



HAL
open science

An evaluation of methods for normalizing diffusion-based covariance operators in variational data assimilation

Anthony Weaver, Marcin Chrust, Benjamin Ménétrier, Andrea Piacentini

► To cite this version:

Anthony Weaver, Marcin Chrust, Benjamin Ménétrier, Andrea Piacentini. An evaluation of methods for normalizing diffusion-based covariance operators in variational data assimilation. Quarterly Journal of the Royal Meteorological Society, 2021, 147 (734), pp.289-320. 10.1002/qj.3918 . hal-04739164

HAL Id: hal-04739164

<https://cnrs.hal.science/hal-04739164v1>

Submitted on 17 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE

An evaluation of methods for normalizing diffusion-based covariance operators in variational data assimilation

Anthony T. Weaver¹  | Marcin Chrust² | Benjamin Ménétrier³ | Andrea Piacentini¹¹CERFACS/CECI CNRS UMR 5318,
Toulouse, France²European Centre for Medium-Range
Weather Forecasts, Reading, UK³INPT-IRIT, Toulouse, France**Correspondence**A.T. Weaver, CERFACS, 42 avenue
Gaspard Coriolis, 31057 Toulouse
Cedex 1, France.

Email: Anthony.Weaver@cerfacs.fr

Funding informationEuropean Union through the Copernicus
Climate Change Service, French national
programme LEFE/INSU, and the JEDI
project of the Joint Center for Satellite
Data Assimilation**Abstract**

Developing effective ways to model and cycle the background-error covariance matrix is an active area of research in data assimilation. An important aspect of this problem when using a filter to model the background-error correlations is the computation of normalization factors to ensure that the diagonal elements of the modelled correlation matrix are all equal to one. Updating the parameters of a flow-dependent correlation model on each assimilation cycle requires updating the normalization factors, which is costly using traditional methods such as randomization. In this article, we discuss the normalization problem within the context of a diffusion filter-based covariance model used for background-error modelling in a variational data assimilation system for the global ocean. We evaluate various methods for estimating normalization factors when the diffusion tensor of the correlation model is derived from an ensemble of ocean states. Our results show that estimates produced using inexpensive methods derived from analytical considerations of the diffusion equation can have significant errors, especially near boundaries. Estimates obtained using randomization with a small sample size (~ 100) are more accurate in a globally averaged sense but are noisy and can have unacceptably large errors locally. Next, we focus on the specific problem of accounting for flow-dependent correlation parameters in the vertical component of the diffusion operator only, which is especially important near the surface for the assimilation of sea surface temperature observations. Remarkably accurate estimates are obtained by approximating the normalization matrix as a separable product of two normalization matrices: one computed using randomization with the horizontal diffusion operator only and the other computed using randomization with the vertical diffusion operator only. If the parameters of the horizontal component of the diffusion operator are static, then only the normalization factors of the flow-dependent vertical component need to be recomputed on each cycle. This result is of significant practical interest

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

since the vertical diffusion operator employs an inexpensive direct solver and thus can be applied on each cycle with a large random sample to obtain a good approximation of the normalization matrix.

KEYWORDS

background error, correlation operators, covariance operators, diffusion operator, ensemble estimation, ocean data assimilation, normalization factors, variational assimilation

1 | INTRODUCTION

An ensemble of model forecasts provides flow-dependent information on forecast error that can be used to enrich background-error covariance models in variational data assimilation (VDA). The benefits of using ensembles to define flow-dependent background-error covariances in atmospheric VDA is well documented (Buehner, 2005; Belo Pereira and Berre, 2006; Bonavita *et al.*, 2012; Clayton *et al.*, 2013; Berre *et al.*, 2015; Kleist and Ide, 2015; Bonavita *et al.*, 2016). There is growing evidence of the benefits of ensembles in ocean VDA as well (Daget *et al.*, 2009; Penny *et al.*, 2015; Storto *et al.*, 2018).

Ensembles can be used to define a sample estimate of the background-error covariance matrix or to calibrate parametric representations of the background-error covariances. A covariance matrix can be factored into a correlation matrix and a standard deviation matrix multiplying either side of the correlation matrix. In VDA, it is common to take advantage of this property in order to estimate the background-error correlations and standard deviations separately. Estimating correlations is inherently more challenging than estimating standard deviations because of the larger number of elements to estimate and because of the added constraints of ensuring symmetry and positive semi-definiteness of the resulting correlation matrix.

In VDA, correlation operators are commonly built from filters that possess positive-definite smoothing kernels. Filters can be used to localize sample covariances of background error or to model the background-error correlations themselves. By construction, a filter maps a space into itself and thus preserves the physical units of its input field. Viewed as an integral operator, such as a convolution for a homogeneous filter, it implies that the amplitude of the smoothing kernel acting on a three-dimensional (3D) input field, for example, must have physical dimensions of inverse volume in order to cancel out the volume elements associated with spatial integration over the 3D domain. In other words, the diagonal elements (variances) of the covariance matrix implied by the spatial filter have physical dimensions of inverse volume. The filter variances are clearly not the background-error variances that we

wish to specify in data assimilation (that would be dimensionally inconsistent). In order to impose the appropriate background-error variances, a filter-based covariance matrix must first be normalized by its diagonal in order to transform it into a dimensionless correlation matrix, which has 1s along its diagonal.

The diagonal elements of a filter-based covariance matrix are not directly accessible, so techniques are needed to estimate them. When the filter parameters are flow-dependent, the estimation procedure must be repeated on each assimilation cycle, which is prohibitively expensive using a brute-force exact method and hardly affordable using a randomization method. More practical methods exploit the properties of the spatial filter and the structure of the associated filter-based covariance matrix in order to obtain an adequate estimate of the diagonal elements at a reasonable cost.

Correlation models can be defined using spectral filters where a scale-selective damping is performed in spectral space, or using spatial (grid-point) filters where a smoothing operation is performed directly in physical space. Spectral filters are common in meteorological applications of VDA for representing horizontal correlations. In its most basic form, a spectral filter can be interpreted as a spatial filter with an isotropic smoothing kernel. Normalization is comparatively straightforward for spectral filters. For the standard spectral filter, the normalization factors are constant and well approximated by the inverse of a truncated series of modulated spectral coefficients (Parrish and Derber, 1992; Courtier *et al.*, 1998; Rabier *et al.*, 1998). Correlation models based on wavelet filters generalize those based on spectral filters by allowing horizontal correlations to be both scale- and location-dependent (Fisher, 2003; Berre *et al.*, 2015). Chabot *et al.* (2017) show how accurate normalization factors for a wavelet-based covariance matrix can be computed by applying a modified version of the inverse wavelet transform to the vector of specified wavelet-coefficient variances.

The recursive filter (Lorenz, 1992; Wu *et al.*, 2002; Purser *et al.*, 2003a; 2003b) and the closely related class of diffusion filters (Derber and Rosati, 1989; Egbert *et al.*, 1994; Weaver and Courtier, 2001; Yaremchuk and Smith, 2011; Weaver and Mirouze, 2013) are examples of

grid-point (spatial) filters. Grid-point filters are especially popular in ocean VDA because they provide a convenient framework for handling land boundaries in ocean models. Techniques for approximately normalizing covariance models built from recursive or diffusion filters can be derived from analytical considerations of diffusion-like equations, but the techniques become increasingly complex when the smoothing kernels are anisotropic and inhomogeneous (Purser *et al.*, 2003b; Purser, 2008a; 2008b; Yaremchuk and Carrier, 2012). Extensions of the analytical normalization methods to accommodate boundary effects have also been proposed (Mirouze and Weaver, 2010; Mirouze and Storto, 2016). In this article, we evaluate different normalization techniques in a global version of the NEMOVAR ocean data assimilation system (Mogensen *et al.*, 2012). Some of the techniques are already documented in the published literature, while others are relatively new. An objective of this article is to compare them in a consistent experimental framework.

The organization of the article is as follows. In Section 2, we outline the mathematical problem of estimating normalization factors within the context of the diffusion-based background-error covariance matrix used in NEMOVAR. We review some fundamental theoretical results from the diffusion equation, which provides the starting point from which we derive the approximate expressions of the normalization factors that are evaluated in this article. In Section 3, we describe how the correlation length-scales are estimated from a climatological ensemble of perturbations and how they are used to specify the diffusion tensor in the correlation model. In Section 4, we describe the different normalization methods that have been implemented in NEMOVAR. In particular, we focus on evaluating the accuracy of the normalization factors estimated using inexpensive approximate methods relative to the actual normalization factors determined from a costly randomization computation. In Section 5, we address the specific problem of updating normalization factors when only the correlation parameters of the vertical component of diffusion are evolved from one assimilation cycle to the next. A summary and discussion are given in Section 6.

2 | DIFFUSION-BASED CORRELATION MODELS

2.1 | Basic notation and statement of the problem

The model state variables in the Nucleus for European Modelling of the Ocean (NEMO; Madec and NEMO team 2008) are temperature, salinity, the two components of

horizontal velocity, and sea surface height (SSH). The state vector $\mathbf{x} \in \mathbb{R}^{N_x}$ contains the discrete values of the state variables on the model grid. In NEMOVAR, the control variables that are estimated through data assimilation are temperature, and the unbalanced parts of salinity, horizontal velocity and SSH. The control vector $\mathbf{w} \in \mathbb{R}^{N_w}$ contains the discrete values of the control variables on the model grid where $N_w = N_x$ if all the unbalanced fields are estimated.

We consider the following standard form of the background-error covariance matrix (Weaver *et al.*, 2005; Mogensen *et al.*, 2012; Balmaseda *et al.*, 2013)

$$\mathbf{B}_{(\mathbf{x})} = \mathbf{K}_b \mathbf{B}_{(\mathbf{w})} \mathbf{K}_b^T,$$

where $\mathbf{K}_b \in \mathbb{R}^{N_x \times N_w}$ is a lower triangular matrix that contains linearized balance constraints that relate the control variables to the state variables. Fundamental to the formulation of $\mathbf{B}_{(\mathbf{x})} \in \mathbb{R}^{N_x \times N_x}$ is the assumption that the control variables are mutually uncorrelated, so that their associated background-error covariance matrix $\mathbf{B}_{(\mathbf{w})} \in \mathbb{R}^{N_w \times N_w}$ can be modelled as a block-diagonal (univariate) matrix:

$$\mathbf{B}_{(\mathbf{w})} = \mathbf{\Sigma}_{(\mathbf{w})} \mathbf{C}_{(\mathbf{w})} \mathbf{\Sigma}_{(\mathbf{w})},$$

where $\mathbf{C}_{(\mathbf{w})}$ is a correlation matrix, block-diagonal with respect to the control variables, and $\mathbf{\Sigma}_{(\mathbf{w})}$ is a diagonal matrix containing estimates of the background-error standard deviations of the control variables.

If the subscript α is used to denote a particular control variable, then each block-diagonal element of $\mathbf{B}_{(\mathbf{w})}$ is

$$\mathbf{B}_\alpha = \mathbf{\Sigma}_\alpha \mathbf{C}_\alpha \mathbf{\Sigma}_\alpha,$$

where

$$\mathbf{\Sigma}_{(\mathbf{w})} = \begin{pmatrix} \ddots & & \\ & \mathbf{\Sigma}_\alpha & \\ & & \ddots \end{pmatrix} \quad \text{and} \quad \mathbf{C}_{(\mathbf{w})} = \begin{pmatrix} \ddots & & \\ & \mathbf{C}_\alpha & \\ & & \ddots \end{pmatrix}.$$

Here, we focus on the three-dimensional variational assimilation system used for the production of ECMWF's Ocean Reanalysis ORAS5 (Zuo *et al.*, 2019). The observations assimilated in that system contain information about temperature, salinity and SSH, but not horizontal velocity. As a consequence, the unbalanced part of horizontal velocity is unaffected by data assimilation and can be ignored. For this particular case, horizontal velocity is determined entirely from the velocity balance relationship (geostrophy) in \mathbf{K}_b , which becomes a rectangular matrix since $N_w < N_x$.

In order to provide flexibility in the shape of the correlation functions, \mathbf{C}_α can be defined as a linear

combination of univariate correlation matrices \mathbf{C}_{α_l} , $l = 1, \dots, L$ (Mirouze *et al.*, 2016):

$$\mathbf{C}_{\alpha} = \sum_{l=1}^L \mathbf{Y}_{\alpha_l} \mathbf{C}_{\alpha_l} \mathbf{Y}_{\alpha_l}^T, \quad (1)$$

where \mathbf{Y}_{α_l} is a diagonal weighting matrix such that

$$\sum_{l=1}^L \mathbf{Y}_{\alpha_l}^2 = \mathbf{I},$$

in order to preserve the property that a correlation matrix has 1s on its diagonal. Without loss of generality, we can focus on a particular component \mathbf{C}_{α_l} of \mathbf{C} . For clarity of notation, we can then drop the subscript α_l in what follows.

Each correlation matrix is modelled using a diffusion operator \mathbf{L} (Weaver and Courtier, 2001). In its standard implicit form, the diffusion operator is

$$\mathbf{L} = \mathbf{F}^M = (\mathbf{A}^{-1})^M, \quad (2)$$

where the elements of the matrix \mathbf{A} are associated with the discrete representation of the elliptic operator $A = I - \nabla \cdot \boldsymbol{\kappa} \nabla$ involving the identity operator I , gradient operator ∇ , divergence operator $\nabla \cdot$, and diffusion tensor $\boldsymbol{\kappa}$. The inverse matrix $\mathbf{F} = \mathbf{A}^{-1}$ is a diffusion operator and the parameter M is an integer that defines the total number of diffusion steps, which we assume to be even for convenience. The diffusion operator acts as a smoothing operator where the parameters $\boldsymbol{\kappa}$ and M control the spatial length-scales and spectral properties of the underlying smoothing functions. The numerical algorithms used in NEMOVAR to apply Equation (2) are outlined in Section 2.2 and described in more detail in Weaver *et al.* (2016) and Weaver *et al.* (2018b).

The elliptic operator A is self-adjoint with respect to the $L^2(\Omega)$ -inner product in the spatial domain Ω under consideration. We assume that this property is preserved after discretization, which it is using the discretization techniques outlined in Section 2.2. The discrete form of the $L^2(\Omega)$ -inner product, which we call the \mathbf{W} -inner product, is defined by a diagonal matrix \mathbf{W} of grid- and geometry-dependent weights. Since \mathbf{A} is self-adjoint with respect to the \mathbf{W} -inner product, the following factorization holds:

$$\mathbf{L} = \mathbf{L}^{1/2} \mathbf{L}^{1/2} = \mathbf{L}^{1/2} \mathbf{W}^{-1} (\mathbf{L}^{1/2})^T \mathbf{W}. \quad (3)$$

The matrix $\mathbf{L}\mathbf{W}^{-1}$ is a covariance matrix that is symmetric in the usual sense (*i.e.*, with respect to the canonical inner product).

As pointed out by Ménétrier and Auligné (2015), the matrix square-root notation that is often used in data

assimilation can be confusing. From Equation (2) the meaning is clear as $\mathbf{L}^{1/2}$ can be interpreted as applying the diffusion operator over $M/2$ steps. However, with other diffusion formulations (*e.g.*, those described in Section 2.2.2 and Section 4.5.2), this notation is ambiguous, so we prefer to avoid it and use the matrix \mathbf{V} to denote the “square-root” factor of a diffusion matrix. For the standard case (Equation (3)), $\mathbf{V} = \mathbf{L}^{1/2}$. The general expression of the factored correlation matrix (symmetric with 1s on its diagonal) is

$$\mathbf{C} = \mathbf{\Gamma} \mathbf{V} \mathbf{W}^{-1} \mathbf{V}^T \mathbf{\Gamma}, \quad (4)$$

where $\mathbf{\Gamma}^2$ is a diagonal matrix of normalization factors γ_n^2 :

$$\mathbf{\Gamma}^2 = \text{diag}(\gamma_1^2, \dots, \gamma_N^2),$$

N being the total number of grid points for the control variable under consideration. To ensure that \mathbf{C} has 1s on its diagonal, the n th element of the normalization matrix should be defined as the *inverse* of element nn of the symmetric diffusion matrix:

$$\gamma_n^{-2} = \{\mathbf{V} \mathbf{W}^{-1} \mathbf{V}^T\}_{nn}.$$

When the correlation length-scales are large compared to the resolution of the grid, it can be computationally advantageous to apply the diffusion operator on a coarser grid. This situation arises when there are “large-scale” terms in a multiple-scale formulation of \mathbf{C} (Equation (1)) or when the diffusion operator is used to perform covariance localization which tends to involve very broad scales. When the diffusion operator is applied on a grid that is coarser than the native grid, the correlation matrix becomes

$$\mathbf{C} = \mathbf{\Gamma} \mathbf{T} \mathbf{V}_c \mathbf{W}_c^{-1} \mathbf{V}_c^T \mathbf{T}^T \mathbf{\Gamma}, \quad (5)$$

where the subscript “c” denotes the coarse-grid operators, and \mathbf{T} and \mathbf{T}^T are transfer operators from coarse-to-fine grid and fine-to-coarse grid, respectively. For the coarse-grid formulation,

$$\gamma_n^{-2} = \{\mathbf{T} \mathbf{V}_c \mathbf{W}_c^{-1} \mathbf{V}_c^T \mathbf{T}^T\}_{nn}.$$

Determining the diagonal elements of the diffusion matrix is not straightforward since, apart from \mathbf{W}^{-1} and \mathbf{W}_c^{-1} which are diagonal matrices, the matrix components have complicated structure and are only available in operator form. If the normalization factors are not accurately specified, then the background-error variances will be effectively modified from their original values, which can lead to a degraded analysis.

2.2 | Theoretical and numerical considerations

The spatial domain under consideration is the 3D global ocean as modelled by NEMO (Madec and NEMO team, 2008). The primitive equations in NEMO are represented in orthogonal curvilinear coordinates (i, j, k) on the sphere where longitude λ and latitude ϕ are continuous and differentiable functions of the horizontal coordinates (i, j) and depth z is a continuous and differentiable function of the vertical coordinate k . Local curvilinear distance in the (i, j, k) coordinate system is defined by the differential elements $(ds_1, ds_2, ds_3) = (e_1 di, e_2 dj, e_3 dk)$ where (e_1, e_2, e_3) are scale factors that represent the local deformation of the curvilinear coordinates.

NEMOVAR inherits the numerical framework of NEMO. The discretization techniques are based on centred, second-order-accurate finite differences where the state variables are arranged on an Arakawa C-grid. The ocean mesh is defined by the transformation from (i, j, k) to geographical coordinates (λ, ϕ, z) . The grid points are located at integer or integer and a half values of (i, j, k) , so that the mesh on which partial derivatives are evaluated is uniform with unit grid size ($\Delta i = \Delta j = \Delta k = 1$). Local curvilinear distance Δs_m on the grid is then equal to the local scalar factor e_m , $m = 1, 2, 3$. The scale factors define elements of distance, area or volume in the weighting matrix \mathbf{W} in Equation (3) depending on whether the diffusion operator is one-dimensional (1D), two-dimensional (2D) or three-dimensional (3D). For the 1D vertical (z), 2D horizontal (h), and 3D horizontal-vertical (hz) diffusion operators, the weighting matrices are, respectively,

$$\left. \begin{aligned} \mathbf{W}_z &= \text{diag}(\{e_3\}_1, \dots, \{e_3\}_{N_z}) \in \mathbb{R}^{N_z \times N_z}, \\ \mathbf{W}_h &= \text{diag}(\{e_1 e_2\}_1, \dots, \{e_1 e_2\}_{N_h}) \in \mathbb{R}^{N_h \times N_h}, \\ \mathbf{W}_{hz} &= \text{diag}(\{e_1 e_2 e_3\}_1, \dots, \{e_1 e_2 e_3\}_N) \in \mathbb{R}^{N \times N}, \end{aligned} \right\} \quad (6)$$

where N_h and N_z denote the total number of horizontal grid-points and vertical levels, respectively, and $N = N_h N_z$.

In global applications of NEMOVAR, the analysis is performed directly on the tri-polar global grid (known as ORCA) used by NEMO, or on a subsampled version of that grid if the diffusion is performed on a coarse grid as in Equation (5).

2.2.1 | 2D implicit diffusion

The 2D implicit diffusion-based correlation matrix is formulated as (cf. Equation (4))

$$\mathbf{C}_h = \mathbf{\Gamma}_h \mathbf{V}_h \mathbf{W}_h^{-1} \mathbf{V}_h^T \mathbf{\Gamma}_h,$$

where $\mathbf{V}_h = \mathbf{F}_h^{M_h/2} = (\mathbf{A}_h^{-1})^{M_h/2}$, M_h is the total number of implicit diffusion steps, and $\mathbf{\Gamma}_h$ is the normalization matrix. The matrix \mathbf{A}_h is the discrete representation of the 2D elliptic operator $A_h = I - \nabla \cdot \boldsymbol{\kappa}_h \nabla$ in the (i, j) curvilinear coordinate system on the sphere where $\boldsymbol{\kappa}_h$ is the (horizontal) diffusion tensor. The ocean–land boundary condition is taken to be of Neumann type; that is, the field derivative in the direction normal to the land boundary is set to zero. At each grid-point, $\boldsymbol{\kappa}_h = (\kappa_{pq}) \in \mathbb{R}^{2 \times 2}$ is a symmetric positive-definite (SPD) matrix. The off-diagonal elements ($\kappa_{12} = \kappa_{21}$) of the tensor allow for anisotropic stretching in directions that are not necessarily aligned with the curvilinear coordinates. In NEMOVAR, the off-diagonal elements are currently neglected in the diffusion operator, which has implications for the estimation of the diffusion tensor, as discussed in Section 3.

