

**Using ensemble-estimated background error variances and
correlation scales in the NEMOVAR system^a**

A. T. WEAVER, M. CHRUST, B. MÉNÉTRIER, A. PIACENTINI, J. TSHIMANGA,
Y. YANG, S. GÜROL, H. ZUO

Technical Report TR/PA/18/15

^aReport prepared for the ERA-CLIM2 EU-FP7 project.

Using ensemble-estimated background error variances and correlation scales in the NEMOVAR system*

A. T. Weaver¹, M. Chrust², B. Ménétrier³, Y. Yang⁴, A. Piacentini¹,
J. Tshimanga¹, S. Gürol¹, and H. Zuo²

¹CERFACS / CECI CNRS UMR 5318, Toulouse, France

²ECMWF, Reading, United Kingdom

³Météo-France / CNRS UMR 3589, Toulouse, France

⁴IRIT / CNRS UMR 5505, Toulouse, France

January 10, 2018

Abstract

We have developed two complementary methods for using ensemble perturbations to define the background-error covariance matrix (\mathbf{B}) in the variational ocean data assimilation system (NEMOVAR) used for the Coupled ECMWF ReAnalysis (CERA). The first method uses ensembles to estimate the variances and correlation length scales (diffusion tensor) of the diffusion-based background-error covariance model. To account for sampling error, the parameters are filtered using an objective method that depends on the ensemble size. The second method uses ensembles to construct a low-rank sample covariance matrix and includes an objective procedure to localize this matrix to eliminate remote correlations associated with sampling error. Hybrid variants of both methods have also been developed. For the first method, the hybrid formulation involves linearly combining the ensemble estimates of the covariance parameters with modelled representations of those parameters. For the second method, the hybrid formulation involves linearly combining the localized sample covariance matrix with the full-rank matrix described by the parameterized covariance model. The correlation model, localization operator and parameter filter are all based on an algorithm that involves solving an implicitly formulated diffusion equation. We have completely revised the diffusion model to make it more general and to improve the computational efficiency of the implicit solver on high-performance computers.

All methods have been fully integrated into the NEMOVAR source code maintained at ECMWF. The developments have been validated technically and scientifically, mainly using single cycle assimilation experiments with the global ocean configuration at 1° horizontal resolution (ORCA1). Results from validation experiments are provided in this report. Multi-cycle ocean reanalysis experiments are ongoing to evaluate the impact of flow-dependent background-error variances derived from ensembles. These are being conducted with a global configuration at $1/4^\circ$ resolution (ORCA025) as well as ORCA1. Extensive experimentation using the ECMWF system will be necessary to evaluate the different options for defining \mathbf{B} from ensembles and to determine the best cost-effective combination. A pragmatic strategy for progressively improving the representation of \mathbf{B} using ensembles is proposed in the concluding section.

*Report prepared for the ERA-CLIM2 EU-FP7 project.

1 Introduction

This deliverable was concerned with developing methods for using ensembles to improve background error covariances in the ocean data assimilation component (NEMOVAR) of CERA. In addition to scientific developments, a major part of this deliverable was to build the Fortran code infrastructure in NEMOVAR to allow for hybrid ensemble-variational data assimilation. The scientific developments are described in Section 2, while the technical changes required to NEMOVAR are discussed in Section 3. Section 4 discusses a strategy for including the developments progressively in the CERA system, as well as areas of future research for improving the computational efficiency of the system.

The baseline for the developments described in this report is the background-error covariance matrix (\mathbf{B}) used in the ocean component of the CERA system. Section 1.1 highlights its main features. The baseline \mathbf{B} has been used for the ocean reanalysis system ORAS4 as well for CERA. Section 1.2 presents some results from ORAS4 that underscores the need to improve \mathbf{B} in a reanalysis framework and where an ensemble of data assimilations can be expected to provide that improvement.

1.1 The background-error covariance matrix (\mathbf{B}) used in CERA

The ocean data assimilation system used in CERA (NEMOVAR version 3.4) employs a purely modelled form of the background-error covariance matrix given by

$$\mathbf{B}_m = \mathbf{K}_b \underbrace{\widehat{\mathbf{D}}_m^{1/2} \widehat{\mathbf{C}}_m \widehat{\mathbf{D}}_m^{1/2}}_{\widehat{\mathbf{B}}_m} \mathbf{K}_b^T \quad (1)$$

where \mathbf{K}_b is a multivariate balance operator that acts on the control variables [Balmaseda *et al.*, 2013; Weaver *et al.*, 2005], $\widehat{\mathbf{D}}_m = \widehat{\mathbf{D}}_m^{1/2} \widehat{\mathbf{D}}_m^{1/2}$ is a diagonal matrix containing estimates of the background-error variances of the control variables, and $\widehat{\mathbf{C}}_m$ is a correlation matrix that is block-diagonal (univariate) with respect to the control variables. The subscript ‘m’ indicates that the covariances are *modelled*. The ‘hat’ on matrices indicates that these quantities are defined for the control variables, not the state variables. For the 3D-Var system used in CERA, the control variables consist of temperature (T), unbalanced salinity (S_u) and unbalanced sea-surface height (η_u). In the CERA system, there is no assimilated observational information about velocity (either direct or indirect observations), so there is no need to introduce control variables for the unbalanced components (u_u and v_u) of horizontal velocity. Horizontal velocity increments are completely determined by the balance operator.

The univariate covariance matrix for the control variables is the matrix denoted by $\widehat{\mathbf{B}}_m$ in (1). The specification of $\widehat{\mathbf{D}}_m$ is somewhat heuristic, involving physically-motivated relations that depend on the background state and empirically-tuned parameters [Balmaseda *et al.*, 2013; Mogensen *et al.*, 2012; Weaver *et al.*, 2003]. The block components of $\widehat{\mathbf{C}}_m$ are modelled using a symmetric product of one-dimensional (1D) implicit diffusion operators to approximate a multi-dimensional Gaussian correlation matrix [Mirouze and Weaver, 2010; Waters *et al.*, 2015]. Like the variances, the correlation length scales are also specified heuristically [Mogensen *et al.*, 2012].

There are three points worth highlighting regarding the balance operator \mathbf{K}_b . First, it is derived from a combination of strictly linear balance relations (e.g., geostrophy) as well as nonlinear balance relations (Temperature-Salinity (T-S) relation and equation of state) that are linearized with respect

to a reference state. The nonlinear T-S relation is not explicitly available, so cannot be used to derive a linearized T-S relation. Instead, the linearized T-S relation has been defined directly, using a reference state-dependent expression designed to adjust salinity in response to changes in temperature, in a way that approximately preserves the T-S water mass properties of the reference state [Ricci *et al.*, 2005]. Not having access to the underlying nonlinear T-S relation has implications on the way we define the ensemble perturbations, as discussed in Section 2.2. Second, the structural form of \mathbf{K}_b , which is lower block triangular, means that the inverse balance operator \mathbf{K}_b^{-1} is readily available and easy to apply. As such, it is possible to estimate statistics of the assimilation control variables from ensembles of the model state vector (temperature, salinity, and sea-surface height (SSH) being the relevant state variables for the CERA 3D-Var system). Third, the fact that the reference state, which is initially taken to be the background state, can be updated on the outer loops of the variational minimization algorithm means that \mathbf{B}_m will change from one outer loop to the next. This has implications on the minimization algorithms used in NEMOVAR [Gürol *et al.*, 2014]. For practical convenience, and to allow for consistent re-linearization, the balance operator is effectively included in the (generalized) observation operator rather than in \mathbf{B}_m (see the cost function formulation in Section 2 of Mogensen *et al.* [2012]). While some parameters of the univariate covariance model ($\hat{\mathbf{B}}_m$) are also state-dependent, they are fixed to their background values when using multiple outer loops. This is done to ensure perfect consistency when using different minimization algorithms available in NEMOVAR.

1.2 Evidence of the need to improve \mathbf{B} in reanalysis

Differences between the observations and their model-estimated counterparts before and after assimilation (the innovations and analysis residuals, respectively) provide valuable information for assessing the performance of the data assimilation system. This information is also important for evaluating the consistency of the error covariances prescribed in the system [Desroziers *et al.*, 2005]. Figure 1 shows an example of a consistency check based on Desroziers diagnostics applied to the ECMWF ORAS4 reanalysis [Balmaseda *et al.*, 2013]. The version of NEMOVAR used for ORAS4 was very similar to that used for the ocean component in CERA. The figure shows a 40-year time series of the globally averaged observation-space representations of the specified and expected temperature background error standard deviations (BESD) in the upper 50 m (blue and red curves, respectively). The specified BESD are parameterized in terms of the vertical temperature gradient in the background state. Discrepancies between the specified and expected BESD are an indication of sub-optimality in the error covariance specifications. Seasonal variations in the specified and expected BESD are reasonably consistent. However, there is a noticeable decreasing trend in the expected BESD that is not present in the specified BESD. This trend roughly mirrors the increasing trend in the number of observations (black curve) and reflects the fact that the background state is becoming steadily more accurate, especially in the final decade as a result of the assimilation of Argo data. This important influence of the observation network on background error is not captured by the state-dependent parameterization used in ORAS4 and suggests that improved background error estimation methods are necessary. Ensemble data assimilation provides the appropriate framework for making such improvements.

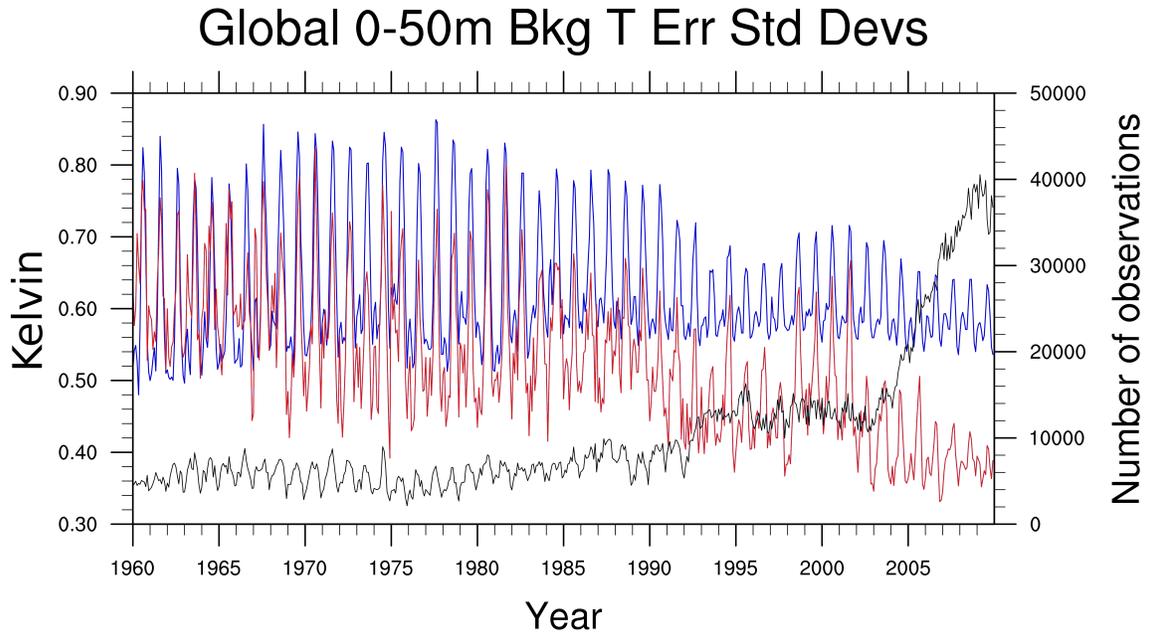


Figure 1: Time series of monthly-averaged temperature background-error standard deviations (BESD) in the ECMWF ORAS4 reanalysis [Balmaseda *et al.*, 2013]. (1) the specified BESD in the assimilation system (blue curve); and (2) the expected BESD as determined using the Desroziers *et al.* [2005] method (red curve). The black curve shows the number of observations as function of time (right axis). (From Martin *et al.* [2015]).

2 Scientific description of the developments

2.1 A revised correlation model

The formulation of the correlation operator has been entirely revised to be more general, to produce more realistic correlations near coastlines, and to employ an implicit solver with improved scalability properties on high-performance computers. This work is documented in detail in two articles that were prepared during the project [Weaver *et al.*, 2017, 2016]. A brief summary is given below.

