INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2014; **75**:365–384
Published online 19 March 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/fld.3900

Discontinuous Galerkin finite element discretization of a strongly anisotropic diffusion operator

A. Pestiaux[1,*,†], S. A. Melchior[2], J. F. Remacle[2], T. Kärnä[3], T. Fichefet[1] and J. Lambrechts[2]

[1] *Earth and Life Institute (ELI), Georges Lemaître Centre for Earth and Climate Research (TECLIM), Université catholique de Louvain, Bte L4.03.08, place Louis Pasteur 3, B-1348 Louvain-la-Neuve, Belgium*
[2] *Institute of Mechanics, Materials and Civil Engineering (IMMC), Université catholique de Louvain, Bte L4.05.02, 4 avenue Georges Lemaître, B-1348 Louvain-la-Neuve, Belgium*
[3] *Science and Technology Center for Coastal Margin Observation & Prediction, Institute of Environmental Health, Oregon Health & Science University, 20000 NW Walker Road, Beaverton, OR 97006, USA*

SUMMARY

The discretization of a diffusion equation with a strong anisotropy by a discontinuous Galerkin finite element method is investigated. This diffusion term is implemented in the tracer equation of an ocean model, thanks to a symmetric tensor that is composed of diapycnal and isopycnal diffusions. The strong anisotropy comes from the difference of magnitude order between both diffusions. As the ocean model uses interior penalty terms to ensure numerical stability, a new penalty factor is required in order to correctly deal with the anisotropy of this diffusion. Two penalty factors from the literature are improved and established from the coercivity property. One of them takes into account the diffusion in the direction normal to the interface between the elements. After comparison, the latter is better because the spurious numerical diffusion is weaker than with the penalty factor proposed in the literature. It is computed with a transformed coordinate system in which the diffusivity tensor is diagonal, using its eigenvalue decomposition. Furthermore, this numerical scheme is validated with the method of manufactured solutions. It is finally applied to simulate the evolution of temperature and salinity due to turbulent processes in an idealized Arctic Ocean. Copyright © 2014 John Wiley & Sons, Ltd.

Received 15 August 2013; Revised 31 January 2014; Accepted 16 February 2014

KEY WORDS: advection–diffusion; discontinuous Galerkin; interior penalty factor

1. INTRODUCTION

In ocean general circulation models, all physical processes cannot be resolved explicitly because of insufficient spatial resolution. Hence, appropriate parameterizations are required in order to account for those processes. Iselin [1] and Montgomery [2] suggested that the mixing of tracers by mesoscale eddies in the stratified ocean mainly occurs along the surfaces of constant density, that is, the isopycnals. It appears that another diffusion, which is called diapycnal, also occurs orthogonally to the isopycnals, but its magnitude is much weaker. This situation creates a strong anisotropy in the diffusion tensor. In 1982, Redi introduced the isopycnal diffusivity tensor as a non-linear function of the active tracers (salinity and temperature) [3]. This operator differs fundamentally from isotropic and homogeneous diffusion because the tensor is not defined as diagonal or constant. But, as in most

[*] Correspondence to: A. Pestiaux, Université catholique de Louvain, Earth and Life Institute (ELI), Georges Lemaître Centre for Earth and Climate Research (TECLIM), Bte L4.03.08, place Louis Pasteur 3, B-1348 Louvain-la-Neuve, Belgium.

[†] E-mail: alice.pestiaux@uclouvain.be

ocean general circulation models used in climate studies, the main directions of the diffusion are not aligned with the mesh, and this can create numerical errors.

In the present work, only the tracer equation from an ocean model is considered. Even if many processes can influence the evolution of oceanic tracers, such as the advection or the vertical mixing, we focus only on the isopycnal diffusion, which is discretized with a discontinuous Galerkin finite element method (DGFEM) because it is developed in the framework of an unstructured grid oceanic model, the Second-generation Louvain-la-Neuve Ice-ocean Model (SLIM[‡], [4–6]). Even if advection is not present in this study, this is an important ocean process that cannot be forgotten for a realistic oceanic simulation. When the advection term is present, DGFEM is better adapted than the continuous Galerkin (CG) method because the numerical dissipation is lower than in CG for an equivalent mesh and the dispersion is optimal [7]. For the last 30 years, DGFEM has been used to solve partial differential equations in engineering applications, but the anisotropy of these models was much less than in the ocean [8, 9]. In a natural way, the numerical fluxes and the slope limiters were introduced [10]. The DGFEM allows to approximate the solution on each element separetely, and some discontinuities, called jumps, can appear at the interface of the elements [11]. For these many reasons, SLIM has been developed with DGFEM, and as this oceanic model is employed and improved, the DGFEM is used instead of the CG. In the framework of SLIM, interior penalty (IP) terms are introduced to yield a compact scheme. Especially, the estimation of the penalty factor is required to stabilize the finite element method.

In the ocean, the anisotropy is quantified, thanks to a factor ϵ, named *anisotropy factor*. Its high magnitude, which comes from the ratio between the maximum and minimum eigenvalues of the diffusivity tensor, λ_M and λ_m, respectively, is not usual in standard engineering analysis, such as in the composite materials or in petroleum geology [12]. The derivation of the penalty factor is not an easy task because it influences the results. If it is too small, the numerical scheme becomes unstable. But, if it is too large, too much numerical diffusion is introduced, and this reduces the quality of the approximate solution. Houston *et al.* [13] analyzed the discretization of the advection–diffusion equation with a discontinuous Galerkin method when the diffusivity is heterogeneous and less anisotropic than in the ocean. As the local and small diffusivity in some parts of the domain can influence the internal layers if there is advection, Gastaldi and Quarteroni [14], Croisille *et al.* [15], and Di Pietro *et al.* [16] investigated the regions where the diffusion vanishes and reappears further. The discontinuity-penalization parameter does not take into account the direction and is thus not appropriate when the diffusivity is anisotropic.

In her book [17], Rivière has proposed a DGFEM IP method that is able to deal with moderate anisotropic diffusion. In practical case, the mesh is usually aligned with the direction of anisotropy. Consider the Laplace problem

$$\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} = 0, \tag{1}$$

on a uniform mesh made of squares where C is a tracer (Figure 1). Consider a change in coordinate $y' \mapsto hy$, which leads to

$$\frac{\partial^2 C}{\partial x^2} + h^2 \frac{\partial^2 C}{\partial y'^2} = 0. \tag{2}$$

An anisotropy of h^2 can exactly be balanced using a mesh that is stretched by a factor h in the direction of anisotropy. One design goal of our approach would be that a numerical solution obtained for Equation (1) on a uniform mesh would be strictly the same as the numerical solution obtained for Equation (2) on a mesh that is stretched by a factor h in the y direction. Rivière's approach deals separately with the anisotropy of the diffusion tensor and with the anisotropy of the mesh. With the kind of anisotropy that is present in ocean modeling, penalty factors computed with Rivière's approach are very high. The corresponding linear systems are so ill conditioned that they cannot be inverted. Actually, the Rivière penalty factor demonstrates its effectiveness when the anisotropy is

[‡] http://www.climate.be/slim/.

DOI: 10.1002/fld

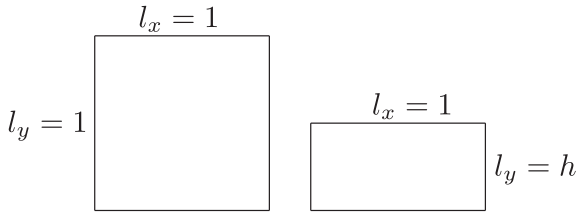


Figure 1. Illustration of two meshes with their respective side lengths l_x and l_y.

Table I. Comparison of an approximation of the penalty term for the Rivière, the eigenvalue, and the oriented methods.

	Rivière	Eigenvalue	Oriented
x	α^{-1}	α^{-1}	1
y	α^{-1}	α	α^{-1}

small and local. In this paper, we want to have a method as well accurate as the Rivière method but practicable with complex simulations. We therefore suggest an improvement to this penalty factor in order to reduce it while preserving the numerical stability. It will be referred to as the *eigenvalue penalty factor*.

Ern [18] suggested another penalty factor. He applied a weighted average method to the diffusivity tensor in the direction normal to the interface between the elements. The use of this factor without the average method is quite intuitive. There is, to the best of our knowledge, no formal demonstration of the use of the penalty factor suggested by Ern. In this paper, we will first prove that this penalty factor without the weighted average, which is called here the *oriented penalty factor*, is sufficient to ensure the coercivity even for strong anisotropic diffusions. The coercivity ensures that the solution is well posed; that is, the uniqueness and continuity properties are satisfied. The latter is defined by rotating the system to align it with the principal axes of the diffusivity tensor. Afterwards, the more appropriate penalty factor for a strong anisotropic diffusion will be determined between the eigenvalue penalty factor and the *oriented penalty factor*. That will allow to have not only the numerical scheme stabilization but also less numerical diffusion and thus a better approximation of the solution. If a simple example is taken where the anisotropic diffusion is defined as

$$\kappa = \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix},$$

where α is assumed to be smaller than 1, a first approximation of each penalty term can be theoretically computed (Table I). As expected, the Rivière penalty factor remains large whatever the direction. The eigenvalue penalty factor is a little better, whereas the oriented penalty factor that changes with the main axes diffusion seems the best.

The paper is organized as follows. The diffusion tensor is defined in Section 2. The DGFEM is applied to the diffusion equation in Section 3. Section 4 presents both penalty factors discussed here and compares them. The method of manufactured solutions (MMS) is applied in Section 5. A physical application is suggested in Section 6. Finally, concluding remarks close the paper in Section 7.