The standard procedure in NEMOVAR for applying \mathbf{V}_h and \mathbf{V}_h^T is as follows. First, the self-adjoint matrices \mathbf{A}_h and \mathbf{A}_h^T are transformed into a common, unit-preserving symmetric matrix:

$$\hat{\mathbf{A}}_h \equiv \mathbf{W}_h^{1/2} \mathbf{A}_h \mathbf{W}_h^{-1/2} = \mathbf{W}_h^{-1/2} \mathbf{A}_h^T \mathbf{W}_h^{1/2} = \hat{\mathbf{A}}_h^T.$$

Then, \mathbf{V}_h and \mathbf{V}_h^T are reformulated in terms of $\hat{\mathbf{A}}_h$ so that it can be inverted using an efficient solver for sparse SPD matrices. In NEMOVAR, we use a linear solver based on the Chebyshev Iteration to build an approximation of $\hat{\mathbf{F}}_h = \hat{\mathbf{A}}_h^{-1} = \hat{\mathbf{F}}_h^T$. Furthermore, we apply the same fixed number of iterations in \mathbf{V}_h and \mathbf{V}_h^T in order to preserve symmetry of \mathbf{C}_h to machine precision (Weaver *et al.*, 2016). The square-root operators in terms of $\hat{\mathbf{F}}_h$ are

$$\left. \begin{aligned} \mathbf{V}_h &= \mathbf{W}_h^{-1/2} \hat{\mathbf{F}}_h^{M_h/2} \mathbf{W}_h^{1/2}, \\ \mathbf{V}_h^T &= \mathbf{W}_h^{1/2} \hat{\mathbf{F}}_h^{M_h/2} \mathbf{W}_h^{-1/2}. \end{aligned} \right\}$$

2.2.2 | 2D×1D implicit diffusion

In NEMOVAR, the conventional diffusion-based correlation matrix for 3D variables is built from a 2D horizontal diffusion operator that is applied separately in each model level k , and a 1D vertical diffusion operator that is applied separately at each horizontal grid-point (i, j) . This separation is justified by the large difference in spatial scales between the horizontal and vertical directions. We denote the horizontal and vertical diffusion operators as $\mathbf{L}_{hk} \in \mathbb{R}^{N_h \times N_h}$ and $\mathbf{L}_{zj} \in \mathbb{R}^{N_z \times N_z}$, respectively. Note that we require the index k on the horizontal diffusion operator since variations in bathymetry result in different land masks in each model level and since the horizontal diffusion tensor can be different in each model level. Likewise, we require the index ij on the vertical diffusion operator

because variations in bathymetry result in different ocean depths at each horizontal grid-point and since the vertical diffusion coefficient (κ_z) can be different at each horizontal grid-point.

For the implicit diffusion operators, $\mathbf{L}_{h_k} = \mathbf{F}_{h_k}^{M_h}$ where, as in the previous section, $\mathbf{F}_{h_k} = \mathbf{A}_{h_k}^{-1}$ is the inverse of the matrix representing the discretized 2D elliptic operator and M_h is the number of 2D implicit diffusion steps. Similarly, $\mathbf{L}_{z_{ij}} = \mathbf{F}_{z_{ij}}^{M_z}$ where $\mathbf{F}_{z_{ij}} = \mathbf{A}_{z_{ij}}^{-1}$ is the inverse of the matrix representing the discretized 1D elliptic operator and M_z is the corresponding number of 1D implicit diffusion steps (which can be different from M_h).

Constructing a self-adjoint 3D diffusion operator $\mathbf{L}_{h \times z} \in \mathbb{R}^{N \times N}$ from the self-adjoint components \mathbf{L}_{h_k} and $\mathbf{L}_{z_{ij}}$ requires some care. The formulation used in NEMOVAR is outlined below. Let $\boldsymbol{\zeta} = (\zeta_1^T, \dots, \zeta_{N_z}^T)^T \in \mathbb{R}^N$ be a vector associated with a 3D variable where $\boldsymbol{\zeta}_k \in \mathbb{R}^{N_h}$ contains the values of the variable in model level k . In order to reconcile the different dimensions of the diffusion matrices, we introduce the following extension matrices to map from the reduced spaces \mathbb{R}^{N_h} and \mathbb{R}^{N_z} to the full space \mathbb{R}^N :

$$\mathbf{E}_{h_k} = \begin{pmatrix} \mathbf{0}_{N_z} \\ \vdots \\ \mathbf{I}_{N_z} \\ \vdots \\ \mathbf{0}_{N_z} \end{pmatrix} \in \mathbb{R}^{N \times N_h} \quad \text{and} \quad \mathbf{E}_{z_{ij}} = \begin{pmatrix} \mathbf{e}_{ij} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{e}_{ij} \end{pmatrix} \in \mathbb{R}^{N \times N_z},$$

where the non-zero matrix \mathbf{I}_{N_z} acts on variables in level k , and $\mathbf{e}_{ij} = (0, \dots, 0, 1, 0, \dots, 0)^T \in \mathbb{R}^{N_h}$, the non-zero element being located at the ij th entry. The transpose of the extension matrices, $\mathbf{E}_{h_k}^T \in \mathbb{R}^{N_h \times N}$ and $\mathbf{E}_{z_{ij}}^T \in \mathbb{R}^{N_z \times N}$, are restriction matrices, which satisfy the properties $\mathbf{E}_{h_k}^T \mathbf{E}_{h_k} = \mathbf{I}_{N_h}$ and $\mathbf{E}_{z_{ij}}^T \mathbf{E}_{z_{ij}} = \mathbf{I}_{N_z}$.

Now, using the above matrices and their properties, we can formulate horizontal and vertical diffusion operators that act on the complete vector $\boldsymbol{\zeta}$:

$$\mathbf{L}_h = \sum_k \mathbf{E}_{h_k} \mathbf{L}_{h_k} \mathbf{E}_{h_k}^T = \mathbf{F}_h^{M_h} \in \mathbb{R}^{N \times N}, \quad (7)$$

$$\mathbf{L}_z = \sum_i \sum_j \mathbf{E}_{z_{ij}} \mathbf{L}_{z_{ij}} \mathbf{E}_{z_{ij}}^T = \mathbf{F}_z^{M_z} \in \mathbb{R}^{N \times N}, \quad (8)$$

where $\mathbf{F}_h = \sum_k \mathbf{E}_{h_k} \mathbf{F}_{h_k} \mathbf{E}_{h_k}^T$ and $\mathbf{F}_z = \sum_i \sum_j \mathbf{E}_{z_{ij}} \mathbf{F}_{z_{ij}} \mathbf{E}_{z_{ij}}^T$. Let $\widehat{\mathbf{W}}_h = \sum_k \mathbf{E}_{h_k} \mathbf{W}_h \mathbf{E}_{h_k}^T \in \mathbb{R}^N$ and let $\widehat{\mathbf{W}}_z = \sum_i \sum_j \mathbf{E}_{z_{ij}} \mathbf{W}_z \mathbf{E}_{z_{ij}}^T \in \mathbb{R}^N$, where \mathbf{W}_h and \mathbf{W}_z are defined in Equation (6). Using the property that \mathbf{F}_h is self-adjoint with respect to the $\widehat{\mathbf{W}}_h$ -inner product and that \mathbf{F}_z is self-adjoint with respect to the $\widehat{\mathbf{W}}_z$ -inner product, we can express Equations (7) and (8) in the following square-root

forms:

$$\begin{aligned} \mathbf{L}_h &= \mathbf{L}_h^{1/2} \widehat{\mathbf{W}}_h^{-1} (\mathbf{L}_h^{1/2})^T \widehat{\mathbf{W}}_h, \\ \mathbf{L}_z &= \mathbf{L}_z^{1/2} \widehat{\mathbf{W}}_z^{-1} (\mathbf{L}_z^{1/2})^T \widehat{\mathbf{W}}_z, \end{aligned}$$

where $\widehat{\mathbf{W}}_h \widehat{\mathbf{W}}_z = \mathbf{W}_{hz}$ (the last relation in Equation (6)).

The product $\mathbf{L}_h \mathbf{L}_z$ is not self-adjoint with respect to the \mathbf{W}_{hz} -inner product, so it cannot be used directly to define a symmetric correlation matrix for 3D variables. Self-adjointness can be imposed by using the square-root forms of \mathbf{L}_h and \mathbf{L}_z , forming their product, and interchanging the terms assuming that they commute (which they do not). This leads to several possible formulations involving different expressions for the square-root matrix, which we denote by $\mathbf{V}_{h \times z}^{(m)}$:

$$\mathbf{L}_{h \times z} = \mathbf{V}_{h \times z}^{(m)} \mathbf{W}_{hz}^{-1} (\mathbf{V}_{h \times z}^{(m)})^T \mathbf{W}_{hz} \approx \mathbf{L}_h \mathbf{L}_z. \quad (9)$$

The corresponding correlation matrix is

$$\mathbf{C}_{h \times z} = \boldsymbol{\Gamma}_{h \times z} \mathbf{V}_{h \times z}^{(m)} \mathbf{W}_{hz}^{-1} (\mathbf{V}_{h \times z}^{(m)})^T \boldsymbol{\Gamma}_{h \times z}, \quad (10)$$

where $\boldsymbol{\Gamma}_{h \times z}^2$ is the normalization matrix. We refer to these formulations as 2D×1D (m) where the value of m will be added in parentheses when we need to distinguish them.

Two obvious formulations for the square-root matrix are

$$\mathbf{V}_{h \times z}^{(1)} = \mathbf{V}_h \mathbf{V}_z \quad \text{and} \quad \mathbf{V}_{h \times z}^{(2)} = \mathbf{V}_z \mathbf{V}_h, \quad (11)$$

where

$$\mathbf{V}_h = \mathbf{F}_h^{M_h/2} \quad \text{and} \quad \mathbf{V}_z = \mathbf{F}_z^{M_z/2},$$

which involve applying the horizontal and vertical diffusion operators separately in the square-root matrix, but in a different order.

Two others formulations are

$$\mathbf{V}_{h \times z}^{(3)} = (\mathbf{F}_h \mathbf{F}_z)^{M_{hz}/2} \quad \text{and} \quad \mathbf{V}_{h \times z}^{(4)} = (\mathbf{F}_z \mathbf{F}_h)^{M_{hz}/2},$$

which can be derived from $\mathbf{V}_{h \times z}^{(1)}$ and $\mathbf{V}_{h \times z}^{(2)}$ by taking $M_h = M_z = M_{hz}$ and moving the step counter outside of the product of $\mathbf{F}_h^{M_{hz}/2}$ and $\mathbf{F}_z^{M_{hz}/2}$. $\mathbf{V}_{h \times z}^{(3)}$ and $\mathbf{V}_{h \times z}^{(4)}$ are preferable to $\mathbf{V}_{h \times z}^{(1)}$ and $\mathbf{V}_{h \times z}^{(2)}$ since they allow more coupling between the horizontal and vertical diffusion operators, which results in better smoothness properties, especially near complex bathymetry. This feature is more important than maintaining flexibility in the choice of M_h and M_z , which is allowed with $\mathbf{V}_{h \times z}^{(1)}$ and $\mathbf{V}_{h \times z}^{(2)}$. We say that $\mathbf{V}_{h \times z}^{(3)}$ and $\mathbf{V}_{h \times z}^{(4)}$ are 2D×1D formulations with *interleaving*. Weaver *et al.* (2016) illustrate the importance of interleaving to avoid numerical artifacts near irregular coastlines

and islands when using a 2D horizontal diffusion model formulated as a product of 1D diffusion operators (their figure 1). Interleaving is important in the 3D formulation for similar reasons, as will be illustrated in Section 5.4. The computational costs of the four formulations above are similar when using the same values of M_h and M_z . We will come back to the formulations in Equation (11) in Section 5 when discussing methods for estimating normalization factors.

The horizontal diffusion operators in $\mathbf{V}_{\text{hz}}^{(m)}$ are applied using the Chebyshev Iteration as outlined in Section 2.2.1. The vertical diffusion operators in $\mathbf{V}_{\text{hz}}^{(m)}$ involve inverting small self-adjoint positive-definite matrices $\mathbf{A}_{z_{ij}}$ of size $N_K \times N_K$, which can be done efficiently using a direct solver. Neumann boundary conditions are imposed at the top and bottom boundaries. We can transform $\mathbf{A}_{z_{ij}}$ into symmetric form $\hat{\mathbf{A}}_{z_{ij}} = \mathbf{W}_z \mathbf{A}_{z_{ij}} = \hat{\mathbf{A}}_{z_{ij}}^T$, precompute the Cholesky decomposition of $\hat{\mathbf{A}}_{z_{ij}}$, and then use standard forward-elimination and backward-substitution procedures to apply $\hat{\mathbf{A}}_{z_{ij}}^{-1}$ (Golub and Van Loan, 1996). In terms of the augmented symmetric matrices, the square-root factors are

$$\left. \begin{aligned} \mathbf{V}_{\text{hz}}^{(1)} &= (\widehat{\mathbf{W}}_h^{-1/2} \hat{\mathbf{F}}_h^{M_h/2} \widehat{\mathbf{W}}_h^{1/2}) (\hat{\mathbf{F}}_z \widehat{\mathbf{W}}_z)^{M_z/2}, \\ (\mathbf{V}_{\text{hz}}^{(1)})^T &= (\widehat{\mathbf{W}}_z \hat{\mathbf{F}}_z)^{M_z/2} (\widehat{\mathbf{W}}_h^{1/2} \hat{\mathbf{F}}_h^{M_h/2} \widehat{\mathbf{W}}_h^{-1/2}), \end{aligned} \right\}$$

and

$$\left. \begin{aligned} \mathbf{V}_{\text{hz}}^{(3)} &= (\widehat{\mathbf{W}}_h^{-1/2} \hat{\mathbf{F}}_h \widehat{\mathbf{W}}_h^{1/2} \hat{\mathbf{F}}_z \widehat{\mathbf{W}}_z)^{M_{\text{hz}}/2}, \\ (\mathbf{V}_{\text{hz}}^{(3)})^T &= (\widehat{\mathbf{W}}_z \hat{\mathbf{F}}_z \widehat{\mathbf{W}}_h^{1/2} \hat{\mathbf{F}}_h \widehat{\mathbf{W}}_h^{-1/2})^{M_{\text{hz}}/2}, \end{aligned} \right\}$$

where $\hat{\mathbf{F}}_z = \mathbf{A}_z^{-1} \widehat{\mathbf{W}}_z^{-1} = \hat{\mathbf{F}}_z^T$. Similar expressions follow for $\mathbf{V}_{\text{hz}}^{(2)}$ and $\mathbf{V}_{\text{hz}}^{(4)}$.

2.2.3 | 3D implicit diffusion

Instead of separating the 3D correlation matrix into horizontal and vertical diffusion operator components, we can model it directly, as in Weaver *et al.* (2018b), as a 3D implicit diffusion operator:

$$\mathbf{C}_{\text{hz}} = \mathbf{\Gamma}_{\text{hz}} \mathbf{V}_{\text{hz}} \mathbf{W}_{\text{hz}}^{-1} \mathbf{V}_{\text{hz}}^T \mathbf{\Gamma}_{\text{hz}}$$

where $\mathbf{V}_{\text{hz}} = \mathbf{F}_{\text{hz}}^{M_{\text{hz}}/2} = (\mathbf{A}_{\text{hz}}^{-1})^{M_{\text{hz}}/2}$, M_{hz} is the total number of 3D implicit diffusion steps, and $\mathbf{\Gamma}_{\text{hz}}$ is the normalization matrix. The matrix \mathbf{A}_{hz} is the discrete representation of the 3D elliptic operator $A_{\text{hz}} = I - \nabla \cdot \boldsymbol{\kappa}_{\text{hz}} \nabla$ in the (i, j, k) coordinate system, and $\boldsymbol{\kappa}_{\text{hz}} = (\kappa_{pq}) \in \mathbb{R}^{3 \times 3}$ is the

horizontal-vertical diffusion tensor where $\kappa_{pq} = \kappa_{qp}$. The horizontal component of the tensor, $\boldsymbol{\kappa}_h$, contains the elements (κ_{pq}) , $p, q = 1, 2$. In the following, we neglect the anisotropic coupling between the horizontal and vertical directions ($\kappa_{13} = \kappa_{23} = 0$). The element κ_{33} then corresponds to the vertical diffusion coefficient κ_z introduced earlier with the 2D×1D formulation.

The procedure for applying \mathbf{V}_{hz} and \mathbf{V}_{hz}^T mirrors that used for the 2D implicit diffusion operator, which involves first transforming \mathbf{A}_{hz} into a symmetric matrix $\hat{\mathbf{A}}_{\text{hz}}$ and then approximating $\hat{\mathbf{F}}_{\text{hz}} = \hat{\mathbf{A}}_{\text{hz}}^{-1}$ using the Chebyshev Iteration. In terms of $\hat{\mathbf{F}}_{\text{hz}}$, the square-root operators are

$$\left. \begin{aligned} \mathbf{V}_{\text{hz}} &= \mathbf{W}_{\text{hz}}^{-1/2} \hat{\mathbf{F}}_{\text{hz}}^{M_{\text{hz}}/2} \mathbf{W}_{\text{hz}}^{1/2} \\ \mathbf{V}_{\text{hz}}^T &= \mathbf{W}_{\text{hz}}^{1/2} \hat{\mathbf{F}}_{\text{hz}}^{M_{\text{hz}}/2} \mathbf{W}_{\text{hz}}^{-1/2} \end{aligned} \right\}$$

The 3D formulation is simpler but more costly than the 2D×1D (m) formulations. A major factor influencing the computational cost of the 3D formulation when using ensemble-estimated correlation parameters is the large condition number of \mathbf{A}_{hz} caused by large vertical length-scales in the mixed-layer region where the vertical resolution is fine. The large condition number affects the convergence rate of the Chebyshev Iteration. In the 2D×1D formulations, the Chebyshev iteration is only used for the horizontal component and the efficiency of the direct solver used for the vertical component is not affected by the condition number of $\hat{\mathbf{A}}_z$.

2.3 | Analytical expressions for the normalization factors

Weaver and Courtier (2001) and Weaver and Mirouze (2013) discuss analytical solutions of the diffusion equation on the sphere \mathbb{S}^2 and derive expressions for the associated positive-definite smoothing kernels that can be interpreted as isotropic covariance functions. Like all isotropic covariance functions on \mathbb{S}^2 , those associated with implicit diffusion are defined as a series expansion of Legendre polynomials (Gaspari and Cohn, 1999). Of particular interest here is the expression for the intrinsic variance of the covariance function as the inverse of this variance is the normalization factor needed to transform the covariance function into a correlation function. For the covariance functions on \mathbb{S}^2 , the normalization factor is the inverse of an infinite series involving the total wave number of the spherical harmonics. A more convenient, closed-form expression of the normalization factor can be derived by considering the analytical solution of the diffusion equation in the Euclidean space \mathbb{R}^2 , which

approximates well the solution on \mathbb{S}^2 on spatial scales that are small compared to the radius of the Earth and thus relevant in ocean data assimilation. Expressions for the normalization factor associated with diffusion in \mathbb{R}^1 and \mathbb{R}^3 can also be derived. They are given below since they provide the basis for the approximate expressions of the normalization factors described in Section 4.5.

The analytical solution of the elliptic equation of implicit diffusion in \mathbb{R}^2 is well known (Whittle, 1963; Guttorp and Gneiting, 2006). If κ_h is constant, then the solution of the elliptic equation can be expressed as an integral equation over \mathbb{R}^2 where its kernel is an isotropic and homogeneous covariance function from the Matérn family. The total number of implicit diffusion steps M_h is related to the standard smoothness parameter of the Matérn covariance function. The inverse of the variance of the Matérn function is the normalization factor and is given by

$$\gamma_h^2 = \omega_h \ell_h^2,$$

where

$$\omega_h = 4\pi (M_h - 1)$$

and

$$\ell_h^2 = \sqrt{\det(\kappa_h)}$$

is the square of the characteristic length-scale of the Matérn function.