Each (univariate) block correlation matrix of $\widehat{\mathbf{C}}_m$, associated with the variable χ , has the symmetric form

$$\mathbf{C}_\chi = \mathbf{\Gamma}^{1/2} \mathbf{L}^{1/2} \mathbf{W}^{-1} (\mathbf{L}^{1/2})^T \mathbf{\Gamma}^{1/2} \quad (2)$$

where $\mathbf{L} = \mathbf{L}^{1/2} \mathbf{L}^{1/2}$ is a self-adjoint, M -step diffusion filter, \mathbf{W} is the matrix of grid-dependent weights associated with the discrete inner product with respect to which \mathbf{L} is self-adjoint (i.e., $\mathbf{L} = \mathbf{W}^{-1} \mathbf{L}^T \mathbf{W}$), and $\mathbf{\Gamma} = \mathbf{\Gamma}^{1/2} \mathbf{\Gamma}^{1/2}$ is a diagonal matrix of normalization factors to ensure that the diagonal elements of \mathbf{C}_χ are approximately equal to one. In continuous form, applying Eq. (2) to a vector ψ_0 , the discrete representation of the d -dimensional continuous variable ψ_0 , involves solving the elliptic equation

$$(1 - \nabla \cdot \boldsymbol{\kappa} \nabla)^M \gamma^{-1/2} \psi_M = \gamma^{1/2} \psi_0, \quad (3)$$

where M is a positive integer, $\gamma^{1/2}$ is a normalization function, $\nabla \cdot$ and ∇ are the d -dimensional divergence and gradient operators, and $\boldsymbol{\kappa}$ is the local diffusion tensor. For constant $\boldsymbol{\kappa}$, the kernel of the integral solution of (3) admits correlation functions from the Matérn class [Guttorp and Gneiting, 2006]:

$$c_d(r) = \gamma_d r^{M-d/2} K_{M-d/2}(r) \quad (4)$$

where $r = \|\mathbf{r}\|_{\boldsymbol{\kappa}^{-1}} = \|\mathbf{s} - \mathbf{s}'\|_{\boldsymbol{\kappa}^{-1}} = \sqrt{(\mathbf{s} - \mathbf{s}')^T \boldsymbol{\kappa}^{-1} (\mathbf{s} - \mathbf{s}')}$ is a non-dimensional distance between points \mathbf{s} and \mathbf{s}' in \mathbb{R}^d , $K_{M-d/2}(\cdot)$ is the Bessel function of the second kind of order $M - d/2$, and γ_d is a normalization constant. The directional length scales of the correlation function are controlled by $\boldsymbol{\kappa}$, which is a $d \times d$ symmetric positive-definite (and hence invertible) matrix. The fatness of the tails of the correlation functions, which in spectral space is related to the decay rate of the correlation spectrum at high wavenumbers, is controlled by the pseudo-time parameter M .

In the CERA system, \mathbf{L} is split into a product of 1D implicit diffusion operators that act separately in each of the coordinate directions defined by the three-dimensional (3D) model grid. This can be considered as an approximation of Eq. (3). It is invoked to simplify the requirements of the implicit solver: each 1D problem involves a small linear system matrix that can be solved using a direct method based on Cholesky decomposition. In the revised correlation model, we do not approximate the elliptic equation, but rather seek an approximate solution using an iterative solver. The solver employed is the Chebyshev iteration; it is a linear solver that requires as input an estimate of the extreme eigenvalues of the implicit diffusion system matrix. The eigenvalues are pre-computed using a Lanczos algorithm. The attractive properties of the Chebyshev solver for implicit diffusion-based correlation modelling in a global ocean model are discussed in Weaver *et al.* [2017, 2016]. An example of how the revised model improves the representation of correlations near complex geometry is illustrated in Fig. 2.

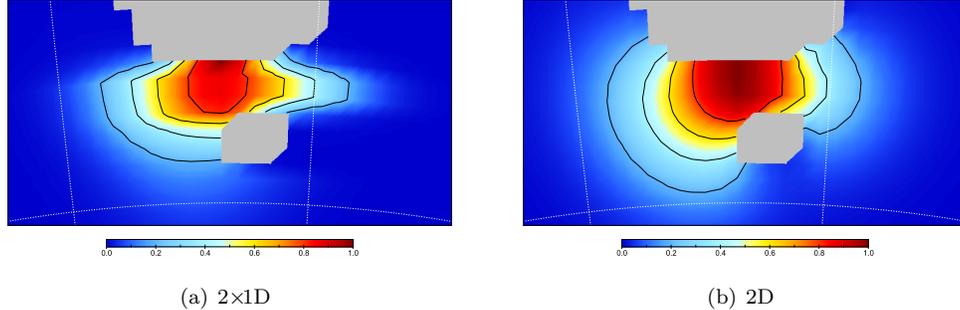


Figure 2: 2D correlations with respect to a point located next to the Australian coastline in the Bass Strait in ORCA1, produced by a diffusion operator with (a) the $2\times 1D$ formulation used by NEMOVAR in CERA; and (b) a newly developed unapproximated 2D formulation for NEMOVAR. The contour interval is 0.2 and the maximum value is 1.0. (From [Weaver *et al.* \[2016\]](#)).

2.2 Constructing ensemble perturbations for estimating \mathbf{B}

ECMWF produces an ensemble of ocean analyses using a set of perturbed background states and set of perturbed observations as input to each 3D-Var analysis. The set of perturbed analysis increments is then used to produce a set of perturbed ocean model forecasts where each forecast is driven by its own set of perturbed forcing fields [[Zuo *et al.*, 2017](#)]. The ocean model forecasts at a particular lead time constitute the ensemble of background states to be used for the next 3D-Var analysis cycle. The ensemble covariance of background states can provide valuable flow-dependent information for specifying \mathbf{B} . Estimating the ensemble covariance of the control variables requires deriving ensemble perturbations for the control variables from the ensemble perturbations of the state variables. The way this has been implemented in NEMOVAR is outlined in this section.

Let $\{\mathbf{x}_p\}$, $p = 0, \dots, N_e$, denote an ensemble of background states where ensemble member $p = 0$ corresponds to a deterministic (unperturbed) background state and the remaining members $p = 1, \dots, N_e$ to the set of perturbed background states. An ensemble of centred control vector perturbations can be generated from the $N_e + 1$ members using the relation

$$\boldsymbol{\epsilon}'_p = K_b^{-1}(\mathbf{x}_p) - \frac{1}{N_e + 1} \sum_{l=0}^{N_e} K_b^{-1}(\mathbf{x}_l), \quad p = 0, \dots, N_e \quad (5)$$

where K_b^{-1} is the inverse of the *nonlinear* balance operator. As mentioned earlier, however, the nonlinear T-S balance relation is not available, so direct use of Eq. (5) is not possible. To circumvent this problem, we remove the unperturbed member \mathbf{x}_0 from the ensemble and use it as the reference state for linearizing K_b . This allows us to define an ensemble of N_e centred control vector perturbations as

$$\boldsymbol{\epsilon}'_p = \mathbf{K}_b^{-1} \boldsymbol{\epsilon}'_p, \quad p = 1, \dots, N_e,$$

where \mathbf{K}_b^{-1} is the inverse of the balance operator (linearized about \mathbf{x}_0), and

$$\boldsymbol{\epsilon}'_p = \mathbf{x}_p - \frac{1}{N_e} \sum_{l=1}^{N_e} \mathbf{x}_l \quad (6)$$

are the centred state-vector perturbations.

2.3 Estimating covariance model parameters from ensemble perturbations

In this section we describe how ensemble perturbations are used to define the variances and correlation length scales of the background-error covariance model. To clarify notation, we rewrite Eq. (1) as

$$\mathbf{B}_m^e = \mathbf{K}_b \underbrace{(\widehat{\mathbf{D}}_m^e)^{1/2} \widehat{\mathbf{C}}_m^e (\widehat{\mathbf{D}}_m^e)^{1/2}}_{\widehat{\mathbf{B}}_m^e} \mathbf{K}_b^T \quad (7)$$

where the superscript ‘e’ indicates that the variance matrix $\widehat{\mathbf{D}}_m^e$ and parameters of the correlation model $\widehat{\mathbf{C}}_m^e$ are specified from ensembles.

2.3.1 Variances

An unbiased estimate of the background-error variances from the ensemble is given by

$$\text{var}(\boldsymbol{\epsilon}') = \frac{1}{N_e - 1} \sum_{p=1}^{N_e} (\boldsymbol{\epsilon}'_p \circ \boldsymbol{\epsilon}'_p) = \overline{\boldsymbol{\epsilon}' \circ \boldsymbol{\epsilon}'}$$

where \circ denotes the Schur (element-by-element) product of two vectors, and $\overline{(\cdot)}$ is shorthand notation for the normalized summation over p . These can be used to define the diagonal elements of $\widehat{\mathbf{D}}_m^e$.

2.3.2 Local correlation tensor

To use the ensemble to estimate the correlation matrix is much more challenging. Rather than trying to estimate the full matrix, here we focus on estimating the local curvature of the diffusion-implied (Matérn) correlation function near its peak. The curvature of a locally homogeneous and at least twice differentiable correlation function $C(\mathbf{r})$ is described by the correlation Hessian tensor [Chorti and Hristopulos, 2008; Hristopulos, 2002], $\mathbf{H} = \mathbf{H}(\mathbf{s})$, where $\mathbf{r} = \mathbf{s} - \mathbf{s}'$ is the separation vector between points \mathbf{s} and \mathbf{s}' in the domain. In \mathbb{R}^d , \mathbf{H} is a symmetric, positive-definite $d \times d$ matrix with elements

$$H_{mn}(\mathbf{s}) = - \left. \frac{\partial^2 C(\mathbf{r})}{\partial r_m \partial r_n} \right|_{\mathbf{r}=\mathbf{0}}, \quad m = 1, \dots, d; \quad n = 1, \dots, d,$$

where $\mathbf{s} = (s_1, \dots, s_d)$ and $\mathbf{r} = (r_1, \dots, r_d) = (s_1 - s'_1, \dots, s_d - s'_d)$. The correlation Hessian tensor has also been referred to as the *local correlation tensor* (LCT) in the data assimilation literature

[Michel *et al.*, 2016]. We will use this alternative terminology in what follows. The LCT is closely related to the local diffusion tensor $\boldsymbol{\kappa}$ of the diffusion-based correlation operator. For constant $\boldsymbol{\kappa}$, Weaver and Mirouze [2013] show that for the Matérn function (4),

$$\boldsymbol{\kappa} = \left(\frac{1}{2M - d - 2} \right) \mathbf{H}^{-1}. \quad (8)$$

In 1D, $\mathbf{H}^{-1} = L^2$ is a scalar where $L = \sqrt{-\partial^2 C(r_1)/\partial r_1^2|_{r_1=0}}$ is the 1D correlation length-scale following the usual definition given in Daley [1991]. For this reason, Weaver and Mirouze [2013] refer to \mathbf{H}^{-1} as the Daley tensor.

In practice, $\mathbf{H} = \mathbf{H}(\mathbf{s})$ can be approximated locally from sample statistics derived from the ensemble. Three formulae have been proposed in the literature. For a locally homogeneous correlation function, it can be shown that [Belo Pereira and Berre, 2006; Weaver and Mirouze, 2013]

$$\mathbf{H}(\mathbf{s}) = \frac{\overline{\nabla \varepsilon'(\mathbf{s}) (\nabla \varepsilon'(\mathbf{s}))^T} - \nabla \sigma(\mathbf{s}) (\nabla \sigma(\mathbf{s}))^T}{(\sigma(\mathbf{s}))^2} \quad (9)$$

where $\nabla \equiv (\partial/\partial s_1, \dots, \partial/\partial s_d)^T$, $\varepsilon'(\mathbf{s})$ is the continuous analogue of the centred random error vector $\boldsymbol{\varepsilon}'$, and $\sigma(\mathbf{s}) = \sqrt{\overline{(\varepsilon'(\mathbf{s}))^2}}$ is the standard deviation of the error. A practical difficulty with Eq. (9) for numerical applications is the minus sign in the numerator, which can destroy the positive definiteness of $\mathbf{H}(\mathbf{s})$ and hence lead to problems computing $(\mathbf{H}(\mathbf{s}))^{-1}$. Assuming that $\sigma(\mathbf{s})$ is slowly varying compared to $\varepsilon'(\mathbf{s})$ then the second term in the numerator can be neglected, resulting in the simpler expression [Sato *et al.*, 2009]

$$\mathbf{H}(\mathbf{s}) \approx \frac{\overline{\nabla \varepsilon'(\mathbf{s}) (\nabla \varepsilon'(\mathbf{s}))^T}}{(\sigma(\mathbf{s}))^2}. \quad (10)$$

However, if $\sigma(\mathbf{s})$ is not slowly varying then neglecting the second term can lead to a significant bias in the estimate $\mathbf{H}(\mathbf{s})$, as illustrated in the idealized experiments of Weaver and Mirouze [2013]. By introducing normalized perturbations $\bar{\varepsilon}'(\mathbf{s}) = \varepsilon'(\mathbf{s})/\sigma(\mathbf{s})$, Michel [2013] remarked that Eq. (9) can be written in the compact form

$$\mathbf{H}(\mathbf{s}) = \overline{\nabla \bar{\varepsilon}'(\mathbf{s}) (\nabla \bar{\varepsilon}'(\mathbf{s}))^T}, \quad (11)$$

which, by construction, ensures positive definiteness of $\mathbf{H}(\mathbf{s})$. Equation (11) is the formulation that has been implemented in NEMOVAR.