2. DIFFUSION EQUATION

In the framework of SLIM, the unstructured meshes are composed of several layers of triangular prisms. As the elements are triangular at the surface, the coastlines can be represented with high geometrical flexibility. Additionally, the mesh is structured along the vertical direction, which preserves the natural stratification of the ocean. Each tracer concentration field $C(\underline{X}, t)$, typically the temperature T or the salinity S, satisfies the following diffusion equation:

$$\frac{\partial C}{\partial t} = \nabla \cdot (\underline{\underline{\kappa}} \cdot \nabla C), \tag{3}$$

where $\underline{\underline{\kappa}}$ is the diffusivity tensor. This symmetric tensor is computed from the density ρ, which is itself a function of T and S through the equation of state. The density is a three-dimensional function implying that the dimension $d = 3$ and $\underline{X} = (x, y, z)$. Initial conditions $C(\underline{X}, 0)$ are supposed to be given, and no normal flux of both temperature and salinity is allowed on the boundary $\partial\Omega$ of the domain $\Omega \subset R^d$. The normal $\underline{n}$ is defined everywhere on the boundary of the elements. From the density ρ, the slope[§] $\tilde{s}$ is obtained as follows:

$$\tilde{s} = [s_x, s_y] = -\frac{\nabla_h \rho}{\partial\rho/\partial z}, \tag{4}$$

where $\nabla_h = (\partial_x, \partial_y)$. Let us now define $\underline{\nu} = -\nabla\rho$ and create the diapycnal unit vector:

$$\underline{\hat{\nu}} = \frac{\underline{\nu}}{\|\underline{\nu}\|} = \frac{[s_x, s_y, -1]}{\sqrt{s_x^2 + s_y^2 + 1}},$$

where $\|\cdot\|$ is the Euclidean norm. The $\tilde{s}$ and $\underline{\nu}$ vectors are important because the anisotropic diffusion in the ocean is oriented along and across the density slope. The diffusivity tensor $\underline{\underline{\kappa}}$, which is made up of isopycnal and diapycnal parts, can then be expressed as follows:

$$\underline{\underline{\kappa}} = A^I(\underline{\underline{\delta}} - \underline{\hat{\nu}}\underline{\hat{\nu}}) + A^D \underline{\hat{\nu}}\underline{\hat{\nu}},$$

where $\underline{\underline{\delta}}$ is the Kronecker delta and A^I and A^D are the non-negative isopycnal and diapycnal diffusivity coefficients, respectively [19]. Using the local density slope $\tilde{s}$, Redi [3] showed that the tensor $\underline{\underline{\kappa}}$ in the (x, y, z) reference frame can be written as follows[¶]:

$$\frac{A^I}{1 + \|\tilde{s}\|^2} \begin{pmatrix} 1 + s_y^2 + \epsilon s_x^2 & (\epsilon - 1)s_x s_y & (1-\epsilon)s_x \\ (\epsilon - 1)s_x s_y & 1 + s_x^2 + \epsilon s_y^2 & (1-\epsilon)s_y \\ (1-\epsilon)s_x & (1-\epsilon)s_y & \epsilon + \|\tilde{s}\|^2 \end{pmatrix}, \tag{5}$$

where $\epsilon = \frac{A^D}{A^I}$ is the ratio of the diapycnal diffusivity to the isopycnal diffusivity.

3. DGFEM FORMULATION

In this study, the elements Ω_e are prisms with vertical faces [20] and P_1^{DG} shape functions; that is, polynomials of degree $p = 1$ are chosen in each element (implying that the number of nodes in the element is $N = 6$). Note that the index e refers to a specific element, whereas the index k refers to an interface between two elements. The usual Galerkin finite element formulation of the tracer equation is as follows:

$$\int_\Omega \left(\frac{\partial C}{\partial t} - \nabla \cdot (\underline{\underline{\kappa}} \cdot \nabla C)\right) \tau d\Omega = 0,$$

$$\Leftrightarrow \int_\Omega \frac{\partial C}{\partial t} \tau d\Omega = \int_\Omega \nabla\tau \cdot \underline{\underline{\kappa}} \cdot \nabla C d\Omega - \int_{\partial\Omega} \underline{n} \cdot \underline{\underline{\kappa}} \cdot \nabla C \tau d\Gamma,$$

[§] The tilde refers to a two-dimensional vector, whereas the underline refers to a three-dimensional vector.

[¶] An equality between $\underline{\underline{\kappa}}$ and its components cannot be written because the former is a tensor, that is, a mathematical object that does not depend on any basis, while the latter is the matrix obtained by expressing the former in a particular basis. The formal link between the tensor and its components κ_{ij} is simply $\underline{\underline{\kappa}} = \kappa_{ij} e_i e_j$ where Einstein convention is used and e_i is the basis vector in direction $i = \{1, 2, 3\}$.

DOI: 10.1002/fld

where τ is the shape function. The integral over the whole domain Ω is decomposed into the sum of integrals over each element Ω_e and each interface $\gamma_k = \Omega_e \cap \Omega_{e'}$. The set of all element interfaces is noted $\Gamma = \bigcup_k \gamma_k$. The interface term is computed over each face:

$$\sum_e \int_{\Omega_e} \frac{\partial C}{\partial t} \tau d\Omega = \sum_e \int_{\Omega_e} \nabla\tau \cdot \underline{\underline{\kappa}} \cdot \nabla C d\Omega - \sum_k \int_{\gamma_k} \underline{n} \cdot \underline{\underline{\kappa}} \cdot \nabla C \tau d\Gamma. \tag{6}$$

Instead of incorporating boundary conditions in the space, Dirichlet boundary conditions are directly weakly imposed through the penalty factor [21]. In DGFEM, the weak formulation consists in finding C such as $a(C,\tau) = b(\tau)$, where a is a bilinear form and b is a linear form. The right-hand side term of Equation (6) is indeed the bilinear form [17], which is defined as follows:

$$\begin{aligned} a(C,\tau) \ = \ & \sum_e \int_{\Omega_e} \nabla\tau \cdot \underline{\underline{\kappa}} \cdot \nabla C \, d\Omega \\ & - \sum_k \int_{\gamma_k} \left([[\tau]] \cdot \{\underline{\underline{\kappa}} \cdot \nabla C\} + [[C]] \cdot \{\underline{\underline{\kappa}} \cdot \nabla \tau\} \right) + \sum_k \int_{\gamma_k} \mu [[C]] \cdot [[\tau]] \, d\Gamma \\ = \ & \underbrace{\sum_e \int_{\Omega_e} \nabla\tau \cdot \underline{\underline{\kappa}} \cdot \nabla C \, d\Omega - \sum_k \int_{\gamma_k} [[\tau]] \cdot \{\underline{\underline{\kappa}} \cdot \nabla C\} \, d\Gamma}_{\textcircled{1}} \\ & \underbrace{- \sum_k \int_{\gamma_k} [[C]] \cdot \{\underline{\underline{\kappa}} \cdot \nabla \tau\} \, d\Gamma}_{\textcircled{2}} + \underbrace{\sum_k \int_{\gamma_k} \mu [[C]] \cdot [[\tau]] \, d\Gamma}_{\textcircled{3}}, \end{aligned} \tag{7}$$

where μ is the penalty factor and $[[.]]$ is the jump vector at the interface such that $[[C]] = \underline{n} \frac{C^+ - C^-}{2}$, with C^+ and C^- being the tracer on the left-hand and right-hand sides, respectively [10]. The term ① comes from the divergence theorem and the integration by parts. The IP terms, that is, the symmetric IP term ② and the penalty term ③, stabilize the diffusion in the DGFEM. The value of μ must be chosen carefully. On the one hand, if μ is not large enough, the bilinear form is not coercive, and the approximate solution is not stable. In this case, numerical artifacts, such as spurious oscillations, deteriorate the quality of the solution appear. On the other hand, if μ is too large, the solution exhibits too much numerical diffusion and modifies the effective value of the diffusivity tensor. Moreover, the numerical schemes will not be efficient. For instance, a large value of μ can have a detrimental effect on the conditioning of the matrix that represents the bilinear form [22].

4. INTERIOR PENALTY FACTOR

The factor μ penalizes the jump of the concentration tracer C over the edge of an element. For scalar diffusivity, Shahbazi [22] derived a penalty factor that is a function of the dimension d, the degree of the polynomial shape function D_p, the area of the interface A, and the volume of the element V:

$$\mu = \frac{(D_p + 1)(D_p + d)}{d} \frac{n_0}{2} \frac{A}{V} \kappa, \tag{8}$$

where n_0 is the number of neighbors of the element, that is, $n_0 = 5$ for prisms, and κ is a scalar diffusivity. As the diffusion is represented by an anisotropic diffusivity tensor, this usual penalty factor cannot be used directly.

In the next sections, two different ways of computing the penalty factor, which take into account the anisotropic diffusivity, are discussed. First, Rivière [17] only used the lower and upper bounds of the eigenvalues of the tensor. As this suggested penalty factor is too large, it is improved and called *eigenvalue penalty factor*. Hence, the maximum eigenvalue is replaced in order to reduce

DOI: 10.1002/fld

the numerical diffusion, which leads thus to a better performance. But the tries with the eigenvalue penalty factor were not convincing. Second, the proof of the oriented penalty factor, which is a function of $\underline{n} \cdot \underline{\underline{\kappa}} \cdot \underline{n}$, is introduced. This factor is defined by rotating the system to align it with the principal axes of the diffusivity tensor, and a value is suggested when strong anisotropy exists in the diffusivity tensor.

4.1. Eigenvalue penalty factor

In the case of a small anisotropic diffusivity, Rivière [17] suggested to replace the penalty factor (8) by the penalty factor μ defined with the eigenvalues of the diffusivity tensor:

$$\mu = \frac{(D_p+1)(D_p+d)}{d}\frac{n_0}{2}\frac{A}{V}\frac{\lambda_M^2}{\lambda_m}, \tag{9}$$

where λ_m and λ_M are the minimum and maximum eigenvalues of the diffusivity tensor, respectively. Even though the anisotropy in the new diffusivity tensor is considered, this penalty factor returns excessive values whenever $\frac{\lambda_M^2}{\lambda_m}$ is large.