For the 1D implicit diffusion operator used in the 2D×1D formulations, we can estimate the normalization factor by examining the solution of the implicit diffusion equation on \mathbb{R}^1 with constant diffusion coefficient (Mirouze and Weaver, 2010). The Matérn kernel of the integral solution is an Auto-Regressive (AR) function of order M_z , with characteristic length-scale

$$\ell_z = \sqrt{\kappa_z}.$$

The normalization factor for the vertical component is the inverse of the intrinsic variance of the AR covariance function:

$$\gamma_z^2 = \omega_z \ell_z,$$

where

$$\omega_z = \frac{2^{2M_z-1} [(M_z - 1)!]^2}{(2M_z - 2)!},$$

which can be evaluated numerically using the recursive formula

$$\left. \begin{aligned} \rho_2 &= 4 \ell_z, \\ \rho_m &= \left(\frac{2m-2}{2m-3} \right) \rho_{m-1} \quad \text{for } m = 3, \dots, M_z, \\ \gamma_z^2 &= \rho_{M_z}. \end{aligned} \right\}$$

As in \mathbb{R}^1 , the covariance functions represented by the implicit diffusion operator in \mathbb{R}^3 are AR functions whose order is equal to the number of 3D implicit diffusion steps M_{hz} and whose characteristic length-scale ℓ_{hz} is

$$\ell_{hz}^3 = \sqrt{\det(\kappa_{hz})} = \sqrt{\kappa_z} \sqrt{\det(\kappa_h)}.$$

The inverse of the intrinsic variance of the M_{hz} -th-order AR covariance function provides an estimate of the normalization factor for the 3D implicit diffusion operator:

$$\gamma_{hz}^2 = \omega_{hz} \ell_{hz}^3,$$

where

$$\omega_{hz} = \frac{2^{2M_{hz}-1} \pi [(M_{hz} - 2)!]^2 (M_{hz} - 1)}{(2M_{hz} - 4)!},$$

which can be evaluated numerically using the recursive formula

$$\left. \begin{aligned} \rho_3 &= 32\pi \ell_{hz}^3, \\ \rho_m &= \left(\frac{2m-2}{2m-5} \right) \rho_{m-1} \quad \text{for } m = 4, \dots, M_{hz}, \\ \gamma_{hz}^2 &= \rho_{M_{hz}}. \end{aligned} \right\}$$

Results can be generalized to \mathbb{R}^d where the expression for the normalization factor after M steps of implicit diffusion with a constant diffusion tensor κ is

$$\gamma^2 = 2^d \pi^{d/2} \frac{\Gamma(M)}{\Gamma(M - d/2)} \ell^d, \quad (12)$$

where $\Gamma(\cdot)$ is the Gamma function and $\ell^d = \sqrt{\det(\kappa)}$.

2.4 | Daley length-scale and tensor

The Daley length-scale D is a standard scale parameter used in data assimilation for characterizing isotropic correlation functions $c(r)$ that are at least twice differentiable (Daley, 1991). It is defined in terms of the local curvature of the correlation function at its peak:

$$D = \sqrt{-\frac{d}{\nabla^2 c|_{r=0}}}, \quad (13)$$

where $\nabla^2 = \text{tr}(\nabla \nabla^T)$ is the d -dimensional Laplacian operator. For Matérn functions in \mathbb{R}^d represented by M steps of an implicit diffusion operator with constant diffusion tensor, it can be shown that

$$D = \sqrt{2M - d - 2} \ell,$$

which is valid for $M > (d + 2)/2$. In the limit as $M \rightarrow \infty$ with D held fixed, the Matérn function converges to the Gaussian function and the normalization factor Equation (14) converges to the more familiar expression

$$\gamma^2 = (4M\pi)^{d/2} \ell^d = (2\pi)^{d/2} D_g^d,$$

where $D_g = \sqrt{2M} \ell$ is the Daley length-scale of the Gaussian function.

For anisotropic correlation functions, the component Daley length-scales are defined through a tensor (a $d \times d$ matrix), $\mathbf{D} = \mathbf{H}^{-1}$ where (Weaver and Mirouze, 2013)

$$\mathbf{H} = -\nabla \nabla^T c|_{r=0} \quad (14)$$

is the Local Correlation (Hessian) Tensor (LCT). In the isotropic case, $\mathbf{D} = D^2 \mathbf{I}_d$ where D is a scalar given by Equation (13). For the Matérn functions represented by the d -dimensional implicit diffusion operator with anisotropic diffusion tensor, it is straightforward to show that

$$\boldsymbol{\kappa} = \left(\frac{1}{2M - d - 2} \right) \mathbf{D}. \quad (15)$$

The Daley tensor is of interest since it can be estimated directly from derivatives of ensemble perturbations. Equation (15) then allows us to relate it to the tensor that needs to be specified in the diffusion operator. This technique has been used for defining the diffusion tensor in the experiments reported here, as described in the next section.

3 | SPECIFICATION OF THE DIFFUSION-MODEL PARAMETERS

In order to evaluate the different methods for estimating normalization factors, we consider a diffusion tensor that is estimated from the five-member (four perturbed + one unperturbed) ORAS5 (Ocean ReAnalysis 5) ensemble of ocean reanalyses for the low-resolution (LR) global configuration ORCA1-Z42 (approximately 1 degree horizontal resolution and 42 vertical levels). The LR version of ORAS5 is denoted ORAS5-LR. A four perturbed-member ensemble is far too small to determine a statistically robust estimate of the diffusion tensor. In order to reduce sampling error, we estimate the tensor from a climatological ensemble for the 5-year period 2010–2014, focusing on seasonal statistics from December–January–February (DJF). Since the assimilation cycle width in ORAS5-LR was 10 days, the total effective ensemble size used to compute the seasonal climatology is 180 (4 perturbed members \times 9 cycles per season \times 5 years).

On each assimilation cycle, the LCT (Equation (14)) is first estimated at each grid point from the tensor product of the local gradient of the ensemble perturbations:

$$\mathbf{H} = (H_{pq}) \approx \overline{\nabla(\epsilon'/\sigma) \{ \nabla(\epsilon'/\sigma) \}^T}, \quad p, q = 1, 2, 3, \quad (16)$$

where ϵ' denotes an ensemble perturbation for a control variable, centred about the ensemble mean, and σ is the ensemble standard deviation (Michel *et al.*, 2016). The overbar denotes the ensemble average. The ensemble perturbation for a control variable is computed by removing the balanced component from the ensemble perturbation of the associated state variable. This is done by applying the inverse of the balance operator (\mathbf{K}_b^{-1}) to the state perturbations where the linearization state for \mathbf{K}_b is provided by the unperturbed member. The unperturbed member is not used to compute the ensemble mean and standard deviation.

The derivatives in Equation (16) are estimated using centred finite differences. The procedure is outlined below. The reader is referred to appendix A of Weaver *et al.* (2018a) for details on the numerics. For the control variables under consideration, we choose to estimate \mathbf{H} at T-points on the Arakawa C-grid. These points correspond to the centre of the grid cells. First, we use a centred difference of the normalized perturbations at adjacent T-points to estimate the first derivatives of the perturbations with respect to i and j . This produces derivative estimates at the ensemble of U- and V-points on the C-grid, which are points located in the middle of the edges of the grid cells. At coastline points, the derivative in the direction normal to the coastline is set to zero. The sample variance of the product of the i -component derivative with itself is then computed directly at the U-point to estimate the diagonal element H_{11} . Likewise, the sample variance of the product of the j -component derivative with itself is computed directly at the V-point to estimate H_{22} .

The off-diagonal elements H_{12} and H_{21} are estimated from the sample variance of the cross-product of the perturbation derivatives with respect to i and j . This requires interpolating one of the component derivatives to the point where the other component derivative is defined. Here, we do this by simply averaging the derivatives at the four nearest surrounding points. We estimate H_{12} (defined at V-points) and H_{21} (defined at U-points) separately.

For 3D variables, the sample variance of the product of the k -component derivative of the normalized perturbations with itself is used to provide an estimate of H_{33} at W-points. These points are located at the interface between model layers. Currently, we do not estimate the off-diagonal elements H_{13} , H_{31} , H_{23} and H_{32} . The tensor elements at U- and V-points of a given cell are then

averaged to the T-point of that cell. Finally, in order to ensure that the resulting tensors are SPD, and hence invertible, the off-diagonal elements H_{12} and H_{21} are averaged, and are set to zero at grid-points where the determinant is not positive.

The estimated LCT elements are filtered in each model level using a 2D diffusion operator with a constant horizontal filtering scale. The filtering procedure proposed by Michel *et al.* (2016) is used to preserve positive definiteness of the LCT. The filtering scale is determined using the optimally based method of Ménétrier *et al.* (2015), where its implementation in NEMOVAR is described in Weaver *et al.* (2018a). The filtered LCT is then averaged over the months DJF from 2010–2014 to produce a climatological LCT and inverted to produce a climatological Daley tensor \mathbf{D} . The climatological estimate of \mathbf{D} is converted into an effective diffusion tensor $\boldsymbol{\kappa}$ by dividing by the factor $(2M - d - 2)$ in Equation (15). Ten implicit diffusion steps ($M = 10$) are used with all diffusion models (1D, 2D and 3D). The covariance function associated with this value of M is approximately Gaussian. Finally, the elements of the estimated diffusion tensor are averaged to the appropriate cell interface points where their values are required for the finite difference formulation of the diffusion operator.

As stated earlier, the current diffusion model in NEMOVAR does not account for the off-diagonal elements of the diffusion tensor. This raises the question of how best to define the diagonal elements, given estimates of both the diagonal and off-diagonal elements. A sensible approach is to rescale the diagonal elements such that the determinant of the adjusted diagonal tensor equals the determinant of the estimated non-diagonal tensor:

$$\left. \begin{aligned} \kappa_{11} &= \hat{\kappa}_{11} \sqrt{1 - \hat{\kappa}_{12}^2 / \hat{\kappa}_{11} \hat{\kappa}_{22}}, \\ \kappa_{22} &= \hat{\kappa}_{22} \sqrt{1 - \hat{\kappa}_{12}^2 / \hat{\kappa}_{11} \hat{\kappa}_{22}}, \end{aligned} \right\} \quad (17)$$

where the elements $\hat{\kappa}_{pq}$ on the right-hand side are those estimated from the ensemble. This method has been applied here.

Various checks are then applied on the estimated diffusion tensor. In the upper 600 m, which roughly corresponds to the maximum mixed-layer depth, the maximum-allowed value of the vertical length-scale $\ell_z = \sqrt{\kappa_z}$ is bounded by the maximum mixed-layer depth itself (600 m). This check avoids excessively large vertical length-scales in the mixed-layer region caused by small vertical derivatives in the tensor estimation method. Another check is used to prevent the vertical length-scales from exceeding the local ocean depth, which can occur in shallow-sea regions. The estimated horizontal tensor elements can be artificially large in semi-closed seas and close to boundaries. To prevent this, those elements are scaled

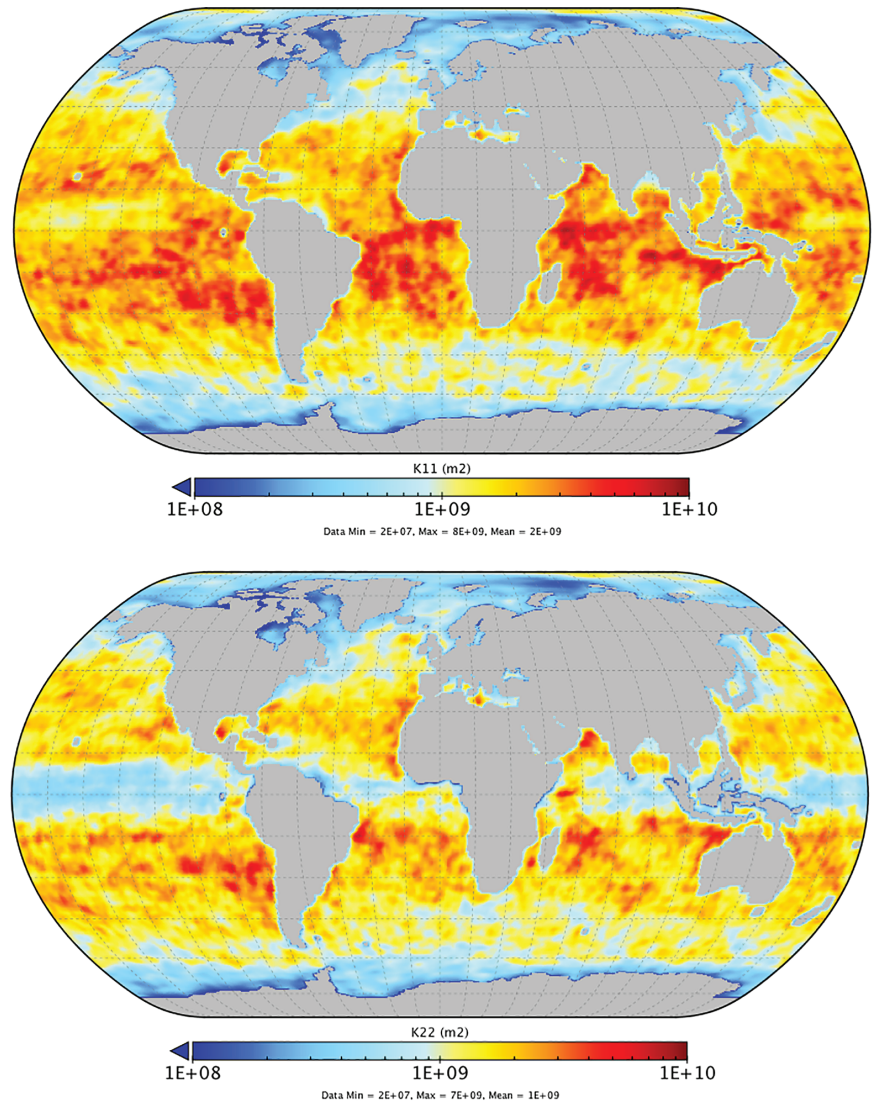
such that, at each grid point, the horizontal length-scale $\ell_h = \{\det(\boldsymbol{\kappa}_h)\}^{1/4}$ does not exceed the Euclidean distance to the nearest coastline point. A final check is made in order to eliminate length-scales that are smaller than the local grid resolution. In particular, if the estimated effective local length-scale is smaller than the geometric mean of the local scale factors, then the tensor elements are set to the scale factors themselves. A similar check is done on the vertical diffusion coefficient.

The estimated horizontal tensor elements for temperature in the surface level are shown in Figure 1. The tensor elements are proportional to the square of the correlation length-scales. While it is not the objective of this article to provide a detailed analysis of the estimates, various basic features can be identified. In the Tropics, κ_{22} is smaller than κ_{11} , reflecting well-known anisotropy related to equatorial dynamics. Smaller scales are also evident in western boundary current regions (Gulf Stream and Kuroshio Current, especially for temperature) where eddies and ocean fronts are prominent.

It is interesting to compare these estimates qualitatively with the parametrized horizontal tensor elements used for the actual production of ORAS5-LR. For ORAS5-LR, the same values were used for all variables and at all levels. The parametrized horizontal tensor elements correspond to values of the Daley length-scale equal to $D \approx 222\text{km}$ everywhere except within 15° of the Equator where the zonal (meridional) components were stretched (reduced) following a cosine dependence (Mogensen *et al.*, 2012). Directly at the Equator, $D_{\text{zonal}} \approx 444\text{ km}$ and $D_{\text{merid}} \approx 111\text{ km}$. Compared with the ensemble estimates in Figure 1, the parametrized tensor elements (not shown) are generally larger everywhere, with the exception of the equatorial region where the parametrized and estimated κ_{22} have some similarities. The horizontal Daley length-scales used in ECMWF's high-resolution ocean reanalysis ORAS5, produced with the global configuration ORCA025-Z75 (1/4 degree horizontal resolution and 75 vertical levels), was based on a different parametrization that depends on the Rossby radius. That parametrization produces horizontal tensor elements that decrease with increasing latitude, similar to what is observed in the ensemble-estimated values with ORAS5-LR.

The estimated vertical diffusion coefficient is shown in Figures 2 and 3. In the top level (Figure 2), the vertical length-scales are much larger in the Northern Hemisphere than in the Southern Hemisphere. The large scales are associated with strongly correlated errors in the mixed layer, which is deep in the Northern Hemisphere winter. The vertical length-scales are also large in the lowest levels of the model where the resolution is coarse, but much smaller in the highly stratified thermocline region (Figure 3).

FIGURE 1 The diagonal elements (a) κ_{11} and (b) κ_{22} (m^2) of the horizontal diffusion tensor for temperature at 5 m depth (the surface level), estimated from ensemble perturbations from months DJF in the period 2010–2014 of the ORAS5-LR reanalysis. The colour palette uses a \log_{10} scale



For comparison, the parametrized vertical diffusion coefficients used in ORAS5-LR were specified by setting the vertical Daley length-scales to be proportional to the vertical scale factors (e_3) where the proportionality constant was taken to be equal to one. This simple resolution-dependent parameterization broadly captures some features observed in the ensemble estimates, except for the mixed-layer region where the parametrized vertical scales are too small. The parametrized vertical scales in the deep ocean are also somewhat underestimated compared to those from the ensemble.

4 | NORMALIZATION FACTOR ESTIMATION METHODS

The constant normalization factors derived from the analytical solutions of the diffusion equation are only crude approximations of the actual normalization factors for diffusion-based correlation models used in practice. There

are three basic reasons for this. First, the actual normalization factors will depend on the numerical approximations used to discretize and solve the diffusion equation. Second, realistic diffusion tensors, such as those estimated from an ensemble, are not constant, but rather have considerable 3D spatial structure. Third, normalization factors are strongly affected by solid-wall boundary conditions, especially in a global ocean model where the boundary geometry is complex.

In the following subsections, we describe the different methods that have been implemented in NEMOVAR for estimating normalization factors, and we compare them in terms of their accuracy and computational cost.

4.1 | Brute-force (exact) method

Let $\mathbf{e}_n = (0, \dots, 1, \dots, 0)^T$ where the value of 1 is located at the n th grid-point. The n th diagonal element of any square matrix can be computed *exactly* by applying that

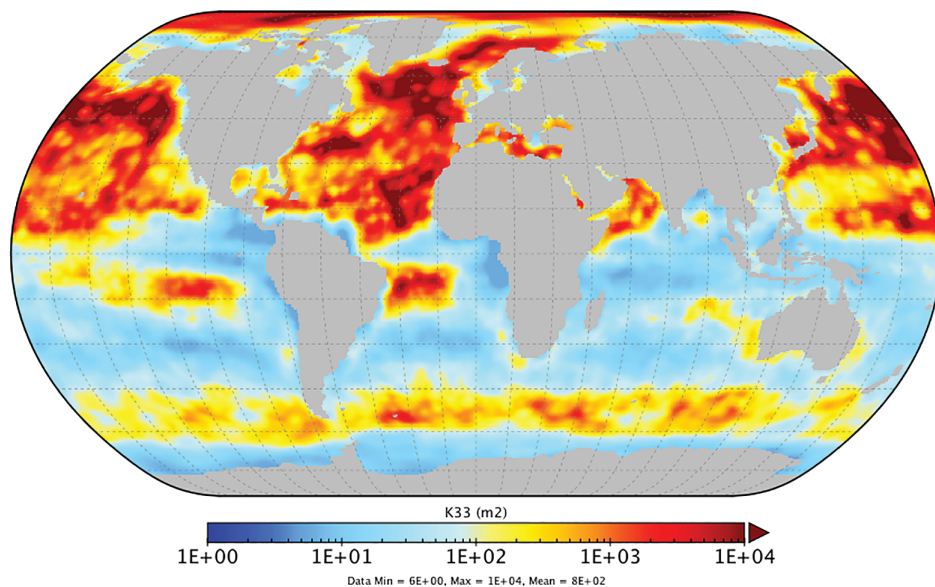


FIGURE 2 The vertical diffusion coefficient κ_z (m^2) for temperature at 5 m depth (the surface level) for months DJF. The colour palette uses a \log_{10} scale

matrix to \mathbf{e}_n and extracting the n th element from the resulting vector. Specifically, the inverse of the normalization factor at the n th grid-point of the symmetric diffusion matrix is given by (cf. Equation (4))

$$\gamma_n^{-2} = \{\mathbf{V} \mathbf{W}^{-1} \mathbf{V}^T \mathbf{e}_n\}_n, \quad (18)$$

where $\{\cdot\}_n$ means the n th element of the vector in curly brackets. For the coarse-grid formulation (Equation (5)), the expression would also involve the transfer operator and its transpose. To compute all normalization factors using Equation (18) requires as many applications of the diffusion operator as there are grid points, which is clearly prohibitive for a high-resolution global configuration where the number of ocean points for each 3D variable is on the order of 10^7 . The cost of applying the diffusion operator can be halved by noting that γ_n^{-2} is equal to the scalar product

$$\begin{aligned} \gamma_n^{-2} &= \mathbf{e}_n^T \mathbf{V} \mathbf{W}^{-1} \mathbf{V}^T \mathbf{e}_n \\ &= (\mathbf{W}^{-1/2} \mathbf{V}^T \mathbf{e}_n)^T (\mathbf{W}^{-1/2} \mathbf{V}^T \mathbf{e}_n). \end{aligned} \quad (19)$$

Equation (19) can be evaluated by applying the adjoint of the diffusion operator over half the number of diffusion steps and then computing the scalar product of the result with itself. In this case, half of the diffusion operations per grid point are replaced by a scalar product (global sum) per grid point. This is generally cheaper than using Equation (18) directly, but still prohibitively expensive to be applied at all grid points. In practice, this method can only be used to compute exact normalization factors at a selected number of points, which can be useful for diagnostic purposes.

