilities, from isopycnal grids to generalized vertical grids (Kasahara 1974) and classical z coordinates.

The normalization of ξ, η is by no means limiting; in the contrary, it encompasses all possible situations,

including nonuniform grids by allowing the general transformation

$$\xi = \xi(x, z) \tag{1}$$

$$\eta = \eta(x, z). \tag{2}$$

In this way, the nodes of the numerical grid are easily referenced. We now assume that diffusion should be expressed in a specific relative coordinate system (s, t):

$$s = s(\xi, \eta) \tag{3}$$

$$t = t(\xi, \eta). \tag{4}$$

Between these two coordinate systems, the Jacobian $\mathcal{J}$ is

$$\mathcal{J} = \frac{\partial s}{\partial \xi}\frac{\partial t}{\partial \eta} - \frac{\partial s}{\partial \eta}\frac{\partial t}{\partial \xi}. \tag{5}$$

The most general formulation of a diffusion operator applied to variable Ψ reads

$$\mathcal{D} = \begin{pmatrix} \frac{\partial}{\partial s} & \frac{\partial}{\partial t} \end{pmatrix} \begin{pmatrix} K_{ss} & K_{st} \\ K_{ts} & K_{tt} \end{pmatrix} \begin{pmatrix} \frac{\partial \Psi}{\partial s} \\ \frac{\partial \Psi}{\partial t} \end{pmatrix}. \tag{6}$$

Since we assume that in the special coordinate system (s, t), the diffusion tensor is diagonal by construction, there should simply be a "vertical" and "isopycnal" diffusion:

$$\mathcal{D} = \frac{\partial}{\partial s}\left(\mathcal{A}\frac{\partial \Psi}{\partial s}\right) + \frac{\partial}{\partial t}\left(\mathcal{A}_t\frac{\partial \Psi}{\partial t}\right). \tag{7}$$

We will analyze the pure diffusion along s because diapycnal or vertical diffusion may become very weak, especially if a turbulent closure scheme is used. Since we want to be sure that the diffusion along s does not create any problem, we cannot rely on the vertical or diapycnal component to compensate for problems related to the diffusion along s since we generally cannot guarantee that diapycnal or vertical diffusion is always present. In other words, we focus on the diffusion along a given coordinate line, regardless of another diffusion in the vertical or diapycnal direction. In this case, any diffusion of a field Ψ along the coordinate line $s(\xi, \eta)$ may be written as

$$\mathcal{D} = \frac{\partial}{\partial s}\left(\mathcal{A}\frac{\partial \Psi}{\partial s}\right). \tag{8}$$

If variations of the distances H between the coordinate curves on which diffusion acts are to be taken into account, this is generally parameterized as

$$H\mathcal{D} = \frac{\partial}{\partial s}\left(H\mathcal{A}\frac{\partial \Psi}{\partial s}\right). \tag{9}$$

In our dicussion, we work with formulation (8) since an appropriate renaming of $\mathcal{A}$ (which remains positive) allows one to switch from one formulation to the other.

We can rewrite the diffusion term in the numerical coordinate system used for discretization:

$$\mathcal{D} = \mathcal{A}\left(\frac{\partial \xi}{\partial s}\right)^2\frac{\partial^2 \Psi}{\partial \xi^2} + \mathcal{A}\left(\frac{\partial \eta}{\partial s}\right)^2\frac{\partial^2 \Psi}{\partial \eta^2} + 2\mathcal{A}\frac{\partial \xi}{\partial s}\frac{\partial \eta}{\partial s}\frac{\partial^2 \Psi}{\partial \xi \partial \eta} + \frac{\partial}{\partial s}\left(\mathcal{A}\frac{\partial \xi}{\partial s}\right)\frac{\partial \Psi}{\partial \xi} + \frac{\partial}{\partial s}\left(\mathcal{A}\frac{\partial \eta}{\partial s}\right)\frac{\partial \Psi}{\partial \eta}. \tag{10}$$

This expression is equivalent to the rotated Cox tensor if coordinate lines are z levels and diffusion is done along isopycnals (s is along constant ρ and $H = \rho_0/\sqrt{\rho_x^2 + \rho_z^2}$).