It is possible to find sharper bounds for μ when anisotropy is very large. The penalty factor must be chosen so that the bilinear form $a(C,\tau)$ is coercive; that is, there exists a positive constant c_1 such that

$$\begin{aligned} c_1\|C\|_S^2 &\leqslant a(C,C), \\ &\leqslant \sum_e \int_{\Omega_e} \nabla C \cdot \underline{\underline{\kappa}} \cdot \nabla C d\Omega - 2\sum_k \int_{\gamma_k} \{\underline{n} \cdot \underline{\underline{\kappa}} \cdot \nabla C\}[[C]] d\Gamma, \\ &\quad + \sum_k \int_{\gamma_k} \mu [[C]]^2 d\Gamma, \end{aligned} \tag{10}$$

where τ has been replaced by C in Equation (7) and the norm associated to the broken Sobolev space is $\|C\|_S^2 = \sum_e \int_{\Omega_e} \|\nabla C\|^2 d\Omega + \sum_k \int_{\gamma_k} [[C]]^2 d\Gamma$. As the aim is to ensure the coercivity, $a(C,C)$ must be limited by something that is smaller. Using the arithmetic–geometric mean inequality $-2\alpha\beta \geqslant -\epsilon_Y^{-1}\beta^2 - \alpha^2\epsilon_Y$ with the strictly positive scalar ϵ_Y, the equation becomes

$$\begin{aligned} a(C,C) \geqslant &\sum_e \int_{\Omega_e} \nabla C \cdot \underline{\underline{\kappa}} \cdot \nabla C d\Omega - \frac{1}{\epsilon_Y}\sum_k \int_{\gamma_k} \{\underline{n} \cdot \underline{\underline{\kappa}} \cdot \nabla C\}^2 d\Gamma, \\ &+ \sum_k \int_{\gamma_k} (\mu - \epsilon_Y)[[C]]^2 d\Gamma, \end{aligned} \tag{11}$$

where α has been replaced by $[[C]]$ and β by $\{\underline{n} \cdot \underline{\underline{\kappa}} \cdot \nabla C\}$. Using the geometric law $(m^+ + m^-)^2 \leqslant 2(m^+)^2 + 2(m^-)^2$, the second term can be bounded as follows:

$$\sum_k \int_{\gamma_k} \{\underline{n} \cdot \underline{\underline{\kappa}} \cdot \nabla C\}^2 d\Gamma \leqslant \frac{1}{2}\sum_k \int_{\gamma_k} \left(\left(\underline{n} \cdot \left(\underline{\underline{\kappa}} \cdot \nabla C\right)^-\right)^2 + \left(\underline{n} \cdot \left(\underline{\underline{\kappa}} \cdot \nabla C\right)^+\right)^2 \right) d\Gamma,$$

where the signs $(\cdot)^+$ and $(\cdot)^-$ refer, respectively, to the values of the variable on the left-hand and right-hand sides of the interface. In order to bound the diffusivity tensor, $\underline{\underline{\kappa}}$ is considered constant on each element so that

$$\begin{aligned} \sum_k \int_{\gamma_k} \{\underline{n} \cdot \underline{\underline{\kappa}} \cdot \nabla C\}^2 d\Gamma \leqslant \frac{1}{2}\sum_k \Bigg(&||\underline{n} \cdot \underline{\underline{\kappa}}^-||^2 \int_{\gamma_k} ||\nabla C^-||^2 d\Gamma, \\ &+ ||\underline{n} \cdot \underline{\underline{\kappa}}^+||^2 \int_{\gamma_k} ||\nabla C^+||^2 d\Gamma \Bigg). \end{aligned} \tag{12}$$

The trace inequality [23]

$$\forall \gamma_k \in \Omega_e \int_{\gamma_k} P^2 d\Gamma \leqslant \int_{\Omega_e} c_s \frac{A_k}{V_e} P^2 d\Omega, \tag{13}$$

where $c_s = \frac{(O_p+1)(O_p+d)}{d}$ and O_p is the number of degrees of freedom of the polynomial P, is now used. Because O_p is related to the gradient of the tracer concentration, it is equal to $D_p - 1$, and thus, $c_s = \frac{D_p(D_p-1+d)}{d}$. The inequality (12) can then be written as follows:

$$\sum_k \int_{\gamma_k} \{\underline{n} \cdot \underline{\underline{\kappa}} \cdot \nabla C\}^2 d\Gamma \leqslant \frac{c_s}{2} \sum_e \left(\sum_{k \in e} \frac{A_k}{V_e} ||\underline{n} \cdot \underline{\underline{\kappa}}||^2 \right) \int_{\Omega_e} ||\nabla C||^2 d\Omega. \tag{14}$$

Moreover, the first part of the inequality (10) can be bounded as follows:

$$\int_{\Omega_e} \nabla C \cdot \underline{\underline{\kappa}} \cdot \nabla C d\Omega \geqslant \int_{\Omega_e} \lambda_m ||\nabla C||^2 d\Omega. \tag{15}$$

Using the inequalities (14) and (15), the bilinear form can be written as follows:

$$\begin{aligned} a(C,C) \geqslant & \sum_e \left(\lambda_m - \frac{c_s}{2\epsilon_Y V_e} \sum_{k \in e} A_k ||\underline{n} \cdot \underline{\underline{\kappa}}||^2 \right) \int_{\Omega_e} ||\nabla C||^2 d\Omega, \\ & + \sum_k (\mu - \epsilon_Y) \int_{\gamma_k} [[C]]^2 d\Gamma. \end{aligned} \tag{16}$$

In order to ensure the coercivity $a(C,C) > c_1 ||C||_S^2$, two conditions are required:

$$\begin{cases} \mu - \epsilon_Y = c_1 > 0, \\ \lambda_m - \frac{c_s}{2\epsilon_Y V_e} \sum_{k \in e} A_k ||\underline{n} \cdot \underline{\underline{\kappa}}||^2 \geqslant 0. \end{cases}$$

These conditions are satisfied if μ is chosen such as follows:

$$\mu > \epsilon_Y \geqslant \frac{c_s}{2\lambda_m V_e} \sum_{k \in e} A_k ||\underline{n} \cdot \underline{\underline{\kappa}}||^2.$$

In order to correctly understand the mean of this new penalty factor, an idealized case is considered. On the one hand, the mesh is supposed to be aligned with the axes, and the horizontal faces are larger than the vertical ones. On the other hand, the vertical diffusivity is chosen smaller than the horizontal one so that $A_k ||\underline{n} \cdot \underline{\underline{\kappa}}||^2$ is constant. Ideally, the same penalty factor must be used on both kinds of face. For the horizontal faces, the minimum eigenvalue λ_m is introduced:

$$\begin{aligned} \mu_H &= \frac{c_s}{2\lambda_m V_e} n_0 A_H \lambda_m^2, \\ &= \frac{c_s}{2V_e} n_0 A_H \lambda_m. \end{aligned} \tag{17}$$

This expression reveals that this penalty factor will introduce less numerical diffusion because the ratio $\frac{\lambda_M^2}{\lambda_m}$ disappears. Thus, it will be well adapted to the anisotropic situations. Besides, for the vertical faces, the penalty factor is as follows:

DOI: 10.1002/fld

$$\mu_V = \frac{c_s}{2\lambda_m V_e} n_0 A_V \lambda_M^2,$$

which exactly corresponds to the penalty factor suggested by Rivière (Equation (9)). In this last case, the factor will still return excessive values, which will lead to too much numerical diffusion.

4.2. Oriented penalty factor

As the factor μ penalizes the jump of the concentration tracer C over the edge of an element, a natural approach to estimate the penalty factor for an anisotropic diffusivity tensor is to consider its normal component on both sides of the interface of the elements [18]. In order to build such a penalizing term, the coordinate system is aligned with the principal axes of the diffusivity tensor, and Ω is expressed in another reference frame Ω'. So that the coercivity criterion is satisfied, Equation (11) is also used:

$$\begin{aligned} a(C,C) \geqslant & \sum_e \int_{\Omega_e} \nabla C \cdot \underline{\underline{\kappa}} \cdot \nabla C d\Omega - \frac{1}{\epsilon_Y} \sum_k \int_{\gamma_k} \{\underline{n} \cdot \underline{\underline{\kappa}} \cdot \nabla C\}^2 d\Gamma, \\ & + \sum_k \int_{\gamma_k} (\mu - \epsilon_Y)[[C]]^2 d\Gamma. \end{aligned}$$

With the geometric law $2(m^+)^2 + 2(m^-)^2 \geqslant (m^+ + m^-)^2$, the integral in the second term of the right-hand side can be bounded as follows:

$$-\int_{\gamma_k} \{\underline{n} \cdot \underline{\underline{\kappa}} \cdot \nabla C\}^2 d\Gamma \geqslant -\int_{\gamma_k} \{(\underline{n} \cdot \underline{\underline{\kappa}} \cdot \nabla C)^2\} d\Gamma,$$

and the inequality becomes

$$\begin{aligned} a(C,C) \geqslant & \sum_e \int_{\Omega_e} \nabla C \cdot \underline{\underline{\kappa}} \cdot \nabla C d\Omega - \frac{1}{\epsilon_Y} \sum_k \int_{\gamma_k} \{(\underline{n} \cdot \underline{\underline{\kappa}} \cdot \nabla C)^2\} d\Gamma, \\ & + \sum_k \int_{\gamma_k} (\mu - \epsilon_Y)[[C]]^2 d\Gamma. \end{aligned} \tag{18}$$