4.2 | Randomization method

The randomization method was proposed by Andersson *et al.* (2000) for diagnosing the variances of any covariance matrix that can be represented in square-root operator form. Weaver and Courtier (2001) discussed how the method could be used specifically for estimating normalization factors, which we summarize below.

Consider a random vector ϵ of normally-distributed noise with mean equal to zero and standard deviation equal to one; that is, $\epsilon \sim N[\mathbf{0}, \mathbf{I}]$ where $E[\epsilon] = \mathbf{0}$ and $E[\epsilon \epsilon^T] = \mathbf{I}$, $E[\cdot]$ being the mathematical expectation operator. We can relate ϵ to a symmetric positive semi-definite (covariance) matrix $\mathbf{A} \mathbf{A}^T$ using the identity

$$\mathbf{A} \mathbf{A}^T = \mathbf{A} \mathbf{I} \mathbf{A}^T = \mathbf{A} E[\epsilon \epsilon^T] \mathbf{A}^T = E[\mathbf{A} \epsilon (\mathbf{A} \epsilon)^T].$$

We can use the above identity to compute a sample estimate of the covariance matrix $\mathbf{A} \mathbf{A}^T$ from a set of independently drawn random vectors $\epsilon_q \sim N[\mathbf{0}, \mathbf{I}]$, $q = 1, \dots, Q$.

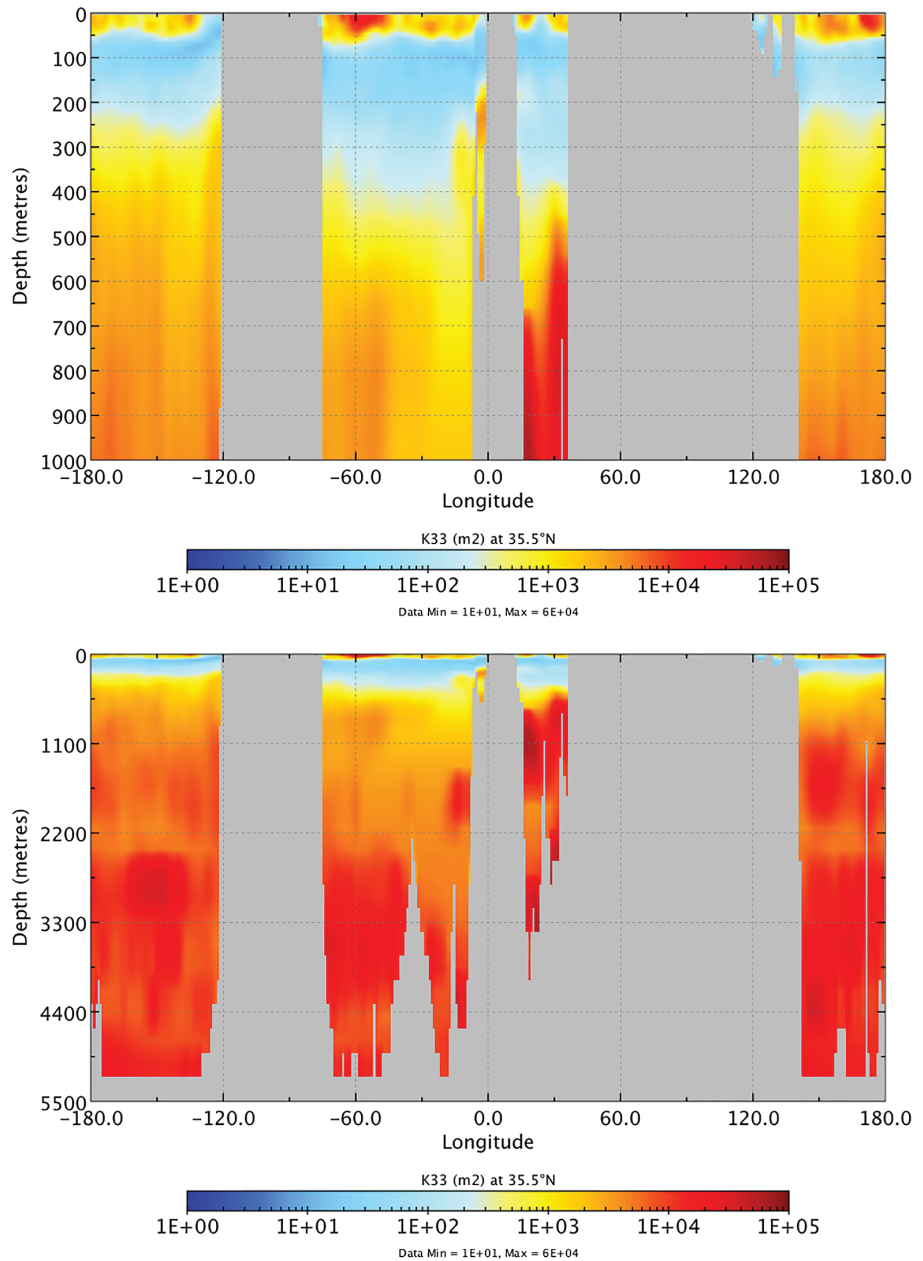
Here, we are interested in estimating the variances (diagonal elements) of $\mathbf{A} \mathbf{A}^T$ for the specific case where $\mathbf{A} = \mathbf{V} \mathbf{W}^{-1/2}$. Let $\mathbf{v} = (v_1, \dots, v_N)^T$ be the vector containing the variances. An unbiased sample estimate $\tilde{\mathbf{v}} = (\tilde{v}_1, \dots, \tilde{v}_N)^T$ of \mathbf{v} is

$$\tilde{\mathbf{v}} = \frac{1}{Q-1} \sum_{q=1}^Q \left(\zeta_q - \frac{1}{Q} \sum_{q=1}^Q \zeta_q \right) \circ \left(\zeta_q - \frac{1}{Q} \sum_{q=1}^Q \zeta_q \right),$$

where \circ denotes the element-by-element (Schur) product of two vectors, and

$$\zeta_q = \mathbf{V} \mathbf{W}^{-1/2} \epsilon_q. \quad (20)$$

FIGURE 3 A vertical section of κ_z (m^2) at approximately 35°N . (b) is an expansion of the top 1,000 m of (a). The colour palette uses a \log_{10} scale



The randomized estimate of the matrix of normalization factors is then

$$\mathbf{\Gamma}^2 \approx \text{diag}(\tilde{v}_1^{-1}, \dots, \tilde{v}_N^{-1}).$$

The sampling error $\tilde{\mathbf{v}}^e = \tilde{\mathbf{v}} - \mathbf{v}$ is unbiased ($E[\tilde{\mathbf{v}}^e] = \mathbf{0}$) and the statistical moments can be expressed in terms of \mathbf{v} (Ménétrier *et al.*, 2015). In particular, for each component v_i , the mean squared error of the estimate is

$$E[(\tilde{v}_i - v_i)^2] = \left(\frac{2}{Q-1} \right) v_i^2, \quad (21)$$

which shows that the relative error of the variance depends on the sample size but not on the true variance,

and hence is independent of the resolution of the model. For example, the relative error is 0.02% with a sample size of $Q = 10^4$. Note that Equation (21) describes the expected mean squared error of the *variance*, not the normalization factor (inverse of the variance) for which there is no simple relationship.

Randomization is also expensive, but not as prohibitively expensive as the exact method. Randomization is practical for computing accurate estimates of the normalization factors when using correlation model parameters that are static or updated infrequently. Furthermore, there is scope for accelerating the convergence of randomization using carefully constructed input vectors obtained using Hadamard matrices or probing techniques (Bekas *et al.*, 2007; Laeuchli, 2016). While

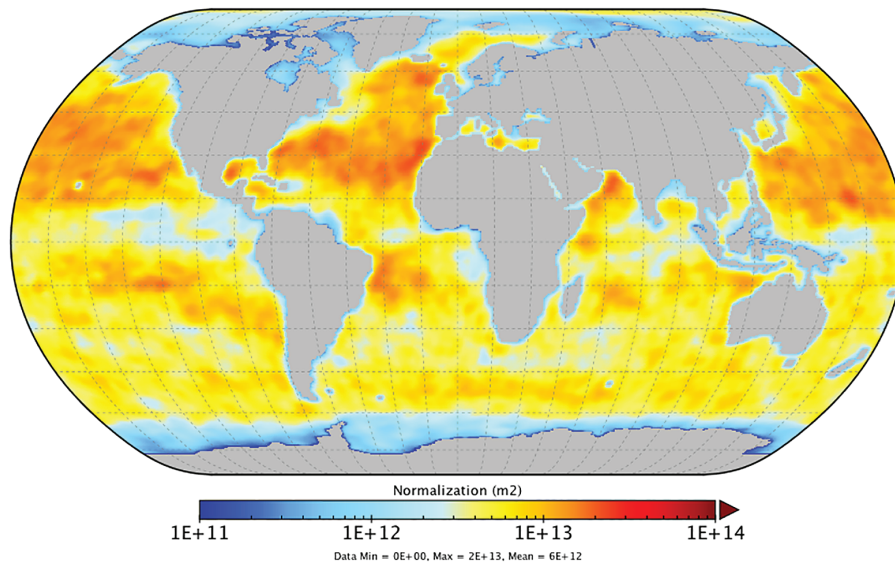


FIGURE 4 The reference normalization factors (m^2) for temperature at 5 m depth (the surface level) with the 2D×1D (3) implicit diffusion operator where the diffusion tensor has been estimated from a climatological ensemble as described in Section 3 and illustrated in Figure 1. The normalization factors have been computed using the randomization method with a sample size of 10^4 . The colour palette uses a \log_{10} scale

this is an interesting possibility, we have not explored it here.

4.3 | Reference field and error diagnostics

In what follows, we will use the normalization factors computed using randomization with a sample size of 10^4 as the reference field to which estimates generated using approximate methods will be compared. The reference normalization factors for temperature with the standard 2D×1D (3) formulation (Equation (10) with $m=3$) are shown in Figures 4 and 5. The spatial structure is similar to that observed in the diffusion tensor elements (Figures 1–3), which illustrates the close connection between the normalization factors and correlation length-scales. Indeed, this is expected from the analytical expression for the normalization factor. Chabot *et al.* (2017) make a similar observation with a wavelet-based covariance model. Also noticeable are the reduced values of the normalization factors close to land boundaries, which counteract an overestimated amplitude caused by the Neumann boundary condition (Mirouze and Weaver, 2010).

The experiments described in the following sections are summarized in Table 1. To assess the quality of the approximate normalization factor $\tilde{\gamma}_n^2$, we will compute the relative error

$$\mathcal{E}_n = \frac{\tilde{\gamma}_n^2 - \gamma_n^2}{\gamma_n^2},$$

where γ_n^2 is the reference. Spatial plots of \mathcal{E}_n will be displayed only for the temperature normalization factors with the 2D×1D (3) formulation. To provide a global measure

of the error, we will also calculate the mean of the absolute value of the relative error (hereafter referred to simply as the mean error),

$$\mathcal{E} = \frac{1}{N} \sum_{n=0}^N |\mathcal{E}_n|, \quad (22)$$

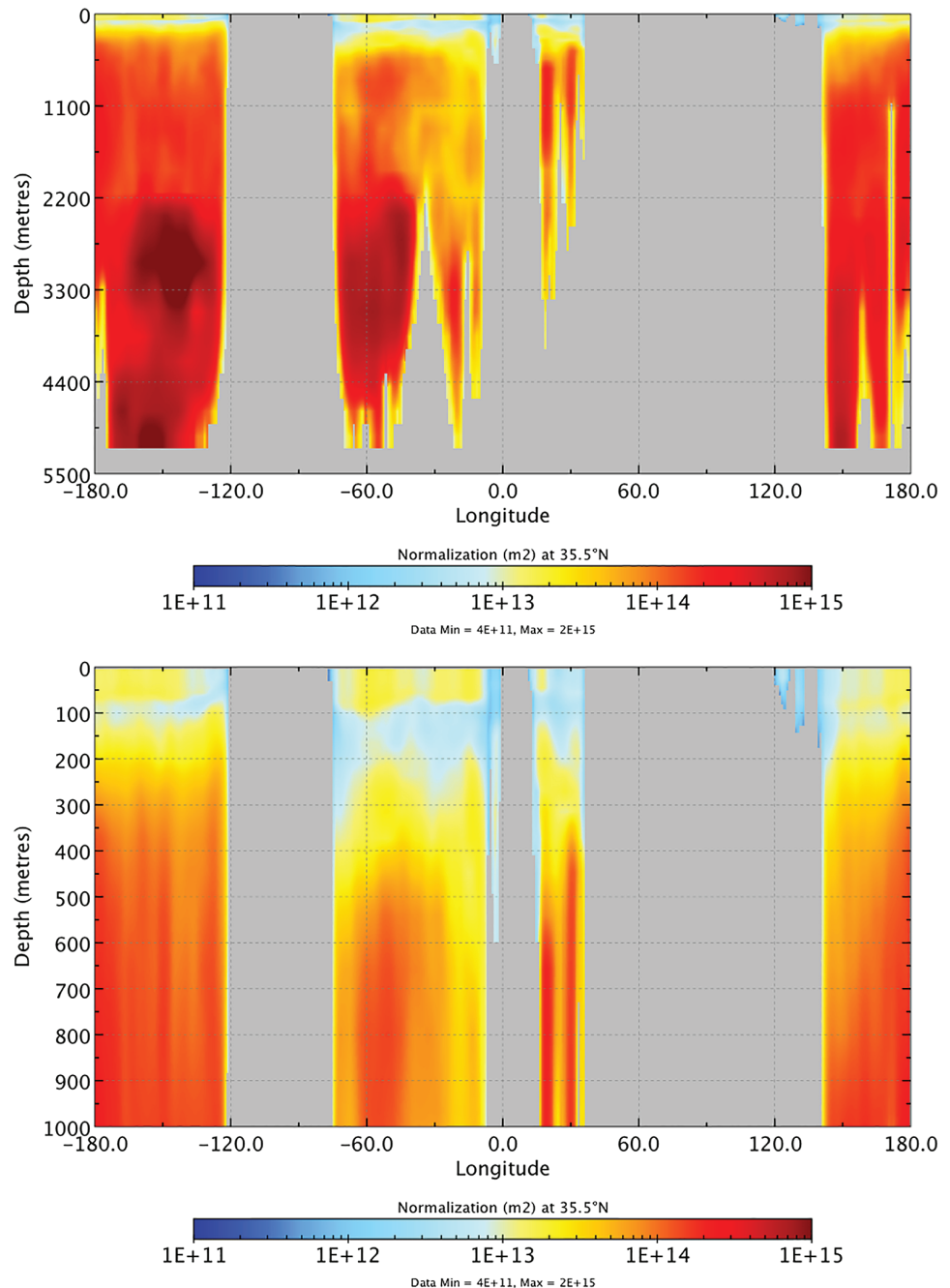
for all the control variables (unbalanced SSH, temperature, unbalanced salinity). The values of \mathcal{E} for the experiments in Table 1 are given in Table 2 where they are expressed as a percentage error.

4.4 | Randomization with a small sample size

A sample size of 100 could be affordable on each assimilation cycle to estimate normalization factors for a fully flow-dependent background-error correlation matrix. The mean relative error with 100 samples is 12% (Experiment 1 in Table 2). However, the randomized estimates are noisy as evident in Figure 6 which displays \mathcal{E}_n for the temperature normalization factors in the surface level. In places, the error exceeds 100%. A spatial filter can be applied to the randomized estimates to suppress the noise to some extent. We can use the diffusion operator as a filter and adjust the diffusion tensor to select a desired smoothing scale. Here, we have applied a horizontal diffusion filter to the randomized estimates of the variances (inverse normalization factors). At each grid point n , the diffusion tensor is taken to be proportional to the diagonal tensor of squared horizontal scale factors:

$$\{\mathbf{\kappa}_h^{\text{ranfil}}\}_n = \left(\frac{\alpha^2}{2M-4} \right) \text{diag}(\{e_1^2\}_n, \{e_2^2\}_n). \quad (23)$$

FIGURE 5 A vertical section at approximately 35°N of the reference normalization factors (m^2) for temperature. (b) is an expansion of (a) in the top 1,000 m. The colour palette uses a \log_{10} scale



As in the correlation operator, we set $M = 10$, which makes the filter approximately Gaussian. While the filter reduces the noise, it has a detrimental effect near land boundaries where the errors are noticeably increased (not shown). The normalization factors are overestimated at grid points closest to the boundary, and underestimated at grid points further from the boundary. The larger errors near the boundaries are responsible for an increase in the average error (Experiments 2 and 3 in Table 2). In Experiment 2, the filter scale is taken to be equal to the local scale factor ($\alpha = 1$). Doubling the filter scale ($\alpha = 2$) suppresses the noise further but exacerbates the problem near land

boundaries and leads to a further increase in the error (Experiment 3).

4.5 | Analytical and empirical methods

4.5.1 | Zeroth-order analytical approximation

An approximate expression for the normalization factors can be derived from the analytical expression for the normalization factors by replacing the constant diffusion

TABLE 1 A summary of experiments for testing approximate methods for estimating normalization factors

Expt	Method	Sample size	Rand. filter	α	Diffusion (m)	BC	Bndy	Smooth	Bias
1	Ran	100	No	–	2D×1D (3)	N	–	–	–
2	Ran	100	Yes	1	2D×1D (3)	N	–	–	–
3	Ran	100	Yes	2	2D×1D (3)	N	–	–	–
4	Ana	–	–	–	2D×1D (3)	N	No	No	No
5	Ana	–	–	–	3D	N	No	No	No
6	Ana	–	–	–	2D×1D (3)	ND	–	No	No
7	Ana	–	–	–	2D×1D (3)	NDz	–	No	No
8	Ana	–	–	–	3D	ND	–	No	No
9	Ana	–	–	–	2D×1D (3)	N	Yes	No	No
10	Ana	–	–	–	3D	N	Yes	No	No
11	Ana	–	–	–	2D×1D (3)	N	Yes	Yes	No
12	Ana	–	–	–	3D	N	Yes	Yes	No
13	Ana	–	–	–	2D×1D (3)	N (NDz)	Yes	Yes	No
14	Ana	–	–	–	3D	N (ND)	Yes	Yes	No
15	Ana	–	–	–	2D×1D (3)	N (NDz)	Yes	Yes	Yes
16	Ran (hxz)	10,000	–	–	2D×1D (2)	N	–	–	–
17	Ran (hxz)	10,000	–	–	2D×1D (3)	N	–	–	–
18	Ran (hxz)	10,000	–	–	2D×1D (3)	NDz	–	–	–

Note: ‘Expt’ is the experiment identifier. ‘Method’ indicates whether the method is randomization (Ran) or analytically-based (Ana). Randomization estimates that are computed separately for the horizontal and vertical components are identified by (hxz). Sample size, whether a horizontal diffusion filter has been applied, and the length-scale factor (α) for the diffusion filter (Equation (23)) are specific to randomization. The sixth column indicates the type of 3D diffusion formulation where the value of m associated with the family of $2D \times 1D$ formulations (Equation (10)) is indicated in parentheses. BC denotes the boundary condition applied in the correlation model: Neumann (N); Dirichlet (D); average of Neumann and Dirichlet for full diffusion (ND) or vertical diffusion only (NDz). The boundary condition used for smoothing the analytical normalization factors is the same as that used for the correlation model, except when an alternative boundary condition is indicated in parentheses. The last three columns indicate whether the boundary correction factor, analytical normalization factor smoothing, and bias correction using linear regression are applied.

tensor with the spatially dependent diffusion tensor. We refer to this as the zeroth-order approximation. For 2D diffusion and the two basic formulations of 3D diffusion, the zeroth-order expressions are

$$\left. \begin{aligned} \{\hat{\gamma}_h^2\}_n &= \omega_h \sqrt{\det(\{\kappa_h\}_n)}, \\ \{\hat{\gamma}_{hxz}^2\}_n &= (\omega_h \sqrt{\det(\{\kappa_h\}_n)}) (\omega_z \sqrt{\{\kappa_z\}_n}) \\ &= \{\hat{\gamma}_h^2\}_n \{\hat{\gamma}_z^2\}_n, \\ \{\hat{\gamma}_{hz}^2\}_n &= \omega_{hz} \sqrt{\{\kappa_z\}_n} \sqrt{\det(\{\kappa_h\}_n)}. \end{aligned} \right\} \quad (24)$$

We can expect these expressions to provide a good approximation of the normalization factors in regions far from land boundaries and where the spatial variations in the diffusion tensor are weak. However, these expressions are unlikely to be adequate in regions close to boundaries and where the diffusion tensor varies substantially

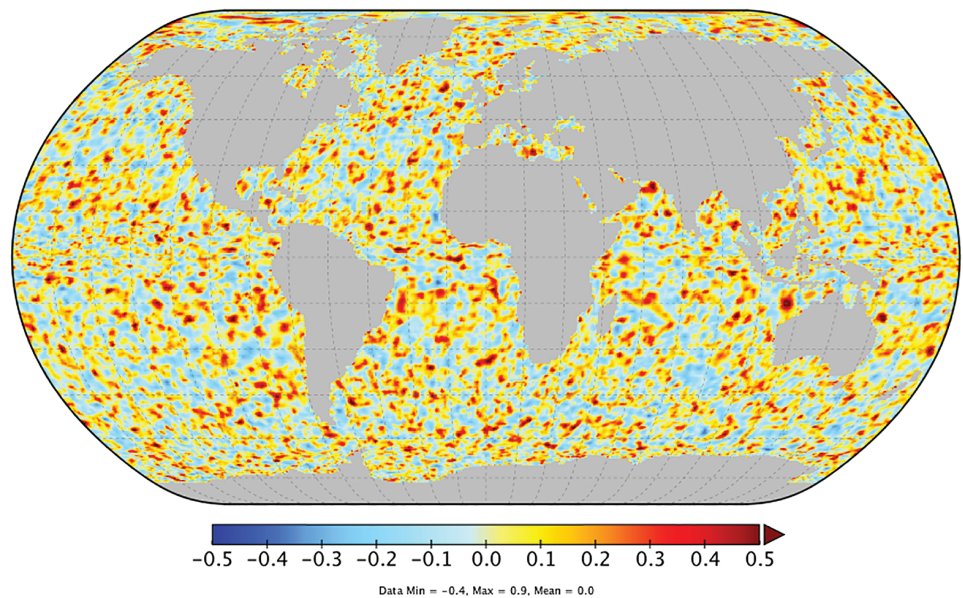
between grid points, as occurs with a tensor estimated from ensembles. This is confirmed by Figures 7 and 8, which show the relative error in the normalization factors for temperature when using Equation (26). The spatial structure of the error roughly matches that of the vertical diffusion coefficient in Figures 2 and 3. The normalization factors are largely overestimated in the upper ocean levels and near bottom bathymetry (the errors are positive) where they are strongly influenced by the Neumann boundary condition, the effects of which are not accounted for in the analytical approximation. The errors are especially large (over 1,000%) in the upper levels in the Northern Hemisphere where the vertical diffusion coefficients are large. The mean errors in Table 2 (Experiments 4 and 5) are significantly larger for temperature and unbalanced salinity (3D fields) than for unbalanced SSH (a 2D field).