It also contains the case of horizontal diffusion in generalized σ coordinates (for which x was conveniently normalized):

$$\xi = x \tag{11}$$

$$\eta = F(\sigma), \qquad \sigma = \frac{z + h}{\zeta + h} \tag{12}$$

$$s = x = \xi \tag{13}$$

$$t = z = F^{-1}(\eta)(\zeta + h) - h \tag{14}$$

$$\frac{\partial \xi}{\partial s} = 1 \tag{15}$$

$$\frac{\partial \eta}{\partial s} = F'\left\{\frac{h_x}{\zeta + h} - \frac{(z + h)(\zeta_x + h_x)}{(\zeta + h)^2}\right\}. \tag{16}$$

For classical σ coordinates with uniform spacing $F(\sigma) = \sigma$, and its inverse function F^{-1} is readily obtained.

3. Discretization

We will now show that any nine-point, linear, consistent discretization of Eq. (10) leads necessarily to a scheme that does not satisfy the min–max principle of real diffusion, except in some simple and well-known degenerated cases. This means that such schemes may generate minima and maxima outside the range of the initial data. At this stage we do not even require precise or conservative schemes, but only a consistent scheme (that is equivalent to the mathematical expression when time steps and grid sizes tend independently to be infinitely small). To demonstrate the conjecture that such a scheme cannot in general satisfy the min–max principle of its physical counterpart, we will use the most general nine-point discretization that can be taken on the stencil around the local point (0, 0). This discrete version reads

$$\mathcal{D} = \sum_{i=-1,1}\sum_{j=-1,1} a_{i,j}\Psi(i, j). \tag{17}$$

If we assume that we have found a consistent discretization of the space operator, one then simply has to focus on the time T tendencies $\partial\Psi/\partial T$ to retrieve some necessary conditions upon the coefficients $a_{i,j}$ when the discretization has to fulfill the min–max principle:

$$\frac{\partial \Psi}{\partial T}(0, 0) = \sum_{i=-1,1} \sum_{j=-1,1} a_{i,j}\Psi(i, j). \tag{18}$$

From there a necessary condition for the min–max principle is

$$a_{0,0} \le 0 \tag{19}$$

$$0 \le a_{i,j}, \qquad (i, j) \ne (0, 0). \tag{20}$$

This is indeed necessary to assure the min–max principle; otherwise, as shown for example in Jameson (1995), assuming a constant field everywhere except a higher value at the point (i, j) for which one of these necessary conditions does not hold, it is easily shown that the time derivative will have a tendency such that $\Psi(0, 0)$ will have a lower value than the constant field or an even higher value than the value at point (i, j), thus violating the min–max principle.

We will now show that, indeed, the rotation of the diffusion can cause the appearance of coefficients violating condition (19) or (20).

Using classical Taylor development of the field Ψ around (0, 0) we have

$$\Psi(\xi, \eta) = \Psi(0, 0) + \xi\frac{\partial \Psi}{\partial \xi} + \eta\frac{\partial \Psi}{\partial \eta} + \frac{\xi^2}{2}\frac{\partial^2 \Psi}{\partial \xi^2} + \frac{\eta^2}{2}\frac{\partial^2 \Psi}{\partial \eta^2} + \xi\eta\frac{\partial^2 \Psi}{\partial \xi \partial \eta} + O\left(\frac{\partial^3 \Psi}{\partial \xi^m \partial \eta^{3-m}}\right). \tag{21}$$

Requiring that the scheme is at least consistent demands that, when replacing each $\Psi(i, j)$ by its Taylor development in Eq. (17) we retrieve Eq. (10), the reformulation of the initial diffusion law (8); consistency requires that this is true by neglecting derivatives of the third order and higher. Since we are dealing with a linear scheme, the coefficients $a_{i,j}$ do not depend upon the function Ψ, and we must impose that the coefficients multiplying the different derivatives of Ψ appearing when doing the actual discretization are equivalent to those emerging from the analytical development. This requirement leads to six equations for the nine coefficients $a_{i,j}$. At this point of reasoning, one could hope that it is possible to find coefficients where only $a_{0,0}$ is negative. One then could even think about taking advantage of the remaining three degrees of freedom to add constraints of tracer conservation, higher-order precision, or variance diminishing. Unfortunately, we will show that at least one other coefficient than $a_{0,0}$ is strictly negative. Indeed, consistency requires by identifying the coefficients that multiply $\Psi(0, 0)$:

$$\sum_{i=-1,1} \sum_{j=-1,1} a_{i,j} = 0. \tag{22}$$

Identification of the coefficients multiplying $\partial\Psi/\partial\xi$ leads to

$$a_{1,1} - a_{-1,1} + a_{1,0} - a_{-1,0} + a_{1,-1} - a_{-1,-1} = \frac{\partial}{\partial s}\left(\mathcal{A}\frac{\partial \xi}{\partial s}\right), \tag{23}$$

and similarly for $\partial\Psi/\partial\eta$

$$a_{1,1} - a_{1,-1} + a_{0,1} - a_{0,-1} + a_{-1,1} - a_{-1,-1} = \frac{\partial}{\partial s}\left(\mathcal{A}\frac{\partial \eta}{\partial s}\right). \tag{24}$$

These equations [(22), (23), and (24)] do not lead to any interesting conclusion here except that the retrieval of $a_{0,0}$ is negative when all other coefficients are positive. Far more interesting are the relationships emerging from the identification of the higher order derivatives:

$$a_{1,1} + a_{-1,1} + a_{1,0} + a_{-1,0} + a_{1,-1} + a_{-1,-1} = 2\mathcal{A}\left(\frac{\partial \xi}{\partial s}\right)^2 \tag{25}$$

$$a_{1,1} + a_{1,-1} + a_{0,1} + a_{0,-1} + a_{-1,1} + a_{-1,-1} = 2\mathcal{A}\left(\frac{\partial \eta}{\partial s}\right)^2 \tag{26}$$

$$a_{1,1} + a_{-1,-1} - a_{1,-1} - a_{-1,1} = 2\mathcal{A}\frac{\partial \eta}{\partial s}\frac{\partial \xi}{\partial s}. \tag{27}$$

By defining

$$r = \frac{\partial \eta/\partial s}{\partial \xi/\partial s} \tag{28}$$

and adding Eq. (25) to (27) and subtracting Eq. (27) from (25) we get two other equations:

$$2a_{1,1} + a_{1,0} + a_{-1,0} + 2a_{-1,-1} = 2\mathcal{A}\left(\frac{\partial \xi}{\partial s}\right)^2 (1 + r) \tag{29}$$

$$2a_{-1,1} + a_{1,0} + a_{-1,0} + 2a_{1,-1} = 2\mathcal{A}\left(\frac{\partial \xi}{\partial s}\right)^2 (1 - r). \tag{30}$$

Similarly, adding Eq. (26) to (27) and subtracting Eq. (27) from (26) we get

$$2a_{1,1} + a_{0,1} + a_{0,-1} + 2a_{-1,-1} = 2\mathcal{A}\left(\frac{\partial \xi}{\partial s}\right)^2 (r + 1)r \tag{31}$$

$$2a_{-1,1} + a_{0,1} + a_{0,-1} + 2a_{1,-1} = 2\mathcal{A}\left(\frac{\partial \xi}{\partial s}\right)^2 (r - 1)r. \tag{32}$$

We have now four equations, (29), (30), (31), and (32), where the unknowns are multiplied by positive coefficients. If the unknowns $a_{i,j}$ other than $a_{0,0}$ are to be positive, then at least all right-hand sides of equations must be nonnegative. But it is easily seen that one of them is always

negative, if we suppose $r \neq 0$, $r \neq \pm 1$, and $r \neq \pm\infty$. The cases excluded are simply the configurations where the line along which diffusion must take place is just crossing the grid points, in which case a direct classical diffusion discretization along these grid points works well. But if the diffusion direction does not lie on the grid, there is *always* one coefficient $a_{i,j}$ that is negative, thus leading to a scheme which violates the min–max principle. It should be noted that, even if in practice the way the variable grid spacing, variable diffusion coefficients, and coordinate transformations are computed may vary, the consistency constraint is simply requiring that the coefficients $a_{i,j}$ satisfy a relationship in which the actual discretization of $\partial\eta/\partial s$ etc. does not matter, provided that it is done also in a consistent way.