Because the diffusivity tensor in the first term in the right-hand side is symmetric positive definite, it can be diagonalized as follows:

$$\int_{\Omega_e} \nabla C \cdot \underline{\underline{\kappa}} \cdot \nabla C d\Omega = \int_{\Omega_e} \nabla C \cdot \underline{\underline{U}} \cdot \underline{\underline{\lambda}}^{1/2} \cdot \underline{\underline{\lambda}}^{1/2} \cdot \underline{\underline{U}} \cdot \nabla C d\Omega, \tag{19}$$

where the unit tensor $\underline{\underline{U}}$ describes the rotation that aligns the reference frame with the eigenvectors and $\underline{\underline{\lambda}}$ is the diagonal tensor scaled by the corresponding eigenvalues $\lambda_i (i = 1, 2, 3)$. The other terms of the equation can be transformed accordingly:

$$\begin{cases} \nabla' C = \nabla C \cdot \underline{\underline{U}} \cdot \underline{\underline{\lambda}}^{1/2}, \\ \underline{n}' = \dfrac{\underline{\underline{\lambda}}^{1/2} \cdot \underline{\underline{U}} \cdot \underline{n}}{||\underline{\underline{\lambda}}^{1/2} \cdot \underline{\underline{U}} \cdot \underline{n}||}, \\ J' = \sqrt{\lambda_1 \lambda_2 \lambda_3}. \end{cases}$$

Note that the symbol $'$ indicates that the variable is expressed in the modified coordinate system. With this frame change, Equation (19) can thus be expressed as follows:

$$\int_{\Omega_e} \nabla C \cdot \underline{\underline{\kappa}} \cdot \nabla C d\Omega = J' \int_{\Omega_e'} \nabla' C \cdot \nabla' C d\Omega'.$$

Using the trace inequality (13) in the reference frame Ω', this equation can be bounded as follows:

$$\sum_e \int_{\Omega_e} \nabla C \cdot \underline{\underline{\kappa}} \cdot \nabla C d\Omega \geqslant \sum_k \frac{J' V'_e}{c_s A'_k} \int_{\gamma'_k} (\nabla' C \cdot \underline{n}')^2 d\Gamma', \tag{20}$$

where V'_e is the element volume and A'_k is the face surface in the reference frame Ω'. Some terms can be expressed in the initial coordinate system:

$$\begin{cases} V'_e J' = V_e, \\ \frac{1}{A_k} \int_{\gamma_k} \cdot d\Gamma = \frac{1}{A'_k S'} S' \int_{\gamma'_k} \cdot d\Gamma' \text{ with } S' = \frac{d\Gamma}{d\Gamma'}, \end{cases}$$

so that the inequality (20) becomes

$$\sum_e \int_{\Omega_e} \nabla C \cdot \underline{\underline{\kappa}} \cdot \nabla C d\Omega \geqslant \sum_k \frac{V_e}{c_s A_k} \int_{\gamma_k} (\nabla' C \cdot \underline{n}')^2 d\Gamma.$$

Eventually, the squared term can be rotated in the usual reference frame Ω so that

$$\begin{aligned} (\nabla' C \cdot \underline{n}')^2 &= \left(\nabla C \cdot \underline{\underline{U}} \cdot \underline{\underline{\lambda}}^{1/2} \cdot \frac{\underline{\underline{U}} \cdot \underline{\underline{\lambda}}^{1/2} \cdot \underline{n}}{||\underline{\underline{\lambda}}^{1/2} \cdot \underline{\underline{U}} \cdot \underline{n}||} \right)^2, \\ &= \frac{(\nabla C \cdot \underline{\underline{\kappa}} \cdot \underline{n})^2}{||\underline{\underline{\lambda}}^{1/2} \cdot \underline{\underline{U}} \cdot \underline{n}||^2}, \\ &= \frac{(\nabla C \cdot \underline{\underline{\kappa}} \cdot \underline{n})^2}{(\underline{n} \cdot \underline{\underline{\kappa}} \cdot \underline{n})}. \end{aligned}$$

With this formulation, the bilinear form is bounded as follows:

$$\begin{aligned} a(C, C) \geqslant & \sum_k \int_{\gamma_k} \frac{2}{n_0 c_s A_k} \left\{ V_e \frac{(\nabla C \cdot \underline{\underline{\kappa}} \cdot \underline{n})^2}{(\underline{n} \cdot \underline{\underline{\kappa}} \cdot \underline{n})} \right\} d\Gamma, \\ & - \frac{1}{\epsilon_Y} \sum_k \int_{\gamma_k} \left\{ (\nabla C \cdot \underline{\underline{\kappa}} \cdot \underline{n})^2 \right\} d\Gamma + \sum_k (\mu - \epsilon_Y) \int_{\gamma_k} [[C]]^2 d\Gamma, \\ \geqslant & \sum_k \int_{\gamma_k} \left\{ \left(\frac{2V_e}{n_0 c_s A_k} \frac{1}{(\underline{n} \cdot \underline{\underline{\kappa}} \cdot \underline{n})} - \frac{1}{\epsilon_Y} \right) (\nabla C \cdot \underline{\underline{\kappa}} \cdot \underline{n})^2 \right\} d\Gamma, \\ & + \sum_k (\mu - \epsilon_Y) \int_{\gamma_k} [[C]]^2 d\Gamma. \end{aligned}$$

In order to ensure the coercivity $a(C, C) > c_1 ||C||_S^2$, two conditions are required:

$$\begin{cases} \mu - \epsilon_Y = c_1 > 0, \\ \frac{2V_e}{n_0 c_s A_k} \frac{1}{(\underline{n} \cdot \underline{\underline{\kappa}} \cdot \underline{n})} - \frac{1}{\epsilon_Y} \geqslant 0. \end{cases}$$

These conditions are satisfied if μ is chosen such as

$$\mu > \epsilon_Y > \frac{A_k c_s n_0}{2V_e} \underline{n} \cdot \underline{\underline{\kappa}} \cdot \underline{n},$$

and this corresponds to the oriented penalty factor. In the same way as for the eigenvalue factor, both kinds of face and then diffusion are studied. For the large horizontal surfaces of the element, and thus smaller diffusion, the penalty factor can be written as follows:

$$\mu_H = \frac{c_s n_0}{2} \frac{A_H}{V_e} \lambda_m,$$

which matches the eigenvalue factor for the same case (Equation (17)). For the vertical faces, the oriented factor becomes

$$\mu_V = \frac{c_s n_0}{2} \frac{A_V}{V_e} \lambda_M.$$

In this case, the ratio $\frac{\lambda_M^2}{\lambda_m}$ also disappears. The oriented penalty factor seems the most appropriate because it will introduce less numerical diffusion. It will now be compared numerically with the eigenvalue penalty factor.

4.3. *Eigenvalue penalty factor versus oriented penalty factor*

In this section, both penalty factors are compared in an oceanic simulation using SLIM and an unstructured mesh. The aim of this experimentation is, on the one hand, to illustrate the effects of the penalty factors on the numerical solution and, on the other hand, to intuitively understand their differences. A square mesh of 100-km side is considered with 50 vertical layers on a total depth of 200 m. To compare the simulations with an analytic solution [24], the isopycnals are supposed to be plane and equally spaced. Hence, the diapycnal vector $\underline{\nu}$ is homogeneous. Then, the isopycnal tensor is constant, and an analytic solution of this boundary value problem can be found:

$$C^h(\underline{X}, t) = \frac{\exp\left(-\frac{\underline{X} \cdot \underline{\underline{\kappa}}^{-1} \cdot \underline{X}}{4t}\right)}{(4\pi t)^{\frac{3}{2}} \sqrt{\det(\underline{\underline{\kappa}})}} \forall t > 0,$$

where $A^I = 1000$ m^2/s in the tensor $\underline{\underline{\kappa}}$ of Equation (5). The analytic concentration field at $t = 1$ day is used as initial condition $C(\underline{X}, 0)$ to replace the delta Dirac function because this function cannot be computed numerically. The tracer only undergoes isopycnal diffusion, and a diagonally implicit Runge–Kutta semi-implicit time integration is chosen [25]. With the eigenvalue penalty factor, the result appears smoothed on the left-hand side of the Figure 2, which is a vertical cross section of the tracer field after $20\Delta t$ of 1000 s, and no strong jump is observed. But even with the new formulation of the penalty factor of Rivière $\mu \sim \sum_{k \in e} A_k ||\underline{n} \cdot \underline{\underline{\kappa}}||^2$, the eigenvalue penalty factor is still too large and induces too much numerical diffusion. Indeed, with the anisotropy of both mesh and diffusivity, it varies in the range of $[2 \cdot 10^4, 10^{10}]$ m/s. In the case of the oriented penalty factor,

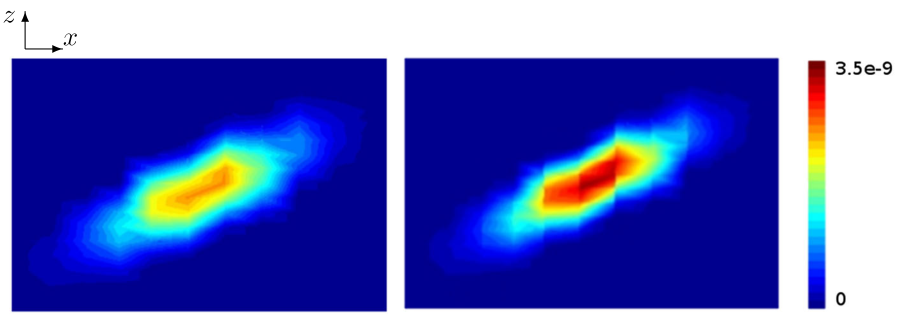


Figure 2. Vertical cross section of a Gaussian tracer field $C_r(\underline{X}, t = 20\Delta t)$ with a time step Δt of 1000 s when the eigenvalue (left-hand side) and oriented (right-hand side) penalty factors are used.