TABLE 2 The spatial average of the absolute value of the errors in the approximate normalization factor normalized by the reference normalization factor (Equation (22)) for the experiments listed in Table 1

Expt	Unbalanced SSH	Temperature	Unbalanced Salinity
1	12	12	12
2	14	15	15
3	14	16	16
4	12	53	72
5	12	53	72
6	12	31	40
7	12	32	40
8	12	32	40
9	9	24	33
10	9	24	33
11	13	26	34
12	13	33	42
13	13	21	23
14	13	27	29
15	7	23	45
16	–	7	8
17	–	4	4
18	–	4	4

Note: The values are displayed as percentage errors.

FIGURE 6 Experiment 1. The relative error between the reference normalization factors for temperature in the surface level (Figure 4) with the 2D×1D (3) implicit diffusion operator and those produced using the randomization method with a sample size of 100



4.5.2 | Diffusion formulation with Neumann and Dirichlet boundary conditions

Near land boundaries, the Neumann boundary condition prevents flux exchanges across the boundary, causing the amplitude of the covariance function to increase there.

The opposite occurs with a Dirichlet boundary condition where the amplitude is diminished near land boundaries in order to satisfy the condition that the field is zero at the boundary. By examining the solution of the continuous, 1D diffusion equation with constant diffusion coefficient, in the presence of an isolated, straight boundary, Mirouze and Weaver (2010) show analytically that the

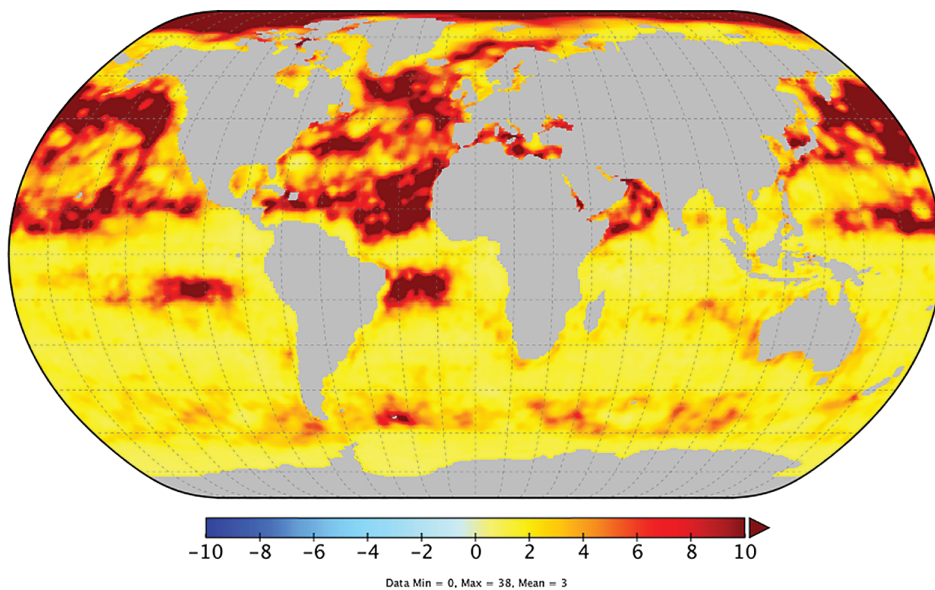


FIGURE 7 Experiment 4. The relative error in the normalization factors for the 2D×1D (3) implicit diffusion operator for temperature in the surface level when using the analytical approximation given by Equation (24)

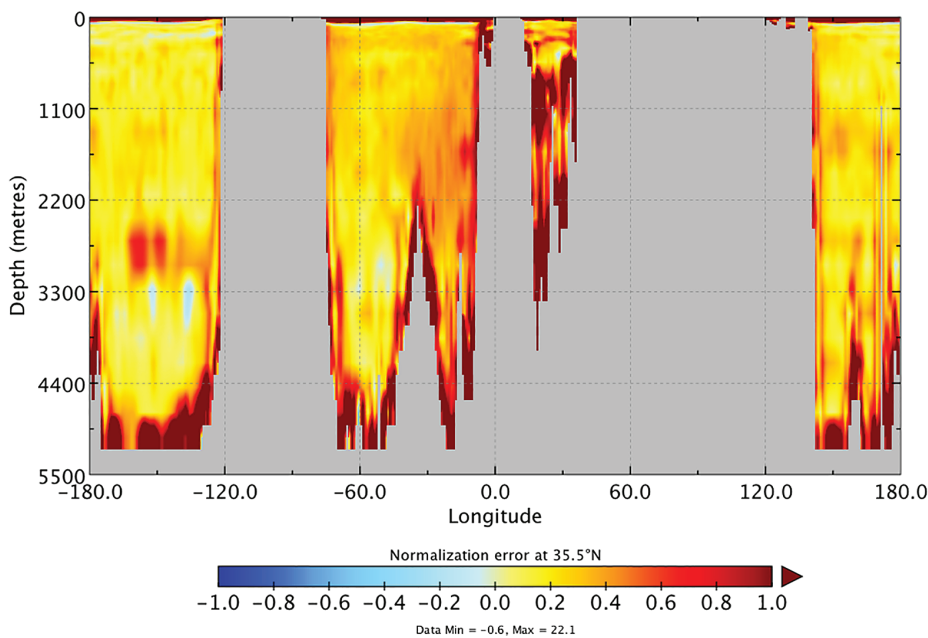


FIGURE 8 Experiment 4. The relative error in the normalization factors for the 2D×1D (3) implicit diffusion operator for temperature at approximately 35°N when using the analytical approximation given by Equation (24). Note that the scale in the colour palette is one order of magnitude smaller than that used in Figure 7

correct amplitude can be obtained by redefining the correlation operator as an average of two diffusion operators, one that employs the Neumann (N) boundary condition and the other that employs the Dirichlet (D) boundary condition. We call it the ND formulation.

The horizontal implicit diffusion operator with the ND formulation can be written as

$$\mathbf{L}_h^{\text{ND}} = \frac{1}{2}(\mathbf{L}_{h,N} + \mathbf{L}_{h,D}) = \frac{1}{2}(\mathbf{F}_{h,N}^{M_h} + \mathbf{F}_{h,D}^{M_h}),$$

where the different boundary conditions are denoted by subscripts N and D. The corresponding ND formulation of the correlation matrix involves a rectangular “square-root”

factor:

$$\mathbf{C}_h = \mathbf{\Gamma}_h \mathbf{V}_h \overline{\mathbf{W}}_h^{-1} \mathbf{V}_h^T \mathbf{\Gamma}_h,$$

where

$$\mathbf{V}_h = \left(\frac{1}{\sqrt{2}} \mathbf{F}_{h,N}^{M_h/2}, \frac{1}{\sqrt{2}} \mathbf{F}_{h,D}^{M_h/2} \right) \text{ and } \overline{\mathbf{W}}_h = \text{diag}(\mathbf{W}_h, \mathbf{W}_h).$$

A similar expression follows for the ND formulation of the correlation matrix with the 3D implicit diffusion operator (Section 2.2.3).

For the 2D×1D formulations, we use the approach outlined in Section 2.2.2 to construct a self-adjoint 3D diffusion operator $\mathbf{L}_{h \times z}^{\text{ND}}$ (cf. Equation (9)) from the product

of \mathbf{L}_h^{ND} and the ND formulation of the vertical diffusion operator:

$$\mathbf{L}_z^{\text{ND}} = \frac{1}{2}(\mathbf{L}_{z,N} + \mathbf{L}_{z,D}) = \frac{1}{2}(\mathbf{F}_{z,N}^{M_z} + \mathbf{F}_{z,D}^{M_z}).$$

This yields a family of 3D correlation matrices ($m = 1, \dots, 4$):

$$\mathbf{C}_{\text{hxz}} = \mathbf{\Gamma}_{\text{hxz}} \mathbf{V}_{\text{hxz}}^{(m)} \overline{\mathbf{W}}_{\text{hz}}^{-1} (\mathbf{V}_{\text{hxz}}^{(m)})^T \mathbf{\Gamma}_{\text{hxz}}$$

where

$$\overline{\mathbf{W}}_{\text{hz}} = \text{diag}(\mathbf{W}_{\text{hz}}, \mathbf{W}_{\text{hz}}, \mathbf{W}_{\text{hz}}, \mathbf{W}_{\text{hz}}).$$

In particular, for the 2D×1D formulations without interleaving ($m = 1$) and with interleaving ($m = 3$), the square-root factors are

$$\mathbf{V}_{\text{hxz}}^{(1)} = \left(\frac{1}{2} \mathbf{F}_{h,N}^{M_h/2} \mathbf{F}_{z,N}^{M_z/2}, \frac{1}{2} \mathbf{F}_{h,N}^{M_h/2} \mathbf{F}_{z,D}^{M_z/2}, \right. \\ \left. \frac{1}{2} \mathbf{F}_{h,D}^{M_h/2} \mathbf{F}_{z,N}^{M_z/2}, \frac{1}{2} \mathbf{F}_{h,D}^{M_h/2} \mathbf{F}_{z,D}^{M_z/2} \right)$$

and

$$\mathbf{V}_{\text{hxz}}^{(3)} = \left(\frac{1}{2} (\mathbf{F}_{h,N} \mathbf{F}_{z,N})^{\frac{M_{\text{hz}}}{2}}, \frac{1}{2} (\mathbf{F}_{h,N} \mathbf{F}_{z,D})^{\frac{M_{\text{hz}}}{2}}, \right. \\ \left. \frac{1}{2} (\mathbf{F}_{h,D} \mathbf{F}_{z,N})^{\frac{M_{\text{hz}}}{2}}, \frac{1}{2} (\mathbf{F}_{h,D} \mathbf{F}_{z,D})^{\frac{M_{\text{hz}}}{2}} \right),$$

which involves four 2D×1D diffusion operations compared to two 3D diffusion operators in the corresponding (non-separable) 3D ND formulation. Similar expressions hold for $m = 2$ and $m = 4$, by switching the order of the horizontal and vertical diffusion operators.

The ND formulation can be restricted to the vertical diffusion operator only, which we call the NDz formulation. Assuming Neumann boundary conditions for the horizontal diffusion operator, this yields for $m = 1$ and $m = 3$

$$\left. \begin{aligned} \mathbf{V}_{\text{hxz}}^{(1)} &= \left(\frac{1}{\sqrt{2}} \mathbf{F}_{h,N}^{M_h/2} \mathbf{F}_{z,N}^{M_z/2}, \frac{1}{\sqrt{2}} \mathbf{F}_{h,N}^{M_h/2} \mathbf{F}_{z,D}^{M_z/2} \right), \\ \mathbf{V}_{\text{hxz}}^{(3)} &= \left(\frac{1}{\sqrt{2}} (\mathbf{F}_{h,N} \mathbf{F}_{z,N})^{M_{\text{hz}}/2}, \frac{1}{\sqrt{2}} (\mathbf{F}_{h,N} \mathbf{F}_{z,D})^{M_{\text{hz}}/2} \right) \end{aligned} \right\} (25)$$

and

$$\overline{\mathbf{W}}_{\text{hz}} = \text{diag}(\mathbf{W}_{\text{hz}}, \mathbf{W}_{\text{hz}}),$$

which involves two diffusion operations as in the 3D ND formulation.

The mean errors with and without the ND formulation are reduced by 22% for temperature and 32% for unbalanced salinity, but are similar for unbalanced SSH

(Experiments 6, 7 and 8 in Table 2). For unbalanced SSH, the errors remain large *near* land boundaries, probably due to a combination of the effects of irregular boundary geometry and large spatial variations in the tensor. For temperature and unbalanced s , the errors are predominantly reduced in the upper ocean and near bathymetry (cf. Figures 7–10). The mean errors for the experiments with the full ND and NDz formulations (cf. Experiments 6 and 7) are very similar, indicating that the ND formulation works best with the vertical diffusion operator. This is likely because the assumptions used to derive the ND formulation are better satisfied by the vertical diffusion operator, especially in open ocean areas near the surface where boundary geometry is simple and where the vertical diffusion coefficient is approximately constant. Overall, however, the errors remain unacceptably large.

4.5.3 | Analytical boundary correction

An obvious drawback with the ND formulation is the additional cost that it entails. Building on the theoretical analysis of Mirouze and Weaver (2010), Mirouze and Storto (2016) proposed a simple analytical correction to the normalization factor near the boundary as an alternative to the more costly ND approach. With the Neumann boundary condition, their analysis suggests that the normalization coefficient should be corrected by a multiplicative factor $\xi = 1/\{1 + c(2 r^{\text{bdy}})\}$ where $c(\cdot)$ is the Matérn correlation function corresponding to the approximate kernel of the normalized diffusion operator and r^{bdy} is the Euclidean distance to the boundary. For example, directly at the boundary ($r^{\text{bdy}} = 0$), ξ equals 1/2 to compensate for the doubling of the amplitude there with the Neumann boundary condition for the 1D case when the diffusion coefficient is constant (appendix B in Mirouze and Weaver (2010)). Far from the boundary ($r^{\text{bdy}} \gg \ell$), the correction factor is approximately equal to one. With the Dirichlet boundary condition, the correction factor is $\xi = 1/\{1 - c(2 r^{\text{bdy}})\}$, which is singular directly at the boundary point and therefore must be regularized in numerical applications by adding a small number to the denominator.

The correction for the Neumann boundary condition in the 2D implicit diffusion operator yields approximate normalization factors

$$\{\tilde{\gamma}_h^2\}_n = \frac{\{\hat{\gamma}_h^2\}_n}{\{1 + c_h(2 r_n^{\text{bdy}})\}},$$

where c_h is the correlation function implied by the diffusion operator, and r_n^{bdy} is the Euclidean distance to the nearest coastline point. For 2D implicit diffusion, the correlation function used to compute the correction factor

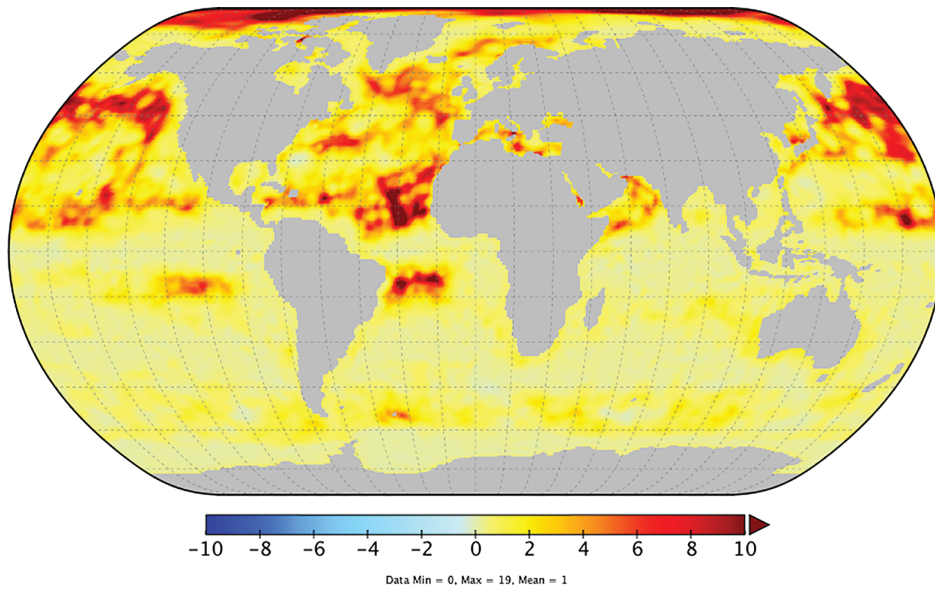


FIGURE 9 Experiment 6. The relative error in the normalization factors for temperature in the surface level when using the analytical approximation given by Equation (24) with a ND formulation of the 2D×1D (3) implicit diffusion operator

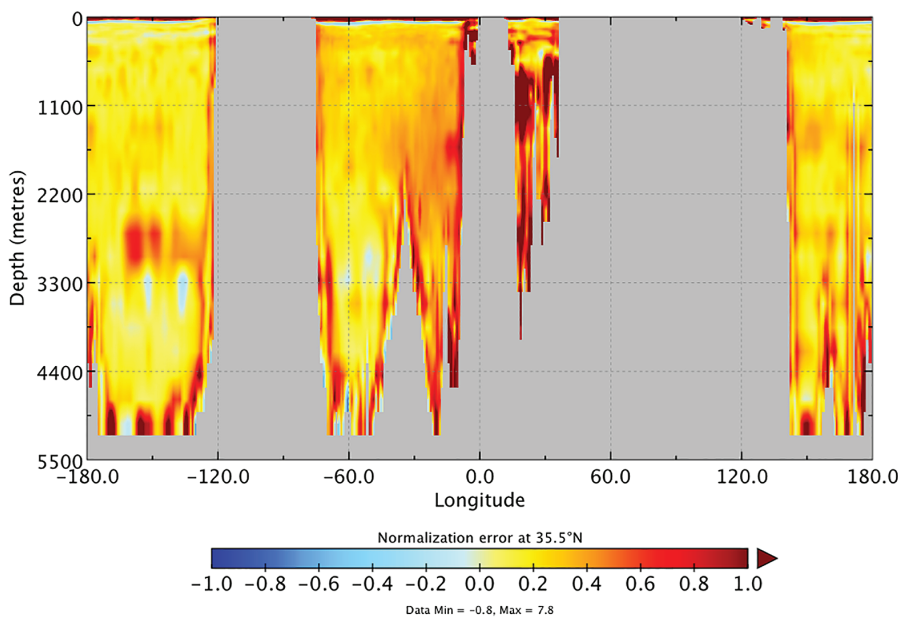


FIGURE 10 Experiment 6. The relative error in the normalization factors for temperature at approximately 35°N when using the analytical approximation given by Equation (24) with a ND formulation of the 2D×1D (3) implicit diffusion operator. Note that the scale in the colour palette is one order of magnitude smaller than that used in Figure 9

should be a Matérn function involving the modified Bessel function of the second kind (Weaver and Mirouze, 2013). For simplicity, we approximate this function by a Gaussian function, which is adequate for the value of $M_h = 10$ considered.