The ensemble average of the product $(\partial \bar{\varepsilon}'_i/\partial s_m)(\partial \bar{\varepsilon}'_j/\partial s_n)$ at a given point \mathbf{s} provides an estimate of $H_{mn}(\mathbf{s})$ for a given control variable. Let \mathbf{G}_m , $m = 1, \dots, d$, be the matrix representation of the discretized component derivative $\partial/\partial s_m$. For each ensemble member, we need to compute the gradient product at each of the N points on the analysis grid. The values can be assembled in an N -dimensional vector

$$\mathbf{g}_{mn} = \mathbf{G}_m \bar{\varepsilon}' \circ \mathbf{G}_n \bar{\varepsilon}'. \quad (12)$$

The ensemble average then provides an estimate of the terms needed to compute (11):

$$\mathbf{h}_{mn} = \overline{\mathbf{g}_{mn}}. \quad (13)$$

For a 3D control variable, there are 6 N -dimensional vectors that need to be estimated to define the full symmetric tensor, and an additional N -dimensional vector to define the variances. The estimation problem is thus of $O(N)$, which is much smaller than the original $O(N^2)$ problem of estimating the entire covariance matrix for the 3D control variable. Furthermore

In NEMOVAR we use centred finite differences to evaluate the derivatives. This requires some care, especially to evaluate the off-diagonal tensor elements. The numerical details are provided in the Appendix. For 3D control variables, only the off-diagonal elements of \mathbf{H} in the horizontal plane (H_{12} and H_{21}) are currently accounted for (i.e., we set H_{13} , H_{23} , H_{31} and H_{32} to zero). Let D_{mn} denote the elements of the $d \times d$ matrix defining the local Daley tensor \mathbf{H}^{-1} . For diagnostics purposes, it is useful to represent the Daley tensor as [Chorti and Hristopulos, 2008; Hristopulos, 2002]

$$\mathbf{H}^{-1} = \mathbf{R}\mathbf{L}\mathbf{R}^T \quad (14)$$

where \mathbf{R} is a rotation matrix, and $\mathbf{L} = \text{diag}(L_1^2, \dots, L_d^2)$ is a diagonal matrix whose components L_1^2 , L_2^2 etc. are the square of the length-scales along the principal axes of the ellipse describing the local anisotropy in the correlations. In particular, consider the 2D case where \mathbf{R} is defined by a rotation angle R_{12} measured positive counter-clockwise from the s_1 axis, and $\mathbf{L} = \text{diag}(L_1^2, L_2^2)$. In terms of the Daley tensor parameters D_{11} , D_{22} and $D_{21} = D_{12}$, we can derive the relations

$$L_1^2 = \frac{1}{2}(D_{11} + D_{22} + F), \quad (15)$$

$$L_2^2 = \frac{1}{2}(D_{11} + D_{22} - F), \quad (16)$$

$$R_{12} = \frac{1}{2} \cos^{-1} \left(\frac{D_{11} - D_{22}}{A_{12}} \right) \quad (17)$$

where

$$F = \sqrt{(D_{11} - D_{22})^2 + 4D_{12}^2}. \quad (18)$$

Given estimated values of D_{11} , D_{22} and $D_{21} = D_{12}$, the degree of anisotropy in the correlations can be determined by plotting the *aspect ratio* $A_{12} = L_1/L_2$ and rotation angle R_{12} as a function of grid-point. This will be illustrated in the next section using ensembles from the ECMWF system. Finally, for 3D fields, we simply have

$$L_3^2 = D_{33} \quad (19)$$

since we have neglected cross-correlations with the vertical coordinate.

Since the diffusion operator does not currently account for general anisotropic diffusion (i.e., non-zero off-diagonal elements in $\boldsymbol{\kappa}$), we cannot use the estimated off-diagonal elements of \mathbf{H}^{-1} . We have allowed for the following three options for defining the diagonal horizontal elements of $\boldsymbol{\kappa}$, via Eq. (8), given \mathbf{H}^{-1} .

1. Ignore the non-diagonal elements in the Daley tensor ($D_{12} = D_{21} = 0$):

$$\left. \begin{array}{l} \kappa_{11} \longrightarrow D_{11} \\ \kappa_{22} \longrightarrow D_{22} \end{array} \right\}.$$

2. Ignore the rotational anisotropic component in the diagonalized form of the Daley tensor ($R_{12} = 0$):

$$\left. \begin{array}{l} \kappa_{11} \longrightarrow L_1^2 \\ \kappa_{22} \longrightarrow L_2^2 \end{array} \right\}.$$

3. Assume the correlations are locally isotropic with scale equal to the geometric mean of L_1 and L_2 :

$$\left. \begin{aligned} \kappa_{11} &\longrightarrow L_1 L_2 \\ \kappa_{22} &\longrightarrow L_1 L_2 \end{aligned} \right\}.$$

The first and second options are adequate approximations only when the principal axes of \mathbf{H}^{-1} are approximately aligned with the coordinate axes of the model. There will clearly be regions where this is not the case (e.g., near coastlines). It is easy to see that by ignoring rotational anisotropy, the spatial extent of the correlations will be underestimated. The third option is a compromise, which will underestimate the length scale in one direction and overestimate it in the other direction, when L_1 and L_2 differ. While accounting for a fully non-diagonal $\boldsymbol{\kappa}$ is desirable, it requires non-trivial extensions to the diffusion operator and would likely substantially increase the computational cost of applying the diffusion operator. The developments described in Section 2.4 are thought to be a better way forward for accounting for fully anisotropic correlations.

2.3.3 Dealing with sampling error

Even though the covariance model allows us to reduce the number of tunable covariance parameters significantly, the estimation problem is still of $O(N)$ where N is typically several orders of magnitude greater than the number of ensemble members N_e . As such, sampling error will still be large and must be filtered to obtain usable estimates for the data assimilation system.

One possibility to reduce sampling error is to time average the ensemble perturbations. For example, [Daget *et al.* \[2009\]](#) employed a 90-day moving average when estimating background-error variances from a 8 perturbed-member ensemble in a low-resolution global ocean configuration (ORCA2). The possibility of using a moving-average time filter has been made available in NEMOVAR. While time-averaging reduces sampling error, it comes at the expense of filtering precious flow-dependent information in the background-error variances.

A better way of reducing sampling error is with an objective spatial filter. [Ménétrier *et al.* \[2015a\]](#) derived optimality criteria for the filtering of sample variances by combining theoretical results from optimal linear filtering and centred moments estimation. In practice, the optimality criteria translate into the following requirements on the design of the spatial filter:

$$\left. \begin{aligned} \mu^S[\widehat{\mathbf{v}}] &= \mu^S[\widetilde{\mathbf{v}}] \\ \mathcal{C} &= 0 \end{aligned} \right\} \quad (20)$$

where $\widetilde{\mathbf{v}}$ is the raw variance, $\widehat{\mathbf{v}}$ is the filtered variance, $\mu^S[\cdot]$ denotes spatial average, which is used as a practical substitute for the mathematical expectation operator, and \mathcal{C} is an optimality function. Assuming a non-Gaussian (NG) error distribution, the optimality function $\mathcal{C} = \mathcal{C}^{\text{NG}}$ is

$$\begin{aligned} \mathcal{C}^{\text{NG}} &= \mu^S[\widetilde{\mathbf{v}} \circ \widetilde{\mathbf{v}}] - \frac{N_e(N_e - 2)(N_e - 3)}{(N_e - 1)(N_e^2 - 3N_e + 3)} \mu^S[\widetilde{\mathbf{v}} \circ \widehat{\mathbf{v}}] \\ &\quad - \frac{N_e^2}{(N_e - 1)(N_e^2 - 3N_e + 3)} \mu^S[\widetilde{\boldsymbol{\xi}}] \end{aligned} \quad (21)$$

where $\widetilde{\boldsymbol{\xi}}$ is the raw fourth-order moment, and \circ denotes Schur product. The optimality function $\mathcal{C} = \mathcal{C}^{\text{G}}$ for Gaussian (G) errors has a simpler form:

$$\mathcal{C}^{\text{G}} = \mu^S[\widetilde{\mathbf{v}} \circ \widetilde{\mathbf{v}}] - \left(\frac{N_e + 1}{N_e - 1} \right) \mu^S[\widetilde{\mathbf{v}} \circ \widehat{\mathbf{v}}]. \quad (22)$$

The first condition in (20) requires that the filter is mean preserving. To satisfy the second condition, we assume that the filter is homogeneous and isotropic with a single, adjustable length-scale parameter L_f ; \mathcal{C}^{NG} or \mathcal{C}^{G} is then considered to be a function of L_f only. As shown in Appendix C of Ménétrier *et al.* [2015a], \mathcal{C}^{NG} and \mathcal{C}^{G} are both monotonically increasing functions of L_f if the spectral coefficients of the filter decrease monotonically when the values of L_f increase. With no filtering ($L_f = 0$), the values of \mathcal{C}^{NG} and \mathcal{C}^{G} are negative. The zero value of \mathcal{C}^{NG} or \mathcal{C}^{G} is thus achieved with a unique value of L_f . The objective filtering scale L_f^* can be found iteratively using a bisection method.

The filtering algorithm has been implemented in NEMOVAR, where we have taken into account both types of optimality functions, \mathcal{C}^{NG} and \mathcal{C}^{G} . For the filter, we use a constant-coefficient 2D implicit diffusion operator. The Matérn-like smoothing kernels implied by this filter satisfy the conditions underlying the optimality criteria (20). For 3D fields, the filter is applied independently in each model level k to determine a globally averaged filtering scale $L_f^*(k)$. As in the correlation operator, the Chebyshev iteration is used as the implicit solver in the diffusion operator. Since the diffusion coefficient (filtering scale) is optimized, the extreme eigenvalues of the system matrix must be recomputed on each iteration of the bisection method. A Lanczos algorithm requires solving the system to estimate the eigenvalues, so is clearly inappropriate for this application where only one solve is required per iteration of the bisection method. Instead, we employ the simpler Power Method to provide an approximate estimate of the maximum eigenvalue. The minimum eigenvalue is approximately equal to one, so does not need to be explicitly computed. Yet another approach is to use an analytical estimate of the maximum value. Weaver *et al.* [2016] present an approximate expression for the maximum eigenvalue of the 2D system matrix in \mathbb{R}^2 under the assumption of uniform resolution and constant diffusion coefficients in the x and y directions. NEMO employs a non-uniform curvilinear grid where the horizontal resolution at grid point (i, j) is defined by scale factors $e_1(i, j)$ and $e_2(i, j)$. For a constant diffusion coefficient $\kappa(k) = (L_f(k))^2 / (2M - 4)$ in level k , we estimate the maximum eigenvalue in each level from the expression (cf. Eq. (22) in Weaver *et al.* [2016])

$$\lambda_{\max}(k) \approx 1 + 4\kappa(k) \max_{i,j} \left(\frac{1}{(e_1(i, j))^2} + \frac{1}{(e_2(i, j))^2} \right). \quad (23)$$

This cheaper approach produces very similar estimates to those of the Power Method.

Figure 3 illustrates the convergence properties of the algorithm when filtering temperature error variances estimated with different ensemble sizes. The ensembles have been produced for the ORCA1 configuration for the date 04/06/2011, using the ensemble-generation procedure developed for ECMWF’s Ocean ReAnalysis System 5 (ORAS5) [Zuo *et al.*, 2017]. The algorithm converges to an acceptably accurate filtering scale in less than 10 iterations. Here, the algorithm is stopped when the length-scale increment is less than 0.01° (1.1 km), although a less stringent criterion would probably be acceptable. Similar results are obtained when using the non-Gaussian (upper panels) and Gaussian (lower panels) optimality functions, except for the small ensemble size ($N_e = 5$) where a slightly larger filtering scale is obtained with the non-Gaussian optimality function (1.4° compared to 1.2°). The relative insensitivity to the Gaussian and non-Gaussian formulations is likely linked to the low resolution of ORCA1. As expected, the objective filtering scale decreases when the ensemble size increases. The raw and filtered temperature standard deviations with the 11-member ensemble are shown in Fig. 4. The standard deviations are largest in the tropics and in the Gulf Stream region. Filtering eliminates small-scale sampling noise present in the raw standard deviation estimates, and results in the maximum value being reduced from 1.9 K to 1.0 K. After filtering, the

standard deviations still exhibit significant spatial variability.

It seems unlikely that the filtering scale in each level should be constant, especially with high resolution (eddy permitting/resolving) configurations and non-uniform grids. This constraint can be relaxed by splitting the domain into smaller regions and estimating the objective filtering scale separately in each region. [Ménétrier *et al.* \[2015b\]](#) used this approach in a convective-scale numerical weather prediction model. While this is a desirable refinement to the algorithm, it is more complicated and has not yet been implemented in NEMOVAR. Our experience with the single filtering scale version of the algorithm has shown that its convergence properties are quite sensitive to grid-points near boundaries. To alleviate this problem, we currently ignore all ocean grid-points within 500 km of land boundaries when optimizing L_f .