DOI: 10.1002/fld

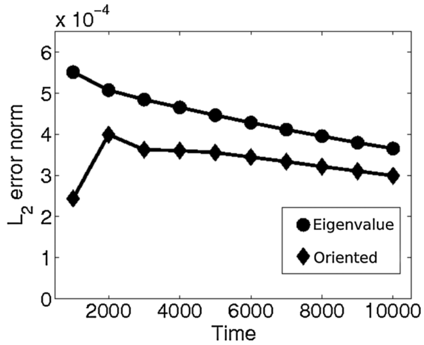


Figure 3. Study of the $\mathcal{L}_2$ error for the eigenvalue penalty factor and the oriented one.

the value of the oriented penalty factor is much smaller than previously, that is, in the range of $[2{\cdot}10^{-3}, 10^3]$ m/s. Nevertheless, the numerical solution on the right-hand side of the Figure 2 reveals large jumps. Even though the numerical solution obtained with the eigenvalue penalty factor looks smoother, the numerical error resulting from this approach is larger than the error made using the oriented penalty factor. Indeed, too much numerical diffusion in this scheme is maybe induced and that could distort the solution. When both figures are compared at the same time step, $C_o(\underline{X}, t)$ has been less diffused because it has larger values than $C_r(\underline{X}, t)$. In order to choose the better penalty factor, the $\mathcal{L}_2$ error, which is defined as $\|C^h - C\|^2_{\mathcal{L}_2} = \int_\Omega (C^h(\underline{X}, t_1) - C(\underline{X}, t_1))^2 d\Omega$ where t_1 refers to the time evolution, is computed for each penalty factor. Figure 3 shows that the $\mathcal{L}_2$ error of the oriented penalty is lower than the error of the eigenvalue penalty factor. That means that the oriented factor, which stabilizes enough the numerical scheme, does not induce too much diffusion, contrary to the eigenvalue penalty factor. This leads us to choose the oriented penalty factor and to pursue the numerical analysis with this one. In large-scale oceanic models, the minimum–maximum principle is often violated [26, 27]. Hence, the tracer concentrations can be negative, which leads to a local undershooting of this tracer. This situation produces unphysical water masses that can be transported and diffused in the world ocean. The monoticity could be discussed as in Mathieu *et al.* (1998) [28] where undershoots/overshoots are highlighted. But, as the diffusion is strongly anisotropic, only one property on the coercivity and the moniticity can be satisfied. Considering that the observed undershoots in the oceanic simulations are quite small, that is, around 10^{-9} under zero, the coercivity property is selected.

5. COMPARISON WITH THE METHOD OF MANUFACTURED SOLUTIONS

In order to know if the oriented penalty factor is well suited numerically, the spatial convergence is investigated. Specifically, the MMS allows to verify the code accuracy. A source term is added to the equation such that the analytic solution is known but non-trivial. Here, a simple anisotropic bidimensional diffusion equation is considered:

$$F(C) = \gamma \frac{\partial^2 C}{\partial x^2} + \alpha \frac{\partial^2 C}{\partial y^2} = 0, \tag{21}$$

where the constant diffusion $\alpha = 10^{-3}$ m^2/s and $\gamma = 1$ m^2/s so that the anisotropy of the diffusion $\epsilon = 1000$. First, an arbitrary manufactured solution is chosen as follows:

$$C_M(\underline{X}) = \frac{\exp\left(-\frac{\beta}{\tau}\left[\frac{x^2}{\gamma} + \frac{y^2}{\alpha}\right]\right)}{4\pi\tau\sqrt{\alpha\gamma}},$$

where the constant $\beta = 10^{-8}$ and the fictional time $\tau = 5$ s. Then, we add $F(C_M) = S$ as a source term of Equation (21) and C_M as boundary condition:

$$
\begin{aligned}
F(C) &= S \quad \underline{X} \in \Omega,\\
C &= C_M \ \underline{X} \in \partial\Omega.
\end{aligned}
$$

By construction, the analytic solution of this problem is $C = C_M$. The error of the numerical solution is an indicator of the quality of the numerical method and allows to estimate the penalty factors' performance. The domain geometry is a square of 100-km side. Several meshes generated with the GMSH software [29, 30] are considered to study the spatial convergence. They are composed of quadrilateral elements with side lengths l_x and l_y linked by this relation: $l_y = \sqrt{\gamma/\alpha} l_x$, so that in the space $x' = x, y' = \sqrt{\gamma/\alpha} y$, the diffusivity tensor, the solution, and the mesh are isotropic. Next, the domain is rotated in order to slightly misalign the elements and the main diffusion axes that stay along the coordinate axes. The rotation angle ω is taken as 0°, 0.5°, and 1° because the oceanic density slope does not exceed 0.01.

In the first phase, the spatial convergence is computed with the norm of the $\mathcal{L}_2$ error defined as follows:

$$
\|C^h - C_M\|^2_{\mathcal{L}_2} = \int_\Omega (C^h(\underline{X}) - C_M(\underline{X}))^2 d\Omega.
$$

The following penalty factors are studied:

1. Rivière : $\mu_R = c_s \max\left(\frac{\sum_{k\in e} A_k}{2V_e}\right)\frac{\lambda_M^2}{\lambda_m}$,
2. eigenvalue : $\mu_E = c_s \max\left(\frac{\sum_{k\in e} A_k \|\underline{n}\cdot\underline{\underline{\kappa}}\|^2}{2V_e}\right)\frac{1}{\lambda_m}$,
3. oriented : $\mu_O = c_s \frac{n_0 A_k}{2\min(V_e)}\underline{n}\cdot\underline{\underline{\kappa}}\cdot\underline{n}$,

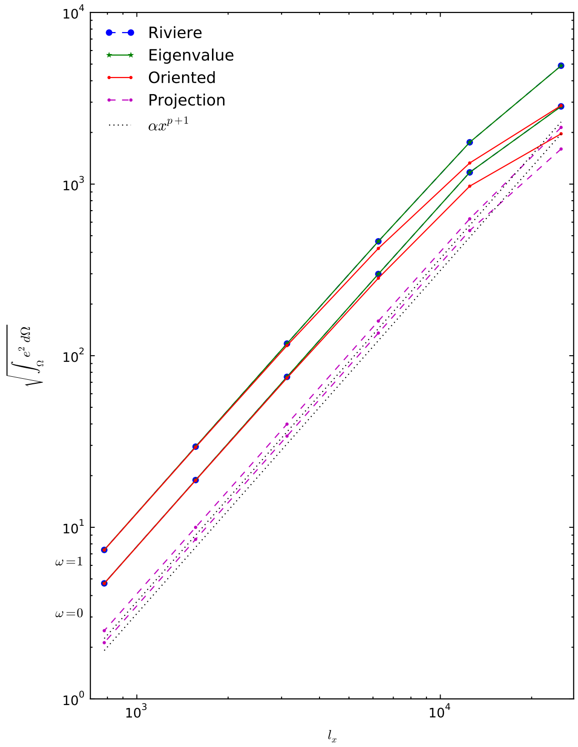


Figure 4. Comparison of the norms of the $\mathcal{L}_2$ error when the eigenvalue, the oriented, and the Rivière penalty factors are used for the method of manufactured solutions at the order 1 and for two rotation angles $\omega = 0 - 1°$. In both cases, the minimum norm of the $\mathcal{L}_2$ error shows that the oriented penalty factor is more appropriate, especially at coarse resolution.

DOI: 10.1002/fld

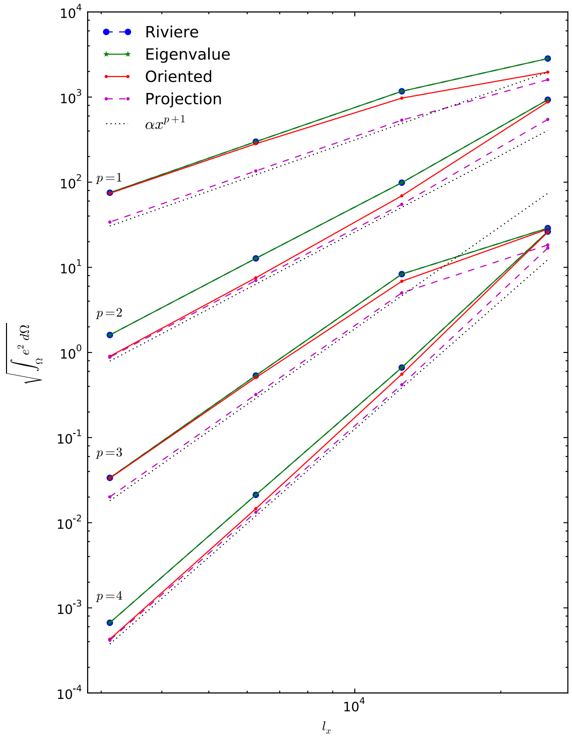


Figure 5. Comparison of the norms of the $\mathcal{L}_2$ error when the eigenvalue, the oriented, and the Rivière penalty factors are used for the method of manufactured solutions without any rotation and for four orders. At the even orders, the oriented penalty factor converges faster.

and their spatial convergence is illustrated in Figure 4 for the order $p = 1$ and the rotation angles $\omega = 0°, 1°$. As expected, the $\mathcal{L}_2$ errors are weaker when there is no misalignement between the diffusion and the mesh. At coarse resolutions, the oriented penalty factor has a smaller error than the other factors, whereas they converge in the same way at the finest resolutions. Besides, the lines for the Rivière and eigenvalue penalty factors cannot be distinguished.

In Figure 5, the spatial convergence is illustrated when there is no rotation angle and the orders 1 to 4 are considered. The lines for the Rivière and eigenvalue penalty factors still cannot be distinguished. At coarse resolution, the error when the oriented penalty factor is used is smaller for the odd orders, whereas it is comparable with the other penalty factor for the even orders. But at fine resolution, the error for the oriented penalty factor is lower. In all cases, there is no major difference in the errors obtained by the different penalty factors.