For the 2D×1D diffusion formulations, the correction gives approximate normalization factors

$$\{\tilde{\gamma}_{h \times z}^2\}_n = \frac{\{\hat{\gamma}_{h \times z}^2\}_n}{\{1 + c_h(2r_n^{\text{bdy}})\}\{1 + c_z(2z_n^{\text{bdy}})\}}, \quad (26)$$

where c_z is an M_z th-order AR function and z_n^{bdy} is the Euclidean distance to the nearest upper or bottom boundary point. The correlation function c_h is the same as the one used in the correction factor for the 2D diffusion

operator. For the 3D diffusion operator, the corrected normalization factors are given by Equation 26 with $\{\hat{\gamma}_{h \times z}^2\}_n$ in place of $\{\hat{\gamma}_{h \times z}^2\}_n$.

For unbalanced SSH, the boundary correction term results in a significant reduction in the error near land boundaries compared to both the zeroth-order and ND formulations (not shown). This improved error reduction is also visible in the mean error (Experiments 9 and 10 in Table 2). The reduction of the mean error is more impressive for temperature and unbalanced salinity where it is up to 8% better than with the ND and NDz formulations (compared to 3% better for unbalanced SSH). Figures 11 and 12 illustrate that the boundary correction term has a similar impact to the ND formulation near the surface, but produces noticeably smaller errors near

FIGURE 11 *Experiment 9.* The relative error in the normalization factors for the 2D×1D (3) implicit diffusion operator for temperature in the surface level when using the analytical approximation (Equation (26)), involving the boundary correction term to account for the effect of the Neumann boundary condition

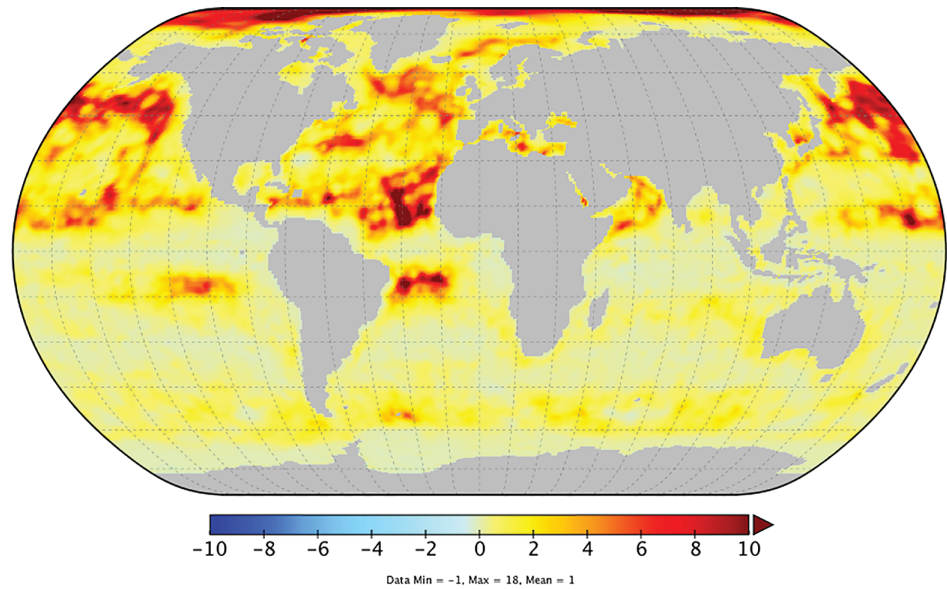
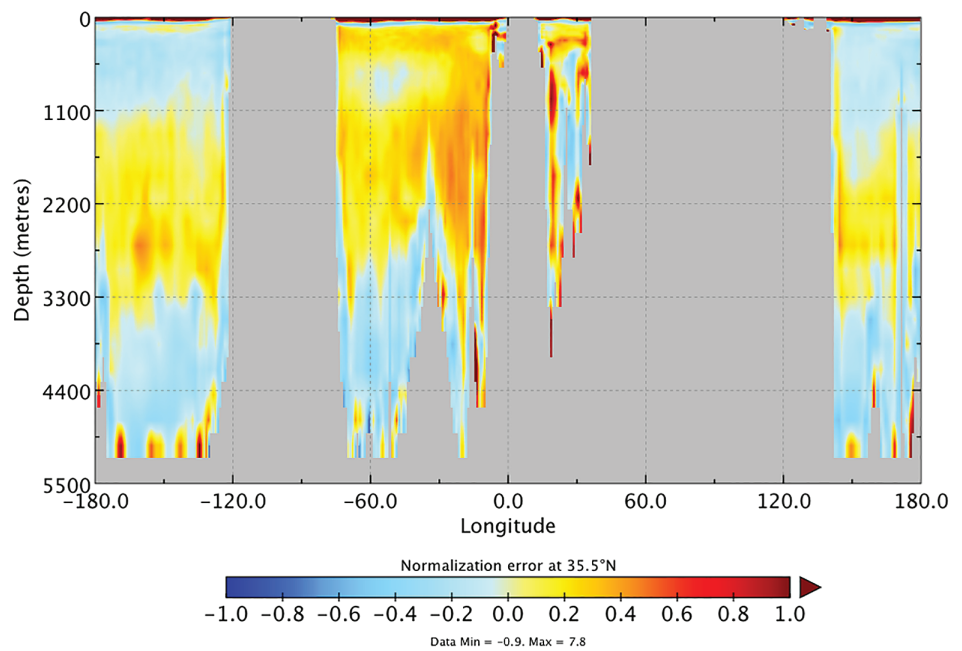


FIGURE 12 *Experiment 9.* The relative error in the normalization factors for the 2D×1D (3) implicit diffusion operator for temperature at approximately 35°N when using the analytical approximation (Equation (26)), involving the boundary correction term to account for the effect of the Neumann boundary condition. Note that the scale in the colour palette is one order of magnitude smaller than that used in Figure 11



bottom bathymetry. The errors also change sign over large areas in the ocean interior, probably due to the large vertical length scales in the mixed layer and bottom boundary regions that are used in the correction term.

4.5.4 | Smoothing the diffusion tensor in the zeroth-order expression

By considering an asymptotic expansion of the solution of the diffusion equation for the case of a slowly and smoothly varying diffusion tensor, Purser *et al.* (2003b) showed that a better approximation can be made by smoothing $\det(\kappa)$ in the analytical expression of the

normalization factor. In particular, their analysis suggests that $\det(\kappa)$ should be smoothed with the square root of the diffusion operator that is used in the correlation operator. For a Gaussian filter, applying the square root of the diffusion operator is equivalent to applying the complete diffusion operator, but with the tensor diffusivity equal to one half of that used in the correlation operator. Rather than using precisely one half of the tensor diffusivity, Yaremchuk and Carrier (2012) suggest treating this factor as a free parameter that can be tuned empirically. Also, they apply the smoothing operator to the inverse of the analytical normalization factor (the filter variance), which is equivalent to smoothing the inverse of the square root of $\det(\kappa)$. We found that this approach can produce

numerical artifacts because the smoothed filter variances could attain excessively small values. Smoothing the analytical normalization factors directly (not their inverse) produces better results in our experiments. This is similar to what was proposed by Purser *et al.* (2003b), but here we smooth the *square root* of $\det(\kappa)$ instead of $\det(\kappa)$.

Taking into account both smoothing of the normalization factors and the Neumann boundary condition correction term, the approximate normalization factors for the 2D implicit diffusion operator are

$$\{\tilde{\gamma}_h^2\}_n = \frac{\{\mathbf{L}_h^s \hat{\mathbf{n}}_h\}_n}{\{1 + c_h(2 r_n^{\text{bdy}})\}}, \quad (27)$$

where $\hat{\mathbf{n}}_h = (\{\hat{\gamma}_h^2\}_1, \dots, \{\hat{\gamma}_h^2\}_N)^T$, \mathbf{L}_h^s is the 2D diffusion operator (without normalization) used in \mathbf{C}_h but with the tensor diffusivity given by

$$\{\kappa_h^s\}_n = \beta \{\kappa_h\}_n, \quad (28)$$

where β is a tunable parameter with $0 < \beta < 1$. Likewise, for the 3D implicit diffusion operator, the approximate normalization factors become

$$\{\tilde{\gamma}_{hz}^2\}_n = \frac{\{\mathbf{L}_{hz}^s \hat{\mathbf{n}}_{hz}\}_n}{\{1 + c_h(2 r_n^{\text{bdy}})\} \{1 + c_z(2 z_n^{\text{bdy}})\}}, \quad (29)$$

where $\hat{\mathbf{n}}_{hz} = (\{\hat{\gamma}_{hz}^2\}_1, \dots, \{\hat{\gamma}_{hz}^2\}_N)^T$, \mathbf{L}_{hz}^s is the 3D diffusion operator (without normalization) used in \mathbf{C}_{hz} , but with the tensor diffusivity reduced according to Equation 28.

For the 2D×1D formulation, we use a slightly different procedure in which the smoothing is performed separately on the horizontal and vertical tensor contributions to the analytical normalization factor:

$$\{\tilde{\gamma}_{hxz}^2\}_n = \frac{\{\mathbf{L}_h^s \hat{\mathbf{n}}_h\}_n \{\mathbf{L}_z^s \hat{\mathbf{n}}_z\}_n}{\{1 + c_h(2 r_n^{\text{bdy}})\} \{1 + c_z(2 z_n^{\text{bdy}})\}}, \quad (30)$$

where $\hat{\mathbf{n}}_z = (\{\hat{\gamma}_z^2\}_1, \dots, \{\hat{\gamma}_z^2\}_N)^T$, \mathbf{L}_h^s and \mathbf{L}_z^s are the horizontal and vertical implicit diffusion operators used in \mathbf{C}_{hxz} but with horizontal tensor diffusivity for \mathbf{L}_h^s reduced according to Equation 28, and

$$\{\kappa_z^s\}_n = \beta \{\kappa_z\}_n$$

for the vertical diffusion operator.

In terms of the mean error, the inclusion of the smoothing operator has a slightly negative impact on the accuracy of the normalization factors for all variables (Experiments 11 and 12) compared to the experiments with only the boundary correction (Experiments 9 and 10). For these

experiments, β has been set to 0.33 for the horizontal diffusion tensor and 0.5 for the vertical diffusion tensor, which are values suggested by the theoretical relation given by Yaremchuk and Carrier (2012) (their equation 12):

$$\beta = \frac{1}{6} + \frac{1}{3d},$$

where d is the dimension of the space. Values of β between 0.2 and 0.5 have also been tested, but result in only modest changes to the mean error.

Figure 13 shows that, near the upper boundary, the errors are still large as in Figure 12 (note that the scales of the vertical axes of these two figures are different). Switching the boundary condition from Neumann to NDz (ND) in the 2D×1D (3D) smoothing operator results in substantially reduced errors in this region as illustrated in Figure 14. The mean errors of the normalization factors for temperature and unbalanced salinity are reduced as well (Experiments 13 and 14 in Table 2) compared to the mean errors from the experiments without smoothing (Experiments 9 and 10) and with smoothing using the Neumann boundary condition (Experiments 11 and 12).

4.5.5 | Bias correction

The approximate expressions for the normalization factors presented thus far in Section 4.5 are derived from purely theoretical considerations of the diffusion equation. As such, they do not account for effects on the normalization factors resulting from numerical approximations. One notable numerical approximation is the modest convergence criterion used with the implicit solver. On each diffusion step, an elliptic equation is solved approximately using a fixed number of iterations of the Chebyshev Iteration algorithm. The number of iterations is chosen such that the 2-norm of the residual is reduced by three to four orders of magnitude compared to the 2-norm of the right-hand side (Weaver *et al.*, 2016; Weaver *et al.*, 2018b). We can expect that the modest accuracy threshold of the solver will bias the normalization factors relative to those deduced from theory.

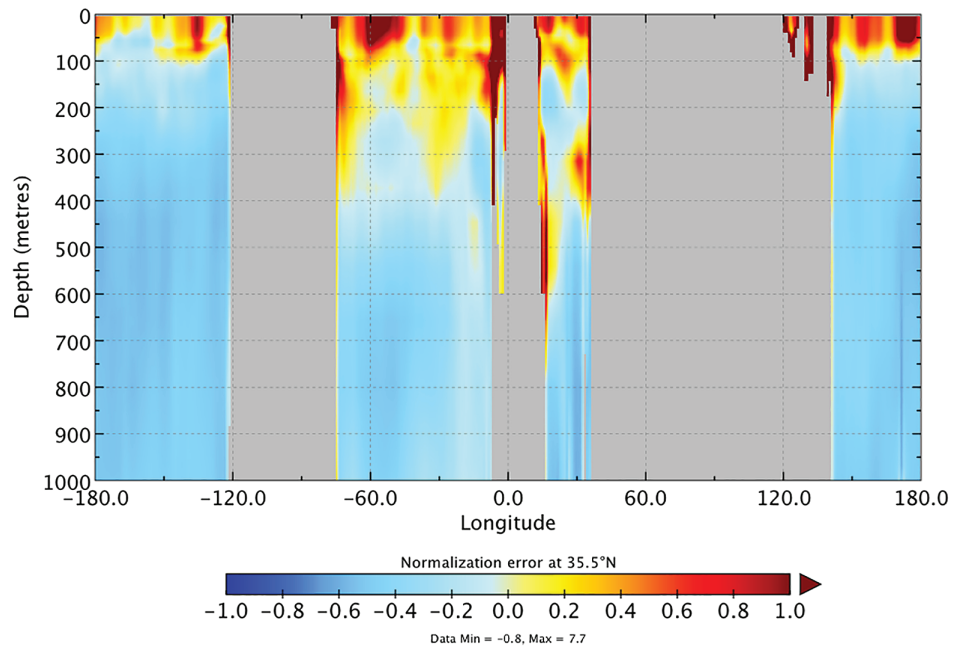
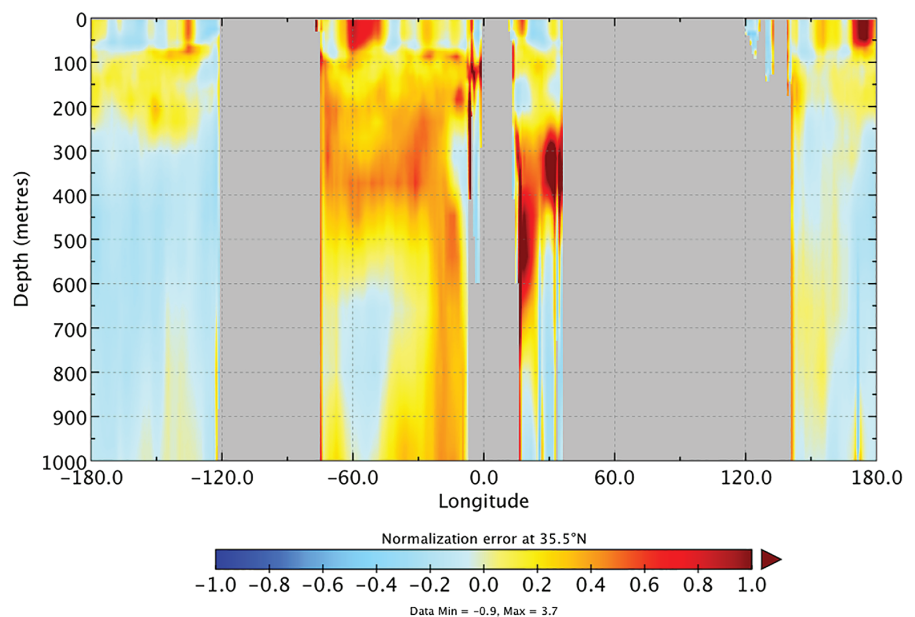
We can attempt to estimate the bias in the normalization factors using linear regression. Let us assume that the normalization factor at each grid point can be described by the linear model

$$\tilde{\gamma}_n^2 = a + b \hat{\gamma}_n^2, \quad (31)$$

where $\hat{\gamma}_n^2$ is our “best” analytical estimate of the normalization factor, and a and b are adjustable constants. Let us also assume that the exact normalization factors γ_k^2 are available at a selected number of $p = 1, \dots, P$ grid points.

FIGURE 13 Experiment 11.

The relative error in the normalization factors for the 2D×1D (3) implicit diffusion operator for temperature at approximately 35°N when using the boundary correction term to account for the effect of the Neumann boundary condition and smoothing of the zeroth-order analytical normalization. The vertical section displays the upper 1,000 m

**FIGURE 14** Experiment 13. As Figure 13, but when the diffusion filter used for smoothing the zeroth-order analytical normalization uses the NDz formulation

We can use the exact method (Equation 18) to compute γ_p^2 at the P points. Using the method of least squares, we can then seek a and b that minimize

$$e = \sum_{p=1}^P (\tilde{\gamma}_p^2 - \gamma_p^2)^2.$$

Carrying out this operation leads to expressions for a and b which, when substituted back in Equation 31, give the bias-corrected estimate of the normalization factors:

$$\tilde{\gamma}_n^2 = \overline{\gamma^2} + \frac{\text{cov}(\hat{\gamma}^2, \gamma^2)}{\text{var}(\hat{\gamma}^2)} (\hat{\gamma}_n^2 - \tilde{\gamma}^2), \quad (32)$$

where $\overline{(\cdot)}$, $\text{var}(\cdot)$ and $\text{cov}(\cdot, \cdot)$ denote sample estimates of the mean, variance and covariance of the quantities in parentheses, computed over the P points.

In Experiment 15, the analytical normalization factors based on Equations (27) and (30) are bias-corrected using Equation (32). The cost of the method is dominated by the computation of the exact normalization factors at the regression points, especially for 3D variables. Here, we choose a sample size of 50 for all variables (2D and 3D), which provides a better sampling of normalization factors for unbalanced SSH than temperature and unbalanced salinity. In order to limit boundary effects, the regression points are selected from open areas of the main ocean

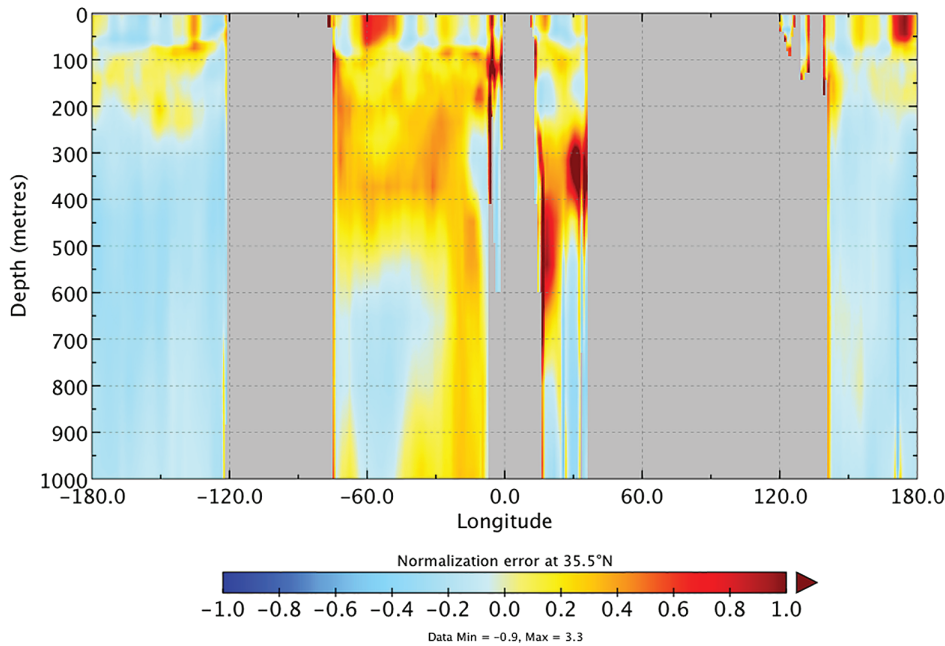


FIGURE 15 *Experiment 15.* The relative error in the normalization factors for the 2D×1D (3) implicit diffusion operator for temperature at approximately 35°N when using bias-corrected analytical normalization factors (Equation (32)). The analytical normalization factors have been computed using Equation (30), which involves both a boundary correction term to account for the effect of the Neumann boundary condition and smoothing of the zeroth-order analytical normalization factors using the NDz formulation.

basins (Pacific, Atlantic, and Indian Ocean basins, as well as the Circumpolar Current region). For temperature and unbalanced salinity, the points are chosen to sample the mixed layer (45 m) and main thermocline (165 m).

Bias correcting the analytical normalization factors leads to further reduction of the mean error for unbalanced SSH compared to all other experiments (Table 2). The improvement is visible in most areas, although in some boundary areas (close to Antarctica and in northern Canada), the relative error is increased and changes sign (not shown). In these areas, it is likely that the bias is dominated by the effects of boundary geometry (not solver precision), which is difficult to capture with the simple regression model Equation (31) and is not sampled by the points used for regression in this experiment. Compared to Experiment 13, results for temperature produce a slight increase in the mean error, while those for unbalanced salinity show a more substantial increase, probably for the reasons already mentioned. Figure 15 shows modest improvement in the temperature normalization factors in the ocean interior (cf. Figure 14).