The objective filtering algorithm can also be applied to the LCT, although this must be done with some care to maintain positive definiteness of the LCT and to avoid violating assumptions in the underlying optimal filtering theory. We follow the basic approach of [Michel *et al.* \[2016\]](#) (their “COV” method). First, as with a standard covariance matrix, we factor the LCT into a product of a matrix of “standard deviations” with diagonal elements $\sqrt{H_{11}}$ and $\sqrt{H_{22}}$, and a “correlation” matrix with diagonal elements equal to one and off-diagonal elements equal to $H_{12}/\sqrt{H_{11}H_{22}}$. This factorization allows us to determine different filtering scales for the diagonal and cross-components of \mathbf{H} , while preserving positive definiteness. The objective filtering algorithm cannot be applied directly to the derivative products involving the normalized perturbations $\hat{\epsilon}'$ because of the presence of σ in the denominator (the sample mean of these terms are not normal “variances”). To counter this problem, [Michel *et al.* \[2016\]](#) suggest a suboptimal procedure that involves first applying the algorithm to the derivative products involving the unnormalized perturbations ϵ' to determine objective filtering scales L_{11}^* and L_{22}^* for the respective elements H_{11} and H_{22} in the approximate expression (10), and then to use these filtering scales to filter H_{11} and H_{22} in the unapproximated expression (9). For $H_{12}/\sqrt{H_{11}H_{22}}$, the filtering scale is set to $L_{12}^* = \sqrt{L_{11}^*L_{22}^*}$, the geometric mean of the objective filtering scales for H_{11} and H_{22} .

Results for the filtered 100 m depth temperature tensor elements with the 11-member ensemble are shown in Fig. 5. The elements are displayed in terms of the principal axes L_1 and L_2 , aspect ratio $A_{12} = L_1/L_2$, and rotation angle R_{12} of the Daley tensor (Eqs (15)–(18)). Largest L_1 scales tend to be in the tropical regions whereas the smallest L_1 scales tend to be in boundary current regions and in the southern oceans (Fig 5a). In these regions where the L_1 scales are small, the L_2 scales also tend to be small (Fig. 5b). The L_2 scales are also small in the equatorial waveguide. In this region the aspect ratio A_{12} is noticeably large (Fig. 5c) and the rotation angle is small (Fig. 5d), reflecting pronounced east-west anisotropy. East-west anisotropy near the equator is accounted for in the parameterized diffusion tensor using a simple analytical function of latitude, with a maximum aspect ratio of 3.3 directly at the equator (not shown). This is roughly consistent with the aspect ratio from the ensemble estimates although the latter has much more spatial variability and can have much larger values locally (the maximum value is 22.6).

While the basic patterns in L_1 and L_2 seem reasonable, the values are still quite noisy. The fact that they are noisier than the ensemble estimates of the standard deviations is perhaps not surprising since the numerical computation of \mathbf{H} , which involves derivatives of perturbations, is inherently noisy. Furthermore, the filtering scales obtained using the objective filtering procedure seem to be underestimated, with values of 0.52° for H11 and 0.38° for H22. At many grid points, these scales fall well below the local resolution. In level 10, the minimum and maximum grid sizes, as defined by the geometric mean of the local horizontal scale factors, is 0.21° and 0.96° , respectively. Besides providing limited filtering, using below-grid filtering scales in the diffusion operator can

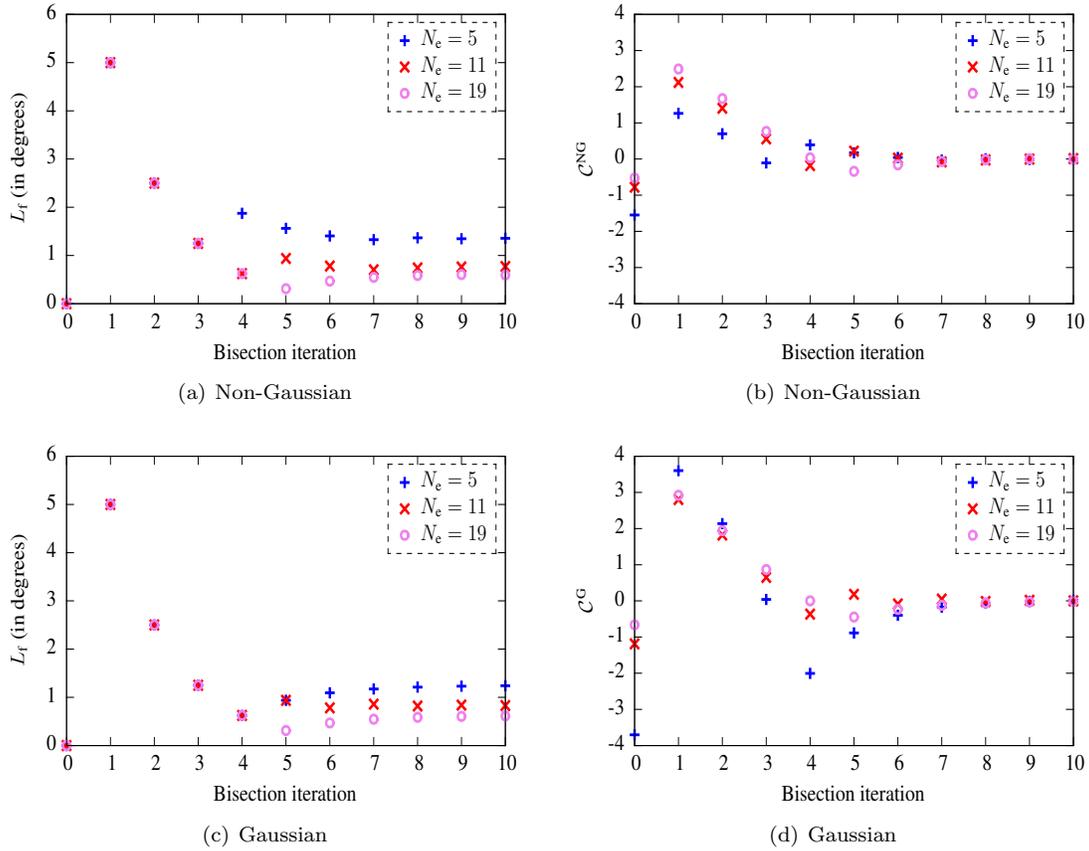


Figure 3: Values of the objective filtering scale (L_f in degrees where $1^\circ \approx 111$ km) and of the optimality function (C) as a function of iteration of the bisection method for temperature error variances at approximately 100 m depth (model level 10) estimated with different ensemble sizes with ORCA1 (see legend). (a)-(b) Non-Gaussian case (Eq. (21)); (c)-(d) Gaussian case (Eq. (22)).

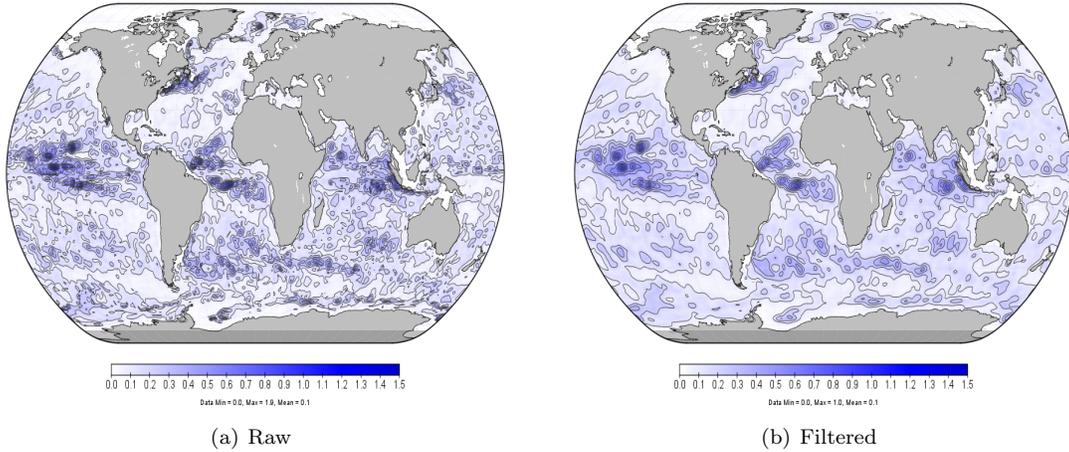


Figure 4: Temperature error standard deviations (K) at approximately 100 m depth (model level 10) estimated from an 11 perturbed-member ensemble with ORCA1. (a) Raw sample estimate; and (b) sample estimate obtained after filtering the variances with an objectively determined filtering scale of 0.83° (92 km). The filter scale is taken from the Gaussian case in Fig. 3c (red crosses). The maximum values of the raw and filtered standard deviations are 1.9 K and 1.0 K, respectively.

be an additional source of numerical noise [Weaver and Courtier, 2001]. It is not clear what is responsible for the apparent under filtering: it could be related to problems with assumptions in the heuristic method of Michel *et al.* [2016], or it could point to the need to estimate geographically dependent filtering scales.

Besides being somewhat noisy, the ensemble-estimated correlation scales can lead to practical difficulties when applying them in the diffusion operator. First, as can be seen in Figs. 5a and b, the scales can be quite large locally. Large scales will increase the condition number of the implicit diffusion system matrix, and can result in slower convergence of the Chebyshev iteration solver [Weaver *et al.*, 2016]. Second, updating the correlation scales on each cycle of the data assimilation system would require re-estimating the normalization factors of the diffusion operator to ensure unit-amplitude correlations. When there is large spatial variability in the scales, as in Figs. 5a and b, the accuracy of analytical methods to approximate the normalization factors will be poor. In such cases, randomization with a possibly large sample of random vectors would be needed to obtain an adequate approximation [Weaver and Courtier, 2001]. Both of these issues may significantly increase the computational cost of the diffusion algorithm. A more practical way of exploiting the ensembles for this problem may be to use a climatological estimate of the scale tensor (i.e., an ensemble average over many cycles of the data assimilation system). The greater effective sample size would reduce sampling error and there would be no need to re-compute normalization factors on each cycle, although this simplification would be achieved at the expense of losing flow-dependent information on the background-error correlation scales.

The ensemble estimate of L_3 for temperature at 5 m depth (level 1) is shown in Fig. 6a. As might

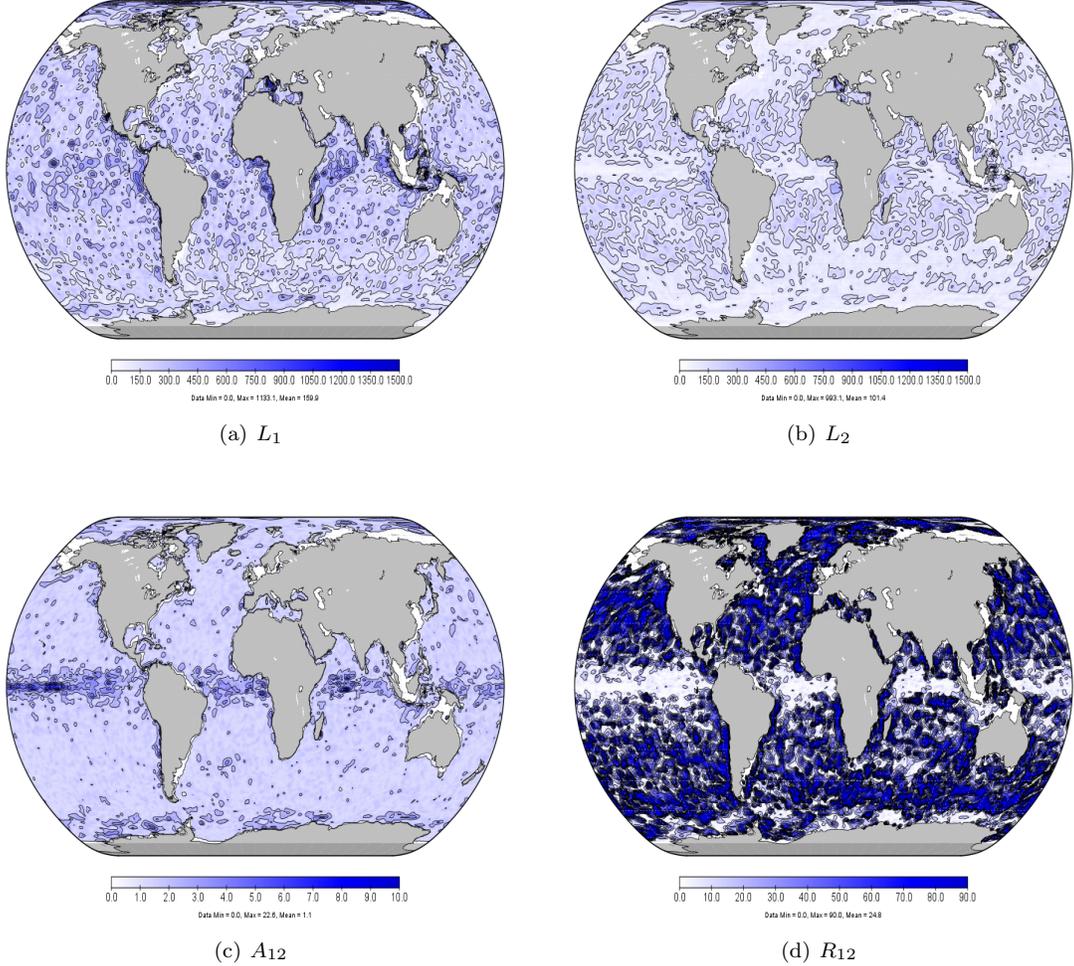


Figure 5: The principal axes (a) L_1 (Eq. (15)) and (b) L_2 (Eq. (16)), (c) the aspect ratio $A_{12} = L_1/L_2$, and (d) the rotation angle R_{12} (Eq. (17)), associated with the local Daley tensor \mathbf{H}^{-1} (Eq. (14)) for background-error temperature at approximately 100 m depth in ORCA1. \mathbf{H} has been estimated from an 11-member ensemble using Eq. (11). The elements of \mathbf{H} have been filtered using the objective procedure of Michel *et al.* [2016]. The objectively-estimated filtering scales were 0.52° for H_{11} , 0.38° for H_{22} , and 0.44° for H_{12} . The scale in panels (a) and (b) is in kilometres; the scale in panel (b) is dimensionless; the scale in panel (c) is in degrees.