However, the main reason for this investigation was not the accuracy of the solution but the inability to solve iteratively the linear system arising from the Rivière approach. In reality, we expect to have a well-conditioned system when the penalty factor is small. For a linear system $Ax = B$, the condition number defined as $\eta = ||A|| \cdot ||A^{-1}||$ allows to give a measure of the accuracy of the system. If the matrix is symmetric, $\eta = \frac{\sigma_M}{\sigma_m}$, where σ_m and σ_M are the minimum and maximum eigenvalues of the system matrix A, respectively. Indeed, the convergence of the iterative methods depends on the cluster of the eigenvalues of the system. The more η is closed to 1, the more the system is well conditioned and thus easy and faster to solve. Moreover, the square root of η gives the number of iterations required to solve the system.

The three penalty factors have been studied at the order $p = 1$ and for the rotation angles $\omega = 0y, 0.5°, 1°$. Table II gives the eigenvalues of the system matrix and the condition number η for each case. When the Rivière penalty factor is used, σ_m cannot be found because the system cannot converge. In order to point out this state, the symbol † has been used, and σ_m has been replaced by the value computed with the oriented method because it does not vary with the method or the rotation angle. But it decreases with the resolution. Moreover, σ_M is independent of the mesh resolution,

DOI: 10.1002/fld

Table II. Study of the condition number η at the order 1 for each penalty factor and three rotation angles ($\omega = 0°, 0.5°, 1°$).

		$\omega = 0°$			$\omega = 0.5°$			$\omega = 1°$		
$l_x[m]$		O	E	R	O	E	R	O	E	R
25,000	σ_m	0.0041	0.0045	†	0.0041	0.0045	†	0.0041	0.0045	†
	σ_M	0.18	1500	52,000	0.19	1500	52,000	0.22	1500	52,000
	η	44	3.4e5	1.3e7	46	3.4e5	1.3e7	53	3.4e5	1.3e7
12,500	σ_m	0.0011	0.0012	†	0.0011	0.0012	†	0.0011	0.0012	†
	σ_M	0.19	1500	52,000	0.20	1500	52,000	0.22	1500	52,000
	η	170	1.3e6	4.7e7	180	1.3e6	4.7e7	200	1.3e6	4.7e7
6250	σ_m	0.00029	†	†	0.00029	†	†	0.00029	†	†
	σ_M	0.19	1500	52,000	0.19	1500	52,000	0.23	1500	52,000
	η	650	5.2e6	1.7e8	680	5.2e6	1.7e8	770	5.2e6	1.7e8
3125	σ_m	7.4e−5	†	†	7.4e−5	†	†	7.4e-5	†	†
	σ_M	0.19	1500	52,000	0.20	1500	52,000	0.23	1500	52,000
	η	2600	2.0e7	7.0e8	2700	2.0e7	7.0e8	3100	2.0e7	7.0e8
1562	σ_m	1.8e−5	†	†	1.8e−5	†	†	1.8e−5	†	†
	σ_M	0.19	1500	52,000	0.20	1500	52,000	0.23	1500	52,000
	η	1.0e4	8.3e7	2.9e9	1.1e4	8.3e7	2.9e9	1.2e4	8.3e7	2.9e9

The symbol † indicates when the value cannot be found because the system cannot converge. In this case, σ_m has been replaced by the value computed with the oriented method in order to compute η.
O, oriented; R, Rivière; E, eigenvalue penalty factor.

Table III. Study of the condition number η for the order $p = 1, 2, 3, 4$ and the rotation angles $\omega = 0°, 1°$ when the oriented penalty factor is used.

		$p = 1$		$p = 2$		$p = 3$		$p = 4$	
$l_x[m]$		$\omega = 0°$	$\omega = 1°$	$\omega = 0°$	$\omega = 1°$	$\omega = 0°$	$\omega = 1°$	$\omega = 0y$	$\omega = 1°$
25,000	σ_m	0.0041	0.0041	0.0020	0.0019	0.0012	0.0011	0.00074	0.00074
	σ_M	0.19	0.22	0.24	0.30	0.61	0.72	1.5	1.7
	η	44	53	120	160	540	640	2000	2000
6250	σ_m	0.00029	0.00029	0.00013	0.00013	7.4e−5	7.4e−5	4.7e−5	4.7e−5
	σ_M	0.19	0.23	0.24	0.30	0.62	0.73	1.5	1.7
	η	650	770	1800	2300	8400	9900	3.2e4	3.7e4

and for the Rivière and eigenvalue penalty factors, it is also independent of the rotation angles. For the oriented penalty factor, σ_M increases quadratically with the angle rotations.

For the orders 1 to 4 and the rotation angles $\omega = 0°, 1°$, Table III gives the eigenvalues of the system matrix and the condition number η, when the oriented penalty factor is used. As previously, σ_M increases with the rotation angle, whereas σ_m remains the same. With the resolution, σ_M does not change but σ_m decreases. In regard to the order, σ_M increases with the order and σ_m decreases in such a way that η becomes larger. This is consistent because there are more nodes in an element but the stabilization remains the same. Actually, the same properties than previously can be observed for the other methods. Besides, the following relations can be established for each method and for each order:

- $p = 1 : \sigma_M^R \approx 30\sigma_M^E \approx 30 * 250000\sigma_M^O$,
- $p = 2 : \sigma_M^R \approx 30\sigma_M^E \approx 30 * 15000\sigma_M^O$,
- $p = 3 : \sigma_M^R \approx 30\sigma_M^E \approx 30 * 5000\sigma_M^O$,
- $p = 4 : \sigma_M^R \approx 30\sigma_M^E \approx 30 * 5\sigma_M^O$,

DOI: 10.1002/fld

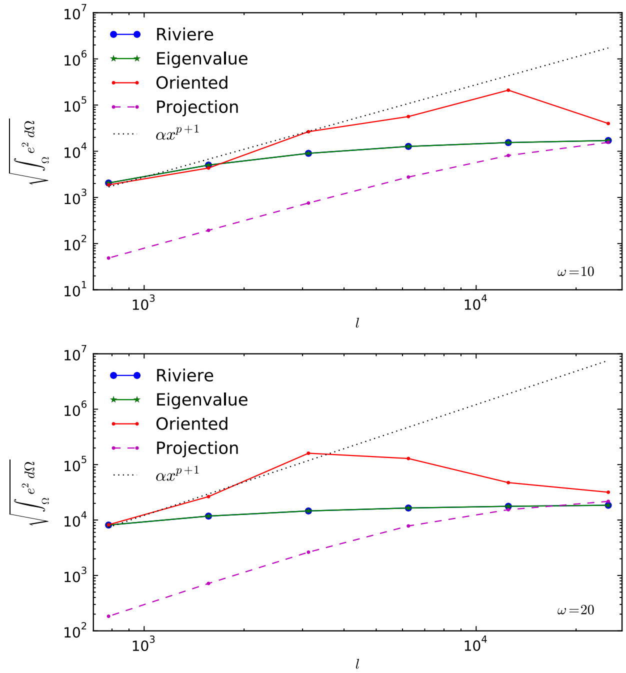


Figure 6. Comparison of the norms of the $\mathcal{L}_2$ error when the eigenvalue, the oriented, and the Rivière penalty factors are used for the method of manufactured solutions at the order 1 and for larger rotation angles: $\omega = 10°$ (top) and $\omega = 20°$ (bottom).

where the superior index gives the method used. The relationship between the σ_M from the Rivière and eigenvalue methods remains the same regardless of the order, but it becomes closer from σ_M computed with the oriented method when the order increases.

In the aim of a complete numerical analysis, larger rotation angles were taken into account in order to simulate larger anisotropies. In Figure 6, the spatial convergence at the order 1 is illustrated for the rotation angles $\omega = 10° - 20°$. For each case, the lines for the Rivière and eigenvalue penalty factors cannot be distinguished, as previously. Unlike the small rotation angles, the difference between the norms of the $\mathcal{L}_2$ error for the Rivière and eigenvalue penalty factors and for the oriented penalty factor is much larger when coarse meshes are used. Indeed, the three lines match further and further when the rotation angle increases. For the oriented penalty factor, the convergence order is reached quite fastly for $\omega = 10'$. For $\omega = 20'$, the asymptotic regime is also achieved but with finest meshes. For larger angles, the convergence will certainly be reached, but the meshes need to be finer, and the asymptotic regime is not really feasible for these cases. However, these large rotation angles require finest meshes, which is not praticable for efficient simulations because such resolution takes too much computational time.

To conclude, the oriented method has a better spatial convergence at coarse resolutions, which is the case for the oceanic meshes and is not worse than the other methods for the finest resolutions. But in terms of efficiency, it gives the better condition number whatever the order. The oriented penalty factor allows to have a well-conditioned system that can be solved rapidly. Its eigenvalues for the system matrix can always be found for all the orders, which is not the case with the Rivière and the eigenvalue methods. The oriented penalty factor is thus the most appropriate to solve problems with strong anisotropic diffusion.

6. PHYSICAL APPLICATION

To complete this study, a more realistic simulation is achieved on an idealized Arctic Ocean. This area is well adapted to investigate the strongly anisotropic diffusion because the density field undergoes high variations, which influence the isopycnals. Even if this diffusion is non-constant in the

DOI: 10.1002/fld

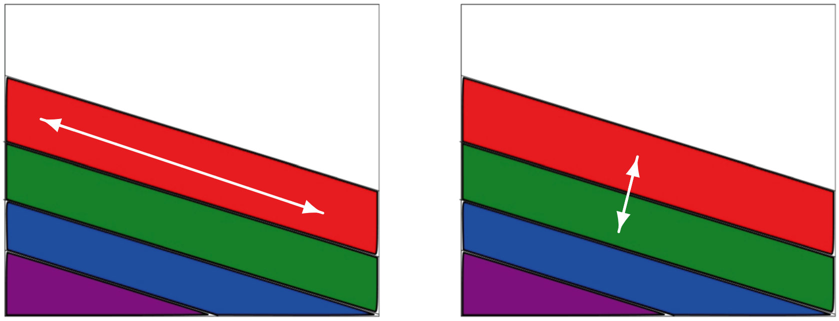


Figure 7. Both pictures represent a vertical cross section of a density field where each color is associated to one density value. The isopycnal direction corresponds to the direction of isopycnals. Hence, an isopycnal diffusivity occurs along the surfaces of constant density (left), whereas the diapycnal one takes place orthogonally to these surfaces (right).