Very disappointing is the result that even for constant slopes, uniform grid spacing, and constant diffusion coefficient [leading to constant right-hand sides of (29), (30), (31), and (32)] no well-behaved scheme can be found. In this case of constant slopes, we have to mention that the variance diminishing method of Griffies et al. (1998) applied to a passive tracer will not assure the min–max principle.

Furthermore, it is also clear that any method limiting the isopycnal slope to a prescribed maximum amplitude a priori is not appropriate to eliminate the min–max violation. In any case it is not the slope that is the important control factor of the "negativeness," but the slope compared to the coordinate slope, as reflected by the parameter r. This parameter can also be interpreted in terms of the hydrostatic consistency requirements in the case of σ-coordinate systems (e.g., Deleersnijder and Beckers 1992; Mellor et al. 1994; Haney 1991; Burchard and Petersen 1997).

One could wonder why the present simple diffusion problem has, to our knowledge, never been analyzed in the scope of computational fluid dynamics. In fact, in classical computational fluid dynamic models in curvilinear grids, diffusion is isotropic so that the right-hand side of Eqs. (29), (30), (31), and (32) contain additional terms that can cancel the negative parts, thus possibly eliminating the problem of negative coefficients. This canceling decreases, however, when grids are strongly distorted. This has been shown by Kershaw (1981), and he concluded that the Laplacian diffusion cannot be both monotonic and linear for arbitrary and, in particular, strongly distorted grids. This is, however, very different from our case: The demonstration of Kershaw holds for a full Laplacian diffusion, not a directional diffusion along a line as the isopycnal diffusion. This is very different since Laplacian diffusion is isotropic and does not know anything about directions, whereas isopycnal diffusion does, which means that the demonstration of Kershaw deals with the problem of expressing an isotropic diffusion in a distorted grid, not the problem of rotating the direction of diffusion. Furthermore, the theorem of (Kershaw 1981) shows that that it is impossible to have a consistent and min–max method on *arbitrary* grids. It does not say that it is impossible in all cases. When grids are sufficiently smooth and sufficiently orthogonal, then isotropic diffusion discretization does work correctly. In particular for orthogonal grids, which may be curvilinear, the five-stencil method does work and reflects again that it is a Laplacian diffusion, which is isotropic and retains its mathematical formulation (and hence well-conditioned discretization) on orthogonal coordinates.

On the contrary, our theorem shows that a *directional* diffusion is *never* both consistent and min–max satisfying, however small the slopes of the rotating.

To summarize, Kershaw (1981) shows that the Laplacian diffusion *may* have problems when grids are too distorted, while we show that *isopycnal* diffusion *always* has problems if linear schemes are used. This has, of course, a very important practical consequence: On the one hand, the problem of consistency and violation of the min–max principle can be controlled in linear Laplacian diffusion discretizations by controlling the grid. This is currently done by appropriate grid generators. On the other hand, isopycnal diffusion discretizations cannot be controlled by slope clipping, unless the slope is forced to be aligned on the grid, which is of course uninteresting.

One may wonder if some other solutions than grid controlling exist in the case of a Laplacian diffusion, which could be useful here. Indeed, for the case of isotropic Laplacian diffusion in strongly distorted grids, Demirdžić et al. (1987) and Zijlema (1996) suggest an ad hoc method how an isotropic conservative diffusion discretization can satisfy the min–max principle, even if grids are strongly distorted. Unfortunately, the nonisotropy of our problem prohibits their approach and we have to search for other remedies to the flaw found in classical oceanic rotated diffusion discretizations.

4. Remedies

We will not describe in detail all possible remedies but mention some possibilities that could be interesting for some specific models.

a. Variance diminishing methods

One possibility is not to impose the min–max principle as a mandatory condition for the numerical discretization, but a decreasing variance characteristic. This approach is used in Griffies et al. (1998), where a linear scheme was designed so as to decrease variances over finite volumes, which are not however the finite volumes used for the model computation. This method seems to work well for diffusion of active T, S fields at large scales, but fails to assure the min–max principle in the case of passive tracers in a fixed uniformly sloped density field. In this case, the numerical stencil (which has to be multiplied by time steps, grid spacing, and a slope factor, all of which are constant) of the diffusion term is given in Fig. 2 and shows that, in