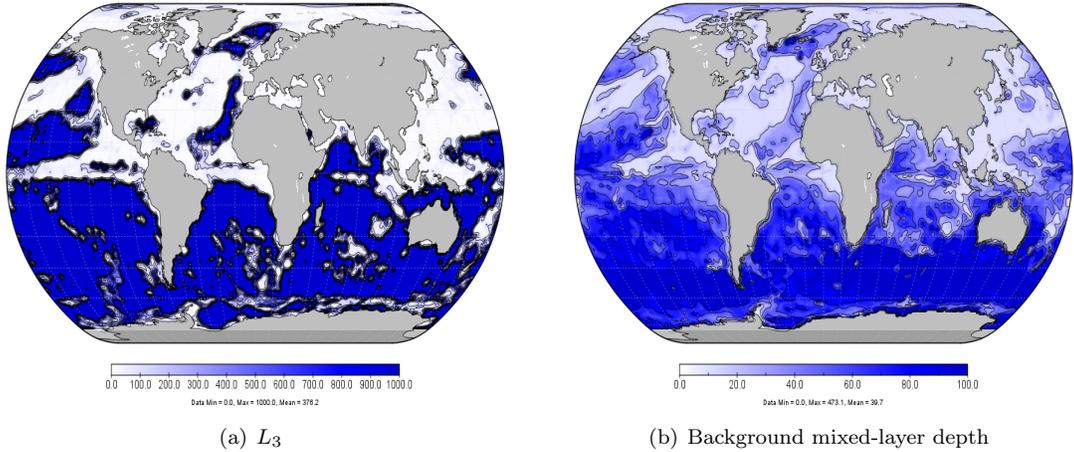


Figure 6: (a) The length of the principal axis L_3 (Eq. (19)) of the local Daley tensor for background-error temperature at approximately 5 m depth in ORCA1. The LCT element H_{33} has been filtered using the objective procedure of Michel *et al.* [2016]. The objectively-estimated filtering scale was 0.75° . For comparison, the background mixed-layer depth is shown in panel (b). The scale in both panels (a) and (b) is in metres.

be expected, the spatial patterns closely match those of the background mixed-layer depth shown in Fig. 6b. The vertical length scales are generally small in the northern hemisphere and large in the southern hemisphere, consistent with shallow mixed-layer depths in the northern hemisphere and deep mixed-layer depths in the southern hemisphere in the boreal spring/summer. In regions where the mixed-layer depth is deep, the magnitude of L_3 can be much larger than the mixed-layer depth itself. However, the magnitude of L_3 can change rapidly with depth, especially near the base of the mixed layer. In the reanalysis used to generate the ensembles, the background-error vertical correlation length scales were parameterized to be proportional to the local vertical grid size. Waters *et al.* [2015] developed an alternative vertical length-scale parameterization in NEMOVAR, in terms of the background mixed-layer depth. Figure 6 indicates that it would be a sensible parameterization to employ in future reanalyses.

2.3.4 Inflation and hybrid covariance parameter formulation

The dispersion of an ensemble tends to be underestimated when N_e is small and when there are deficiencies in the ensemble-generation procedure (e.g., neglected model error sources). To compensate for this problem, we allow for the ensemble perturbations to be inflated by a small amount with respect to their mean. That is, we let

$$\varepsilon'_i \rightarrow r \varepsilon'_i$$

where r is a global *inflation factor* greater than one. The operation \rightarrow means “replace with”. In NEMOVAR, the inflation factor is implemented as a factor multiplying the ensemble estimate of the

background-error standard deviations. Furthermore, for 3D fields, the inflation factor is allowed to be specified differently in each model level. The inflation factor can be estimated in different ways; e.g., by using observation-space diagnostics to compare the ensemble spread with the expected background-error standard deviations [Desroziers *et al.*, 2005], or by comparing the spread to the root-mean-squared error of the ensemble mean [Leutbecher, 2009].

Let \mathbf{v}_m and \mathbf{v}_e be vectors containing the parameterized (modelled) and ensemble-estimated variances, respectively. In order to improve the robustness of the ensemble variances, we can combine them with the parameterized variances in a hybrid variance formulation:

$$\mathbf{v}_h = \alpha_m^2 \mathbf{v}_m + \alpha_e^2 \mathbf{v}_e \quad (24)$$

where α_m and α_e are positive weights, with possibly different values for each field and in each model level. This hybrid variance parameter formulation has been implemented in NEMOVAR.

The specification of the weights α_m and α_e is more meaningful if the spatially averaged hybrid variance v_h was anchored to a reference variance v_r . We can do this by normalizing the hybrid variance:

$$\mathbf{v}_h \rightarrow \left(\frac{v_r}{v_h} \right) \mathbf{v}_h.$$

In NEMOVAR, we define v_r as either the spatially averaged parameterized variance (v_m) or the spatially averaged (inflated) ensemble variance (v_e). The latter would seem preferable given a well-tuned ensemble.

Figure 7 compares the parameterized temperature error standard deviations near 100 m depth (left panel) and those obtained using a hybrid formulation (right panel), where the inflation factor, hybrid weights and normalization factor are specified as indicated in the figure caption. The parameterized standard deviations are defined in terms of the vertical gradient of the background temperature field, to focus large errors in regions where there is large variability in the thermocline. The parameterized standard deviations have been smoothed with a 3D diffusion operator, using smoothing scales identical to those used in the correlation model \mathbf{C}_m . As can be seen in Fig. 7a, the parameterized standard deviations are largest in the tropics, and have large-scale spatial structure that is reasonably consistent with that of the standard deviation estimates from the ensemble (Fig. 4). However, the large standard deviations in western boundary current regions, especially the Gulf Stream, that are present in the ensemble estimate are not at all captured by the parameterization. As expected, the hybrid estimate captures the dominant spatial structure of the ensemble standard deviation estimates and increases the standard deviations in regions where the ensemble estimates are small.

In a similar way, we can combine the elements of the parameterized and ensemble-estimated LCTs in a hybrid tensor formulation [Sato *et al.*, 2009; Weaver and Mirouze, 2013]:

$$\mathbf{h}_{mn}^h = \gamma_m^2 \mathbf{h}_{mn}^m + \gamma_e^2 \mathbf{h}_{mn}^e$$

where γ_m and γ_e are positive weights analogous to α_m and α_e for the hybrid variances. The vector \mathbf{h}_{mn}^e contains the elements of the ensemble-estimated LCT (Eq. (13)). The values of \mathbf{h}_{mn}^e for different pairs (m, n) are first collocated to the variable grid-point of interest, following the procedure described at the end of Appendix A. The values of \mathbf{h}_{mn}^m correspond to the elements of the inverse of the parameterized diffusion tensor, also collocated to the variable grid-point of interest for different pairs (m, n) . As with the hybrid variances, the spatial average of the hybrid tensor elements h_{mn}^h

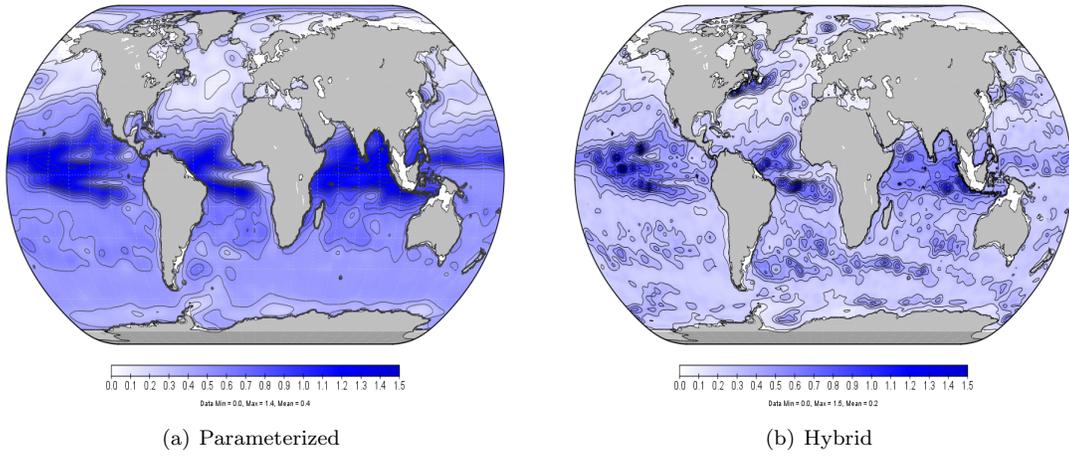


Figure 7: Temperature error standard deviations (K) at approximately 100 m depth (model level 10) in ORCA1 obtained using (a) the standard NEMOVAR parameterization in terms of the background vertical temperature gradient; and (b) a hybrid estimate (Eq. (24)) with weights $\alpha_m = 0.25$ and $\alpha_e = 0.75$, inflation factor $r = 2$, and reference value for normalization $v_r = v_m$. The non-inflated filtered ensemble standard deviations used in the hybrid estimate are based on those displayed in Fig. 4b.

can be fixed to a reference value h_{mn}^r through an appropriate normalization:

$$\mathbf{h}_{mn}^h \rightarrow \left(\frac{h_{mn}^r}{h_{mn}^h} \right) \mathbf{h}_{mn}^h,$$

where h_{mn}^r is defined as either the spatially averaged parameterized tensor element (h_{mn}^m) or the spatially averaged ensemble-estimated tensor element (h_{mn}^e). Together, the collocated elements of \mathbf{h}_{mn}^h define the hybrid LCT, \mathbf{H}_h . The scaled inverse of the LCT then provides the hybrid diffusion tensor via Eq. (8). This hybrid tensor formulation has been implemented in NEMOVAR.

2.4 Pure and hybrid EnVar B formulations

Rather than using ensemble perturbations to estimate parameters of the covariance model, an alternative approach in variational assimilation is to use the ensemble perturbations directly to provide a low-rank sample estimate of the covariance matrix, and then to localize this matrix in order to damp covariances at remote separation distances, which are attributed to sampling error. This **B** formulation is usually referred to as (pure) EnVar (e.g., see Bannister [2017] for a review). The localized sample covariance matrix can be linearly combined with the modelled covariance matrix to provide a more robust covariance matrix. This **B** formulation is usually called hybrid EnVar [Bannister, 2017]. Both these formulations have been developed for NEMOVAR as described below.

2.4.1 Pure EnVar

Let

$$\mathbf{X} = \frac{1}{\sqrt{N_e - 1}} \left(\epsilon'_1 \quad \dots \quad \epsilon'_{N_e} \right)$$

where ϵ'_p , $p = 1, \dots, N_e$, are the centred state-vector perturbations defined by Eq. (6). A sample estimate of the background-error covariance matrix is then

$$\mathbf{B}_e = \mathbf{X} \mathbf{X}^T$$

where we use the subscript ‘e’ to indicate that \mathbf{B}_e is based on a raw *ensemble* estimate. The rank of \mathbf{B}_e is $N_e \ll N$. In data assimilation applications, rank deficiency in \mathbf{B}_e can lead to problems with fitting the data and to spurious analysis increments because of unrealistic covariances in \mathbf{B}_e associated with sampling error. Localization of the sample covariance matrix is necessary to reduce the effects of sampling error on the covariances. This can be done by forming the Schur (element-by-element) product of the sample covariance matrix with a specified spatially-limited correlation matrix \mathbf{C}_L :

$$\mathbf{B}_e = \mathbf{C}_L \circ \mathbf{X} \mathbf{X}^T. \tag{25}$$

Sample covariances will be damped in areas where the elements of \mathbf{C}_L are small.