5 | NORMALIZATION TO ACCOUNT FOR FLOW-DEPENDENT VERTICAL CORRELATIONS

Correct specification of the vertical correlations of background error in the upper ocean is important for properly assimilating surface observations such as Sea Surface Temperature (SST) and Sea Surface Salinity (SSS). Near the surface, the vertical correlations are sensitive to

the mixed-layer depth which can exhibit large variations on time-scales comparable to the width of the assimilation window. Capturing flow dependence in the background-error vertical correlations is highly desirable. Lalouaux *et al.* (2018) illustrate this point in a coupled data assimilation framework.

Randomization is too costly to be used with the complete 2D×1D diffusion matrix to recompute normalization factors on each assimilation cycle. Furthermore, results from the previous section indicate that the analytical-based normalization methods are not sufficiently accurate. In this section, we investigate variants of the randomization procedure that would allow us to update the normalization factors when the vertical correlation parameters change from one cycle to the next but the horizontal correlation parameters are held fixed to their climatological values.

5.1 | Look-up table: mixed-layer dependent vertical length-scales

One way to define flow-dependent background-error correlations, without the need for an ensemble, is to parametrize them in terms of the (flow-dependent) background state. This approach ignores uncertainty in the background state in defining the correlations. The Met Office applies this technique in NEMOVAR to define flow-dependent vertical correlation length-scales in the mixed-layer region. There, the vertical correlation length-scales are assumed to be equal to the mixed-layer depth itself (Waters *et al.*, 2015), a parametrization that is supported by the ensemble-estimated vertical diffusion

coefficients shown in Figure 3. Below the mixed-layer region, the vertical length-scales are taken to be proportional to the local vertical scale factor, with a smooth transition between the two parametrizations near the base of the mixed layer.

Waters *et al.* (2015) describe an approximate method to estimate the normalization factors for this specific parametrization, which relies on a pre-computation of a look-up table of normalization factors. The look-up table is constructed by applying randomization with a large sample size to a set of diffusion models where each model uses a parametrized vertical diffusion tensor with a different mixed-layer depth. In constructing the look-up table, the mixed-layer depth is taken to be the same at all horizontal grid points. On each assimilation cycle, the normalization factors at each grid point are then selected from the look-up table according to the value of the mixed-layer depth at that grid point. The method is approximate since the look-up table is constructed assuming a constant mixed-layer depth whereas the mixed-layer depths in the background state will vary between horizontal grid points. In order to reduce the error in this approximation, the mixed-layer depths are smoothed in each level using several iterations of a Shapiro filter.

In principle, we could apply a similar technique when using flow-dependent vertical correlations estimated from an ensemble, where normalization factors at different levels are computed for a range of typical vertical length-scales. However, the requirement that the horizontal variations of the vertical length-scales be sufficiently smooth for this to be an accurate approximation is a significant drawback, as is the system overhead of having to manipulate and access information from many large files that constitute the look-up table. The alternative methods described in the next subsection could also be applied to the parametrization described in this subsection.

5.2 | 2D×1D diffusion with accurate normalization using randomization

The computational cost of applying the vertical diffusion operator in \mathbf{V}_z is significantly cheaper than applying the horizontal diffusion operator in \mathbf{V}_h . For the global configuration used here, the difference in wall-clock time is more than a factor of ten when using a massively parallel domain decomposition with 12×12 horizontal subdomains (run on four nodes × 36 processors per node) and Message Passing Interface (MPI) communications. There are no MPI communications required when applying \mathbf{V}_z as each subdomain contains the entire water column. On the other hand, local MPI communications are required before each

diffusion step in \mathbf{V}_h , to update the halos on the lateral boundaries that connect adjacent subdomains.

We can exploit the difference in cost between \mathbf{V}_z and \mathbf{V}_h to reduce the cost of randomization for the specific formulation corresponding to $m = 2$ in Equation (11) (2D×1D (2)). The random vectors are generated using Equation (20):

$$\zeta_q = \mathbf{V}_{h \times z}^{(2)} \mathbf{W}_{hz}^{-1/2} \epsilon_q = \mathbf{V}_z^f \hat{\zeta}_q, \quad (33)$$

where $\hat{\zeta}_q = \mathbf{V}_h \mathbf{W}_{hz}^{-1/2} \epsilon_q$ and $q = 1, \dots, Q$ (with Q large). We have added the subscript f to \mathbf{V}_z^f to emphasize that the vertical diffusion operator, unlike \mathbf{V}_h , is flow-dependent. From the order of the operations in Equation (33), it is clear that the static vectors $\hat{\zeta}_q$, associated with the costly horizontal diffusion operator, need only be computed once as they can be reused as the input for \mathbf{V}_z^f when it is updated. This means that the cost of randomization on each cycle is determined by \mathbf{V}_z^f , which is affordable even with a very large sample size ($Q \sim 10^4$) to obtain accurate normalization factors.

Nevertheless, there are two drawbacks with this approach. First, in order to produce accurate normalization factors, a large number of random vectors $\hat{\zeta}_q$ must be precomputed and stored to file for each 3D control variable, and then retrieved on each assimilation cycle. With 10^4 random vectors, two 3D control variables (temperature and unbalanced salinity) and storage in single precision, this amounts to around 350 gigabytes of file-space that is required for ORCA1-Z42. With ORCA025-Z75, the required file-space is about 10 terabytes, which can be further reduced to about 5 terabytes using NetCDF compression tools. There is clearly a penalty to be paid when having to manipulate files of this size in an operational suite, even though it is generally feasible with modern file systems. Another drawback is that we are forced to sacrifice interleaving in order to isolate \mathbf{V}_z^f in Equation (33). The impact of interleaving will be assessed in Section 5.4.

5.3 | Separate randomization for 1D and 2D diffusion

Another approach is to assume that the normalization matrix for *any* of the 2D×1D and 3D formulations can be specified as a product of a normalization matrix for the horizontal diffusion component and a normalization matrix for the vertical diffusion component where each matrix is estimated separately using randomization. As with the previous approach, this would allow us to update only the normalization factors for the time-evolving vertical component to reduce computational cost. However,

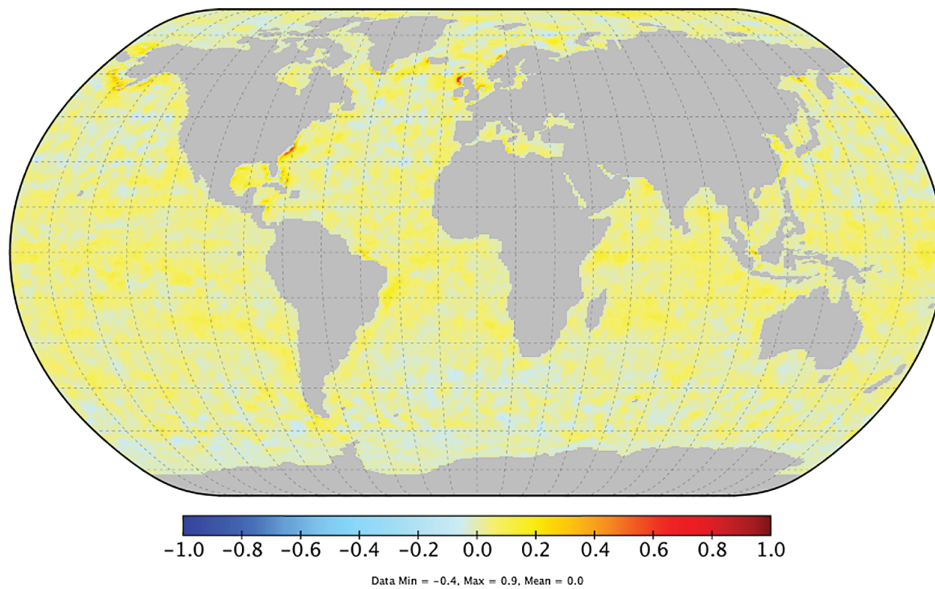


FIGURE 16 *Experiment 17.* The relative error in the normalization factors for the 2D×1D (3) implicit diffusion operator for temperature in the surface level when using normalization factors approximated as a product of factors estimated separately for the horizontal and vertical diffusion components using randomization with a sample size of 10^4 . Note that the error scale is a factor of 10 smaller than the one used in the corresponding error plots for the analytical estimates

unlike the previous approach, there is no need to store a large number of random vectors, which is clearly an advantage. The purpose here is to evaluate the accuracy of this approximation.

Using the notation from Section 2.2.2, we can formalize the estimation procedure as follows. First, we separate the randomization algorithm into two steps: one for computing random vectors ζ_q^h using the horizontal diffusion operator, and one for computing random vectors ζ_q^z using the vertical diffusion operator:

$$\left. \begin{aligned} \zeta_q^h &= \mathbf{V}_h \mathbf{W}_{hz}^{-1/2} \epsilon_q, \\ \zeta_q^z &= \mathbf{V}_z^f \mathbf{W}_{hz}^{-1/2} \epsilon_q. \end{aligned} \right\} \quad (34)$$

The estimates of the normalization factors for the horizontal and vertical diffusion components are $\Gamma_h^2 = \tilde{\Gamma}_h^2 \mathbf{W}_z$ and $\Gamma_z^2 = \tilde{\Gamma}_z^2 \mathbf{W}_h$ where $\tilde{\Gamma}_h^2$ and $\tilde{\Gamma}_z^2$ are the sample estimates built from ζ_q^h and ζ_q^z , respectively. The term \mathbf{W}_z in Γ_h^2 compensates for the vertical scale factors in \mathbf{W}_{hz} in Equation (34), while the term \mathbf{W}_h in Γ_z^2 compensates for the horizontal scale factors in \mathbf{W}_{hz} . These terms are necessary to avoid double-counting of the volume elements. The product

$$\Gamma_h^2 \Gamma_z^2 = \tilde{\Gamma}_h^2 \tilde{\Gamma}_z^2 \mathbf{W}_{hz} \approx \Gamma_{h \times z}^2$$

then provides an estimate of the normalization matrix for $\mathbf{C}_{h \times z}$ (Equation (10)).

Figures 16 and 17 show that this procedure leads to remarkably accurate estimates, except in a few isolated areas near coastlines and bottom bathymetry. Furthermore, the mean errors for both temperature and unbalanced salinity are notably smaller for the 2D×1D formulation with interleaving than without interleaving (4% compared to 7 to 8%). For the latter, the largest errors are

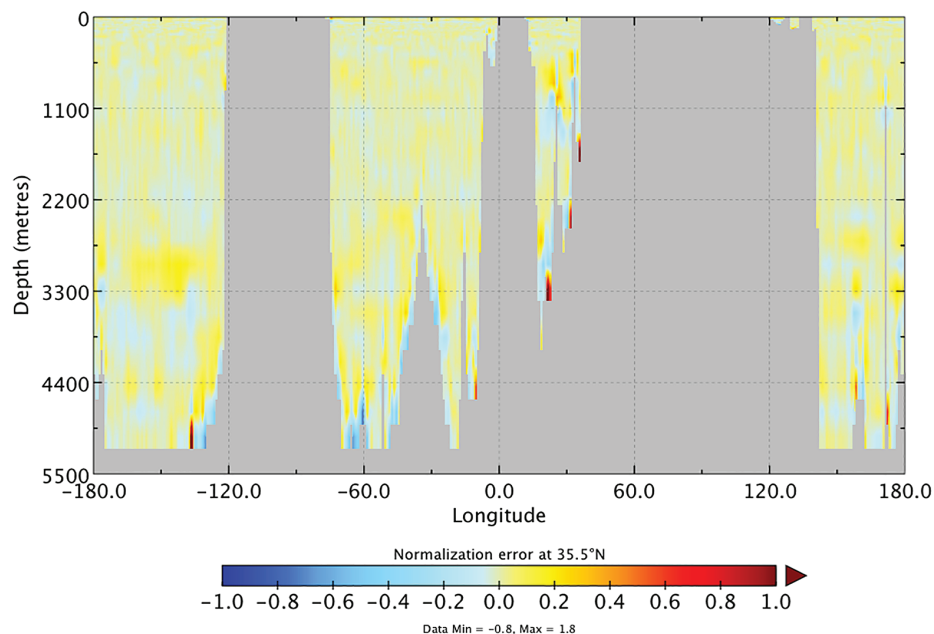
also confined to coastlines and bottom bathymetry, but are more widespread than those obtained with interleaving (not shown). This is a positive result as interleaving is an important feature that we wish to retain, as illustrated in the next subsection.

5.4 | Correlation structures

Up until now, we have focused on evaluating the accuracy of the amplitude of the correlation functions for different formulations of the correlation matrix. In this subsection, we examine the correlation structures themselves.

All formulations produce similar structures in the open ocean (not shown). Near bathymetry, however, they can produce quite different structures as illustrated in Figure 18. Here, we have applied different formulations of the correlation operator to a delta function located at the tip of a sea mount in the Northeast Pacific Ocean (coordinates $(i, k) = (99, 25)$). It is worth highlighting a number of features. First, the 2D×1D formulation with interleaving produces very similar results to the non-separable 3D formulation (cf. Figure 18a and b), but is substantially cheaper. The wall-clock time required to apply the 2D×1D diffusion operator is 20% less when using a 12×12 decomposition run on 4 nodes × 36 processors/node. The 3D formulation would have advantages over the 2D×1D formulation if rotational anisotropy between the horizontal and vertical directions is accounted for in the diffusion tensor, which is currently not the case. Second, the 2D×1D formulation without interleaving (middle left panel) produces unphysical correlations on either side of the sea mount, with the correlations one level up from the source point (points (98, 24) and (100, 24)) being distinctly

FIGURE 17 *Experiment 17.* The relative error in the normalization factors for the 2D×1D (3) implicit diffusion operator for temperature at approximately 35°N when using normalization factors approximated as a product of factors estimated separately for the horizontal and vertical diffusion components using randomization with a sample size of 10^4



larger than those directly adjacent to the source point (points (98, 25) and (100, 25)). Interestingly, this artifact is corrected using approximate normalization (Figure 18d). The correlation structures produced with interleaving and approximate normalization are physically sensible but somewhat underestimated compared to those with accurate normalization (Figure 18e, f).

Next, we examine the sensitivity of the correlations produced near the surface to the upper boundary condition. Figure 19 illustrates the response of the 2D×1D correlation operator to a delta function applied in the uppermost model level. The large vertical extent of the correlations throughout the mixed-layer region is clearly visible. Figure 19a, b compare the correlations produced with Neumann boundary conditions (a) and the ND formulation using Equation (25) where it is applied with the vertical diffusion operator only (NDz; (b)). In both cases, the normalization has been computed accurately. The correlations with Neumann boundary conditions penetrate deeper than those with the NDz formulation. This is an undesirable consequence of normalization, which corrects the amplitude of the correlation function at the expense of affecting the shape of the correlation function. This effect was illustrated by Mirouze and Weaver (2010) in a simple 1D framework (their figure 8). The NDz formulation prevents this behaviour but doubles the cost of the correlation operator. However, this extra cost may be warranted given the importance of the vertical correlations near the surface for correctly assimilating surface observations. Finally, the NDz formulation with approximate normalization does not change the spatial structure of the correlations compared to those obtained with accurate normalization, but does tend to overestimate them at most grid points

(Figure 19c, d). The mean errors for temperature and unbalanced salinity are slightly larger using approximate normalization with the NDz formulation than with the standard formulation that employs Neumann boundary conditions (cf. Experiments 17 and 18 in Tables 1 and 2).

6 | SUMMARY AND DISCUSSION

Accounting for flow-dependent background-error correlations with a diffusion operator requires an efficient method to estimate the diagonal of the diffusion-based covariance matrix in order to compute normalization factors. In this article, we have described several general techniques for this purpose. We have evaluated them in a global ocean configuration of the NEMOVAR data assimilation system. The principal parameters of the diffusion operator are the elements of the diffusion tensor, which are related to the spatial length-scales of the underlying covariance functions. These are key parameters that largely influence the amplitude and spatial structure of the normalization factors. Here, they have been estimated from a climatological ensemble of perturbations from ECMWF's low-resolution ocean reanalysis ORAS5-LR.

In the first part of the article, we evaluated inexpensive methods derived from analytical considerations of the diffusion equation or based on randomization with a small sample size. We can summarize the main results from those experiments as follows.

- The normalization factors estimated from the analytically-based methods can have significant errors and are generally not accurate enough for operational

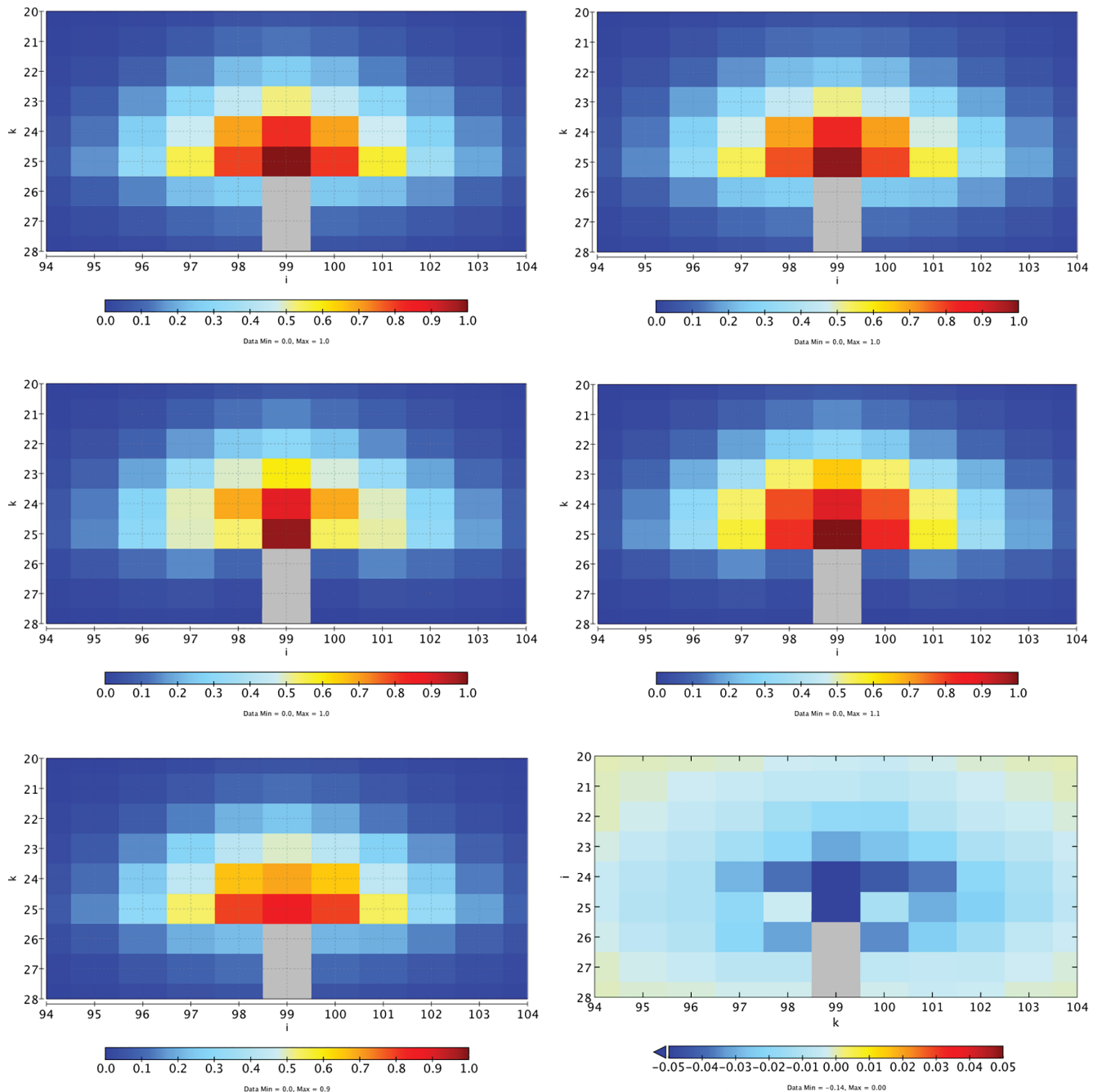


FIGURE 18 The response of different diffusion-based formulations of the correlation matrix to a delta function located at the tip of a sea mount (shaded grey). (a) $2D \times 1D$ (3) with accurate normalization, (b) non-separable $3D$ with accurate normalization, (c) $2D \times 1D$ (2) with accurate normalization, (d) $2D \times 1D$ (2) with approximate normalization, (e) $2D \times 1D$ (3) with approximate normalization, and (f) $2D \times 1D$ (3) with approximate normalization minus $2D \times 1D$ (3) with accurate normalization. The horizontal and vertical axes represent the horizontal grid-point i and model level k index

applications with global ocean models. Randomization estimates with a small sample (~ 100) are more accurate in a globally averaged sense but are too noisy and can attain large errors locally.

- The analytically-based methods are useful for quick testing of developments to the correlation model where high accuracy is generally not required. Among the

different approximations that were tested with the ensemble-derived tensor, overall best results were obtained using smoothed versions of the analytical expressions of the normalization factors together with a correction term near boundaries to compensate for the amplification effect of the Neumann boundary condition (Equations (27), (29) and (30)).