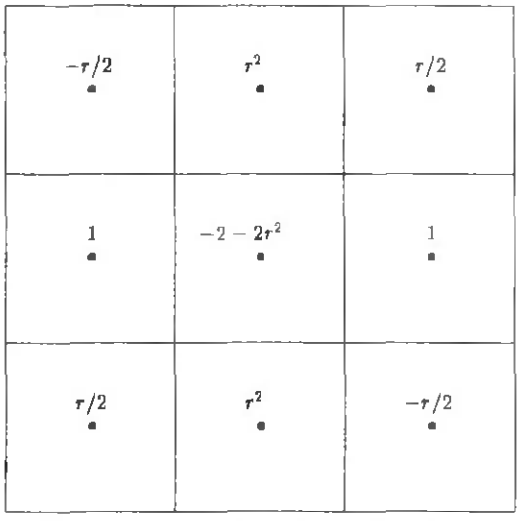

FIG. 2. Stencil for the new method of (Griffies et al. 1997) in the case of constant slopes and grid spacings.

this case, the discretization is just a straightforward linear centered finite differencing of the rotated operator. This, however, exhibits unsatisfactory results if we diffuse, for example, a tracer in a vertical plane that is zero everywhere except in the center. By running the so-obtained pure diffusion scheme for different values of the relative slope r, the global variance indeed continously diminishes, but there is an appearance of local minima, which after relatively few iterations already reach absolute values of 10% of the maximum signal of the tracer field at that moment. This is certainly undesirable in case of tracing CFCs or other variables in climate models or biological components in shelf models. Global variance diminishing methods are thus not the final solution for all applications. Our understanding is that the reason for the important improvement obtained by Griffies et al. (1998) of the original Cox method is mostly due to the balance of isoneutral diffusive fluxes of the active tracers. Indeed their experiments showed that balancing the fluxes had a major effect, while subsequent addition of the new triad diffusion scheme showed no significant differences in the switching experiment. Of course they showed that the triad method also improved the diffusion of a passive tracer in a fixed shortwave density field, but this is to be expected since the whole design of their scheme is based on a more appropriate averaging technique for slopes and tracer gardients, which occurs in such a situation as short waves. But this misses the point that we just showed: that in smooth density fields, passive tracers will not behave according to the min–max principle.

In contrast to the case of active tracers at large scales, which was shown by Griffies et al. (1998) to be well reproduced by their new method, other problems would arise in regional model where the geostrophic adjustment does not arrange the velocity field to be tangent to density surfaces. In those models, rotated diffusion of active tracers (for example, geopotential diffusion in σ-coordinate models) may cause other problems due to a possible amplification of perturbations by the feedback between density fields and advection of these perturbations in the direction of the pressure gradient rather then perpendicularly.

b. Min–max methods

In order to assure a min–max satisfying method, one of the requirements which we used in our demonstration needs to be eliminated: we assumed a nine-point stencil, a linear method (discretization not changing depending on the solution), and a consistent scheme. At least one of these conditions cannot be satisfied by a min–max satisfying algorithm. On the other hand, a conservative scheme would be more than appropriate for long-term climatic calculations and tracer dispersion. If we decide to sacrifice one of the conditions just mentioned, we would at least want to improve the scheme by designing a conservative method. For this purpose, the diffusion term can be reformulated as follows:

$$\mathcal{D} = \frac{\partial}{\partial s}(-\Phi), \tag{33}$$

where Φ is the diffusion flux

$$\Phi = -\mathcal{A}\frac{\partial \Psi}{\partial s}. \tag{34}$$

Equation (33) is readily written as

$$-\mathcal{J}\mathcal{D} = \frac{\partial}{\partial \xi}(\mathcal{J}a_s^{(\xi)}\Phi) + \frac{\partial}{\partial \eta}(\mathcal{J}a_s^{(\eta)}\Phi) \tag{35}$$

$$a_s^{(\xi)} = \frac{\partial \xi}{\partial s} \tag{36}$$

$$a_s^{(\eta)} = \frac{\partial \eta}{\partial s}. \tag{37}$$

We now have a formulation in which the derivatives are expressed in different coordinate systems, but it has the advantage that by using a classical integration over the finite volume box this mixed formulation can be translated into a conservative finite-difference scheme, provided that the quantities $\mathcal{J}\, a_s^{(*)}\Phi$ are known at the interfaces.