In NEMOVAR, localization is performed on the sample *correlation* matrix of the control variables. Let

$$\hat{\mathbf{X}} = \mathbf{K}_b^{-1} \mathbf{X} = \frac{1}{\sqrt{N_e - 1}} \left(\epsilon'_1 \quad \dots \quad \epsilon'_{N_e} \right)$$

where $\boldsymbol{\varepsilon}'_p$ are the transformed centred perturbations (Eq. (6)). Furthermore, let $\widehat{\mathbf{D}}_e$ be the diagonal matrix containing the sample variances of $\widehat{\boldsymbol{\varepsilon}}'_p$; i.e., the diagonal of $\widehat{\mathbf{D}}_e$ is the diagonal of the matrix $\widehat{\mathbf{X}}\widehat{\mathbf{X}}^T$. Now, define a transformed perturbation matrix,

$$\widetilde{\mathbf{X}} = \widehat{\mathbf{D}}_e^{-1/2} \widehat{\mathbf{X}} = \left(\widetilde{\boldsymbol{\varepsilon}}'_1 \quad \dots \quad \widetilde{\boldsymbol{\varepsilon}}'_{N_e} \right) \quad (26)$$

where $\widetilde{\boldsymbol{\varepsilon}}'_p = \widehat{\mathbf{D}}_e^{-1/2} \boldsymbol{\varepsilon}'_p$ are normalized transformed centred perturbations. The sample covariance matrix $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T$ is thus a correlation matrix since all of its diagonal elements are equal to one. We then define the localized sample covariance matrix as

$$\mathbf{B}_e = \mathbf{K}_b \underbrace{\widehat{\mathbf{D}}_e^{1/2} \left(\mathbf{C}_L \circ \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T \right) \widehat{\mathbf{D}}_e^{1/2}}_{\widehat{\mathbf{B}}_e} \mathbf{K}_b^T. \quad (27)$$

We prefer to define \mathbf{B}_e by Eq. (27) rather than by Eq. (25) for two reasons. First, Eq. (27) gives us more control to filter sampling error by allowing us to spatially filter the ensemble variances in addition to localizing the correlations. The variance filtering can be done using the objectively-based techniques described in the previous section for the covariance model. Note that variance filtering is performed on the variances in $\widehat{\mathbf{D}}_e$ that are used after localization in Eq. (27), but not on the variances used to define the perturbation matrix (26). Second, Eq. (27) gives us more flexibility in defining multivariate covariances as they can be defined as a combination of those implied by the balance operator and those estimated from the ensemble. The use of the balance operator in Eq. (27) can compensate for possible disruptions to balance due to localization [Kepert, 2009]. Furthermore, when using multiple outer loops in the variational minimization algorithm together with a hybrid \mathbf{B} formulation (see Section 2.4.3), \mathbf{K}_b must be applied in the (generalized) observation operator to enable it to be re-linearized in a consistent manner. This is possible when combining the covariance model \mathbf{B}_m (Eq. (1)), with \mathbf{B}_e defined by Eq. (27), but is not possible with Eq. (25). Multiple outer loops are central to the CERA framework for allowing coupling of the atmospheric and ocean data assimilation systems, so maintaining a consistent outer-loop capability is viewed as essential.

2.4.2 Localization

The direct element-by-element computation required by the Schur product in Eq. (27) is prohibitively expensive with large matrices. It is straightforward to show that Eq. (27) can be written in the equivalent form

$$\mathbf{C}_L \circ \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T = \sum_{p=1}^{N_e} \boldsymbol{\Lambda}_p \mathbf{C}_L \boldsymbol{\Lambda}_p \quad (28)$$

where $\boldsymbol{\Lambda}_p$ is a diagonal matrix whose diagonal is defined by $\widetilde{\boldsymbol{\varepsilon}}'_p$ (the p th column of $\widetilde{\mathbf{X}}$). The right-hand side of Eq. (28) can also be expressed in “square-root” form as

$$\sum_{p=1}^{N_e} \boldsymbol{\Lambda}_p \mathbf{C}_L \boldsymbol{\Lambda}_p = \widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^T \quad (29)$$

where

$$\widetilde{\mathbf{U}} = \left(\boldsymbol{\Lambda}_1^{1/2} \mathbf{U}_L \quad \dots \quad \boldsymbol{\Lambda}_{N_e}^{1/2} \mathbf{U}_L \right) \quad (30)$$

and $\mathbf{C}_L = \mathbf{U}_L \mathbf{U}_L^T$. Equations (29) and (30) involve standard matrix-matrix computations. In variational assimilation, these matrices can be defined as operators in order to perform the matrix-vector computations required by the minimization algorithm. In NEMOVAR, localization via \mathbf{C}_L is done using a diffusion-based correlation operator. Applying Eq. (29) to a vector thus involves applying a diffusion-based correlation operator N_e times. While this is potentially costly, it is still tractable in contrast to the direct evaluation of a Schur product of two large matrices. Possible ways of reducing the cost localization are discussed in Section 4.

A localized sample background-error covariance matrix based on Eqs (27), (29) and (30) has been implemented in NEMOVAR. We have implemented the “square-root” form of the localization (the right-hand side of Eq. (29)), rather than the full form (the left-hand side of Eq. (29)), to allow for randomization applications and to provide more flexibility in NEMOVAR with regard to the choice of preconditioned minimization algorithm.

The specific formulation of \mathbf{C}_L is important in terms of controlling the computational cost and flexibility of the localization operator. Five formulations of \mathbf{C}_L have been implemented in NEMOVAR, which we present below in terms of the choice of “square-root” factor \mathbf{U}_L . From Eq. (27) we note that \mathbf{U}_L acts on vectors whose components correspond to the control variables T , S_u , η_u , u_u and v_u .

1. *No localization.*

When no localization is employed, \mathbf{C}_L is a rank-one matrix of 1s and \mathbf{U}_L is an N -dimensional vector of 1s:

$$\mathbf{U}_L = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

While this formulation is not recommended when using a small ensemble, it has been implemented as a limiting case for academic purposes. The non-localized sample covariance matrix has a low-rank matrix representation $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$. A low-rank \mathbf{B} matrix derived from a truncated set of empirical orthogonal functions (EOFs) was developed for NEMOVAR by the Met Office as part of deliverable D2.1. The EOF-based covariance matrix also has the basic form $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$, but with $\tilde{\mathbf{X}} \rightarrow \mathbf{E}\mathbf{P}^{1/2}$ where \mathbf{E} is a rectangular matrix whose columns are the dominant eigenvectors (EOFs) and \mathbf{P} is a diagonal matrix containing the corresponding eigenvalues. The developments for the EOF-based \mathbf{B} matrix exploited the code infrastructure developed here for the non-localized ensemble-based \mathbf{B} .

2. *Univariate and separate localization for each control variable.*

\mathbf{C}_L is a full-rank matrix where \mathbf{U}_L is an $N \times N$ block-diagonal matrix

$$\mathbf{U}_L = \text{diag}(\mathbf{U}_T, \mathbf{U}_{S_u}, \mathbf{U}_{\eta_u}, \mathbf{U}_{u_u}, \mathbf{U}_{v_u}) \quad (31)$$

where $\mathbf{C}_\chi = \mathbf{U}_\chi \mathbf{U}_\chi^T$ is a diffusion-based (unit-amplitude) localization operator for the control variable χ . This formulation gives full flexibility for defining different localization matrices for different variables but completely filters out cross-covariance information between control variables. As such, multivariate covariances are described entirely by the balance operator.

3. *Multivariate and common localization for each control variable.*

We define \mathbf{C}_L from the factor

$$\mathbf{U}_L = \begin{pmatrix} \mathbf{I} \\ \mathbf{I} \\ \mathbf{R}_{T \rightarrow \eta} \\ \mathbf{A}_{T \rightarrow u} \\ \mathbf{A}_{T \rightarrow v} \end{pmatrix} \mathbf{U}_T \quad (32)$$

where $\mathbf{C}_T = \mathbf{U}_T \mathbf{U}_T^T$ is a 3D localization operator defined at T -points on the staggered Arakawa C-grid used in NEMO, and \mathbf{U}_L is an $N \times N_T$ matrix where N_T is the number of 3D T points. \mathbf{C}_L is then a matrix of rank N_T . The matrix \mathbf{I} is the $N_T \times N_T$ identity matrix. The matrix $\mathbf{R}_{T \rightarrow \eta}$ is a restriction operator from 3D T -points to 2D η -points, which are co-located with T -points in the first level. $\mathbf{R}_{T \rightarrow \eta}^T = \mathbf{E}_{\eta \rightarrow T}$, which appears in the factor \mathbf{U}_L^T , is an extension operator from 2D η -points to 3D T -points. The matrices $\mathbf{A}_{T \rightarrow u}$ and $\mathbf{A}_{T \rightarrow v}$ represent local averaging operators from T -points to u - and v -points, respectively. Similarly, the matrices $\mathbf{A}_{T \rightarrow u}^T = \mathbf{A}_{u \rightarrow T}$ and $\mathbf{A}_{T \rightarrow v}^T = \mathbf{A}_{v \rightarrow T}$, which appear in the factor \mathbf{U}_L^T , represent local averaging operators from u - and v -points to T -points, respectively.

Contrary to Eq. (31), Eq. (32) does not allow different localization matrices for different variables. However, it does have advantages over Eq. (31) in other respects. First, it is cheaper than Eq. (31) since only one application of a 3D diffusion operator is required per ensemble member, instead of four applications of a 3D diffusion operator and one application of a 2D diffusion operator per ensemble member. Second, the memory requirements with the ‘‘square-root’’ \mathbf{B} -preconditioned conjugate gradient algorithm available in NEMOVAR are also smaller with Eq. (32) since they depend on the number of columns of the first operator component of \mathbf{U}_L (in this case \mathbf{U}_T). Third, the localization operator corresponding to Eq. (32) has non-zero blocks associated with the cross-variable blocks in the sample covariance matrix. This means that multivariate covariance information from the ensemble is retained (and localized), unlike with Eq. (31) where it is discarded entirely. Since the cross-variable covariances in the ensemble are associated with the unbalanced fields, they should be small (respectively, large) in areas where the variance of the balanced perturbations explains a large (respectively, small) amount of the total variance.

This reduced-rank formulation of the localization matrix, where the same localization is applied to each variable, is the most commonly used formulation in EnVar applications in meteorology (e.g., see Desroziers *et al.* [2014]).

4. *Multivariate and separate localization for each control variable.*

It is possible to relax the common localization constraint using the following formulation (also implemented in NEMOVAR):

$$\mathbf{U}_L = \begin{pmatrix} \mathbf{U}_T \\ \mathbf{U}_{S_u} \\ \mathbf{U}_{\eta_u} \mathbf{R}_{T \rightarrow \eta} \\ \mathbf{U}_{u_u} \mathbf{A}_{T \rightarrow u} \\ \mathbf{U}_{v_u} \mathbf{A}_{T \rightarrow v} \end{pmatrix}, \quad (33)$$

which, like Eq. (32), is an $N \times N_T$ matrix. However, the computational cost of applying Eq. (33) is greater than applying Eq. (32) since separate applications of the diffusion operator must be performed for each control variable. It is also worth remarking that the cross-variable localization blocks associated with Eq. (33) do not have unit amplitude, contrary to those associated with Eq. (32). It is not clear, however, whether this is an important requirement in practice.

5. Full localization.

Full localization corresponds to setting \mathbf{C}_L to the identity matrix. In this case, the localized sample correlation matrix in Eq. (27) is also equal to the identity matrix:

$$\mathbf{C}_L \circ \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T = \mathbf{I}.$$

With full localization, only the ensemble variances are retained in $\hat{\mathbf{B}}_e$. This limiting case with pure EnVar has not been implemented since a diagonal $\hat{\mathbf{B}}_e$ matrix is of no practical interest. It is worth remarking, however, that when a diagonal $\hat{\mathbf{B}}_e$ matrix is linearly combined with $\hat{\mathbf{B}}_m$ in hybrid EnVar (discussed in the next section), the resulting hybrid $\hat{\mathbf{B}}$ matrix is equivalent to the hybrid variance formulation presented earlier in Section 2.3.4.

2.4.3 Hybrid EnVar

A hybrid \mathbf{B} matrix is formed by linearly combining the localized sample covariance matrix with the modelled covariance matrix (e.g., see Bannister [2017] for a review). This results in a more robust covariance formulation than the pure EnVar \mathbf{B} matrix. For the control variables, the hybrid covariance matrix is given by

$$\hat{\mathbf{B}}_h = \beta_m^2 \hat{\mathbf{B}}_m + \beta_e^2 \hat{\mathbf{B}}_e \quad (34)$$

where β_m and β_e are positive weights, and $\hat{\mathbf{B}}_m$ and $\hat{\mathbf{B}}_e$ are the matrices highlighted by the underbrace in Eqs (1) and (27). As with the hybrid parameter formulation, it is convenient to allow the hybrid weights to depend (at least) on the control variable and model level. The hybrid weights must then be defined as vectors, β_m and β_e , and introduced symmetrically in the hybrid \mathbf{B} matrix as [Ménétrier and Auligné, 2015]

$$\hat{\mathbf{B}}_h = \beta_m \beta_m^T \circ \hat{\mathbf{B}}_m + \beta_e \beta_e^T \circ \hat{\mathbf{B}}_e. \quad (35)$$

Equation (35) can also be written as

$$\hat{\mathbf{B}}_h = \Upsilon_m^{1/2} \hat{\mathbf{B}}_m \Upsilon_m^{1/2} + \Upsilon_e^{1/2} \hat{\mathbf{B}}_e \Upsilon_e^{1/2} \quad (36)$$

where $\Upsilon_m^{1/2}$ (respectively, $\Upsilon_e^{1/2}$) is a diagonal matrix whose diagonal is defined by β_m (respectively, β_e). Equation (36) shows that the matrices $\Upsilon_m^{1/2}$ and $\Upsilon_e^{1/2}$ act to modulate the matrix of standard deviations $\hat{\mathbf{D}}_m^{1/2}$ and $\hat{\mathbf{D}}_e^{1/2}$, respectively.