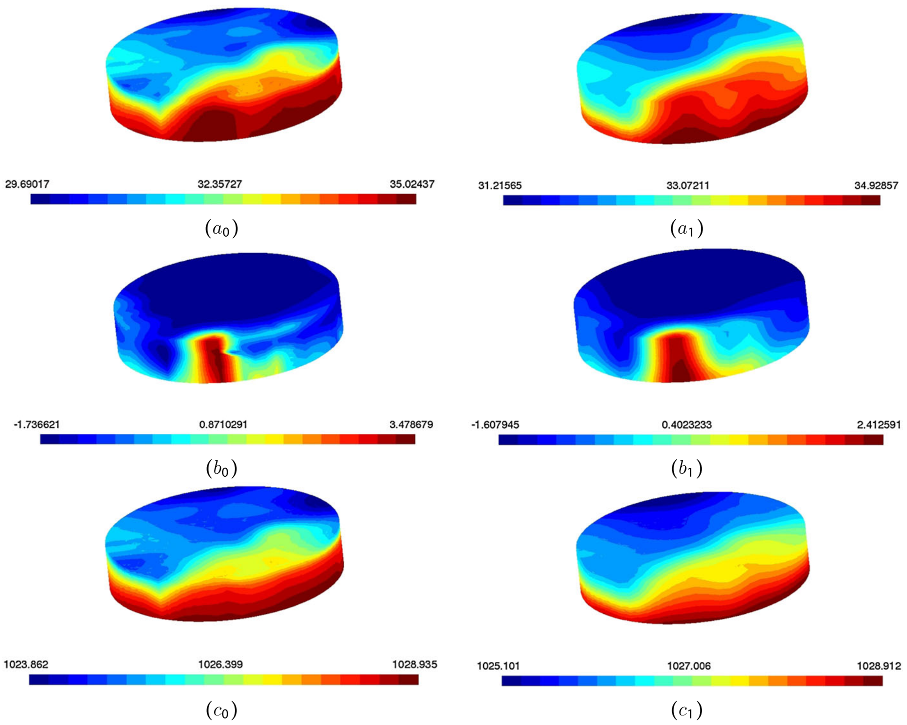
time and in the space and thus that no convergence study will be possible, this application allows to highlight the importance, on the one hand, of a well-conditioned system and, on the other hand, of less numerical diffusion that could distort the solutions. This isopycnal diffusion is a part of a well-known process: the mesoscale eddies. Mesoscale eddies, which have length scale from 10 to 100 km, are found almost everywhere in the ocean. Their kinetic energy is much larger than that of the time-average circulation. They are formed as a result of instabilities and are highly influenced by the rotation of the Earth; they stir and mix the salt and other tracers, transport quantities, and influence the density field and the general ocean circulation [31].

More than half a century ago, Iselin and Montgomery [1, 2] suggested that the mixing of tracers by eddies in the stratified ocean occurs along the isopycnals. This is why Redi [3] introduced the isopycnal diffusivity tensor as a non-linear function of the active tracers because the diffusion operator depends on the density that, in turn, is a function of temperature and salinity [32]. McDougall [33] emphasized that the neutral directions are relevant for the diffusive fluxes of the tracers (Figure 7). Gent *et al.* [34] suggested to use a special closure in ocean models to incorporate the baroclinic effects of mesoscale eddies. The available potential energy is then transformed into eddy kinetic energy. This extra non-divergential velocity, called the *Gent–McWilliams velocity*, yields some improvements in climate simulations because it relaxes the density slopes and thus releases potential energy [35].

To observe the effects of the Gent–McWilliams velocity and isopycnal diffusion on a closed domain, a cylindrical geometry modelling the upper central Arctic Ocean, with a 200-m depth and a radius of 10° of latitude, is meshed with 30 layers of prismatic elements whose horizontal characteristic length is about 10^5 m. The temperature and salinity are initialized on this mesh using the PHC data (*Polar science center Hydrographic Climatology* [36]). In order to remove the effects of compressibility of the ocean water, the considered parcel of water is raised adiabatically from its depth to the sea surface ($p = 0$) before computing the density, which in this case is called *potential density*. The latter is obtained from the Jackett and McDougall [37] equation of state:

$$\rho(S,\vartheta,p) = \frac{P_1(S,\vartheta,p)}{P_2(S,\vartheta,p)},$$

where ϑ is the potential temperature and P_1, P_2 are both polynomial functions of 12 and 13 terms, respectively. Because the potential density field is more complex, some static instabilities can appear during the simulation. Hence, when a parcel of water with a potential density ρ_1 is below another parcel of potential density ρ_2 such that $\rho_1 < \rho_2$, the column of water is unstable. In nature, convective processes quickly re-establish the static stability of the column. Because these processes are not included into the ocean model because of the hydrostratic assumption, a convective adjustement scheme is added to counteract these undesirable effects [38]. Various techniques can be used such as a non-penetrative convective adjustment, a turbulent closure scheme, or an enhanced



(a_0) (a_1)

(b_0) (b_1)

(c_0) (c_1)

Figure 8. (a_0) Initial condition of the tracer S $[psu]$; (b_0) initial condition of the tracer T $[°C]$; (c_0) potential density computed with the equation of state $\rho = \rho(S, T, p)$ [kg/m^3]; (a_1) tracer S after 520 days $[psu]$; (b_1) tracer T after 520 days $[°C]$; (c_1) potential density after 520 days [kg/m^3].

vertical diffusion. In this work, the latter is used in the tracer equation. It consists to enlarge the vertical diffusivity coefficient to $1[m^2/s]$ when the stratification is unstable, that is, when the Brunt–Vaïsälä frequency N_b^2 is negative [39].

The tracer equation with both advection and diffusion terms is now considered:

$$\frac{\partial C}{\partial t} + \nabla \cdot (\underline{u}_{ed} C) = \frac{\partial}{\partial z}\left(\nu_v \frac{\partial C}{\partial z}\right) + \nabla \cdot (\underline{\underline{\kappa}}^s \cdot \nabla C),$$

where $\underline{u}_{ed}$ is the Gent–McWilliams velocity, ν_v the scalar vertical diffusion coefficient that can be enhanced by the convective adjustement, and $\underline{\underline{\kappa}}^s$ the diapycnal–isopycnal diffusivity tensor. The vertical diffusivity decreases with depth from 10^{-3} m^2/s to 10^{-5} m^2/s over the upper 200 m. In Figure 8, the initial states are in the left column, while the states after 520 days are in the right column. Note that the range of each tracer at the initial time differs from the range at the final time. As the density field is created from temperature and salinity, it is influenced by them during the tracers evolution. But, as the domain is situated in the Arctic, and thus in a cold area, the density is more influenced by salinity than by temperature. Both temperature and salinity are diffused in order to align themselves along the isopycnals. But as the temperature has an initial field much more different than the salinity, it evolves faster. Indeed, the difference between the initial and the final maximum values is around 1.06 for temperature, whereas it is around 0.09 for salinity. As expected, each tracer tends toward its mean value in the time. In fact, their minimum and maximum values increase and decrease respectively, which means that the minimum–maximum principle is kept.

The Gent–McWilliams velocity obtained at the end of the simulation is shown in Figure 9. Even if its maximum value is quite small, this velocity really has an impact on the global oceanic circulation but in the long run. As expected, the velocity field never crosses the boundaries of the domain (Figure 9(a)) because it is a divergence-free velocity. A small closed circulation is thus created and can be easily observed. On the front of the middle of the domain in Figure 9(b), the velocity is larger than in other places. This situation points out that the spatial variation of the density is strong

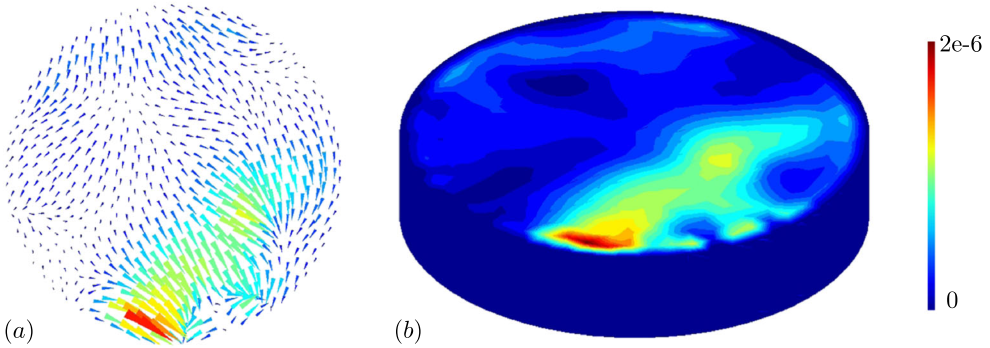


Figure 9. Gent–McWilliams velocity [m/s] (a) after 520 days at the surface and (b) its norm on the whole domain.

at this place (Figure 8). Furthermore, this velocity tends to reduce the slope of the density field (Figure 8(c1)) where the isopycnals have been flattened and smoothened. Finally, all these features show that both isopycnal diffusion and Gent–McWilliams velocity significantly influence the large-scale transport of the oceanic tracers, as discussed by Cox (1987) [26] and Gent *et al.* (1990,1995) [32, 34]. We conclude thus that the discretization of this strong anisotropy diffusion with DGFEM is well adapted for the ocean.