1) NONLINEAR FLUX COMPUTATION

An interesting possibility would be to use the approach of Stelling and Van Kester (1994). In their work, the authors tackle the problem of real horizontal diffusion in a σ-coordinate-like model. Their approach is based on a "back to z" paradigm, in the sense that the finite volumes are first rotated so as to have rectangular horizontal boxes. Then, since these boxes are not nicely connected to their neighbors, a z interpolation of scalars is needed to compute the fluxes at interfaces. For small

slopes this involves only the classical nine points and can be efficient but, when slopes are arbitrary, the interpolation method requires the scanning of the whole water column for each flux computation, a heavy burden in terms of CPU requirements. On the other hand, the authors prove their scheme to be min–max, in the case the fluxes are computed nonlinearly by choosing the minimal amplitude flux of two possible interpolations at the interface if they have the same sign or a zero flux otherwise. This consistent method is unfortunately time consuming in a general case. As the computational burden of a hopefully small effect ("horizontal" diffusion) should not penalize the whole ocean model, we could adopt the approach of Stelling and Van Kester (1994) if relative slopes are small (which could eventually be enforced by slope limiting in the code) or if we find another similar nonlinear interpolation method limited to the local stencil rather then the whole water column.

2) Nonconsistent diffusion

When a grid has regions for which $r \geq 1$ and one wants to avoid the previous method, a nonconsistent combination of diffusion along the grid lines can be envisaged: In a nonflux form this can be written as

$$\mathcal{D} = \alpha\mathcal{D}^{\prime} + \beta\mathcal{D}^{-} + \gamma\mathcal{D}^{\prime} + \delta\mathcal{D}^{\backslash} \tag{38}$$

$$\mathcal{D}^{\prime} = \Psi(0, 1) + \Psi(0, -1) - 2\Psi(0, 0) \tag{39}$$

$$\mathcal{D}^{-} = \Psi(1, 0) + \Psi(-1, 0) - 2\Psi(0, 0) \tag{40}$$

$$\mathcal{D}^{\prime} = \Psi(1, 1) + \Psi(-1, -1) - 2\Psi(0, 0) \tag{41}$$

$$\mathcal{D}^{\backslash} = \Psi(1, -1) + \Psi(-1, 1) - 2\Psi(0, 0). \tag{42}$$

This discretization is generally not consistent, but if the coefficients α, β, γ, δ are nonnegative, then the min–max principle is easily satisfied for small time steps. The choice of these coefficients can then at least be done so as to have a discretization that mimics at best the real diffusion. It can also be reformulated in a conservative fashion and the strategy would probably impose weightings that favor the grid direction closest to the s line.

3) Equivalent nonlinear diffusion along grid lines

Another possibility seems to be the use of a scheme as in Harvey (1995), but this is neither easily implemented into existing GCMs nor very efficient in terms of CPU resources since it is fully implicit. But Harvey (1995) leads to another way of thinking: One can indeed rewrite

$$\mathcal{J}\mathcal{D} = \frac{\partial}{\partial\xi}\left[\mathcal{J}\left(\frac{\partial\xi}{\partial s}\right)^2\mathcal{A}\left(1 - \frac{r}{R}\right)\frac{\partial\Psi}{\partial\xi}\right] + \frac{\partial}{\partial\eta}\left[\mathcal{J}\left(\frac{\partial\eta}{\partial s}\right)^2\mathcal{A}\left(1 - \frac{R}{r}\right)\frac{\partial\Psi}{\partial\eta}\right], \tag{43}$$