Note that $\hat{\mathbf{B}}_m$ in Eqs (34) and (36) can be replaced by the matrix $\hat{\mathbf{B}}_m^e$ (Eq. (7)), which itself is specified using ensembles. This would be reasonable if $\hat{\mathbf{B}}_m^e$ was defined in terms of a climatological ensemble, but could be problematic if both $\hat{\mathbf{B}}_m^e$ and $\hat{\mathbf{B}}_e$ were specified using the same flow-dependent ensemble, as this would result in double counting of the ensemble information and hence a break down in the assumption that these two matrices can be decoupled.

2.4.4 Estimating the localization matrix and hybridization weights

We have adopted the objective procedure of [Ménétrier and Auligné \[2015\]](#) to estimate the localization matrix and hybridization weights. The procedure optimizes the elements of the localization matrix and hybridization weights in order to minimize the effects of random sampling error in $\hat{\mathbf{B}}_h$. This is done by minimizing the expected squared error between $\hat{\mathbf{B}}_h$ and the asymptotic sample covariance matrix that would be obtained with a hypothetical infinite ensemble size. We denote this (unknown) matrix by $\tilde{\mathbf{B}}^* = \lim_{N_e \rightarrow \infty} \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$, and formally write the optimality criterion as finding the minimum of

$$e^h = E \left[\left\| \hat{\mathbf{B}}_h - \tilde{\mathbf{B}}^* \right\|_F^2 \right]$$

where $E[\cdot]$ is the expectation operator and $\|\cdot\|_F$ is the Frobenius norm. This optimality criterion does not account for the possibility of systematic errors in the ensemble.

For the optimization problem, the hybridization matrix $\beta_e \beta_e^T$ and localization matrix \mathbf{C}_L in (35) can be merged into an effective localization matrix \mathbf{C}_L^h . The optimal localization and hybridization parameters are found by setting to zero the derivatives of e^h with respect to the elements C_{ij}^h of \mathbf{C}_L^h , and with respect to the independent parameters in the weighting matrix $\beta_m \beta_m^T$. For simplicity, we assume here that $\beta_m \beta_m^T$ is given by a constant weight β_m^2 . Under the mild assumption that the sampling error and asymptotic covariances are uncorrelated, it can be shown that [\[Ménétrier and Auligné, 2015\]](#)

$$\beta_m^2 = \frac{\sum_{ij} (1 - C_{i,j}) E[\tilde{B}_{ij}] \hat{B}_{ij}}{\sum_{ij} \left(\frac{E[\tilde{B}_{ij}^2] - E[\tilde{B}_{ij}]^2}{E[\tilde{B}_{ij}^2]} \right) \hat{B}_{ij}^2}, \quad (37)$$

$$C_{ij}^h = C_{ij} - \frac{E[\tilde{B}_{ij}]}{E[\tilde{B}_{ij}^2]} \hat{B}_{ij} \beta_m^2 \quad (38)$$

where \tilde{B}_{ij} denotes an element of $\tilde{\mathbf{B}} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$, \hat{B}_{ij} denotes an element of $\hat{\mathbf{B}}_m$, and C_{ij} denotes an element of the optional localization matrix obtained with the pure EnVar formulation ($\beta_m = 0$). Assuming a non-Gaussian (NG) error distribution then $C_{ij} = C_{ij}^{\text{NG}}$ where

$$C_{ij}^{\text{NG}} = \frac{(N_e - 1)^2}{N_e(N_e - 3)} - \frac{N_e}{(N_e - 2)(N_e - 3)} \frac{E[\tilde{\Xi}_{ijij}]}{E[\tilde{B}_{ij}^2]} + \frac{N_e - 1}{N_e(N_e - 2)(N_e - 3)} \frac{E[\tilde{B}_{ii}\tilde{B}_{jj}]}{E[\tilde{B}_{ij}^2]}. \quad (39)$$

Assuming a Gaussian (G) error distribution then $C_{ij} = C_{ij}^{\text{G}}$ where

$$C_{ij}^{\text{G}} = \left[(N_e - 1) - \frac{E[\tilde{B}_{ii}\tilde{B}_{jj}]}{E[\tilde{B}_{ij}^2]} \right]. \quad (40)$$

Equations (37)–(40) depend on the ensemble size, modelled covariances and sample covariances. These are all quantities that are known or can be estimated. In practice, an ergodicity assumption is invoked so that the expectation operator can be replaced by a spatial or angular averaging operator over a particular domain. An important theoretical result shown by [Ménétrier and Auligné \[2015\]](#)

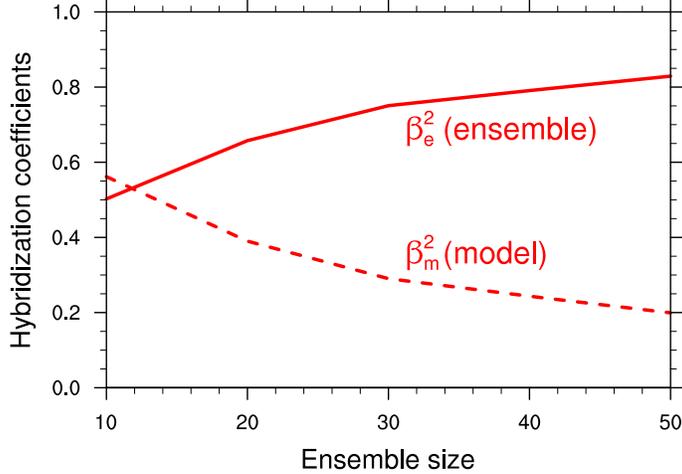


Figure 8: Estimated hybridization weights in Eq. (34) as a function of ensemble size using idealized experiments with ORCA1. The ensembles are defined as randomized vectors with a specified covariance matrix. The hybridization weights have been estimated using the optimal formulae (37), (38) and (40).

is that $\widehat{\mathbf{B}}_h$, with parameters optimized using the procedure above, is always more accurate than $\widehat{\mathbf{B}}_e$ taken alone, regardless of the quality of $\widehat{\mathbf{B}}_m$.

The methods described above have been implemented by B. Ménétrier in an open source diagnostic software package called `hdiag_nicas`¹. `hdiag_nicas` is applied offline from NEMOVAR. To interface `hdiag_nicas` with NEMOVAR has required adapting it to support the global ORCA grids used by NEMO. In addition to the model coordinates, `hdiag_nicas` requires as input (in NetCDF file format), the ensemble perturbations and, in the case of hybridization, a limited number of randomized vectors $\mathbf{b}_p \sim N(\mathbf{0}, \widehat{\mathbf{B}}_m)$, $p = 1, \dots, N_r$, which can be generated using the “square-root” operator of $\widehat{\mathbf{B}}_m$. The output from `hdiag_nicas` used by NEMOVAR consists of localization scales, which are used by the diffusion operator, and hybridization weights, as a function of variable and, possibly, model level.

The algorithm has been validated with NEMOVAR using an ensemble of randomized vectors whose covariance matrix is given by the modelled covariance matrix with specific (“true”) parameters, different from those specified in $\widehat{\mathbf{B}}_m$. Figure 8 illustrates the values of the estimated hybridization weights β_m^2 and β_e^2 as a function of ensemble size. The sum of the estimated weights is close to one in all cases. With an ensemble size of 10, $\widehat{\mathbf{B}}_m$ and $\widehat{\mathbf{B}}_e$ are given approximately equal weight. As expected, the weight to $\widehat{\mathbf{B}}_e$ then increases steadily with increasing ensemble size, reaching over 80% of the total weight with an ensemble size of 50.

Figure 9 shows the estimated localization functions for different ensemble sizes. The average sample correlation function (black curve) is also plotted for comparison. The amplitude of the localization function at zero separation gives the hybridization weight β_e^2 (cf. Fig. 8). As expected, the scale of the localization function increases with increasing ensemble size. The localization

¹https://github.com/benjaminmenetrier/hdiag_nicas

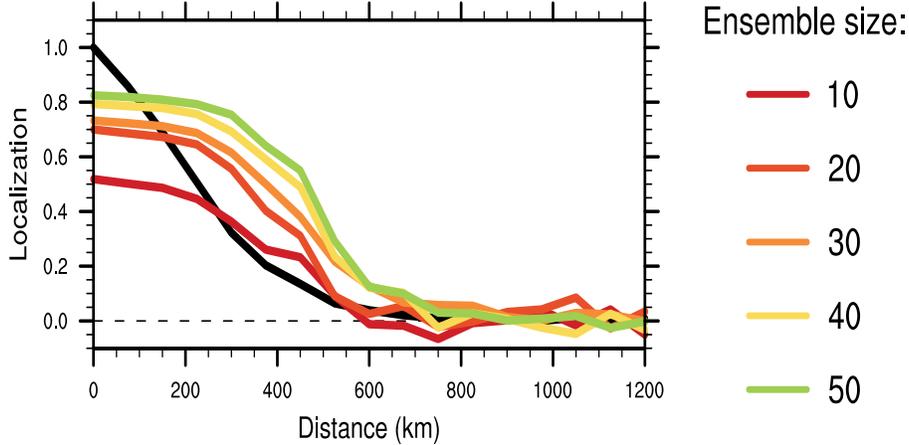


Figure 9: Estimated localization functions in Eq. (34) as a function of ensemble size using idealized experiments with ORCA1 (coloured curves). The ensembles are defined as randomized vectors with a specified covariance matrix. The localization functions have been estimated using the optimal formulae (37), (38) and (40). The average sample correlation function is also shown (black curve).

functions are also noticeably flat near the origin, which makes them difficult to model accurately with standard correlation functions such as Matérn or Gaussian. For NEMOVAR, these curves are first fit to a Gaussian function and the best-fit localization scale is then used to define the diffusion tensor of the diffusion-based localization operator.

Figure 10 gives an example of background error correlations generated by $\hat{\mathbf{B}}_m$, $\hat{\mathbf{B}}_e$ and $\hat{\mathbf{B}}_h$ in NEMOVAR. The ensembles are again based on randomized vectors from a “true” modelled \mathbf{B}_m . The correlations are shown for temperature in level 1 at an arbitrary horizontal grid point in the global domain. The true correlations are noticeably anisotropic, being elongated in the north-south direction. This feature is not represented by the modelled correlation matrix (\mathbf{B}_m), which is specified using a purely isotropic formulation. The correlations (notably the anisotropic features) are reasonably well recovered by the localized ensemble correlation matrix (\mathbf{B}_e), although their spatial extent is somewhat underestimated. The correlations are clearly improved with hybridization (\mathbf{B}_h).

3 Technical description of the code and its review

The developments have required a major rewrite of the existing Fortran code dedicated to \mathbf{B} . One of the main structural changes of the code has been to introduce nested derived types to group covariance and ensemble parameters associated with a specific control variable. This has resulted in a much more compact and flexible code, and has allowed for better control of memory allocation. There are five main derived types, which are defined separately for 2D and 3D fields: a covariance type (‘cov’), correlation type (‘cor’), diffusion type (‘dif’), implicit solver type (‘sol’), and ensemble

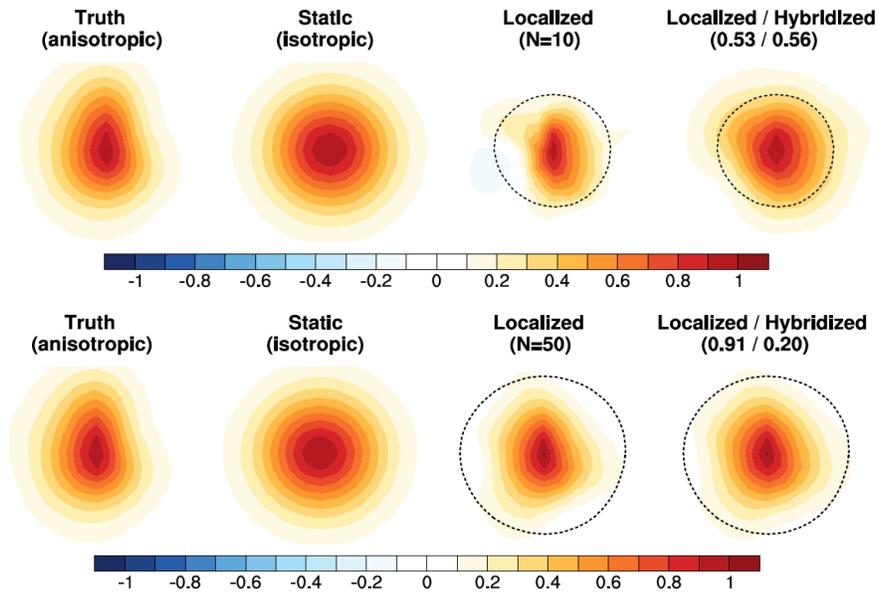
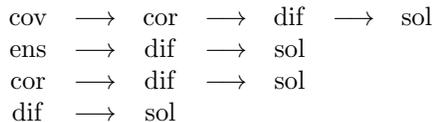
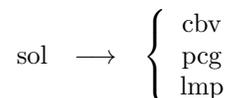


Figure 10: An example of temperature background-error correlations represented in ORCA1 at an arbitrary grid point. From left to right: the “true” correlation; the modelled correlation in $\hat{\mathbf{B}}_m$; the localized ensemble correlation in $\hat{\mathbf{B}}_e$; and the hybrid correlation in $\hat{\mathbf{B}}_h$. Two different ensemble sizes are considered: 10 (top panels) and 50 (bottom panels). The radius of the dotted circles corresponds to the localization scale.

type ('ens'). Their nested arrangement can be represented schematically as:



Each control variable has a 'cov' type when the covariances are modelled and an 'ens' type when the covariances are estimated from ensembles. The 'ens' type depends on a 'dif' type for spatial filtering. With a sample covariance matrix, each control variable has a 'cor' type for localization. The 'dif' type is used for any variable or parameter employing spatial filtering with the diffusion operator. The 'sol' type assembles information for the different solvers and preconditioners used with the implicit diffusion operator. Solver parameters for the Chebyshev iteration and conjugate gradient methods are contained in 'cbv' and 'pcg' types, respectively, and limited-memory (Ritz) preconditioning parameters are contained in an 'lmp' type:



The restructuring of the code in this manner has greatly facilitated the introduction of the new methods for defining \mathbf{B} that have been described in this report. The main Fortran modules that have been either modified or added are located in the NEMOVAR source code directories BGE, COV, COR, DIF and ENS.