7. CONCLUDING REMARKS

In this paper, the discretization of a diffusion equation with a strong anisotropy by a DGFEM is investigated. The standard discontinuous Galerkin discretization required a special attention to the penalty factor in order to deal correctly with the jumps between the elements and ensure the numerical stability.

Two penalty factors have been proposed and compared. On the one hand, the penalty factor suggested by Rivière is a function of the anisotropy factor and can sometimes be very large. In this case, the numerical solution is too much diffused and thus more approximate. This penalty factor was then improved and renamed *eigenvalue penalty factor*. On the other hand, in order to take into account the diffusion in the direction normal to the interface between the elements, the *oriented penalty factor* is defined by rotating the system to align it with the principal axes of the diffusivity tensor. When strong anisotropy exists in the diffusivity tensor, a value is suggested.

The comparison between both factors shows that the oriented factor provides less numerical diffusion than the eigenvalue factor and still stabilizes enough the numerical scheme. Moreover, the MMS revealed that the oriented penalty factor has a better spatial convergence at coarse resolutions, which is the case for the oceanic meshes. But, in terms of efficiency, it gives the best condition number whatever the order and thus allows to have a well-conditioned system that can be solved rapidly. Finally, this factor is used in a physical application (an idealized Arctic Ocean) where the density field can undergo large variations. Hence, the main features of the isopycnal diffusion as well as that of the Gent–McWilliams velocity are observed: the tracers tend to follow the isopycnals, and the slopes of the density field are progressively reduced.

To the best of our knowledge, it is the first time that a strong anisotropic diffusion is discretized with the DGFEM. The numerical simulations carried out show that the choice of the oriented penalty factor is well adapted to this anisotropy and the conclusions from the physical application go on the same track. Next work will be devoted to the inclusion of the isopycnal diffusion in the complete three-dimensional oceanic model SLIM where all the governing equations are considered. In this study, the impacts of the mesoscale eddies will be analyzed on the long run. Such a study will allow a better understanding of these complex and still not well-known processes.

ACKNOWLEDGEMENTS

We thank Eric Deleersnijder (Université catholique de Louvain, Institute of Mechanics, Materials and Civil Engineering (IMMC), Earth and Life Institute (ELI), Belgium) for his help and scientific discussions. Alice Pestiaux and Tuomas Kärnä are research fellows with the Belgian Fund for Research in Industry and Agriculture (FRIA). Samuel Melchior and Jonathan Lambrechts are research fellows with the Belgian National Fund for Scientific Research (FNRS). The present study was carried out in the framework of the project 'Taking up the challenges of multi-scale marine modelling', which is funded by the Communauté Française de Belgique under contract ARC 10/15-028 (Actions de recherche concertées) with the aim of developing and using SLIM (www.climate.be/slim).

REFERENCES

1. Iselin CO. The influence of vertical and lateral turbulence on the characteristics of the waters at mid-depth. *Transactions of the American Geophysical Union* 1939; **20**:414–417.
2. Montgomery RB. The present evidence on the importance of lateral mixing processes in the ocean. *American Meteor Society* 1940; **21**:87–94.
3. Redi MH. Oceanic isopycnal mixing by coordinate rotation. *Journal of Physical Oceanography* 1982; **12**:1154–1158.
4. Blaise S, Comblen R, Legat V, Remacle JF, Deleersnider E, Lambrechts J. A discontinuous finite element baroclinic marine model on unstructured prismatic meshes. Part I: space discretization-. *Ocean Dynamics* 2010; **60**:1371–1393.
5. Kärnä T, Legat V, Deleersnijder E. A baroclinic discontinuous Galerkin finite element model for coastal flows. *Ocean Modelling* 2013; **61**:1–20.
6. White L, Deleersnijder E, Legat V. A three-dimensional unstructured mesh finite element shallow-water model, with application to the flows around an island and in a wind-driven, elongated basin. *Ocean Modelling* 2008; **22**:26–47.
7. Ainsworth M. Dispersive and dissipative behaviour of high order discontinuous Galerkin finite element methods. *Journal of Computational Physics* 2004; **198**(1):106–130.
8. Lesaint P, Raviart PA. *On a Finite Element Method for Solving the Neutron Transport Equation*. Academic Press: USA, 1974.
9. Reed WH, Hill TR. Triangular mesh methods for the neutron transport equation. *Los Alamos Scientific Laboratory Report LA-UR-73-479*, Los Alamos, NM, 1973.
10. Cockburn B, Karniadakis G, Shu C. *Discontinuous Galerkin Methods: Theory, Computation and Applications*. Springer: USA, 1–23, 2000.
11. Douglas NA. An interior penalty finite element method with discontinuous elements. *SIAM Journal on Numerical Analysis* 1982; **19**:742–760.
12. Hohn ME. *Geostatistics and Petroleum Geology*. Kluwer Academics Publishers: USA, 159, 1999.
13. Houston P, Schwab C, Suli E. Discontinuous hp-finite element methods for advection-diffusion-reaction problems. *SIAM Journal on Numerical Analysis* 2002; **39**:2133–2163.
14. Gastaldi F, Quarteroni A. On the coupling of hyperbolic and parabolic systems: analytical and numerical approach. *Applied Numerical Mathematics* 1990; **6**:3–31.
15. Croisille JP, Ern A, Lelièvre T, Proft J. Analysis and simultion of a coupled hyperbolic-parabolic model problem. *Journal of Numerical Mathematics* 2005; **13**:81–103.
16. Pietro DD, Ern A, Guermond JL. Discontinuous Galerkin methods for anisotropic semidefinite diffusion with advection. *SIAM Journal on Numerical Analysis* 2008; **46**:805–831.
17. Rivière B. *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation*. Frontiers in Applied Mathematics, SIAM: Texas, 2008.
18. Ern A, Stephansen AF, Zunino P. A discontinuous Galerkin method with weighted average for advection-diffusion equations with locally small and anisotropic diffusivity. *Society for Industrial and Applied Mathematics (SIAM), Journal on Numerical Analysis* 2008; **29**:235–256.
19. Griffies SM, Gnanadesikan A, Pacanowski RC, Larichev VD, Dukowicz JK, Smith RD. Isoneutral diffusion in a z-coordinate ocean model. *American Meteor Society* 1998; **28**:805–830.
20. Wang Q, Danilov S, Schröter J. Finite element ocean circulation model based on triangular prismatic elements, with application in studying the effect of topography representation. *Journal of Geophysical Research* 2008; **13**(C5). DOI: 10.1029/2007JC004482.
21. Arnold DN, Brezzi F, Cockburn B, Marini D. Discontinuous Galerkin methods for elliptic problems. In *Discontinuous Galerkin Methods*, Cockburn B, Karniadakis GE, Shu C-W (eds). Springer: London, 2000; 89–101.
22. Shahbazi K. An explicit expression for the penalty parameter of the interior penalty method. *Journal of Computational Physics* 2004; **205**:401–407.
23. Warburton T, Hesthaven JS. On the constants hp-finite element trace inverse inequalities. *Computer Methods in Applied Mechanics and Engineering* 2003; **192**:2765–2773.
24. Spivakovskaya D, Heemink AW, Deleersnijder E. Lagrangian modelling of multi-dimensional advection-diffusion with space-varying diffusivities: theory and idealized test cases. *Ocean Dynamics* 2007; **57**:189–203.

25. Ascher UM, Ruuth SJ, Spiteri RJ. Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations. *Applied Numerical Mathematics* 1997; **25**:151–167.
26. Cox MD. Isopycnal diffusion in a z-coordinate ocean model. *Ocean Modelling* 1987; **74**:1–5.
27. Harvey D. Impact of isopycnal diffusion on heat fluxes and the transient response of a two-dimensional ocean model. *Journal of Physical Oceanography* 1995; **25**:2166–2176.
28. Mathieu PP, Deleersnijder E. What is wrong with isopycnal diffusion in world ocean models? *Applied Mathematical Modelling* 1998; **22**:367–378.
29. Geuzaine C, Remacle JF. Gmsh: a 3-d finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering* 2009; **79**:1309–1331.
30. Lambrechts J, Comblen R, Legat V, Geuzaine C, Remacle JF. Multiscale mesh generation on the sphere. *Ocean Dynamics* 2008; **58**:461–473.
31. Rhines PB. *Mesoscale Eddies, Encyclopedia of Ocean Sciences*. Elsevier: USA, 2009.
32. Gent PR, McWilliams JC. Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography* 1990; **20**:150–155.
33. McDougall TJ. Neutral surfaces. *Journal of Physical Oceanography* 1987; **17**:1950–1967.
34. Gent PR, Willegrand J, McDougall TJ, McWilliams J.C. Parameterizing eddy-induced tracer transports in ocean circulation models. *Journal of Physical Oceanography* 1995; **25**:463–474.
35. Griffies SM. The Gent–McWilliams skew flux. *Journal of Physical Oceanography* 1998; **28**:831–841.
36. Steele M, Morley R, Ermold W. PHC: a global ocean hydrography with a high quality arctic ocean. *Journal of Climate* 2001; **14**:2079–2087.
37. McDougall TJ, Jackett DR, Wright DG, Feistel R. Algorithms for density, potential temperature, conservative temperature, and the freezing temperature of seawater. *Journal of Atmospheric and Oceanic Technology* 2006; **23**: 1709–1728.
38. Marotzke J. Influence of convective adjustment on the stability of the thermohaline circulation. *Journal of Physical Oceanography* 1991; **21**:903–907.
39. Madec G, the NEMO team. NEMO ocean engine. *Technical Report 27*, Institut Pierre-Simon Laplace (IPSL): France, 2008.

DOI: 10.1002/fld