where

$$R = \frac{\partial\Psi/\partial\xi}{\partial\Psi/\partial\eta} \tag{44}$$

is the relative slope of the field to be diffused, compared to the aspect ratio of the numerical coordinates grid. The problem is thus formally equivalent to a diffusion along the grid lines with diffusion coefficients that depend on the solution and can be negative locally in time and space. A sufficient condition to assure a min–max principle is the use of a positive apparent diffusion coefficient along the grid lines. This is, of course, not necessary since the diffusion along s satisfies the min–max principle and Eq. (43) is just a reformulation of it, showing that "negative" or upgradient diffusion along the grid lines can be necessary. This also means that judging a scheme only by looking at local upgradient fluxes at an interface is not satisfactory since the physical fluxes may indeed be upgradient at a given interface. What really matters is the subsequent compensation of these upgradient fluxes by downgradient fluxes at other interfaces of the grid box. Downgradient fluxes at the interface are sufficient to assure a positive definite scheme, but they are not necessary. So limiting the apparent (solution dependent) diffusion coefficients at the interfaces to positive values is sufficient to satisfy the min–max principle if the time step is chosen short enough (otherwise one could also impose an upper limit on the apparent diffusion coefficient), but it is not necessary. This scheme is also conservative but not consistent when limitations are introduced. It should also be relatively easy to avoid the problems shown by Griffies et al. (1998) concerning the mismatch of computed neutral directions and the isolines of locally referenced potential density. Indeed, the values can be computed at the same point by the same stencil. There are thus numerous advantages, but unfortunately, when the s line and the solution have a slope of the same sign, imposing a local downgradient flux at the interface always requires a limiting of the apparent diffusion coefficient. In this case, when disregarding negative apparent diffusion coefficients, the scheme is not consistent.

Some other remedies could be searched, including flux limiter approaches or larger stencils. However, the local stencil used here limits the computational cost, and it is not expected that a linear model will behave better by using more points since the risk of introducing other negative coefficients increases. On the other hand, for nonlinear schemes larger stencils can eliminate the problem of strong slopes at the price of an increased complexity in the numerical algorithm and problems in designing vertically implicit schemes. Some kind of flux limiter scheme could, however, be interesting since it would modify the upgradient interface fluxes only if they are not compensated by downgradient fluxes.

5. Conclusions

Griffies et al. (1998) presented some important improvements to the discretization of the Cox rotated

diffusion. However, we argue that their improvements are mostly due to the balance of isoneutral diffusive fluxes of the active tracers, while subsequent addition of the new triad diffusion scheme improves the diffusion of a passive tracer in a fixed short-wave density field. But we showed that even in smooth density fields, passive tracers will not behave according to the min–max principle with their scheme. Our demonstration shows even that any linear consistent nine-point scheme cannot satisfy the min–max principle, which is rather annoying. When designing a rotated diffusion discretization, this has to be taken into account and, in our opinion, the operator should have the following properties:

- be conservative,
- satisfy the min–max principle or at least a reduced variance property,
- extend only over a nine-point stencil [for the (x, z) case],
- reduce to the classical horizontal stencil $1, -2, 1$ when no transformation is present,
- be computational efficient since diffusion should be small anyway and a scheme should not be penalized by a second-order term.

These requirements are somehow subjective and could seem easy to satisfy, but we showed that even a consistent nonconservative linear discretization on a nine-point stencil cannot meet the min–max requirement. In addition to the problems identified by Griffies et al., this is another explanation why classical rotated diffusion operators are used with additional filtering along the grid coordinates and/or time. It is also clear that, even if the new method improves drastically active tracer distributions, passive tracer simulation still ask for a more appropriate method. To cope with this problem, we proposed some partial remedies that hopefully can be tested and improved in different models and configurations. Finally, we expect even more severe problems in case of biharmonic diffusion formulations in rotated coordinate frames.

Acknowledgments. A model intercomparison was supported by the European Union (MEDMEX Contract MAS2-CT94-0107) and lead to some questions addressed in the present paper. J. M. Campin, E. Deleersnijder, and P. P. Mathieu work in the framework of Contract GC/DD/09A with Belgium's Federal Office for Scientific, Technical and Cultural Affairs and the Convention N 92/97-154 d'Actions de Recherches Concertées with the Communauté Francaise de Belgique.

The National Fund for Scientific Research (Belgium) has given J. M. Beckers and E. Deleersnijder the opportunity to be active in oceanographic research. P. P. Mathieu is a fellow of FRIA (Fonds pour la formation à la recherche dans l'industrie et l'agriculture).

Stephen Griffies, Anand Gnanadesikan, and John Dukowicz took part in very fruitful and animated discussions and helped to improve our manuscript. Encho Demirov made some helpful remarks.