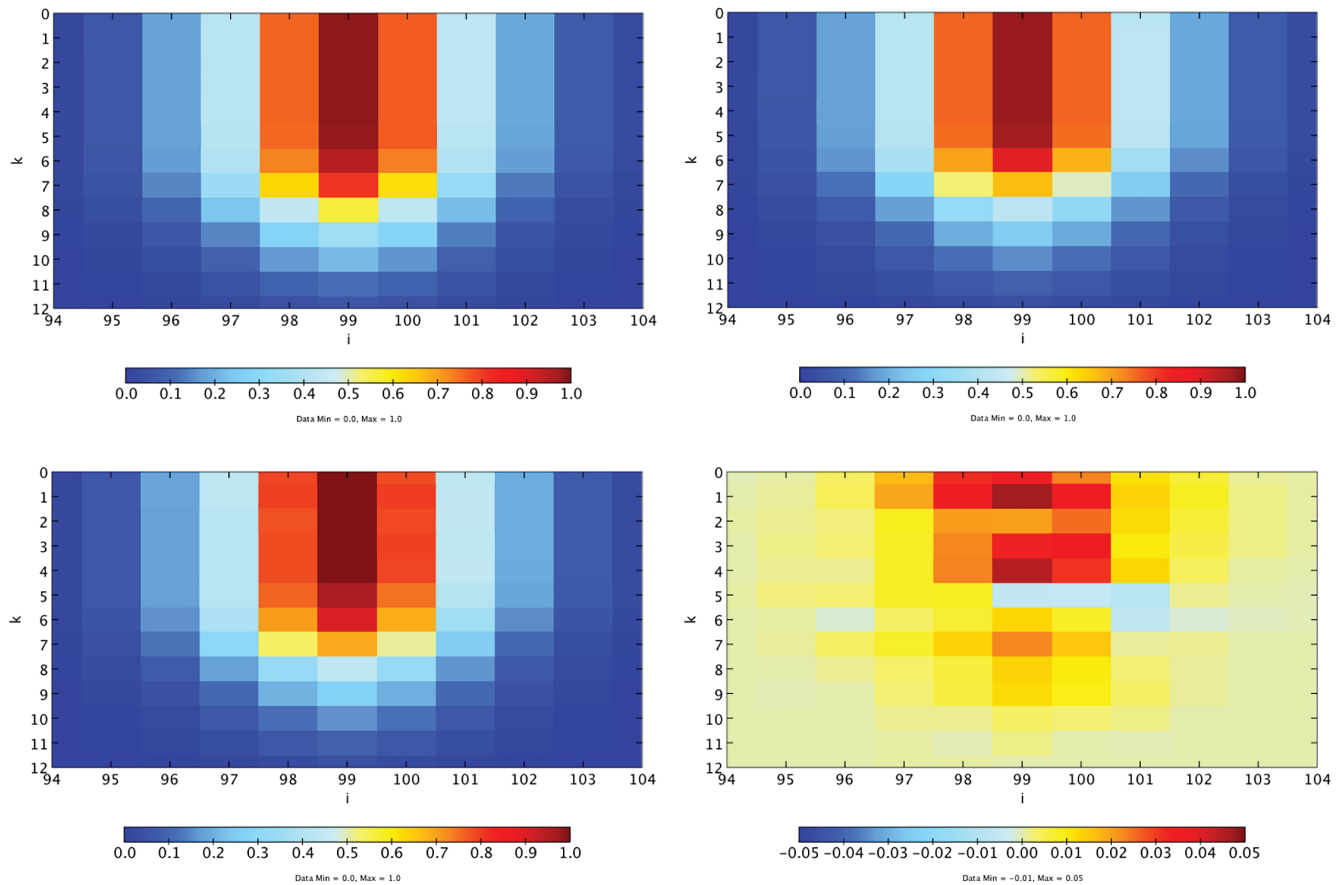


FIGURE 19 The response of different diffusion-based formulations of the correlation matrix to a delta function located in the uppermost level of the model. (a) $2D \times 1D$ (3) with Neumann boundary conditions and accurate normalization, (b) $2D \times 1D$ (3) with ND formulation and accurate normalization, (c) $2D \times 1D$ (3) with ND formulation and approximate normalization and (d) $2D \times 1D$ (3) with ND formulation and approximate normalization minus $2D \times 1D$ (3) with ND formulation and accurate normalization. The horizontal and vertical axes represent the horizontal grid-point i and model level index k where the latter is shifted by one (the uppermost level corresponds to $k = 0$)

- The main limitation of the analytically-based methods for global ocean data assimilation is their inability to produce sufficiently accurate normalization factors near land boundaries, especially where the boundary geometry is complex. The more costly ND formulation of the correlation model, involving a combination of Neumann and Dirichlet boundary conditions, provided some improvement but not enough to justify the additional computational cost that the ND formulation entails. The ND formulation is most effective near the upper boundary in the open ocean where the assumptions underlying the method are best satisfied.
- Coastlines and bathymetry are also problematic when using the diffusion operator as a filter for either suppressing sampling error from randomization estimates or for smoothing the diffusion tensor in the analytical approximation of the normalization factors. Further research aimed at improving the filter response near irregular land boundaries is desirable.
- Results using linear regression to bias correct the analytical normalization factors were positive for unbalanced SSH but somewhat mixed for temperature and unbalanced salinity. This is perhaps not surprising given the simplicity of the regression model that was used, especially for 3D variables. There is clearly scope for using more sophisticated regression models, which could include, for example, predictors based on the elements of the diffusion tensor and their derivatives. This approach has been tested in an idealized 2D configuration, with positive results. Additional predictors to handle boundary geometry and possibly other factors would be needed to apply this method to global ocean configurations. In general, the problem lends itself to a machine learning strategy whereby a regression model is trained using a set of accurate normalization factors (computed using randomization) for different diffusion tensors estimated from an ensemble.
- The normalization methods described in this article are likely to provide better accuracy for diffusion-based

covariance operators in atmospheric models where there are no complicating effects from irregular land boundaries.

Based on the above results, we conclude that, for global ocean applications, only randomization with a large sample (on the order of 10^4) provides a robust method for accurate normalization, albeit at a substantial cost especially with 3D diffusion operators. This clearly has implications on the degree of flow dependence that can be affordably accounted for by a background-error correlation model in an operational high-resolution global data assimilation system for which there are tight constraints on computer resources and on timeliness for the delivery of analyses. A reasonable compromise could be to specify the background-error correlations from a seasonal climatology of ensemble perturbations, which could be updated occasionally to capture the slow evolution of background error associated with, for example, changes in the observing system.

In the second part of the article, we examined methods for updating normalization factors more frequently for the special case where only the vertical component of the diffusion operator is made fully flow-dependent. Capturing flow dependence in the vertical correlations of background error is important for correctly assimilating surface observations such as SST and SSS which are strongly influenced by surface mixed-layer processes. With a moderate number of ensemble members, we can use the sample covariance of the vertical derivative of the ensemble perturbations to produce a relatively robust statistical estimate of the vertical correlation length-scales on each cycle.


We showed that normalization factors can be estimated with good accuracy by approximating them at each grid point as a product of two normalization factors: one computed using randomization with the horizontal diffusion operator only and the other computed using randomization with the vertical diffusion operator only. With this approximation, the mean relative error was only 4%, with the largest errors being concentrated in a few isolated regions near bottom bathymetry and coastlines. The magnitude of the errors is likely well within the uncertainty of our best estimates of the true background-error variances. This result is of significant practical interest as the vertical diffusion operator is inexpensive and thus can be applied on each cycle with a large sample of random vectors to obtain an accurate estimate of the normalization factors for the vertical component. These factors can then be combined with the normalization factors for the static horizontal component to obtain a good approximation of the complete normalization matrix. Recent experiments indicate that this approximation

also works well with the higher-resolution configuration ORCA025-Z75.

ACKNOWLEDGEMENTS

This work has been funded by the Copernicus Climate Change Service (C3S), one of six services of the European Union's Copernicus Programme, which is implemented by ECMWF on behalf of the European Commission. It is a contribution to contract C3S_321_INRIA: "Technical preparations for C3S Seasonal Initialization and Global Reanalysis: Enabling an Ensemble of Data Assimilations for the Ocean". Additional support has been provided by the French national programme LEFE/INSU. The development of NEMOVAR is a collaborative effort with contributions from numerous people from CERFACS, ECMWF, INRIA and Met Office. B. Ménétrier is funded by the JEDI project of the Joint Center for Satellite Data Assimilation. The authors thank two anonymous referees whose comments helped improve an earlier version of the article.

ORCID

Anthony T. Weaver  <https://orcid.org/0000-0002-8952-338X>

REFERENCES

- Andersson, E., Fisher, M., Munro, R. and McNally, A.P. (2000) Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme and the explanation of a case of poor convergence. *Quarterly Journal of the Royal Meteorological Society*, 126, 1455–1472. <https://doi.org/10.1002/qj.49712656512>.
- Balmaseda, M.A., Mogensen, K. and Weaver, A.T. (2013) Evaluation of the ECMWF Ocean Reanalysis ORAS4. *Quarterly Journal of the Royal Meteorological Society*, 139, 1132–1161. <https://doi.org/10.1002/qj.2063>.
- Bekas, C.F., Kokiopoulou, E. and Saad, Y. (2007) An estimator for the diagonal of a matrix. *Applied Numerical Mathematics*, 57, 1214–1229. <https://doi.org/10.1016/j.apnum.2007.01.003>.
- Belo Pereira, M. and Berre, L. (2006) The use of an ensemble approach to study the background error covariances in a global NWP model. *Monthly Weather Review*, 134, 2466–2489. <https://doi.org/10.1175/MWR3189.1>.
- Berre, L., Varella, H. and Desroziers, G. (2015) Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4D-Var at Météo-France. *Quarterly Journal of the Royal Meteorological Society*, 141, 2803–2812. <https://doi.org/10.1002/qj.2565>.
- Bonavita, M., Isaksen, L. and Hólm, E.V. (2012) On the use of EDA background-error variances in the ECMWF 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 138, 1540–1559. <https://doi.org/10.1002/qj.1899>.
- Bonavita, M., Hólm, E.V., Isaksen, L. and Fisher, M. (2016) The evolution of the ECMWF hybrid data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 142, 142–303. <https://doi.org/10.1002/qj.2652>.
- Buehner, M. (2005) Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a

- quasi-operational setting. *Quarterly Journal of the Royal Meteorological Society*, 131, 1013–1043. <https://doi.org/10.1256/qj.04.15>.
- Chabot, V., Berre, L. and Desroziers, G. (2017) Diagnosis and normalization of gridpoint background-error variances induced by a block-diagonal wavelet covariance matrix. *Quarterly Journal of the Royal Meteorological Society*, 143, 1268–1279. <https://doi.org/10.1002/qj.3003>.
- Clayton, A.M., Lorenc, A.C. and Barker, D.M. (2013) Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, 139, 1445–1461. <https://doi.org/10.1002/qj.2054>.
- Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., Fisher, M. and Pailleux, J. (1998) The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124, 1783–1807. <https://doi.org/10.1002/qj.49712455002>.
- Daget, N., Weaver, A.T. and Balmaseda, M.A. (2009) Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Quarterly Journal of the Royal Meteorological Society*, 135, 1071–1094. <https://doi.org/10.1002/qj.412>.
- Daley, R. (1991) *Atmospheric Data Analysis*. Cambridge: Cambridge University Press.
- Derber, J.C. and Rosati, A. (1989) A global oceanic data assimilation system. *Journal of Physical Oceanography*, 19, 1333–1347. [https://doi.org/10.1175/1520-0485\(1989\)019<1333:AGODAS>2.0.CO;2](https://doi.org/10.1175/1520-0485(1989)019<1333:AGODAS>2.0.CO;2).
- Egbert, G., Bennett, A. and Foreman, M. (1994) Topex/Poseidon tides estimated using a global inverse model. *Journal of Geophysical Research*, 99, 24821–24852. <https://doi.org/10.1029/94JC01894>.
- Fisher, M. (2003). Background-error covariance modelling, pp. 45–63 in Proceedings of Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, 8–12 September 2003. ECMWF, Reading, UK. <https://www.ecmwf.int/node/9404>.
- Gaspari, G. and Cohn, S.E. (1999) Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757. <https://doi.org/10.1002/qj.49712555417>.
- Golub, G.H. and Van Loan, C.F. (1996) *Matrix Computations* (3rd edition). Baltimore, MD: Johns Hopkins University Press.
- Guttorp, P. and Gneiting, T. (2006) Miscellanea studies in the history of probability and statistics XLIX: On the Matérn correlation family. *Biometrika*, 93, 989–995. <https://www.jstor.org/stable/20441340>.
- Kleist, D.T. and Ide, K. (2015) An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS, Part I: System description and 3D-hybrid results. *Monthly Weather Review*, 143, 433–451. <https://doi.org/10.1175/MWR-D-13-00351.1>.
- Laeuchli, J.H. (2016) *Methods for Estimating the Diagonal of Matrix Functions*. Paper 1477067934 in theses, dissertations, and Masters' projects, College of William and Mary, Williamsburg, VA. <http://doi.org/10.21220/S2CC7X>.
- Lalouaux, P., Frolov, S., Ménétrier, B. and Bonavita, M. (2018) Implicit and explicit cross-correlations in coupled data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144, 1851–1863. <https://doi.org/10.1002/qj.3373>.
- Lorenc, A.C. (1992) Iterative analysis using covariance functions and filters. *Quarterly Journal of the Royal Meteorological Society*, 118, 569–591. <https://doi.org/10.1002/qj.49711850509>.
- Madec, G. and NEMO team (2008). NEMO ocean engine. Note du Pôle de modélisation No 27, Institut Pierre-Simon Laplace, Paris.
- Ménétrier, B. and Auligné, T. (2015) An overlooked issue of variational data assimilation. *Monthly Weather Review*, 143, 3925–3930. <https://doi.org/10.1175/MWR-D-14-00404.1>.
- Ménétrier, B., Montmerle, T., Michel, Y. and Berre, L. (2015) Linear filtering of sample covariances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization. *Monthly Weather Review*, 143, 1622–1643. <https://doi.org/10.1175/MWR-D-14-00157.1>.
- Michel, Y., Ménétrier, B. and Montmerle, T. (2016) Objective filtering of the local correlation tensor. *Quarterly Journal of the Royal Meteorological Society*, 142, 2314–2323. <https://doi.org/10.1002/qj.2824>.
- Mirouze, I. and Storto, A. (2016) Handling boundaries with the one-dimensional first-order recursive filter. *Quarterly Journal of the Royal Meteorological Society*, 142, 2478–2487. <https://onlinelibrary.wiley.com/doi/10.1002/qj.2840>.
- Mirouze, I. and Weaver, A.T. (2010) Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Quarterly Journal of the Royal Meteorological Society*, 136, 1421–1443. <https://doi.org/10.1002/qj.643>.
- Mirouze, I., Blockley, E.W., Lea, D.J., Martin, M.J. and Bell, M.J. (2016) A multiple length-scale correlation operator for ocean data assimilation. *Tellus A*, 68, 29744. <https://doi.org/10.3402/tellusa.v68.29744>.
- Mogensen, K., Balmaseda, M.A. and Weaver, A.T. (2012) *The NEMOVAR ocean assimilation system and its implementation for ECMWF's ocean analysis for System 4*. Technical Memorandum 668, ECMWF, Reading, UK. Also CERFACS Technical Report No. TR-CMGC-12-30. <https://www.ecmwf.int/node/11174>.
- Parrish, D.F. and Derber, J.C. (1992) The National Meteorological Center's Spectral Statistical-Interpolation analysis system. *Monthly Weather Review*, 130, 1747–1763. [https://doi.org/10.1175/1520-0493\(1992\)120<1747:TNMCS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCS>2.0.CO;2).
- Penny, S.G., Behringer, D.W., Carton, J.A. and Kalnay, E. (2015) A hybrid global ocean data assimilation system at NCEP. *Monthly Weather Review*, 143, 4660–4677. <https://doi.org/10.1175/MWR-D-14-00376.1>.
- Purser, R.J. (2008a). Normalization of the diffusive filters that represent the inhomogeneous covariance operators of variational assimilation, using asymptotic expansions and techniques of non-Euclidean geometry; Part I: Analytic solutions for symmetrical configurations and the validation of practical algorithms. Office Note 456, National Centers for Environmental Prediction, Camp Springs, MD.
- Purser, R.J. (2008b). Normalization of the diffusive filters that represent the inhomogeneous covariance operators of variational assimilation, using asymptotic expansions and techniques of non-Euclidean geometry; Part II: Riemannian geometry and the generic parametric expansion method. Office Note 457, National Centers for Environmental Prediction, Camp Springs, MD.
- Purser, R.J., Wu, W.S., Parrish, D.F. and Roberts, N.M. (2003a) Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: spatially homogeneous and isotropic Gaussian covariances. *Monthly Weather Review*,

- 131, 1524–1535. [https://doi.org/10.1175//1520-0493\(2003\)131<1524:NAOTAO>2.0.CO;2](https://doi.org/10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2).
- Purser, R.J., Wu, W.S., Parrish, D.F. and Roberts, N.M. (2003b) Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: spatially inhomogeneous and anisotropic general covariances. *Monthly Weather Review*, 131, 1536–1548. <https://doi.org/10.1175//2543.1>.
- Rabier, F., McNally, A.P., Andersson, E., Courtier, P., Undén, P., Eyre, J.R., Hollingsworth, A. and Bouttier, F. (1998) The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: structure functions. *Quarterly Journal of the Royal Meteorological Society*, 124, 1809–1829. <https://doi.org/10.1002/qj.49712455003>.
- Storto, A., Oddo, P., Cipollone, A., Mirouze, I. and Lemieux-Dudon, B. (2018) Extending an oceanographic variational scheme to allow for affordable hybrid and four-dimensional data assimilation. *Ocean Modelling*, 128, 67–86. <https://doi.org/10.1016/j.ocemod.2018.06.005>.
- Waters, J., Lea, D.J., Martin, M.J., Mirouze, I., Weaver, A.T. and While, J. (2015) Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. *Quarterly Journal of the Royal Meteorological Society*, 141, 333–349. <https://doi.org/10.1002/qj.2388>.
- Weaver, A.T. and Courtier, P. (2001) Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, 127, 1815–1846. <https://doi.org/10.1002/qj.49712757518C>.
- Weaver, A.T. and Mirouze, I. (2013) On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 139, 242–260. <https://doi.org/10.1002/qj.1955>.
- Weaver, A.T., Deltel, C., Machu, E., Ricci, S. and Daget, N. (2005) A multivariate balance operator for variational ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131, 3605–3625. <https://doi.org/10.1256/qj.05.119>.
- Weaver, A.T., Tshimanga, J. and Piacentini, A. (2016) Correlation operators based on an implicitly formulated diffusion equation solved with the Chebyshev iteration. *Quarterly Journal of the Royal Meteorological Society*, 142, 455–471. <https://doi.org/10.1002/qj.2664>.
- Weaver, A.T., Chrust, M., Ménérier, B., Piacentini, A., Tshimanga, J., Yang, Y., Gürol, S. and Zuo, H. (2018a) *Using ensemble-estimated background-error variances and correlation scales in the NEMOVAR system*. Technical Report TR-PA-18-15, CERFACS, Toulouse, France.
- Weaver, A.T., Gürol, S., Tshimanga, J., Chrust, M. and Piacentini, A. (2018b) ‘Time’-parallel diffusion-based correlation operators. *Quarterly Journal of the Royal Meteorological Society*, 144, 2067–2088. <https://doi.org/10.1002/qj.3302>.
- Whittle, P. (1963) Stochastic processes in several dimensions. *Bulletin of the International Statistical Institute*, 40, 974–994.
- Wu, W.S., Purser, R.J. and Parrish, D.F. (2002) Three-dimensional variational analysis with spatial inhomogeneous covariances. *Monthly Weather Review*, 130, 2905–2916. [https://doi.org/10.1175/1520-0493\(2002\)130<2905:TDVAWS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2).
- Yaremchuk, M. and Carrier, M. (2012) On the renormalization of the covariance operators. *Quarterly Journal of the Royal Meteorological Society*, 140, 637–649. <https://doi.org/10.1175/MWR-D-11-00139.1>.
- Yaremchuk, M. and Smith, S. (2011) On the correlation functions associated with polynomials of the diffusion operator. *Quarterly Journal of the Royal Meteorological Society*, 137, 1927–1932. <https://doi.org/10.1002/qj.893>.
- Zuo, H., Balmaseda, M.A., Tietsche, S. and Mogensen, K. (2019) The ECMWF operational ensemble reanalysis-analysis system for ocean and sea ice: A description of the system and assessment. *Ocean Science*, 15, 779–808. <https://doi.org/10.5194/os-15-779-2019>.

How to cite this article: Weaver AT, Chrust M, Ménérier B, Piacentini A. An evaluation of methods for normalizing diffusion-based covariance operators in variational data assimilation. *Q.J.R. Meteorol. Soc.* 2021;147:289–320. <https://doi.org/10.1002/qj.3918>