REFERENCES

Burchard, H., and O. Petersen, 1997: Hybridization between σ and z coordinates for improving the internal pressure gradient calculation in marine models with steep bottom slopes. *Int. J. Numer. Methods Fluids,* **25,** 1003–1023.

Cox, M. D., 1987. Isopycnal diffusion in a z-coordinate ocean model. *Ocean Modelling* (unpublished manuscripts), **74,** 1–5.

Deleersnijder, E., and J.-M. Beckers, 1992: On the use of the σ-coordinate system in regions of large bathymetric variations. *J. Mar. Syst.,* **3,** 381–390.

Demirdžić, I., A. D. Gosman, R. I. Issa, and M. Peric, 1987: A calculation procedure for turbulent flow in complex geometries. *Comput. Fluids,* **15,** 251–273.

Gent, P. R., and J. C. McWilliams, 1990: Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.,* **20,** 150–155.

Gerdes, R., C. Köberle, and J. Willebrand, 1991: The influence of numerical advection schemes on the results of ocean general circulation models. *Climate Dyn.,* **5,** 211–226.

Griffies, S. M., 1998: The Gent–McWilliams skew flux. *J. Phys. Oceanogr.,* **28,** 831–841.

——, A. Gnanadesikan, R. Pacanowski, V. Larichev, J. Dukowicz, and R. Smith, 1998: Isoneutral diffusion in a z-coordinate ocean model. *J. Phys. Oceanogr.,* **28,** 805–830.

Haidvogel, D. B., J. L. Wilkin, and R. E. Young, 1991: A semi-spectral primitive equation ocean circulation model using vertical sigma and orthogonal curvilinear horizontal coordinates. *J. Comput. Phys.,* **94,** 151–185.

Haney, R. L., 1991: On the pressure gradient force over steep topography in sigma coordinate ocean models. *J. Phys. Oceanogr.,* **21,** 610–619.

Harvey, L. D. D., 1995: Impact of isopycnal diffusion on heat fluxes and the transient response of a two-dimensional ocean model. *J. Phys. Oceanogr.,* **25,** 2166–2176.

Hedström, K., 1994: *User's Manual for a Semi-Spectral Primitive Equation Ocean Circulation Model.* Institute of Marine and Coastal Sciences, Rutgers University, 130 pp.

——, 1996: SPEM 5.1 README file. Institute of Marine and Coastal Sciences, Rutgers University.

Jameson, A. R., 1995: A positive shock modelling for compressible flows. *Int. J. Numer. Methods Fluids,* **20,** 743–776.

Kasahara, A., 1974: Various vertical coordinate systems used for numerical weather prediction. *Mon. Wea. Rev.,* **102,** 509–522.

Kershaw, D., 1981: Differencing of the diffusion equation in Lagrangian hydrodynamical codes. *J. Comput. Phys.,* **39,** 375–395.

Mathieu, P. P., 1996: *Introduction of Isopycnal Mixing in a World Ocean Model.* Travail de DEA. University of Liège, 75 pp.

McDougall, T. J., and J. A. Church, 1986: Pitfalls with the numerical representation of isopycnal and diapycnal mixing. *J. Phys. Oceanogr.,* **16,** 196–199.

Mellor, G. L., T. Ezer, and L. Y. Oey, 1994: The pressure gradient conundrum of sigma coordinate ocean models. *J. Atmos. Oceanic Technol.,* **11,** 1126–1134.

Pacanowski, R. C., 1995: MOM2 Documentation. User's guide and reference manual version 1.0. GFDL Ocean Tech. Rep., 223 pp.

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

Stelling, G. S., and T. Van Kester, 1994: On the approximation of horizontal gradients is σ-coordinates for bathymetry with steep bottom slopes. *Int. J. Numer. Methods Fluids* **18,** 915–935.

Visbeck, M., J. Marshall, T. Haine, and M. Spall, 1997: Specification of eddy transfer coefficients in coarse-resolution ocean circulation models. *J. Phys. Oceanogr.,* **27,** 381–402.

Zijlema, M., 1996: *Computational Modeling of Turbulent Flow in General Domains.* Cip Data Koninklijke Bibliotheek. 185 pp.