The new code has been fully integrated into the central NEMOVAR code repository that is hosted on a Git server at ECMWF. The code integration has been done progressively during the course of the project, using development branches dedicated to specific tasks. Before being integrated into the central development branch (the trunk), the branches were peer-reviewed by members of the NEMOVAR consortium (typically two reviewers for each branch) and revised accordingly. Code review and integration into the trunk has been done in an efficient and coordinated manner using software collaboration tools (JIRA, Bitbucket, Bamboo) made available at ECMWF. The latest version of NEMOVAR, which includes all the developments described here, will be released as version 5 in the coming weeks.

In addition to the Fortran source code development, substantial work has been made by ECMWF to adapt the operational scripts (prepIFS and SMS) to the new version of NEMOVAR and to be able to apply it within an EDA framework. Four one-week-long visits made by A. Weaver to ECMWF in 2017 were dedicated to validating the new code. This required exhaustive testing and debugging to ensure that the new version of NEMOVAR (soon to be labelled version 5) produced near-identical results to the operational version (version 3.4) using similar algorithmic and parameter choices (see Fig. 11). The new developments have been validated primarily using single cycle experiments (as illustrated in the previous sections), and using standard tests and diagnostics: adjoint operator tests, cost function and gradient tests, minimization diagnostics, single observation experiments, etc.

4 Discussion and conclusions

In this report we have discussed developments that have been made to NEMOVAR to exploit ensemble information in the specification of the background-error covariance matrix (\mathbf{B}). Several new

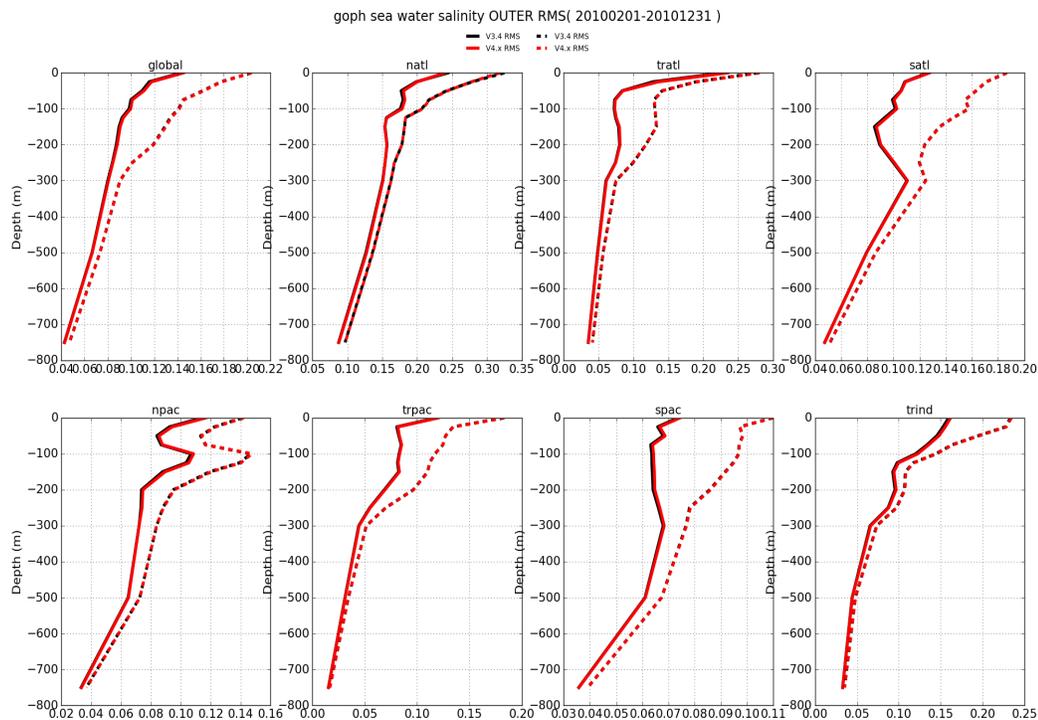
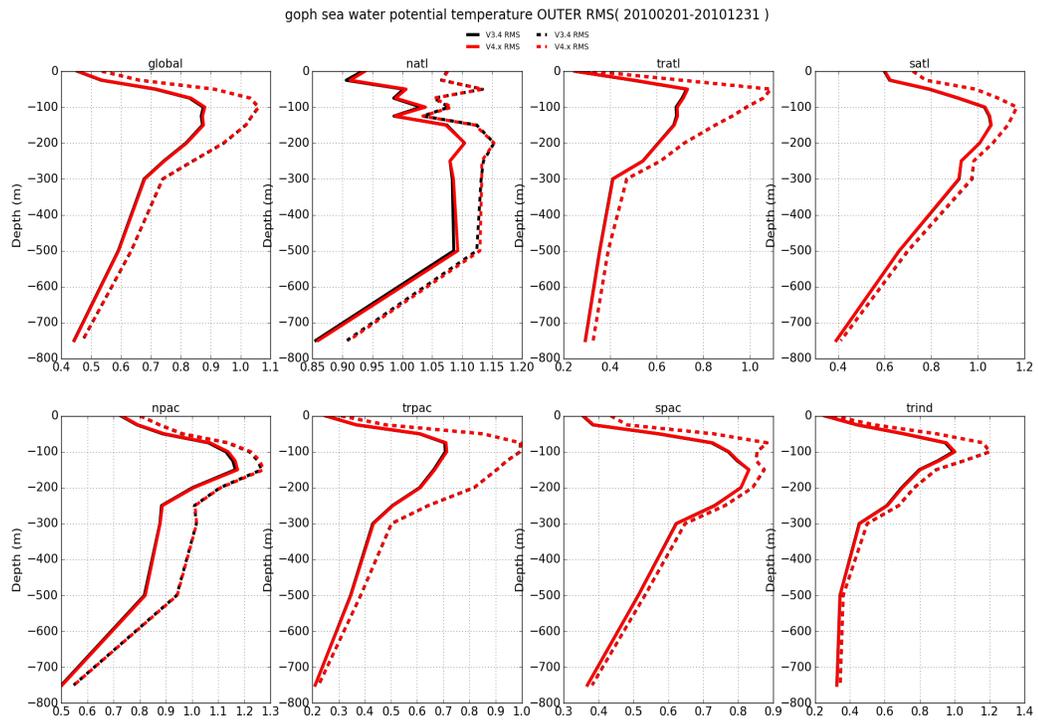


Figure 11: The root-mean-square (RMS) of the analysis-minus-observations (dashed curves) and background-minus-observations (solid curves) obtained with ORCA1 with the current operational version of NEMOVAR (version 3.4; black curves) and with the new version of NEMOVAR developed during ERA-CLIM2 (labelled as version 4.x in the legend; red curves) for an ocean reanalysis experiment from 01/02/2010 to 31/12/2010. Similar parameter settings have been chosen in both experiments to validate the new version of the code. The top (respectively, bottom) 6 panels show the RMS errors for temperature (respectively, salinity) in different regions.

features have been developed and made available in a new version of NEMOVAR. The new version has been integrated into the central source code repository and operational running environment at ECMWF. Preliminary validation experiments have been conducted with the new version in an ocean reanalysis framework using an ensemble of data assimilations (EDA) approach to cycle the analysis-forecast system. The perturbation strategy is based on that developed for ORAS5 [Zuo *et al.*, 2017]. The ensemble developments described in this report have been evaluated with a 3D-Var version of NEMOVAR. An obvious extension of this work would be to combine the ensemble methods with the simplified 4D-Var version of NEMOVAR developed as part of deliverable D2.5. This would bring the ocean data assimilation system to the same basic level of sophistication as the atmospheric data assimilation system to which it is coupled in CERA [Laloyaux *et al.*, 2016].

The options that have been developed for defining \mathbf{B} from ensembles are numerous and have various levels of complexity. Computational cost and pragmatism will be determining factors in choosing the most appropriate options in an operational framework. To this end, a three-step strategy is proposed below for progressively improving the representation of \mathbf{B} in CERA using ensembles.

1. *Cycle-dependent ensemble-based variances.*

Specifying the variances in \mathbf{B} is the most straightforward use of the ensembles and can be done with limited cost overhead. Cycle-dependent background-error variances, either used alone or in hybrid form, would give the ocean data assimilation system leverage to adapt to the time-evolving flow field and to changes in the observing system (see Fig. 1). Focusing initially on the variances would help to identify problems in ensemble spread, which would need to be rectified by improving the ensemble-generation procedure, inflating the variances, and/or calibrating the weights in the hybrid formulation.

The objective filtering algorithm needs to be generalized to allow for different filtering scales to be computed in subdomains of the global model. Adapting the diffusion filter to work in arbitrary subdomains in NEMO would require some non-trivial technical developments. Without these developments, it is possible that the use of subjectively specified filtering scales (e.g., taken to be proportional to the local grid steps) would be as effective as, and a cheaper alternative to, the optimally-based single-scale algorithm.

The NEMOVAR system should also be extended to include ensemble variance estimates of sea-ice concentration, which is a fundamental control variable in the coupled sea-ice ocean data assimilation system.

2. *Climatological ensemble-based horizontal correlation scales.*

In the background-error covariance model, the correlation scales are controlled by the diffusion tensor. While it would be desirable to use ensembles to specify cycle-dependent diffusion tensors as well as cycle-dependent variances, computational cost would preclude the former without further improvements to the numerical efficiency of the diffusion algorithm and to the accuracy of methods for approximating normalization factors on the fly. Avenues for improving the efficiency of the diffusion algorithm are proposed later in this section. To avoid the expensive task of recomputing normalization factors on each cycle, the ensemble could be used to provide a climatological estimate of the diffusion tensor. This should already be a vast improvement over the ad hoc approach currently used in CERA. One possibility would be to use the climatological ensemble to estimate the horizontal diffusion tensor only, and to use a flow-dependent parameterization based on the background mixed-layer depth to specify

the vertical diffusion tensor. An attractive feature of this parameterization, in addition to being flow dependent, is that it can be employed with an accurate normalization procedure [Waters *et al.*, 2015].

In this project, a method was developed to estimate the local diffusion tensor based on the ensemble average of the tensor product of the local gradient of the ensemble perturbations. This approach could be used to provide a climatological estimate of the diffusion tensor by averaging the gradient tensor product over many assimilation cycles. An alternative and more general approach would be to estimate the local diffusion tensor at a given grid point by fitting a locally homogeneous Matérn or Gaussian function to sample correlations available in a neighbourhood of that point [Pannekoucke and Massart, 2008; Yaremchuk and Nechaev, 2013]. The `hdiag_nicas` software has recently been extended by B. Ménétrier to include this possibility and has been tested with ensembles from ORAS5, as illustrated in Fig. 12. The top panel shows raw sample correlations at selected points, computed using a 19-member ensemble. The middle and bottom panels show the correlations obtained by fitting the LCT of a Gaussian function to the sample correlations using a nonlinear least-squares optimization method. The middle panel accounts for fully anisotropic correlations (non-diagonal LCT), whereas the bottom panel accounts for coordinate stretching only (diagonal LCT). The diffusion operator currently used in NEMOVAR would be able to account for the latter only. These results are very encouraging and suggest that this approach should be further investigated and compared with the gradient-based approach in terms of cost, accuracy and robustness. Moreover, the function-fitting approach is better suited to the general multiple-scale background-error covariance model available in NEMOVAR, which is formed from a linear combination of Matérn or Gaussian functions and requires additional parameters to be estimated [Mirouze *et al.*, 2016].

3. *Hybrid EnVar.*

The hybrid **B** formulation described in section 2.4.3 provides a rich and robust framework for using ensembles to define fully cycle-dependent background-error covariances (i.e., variances, correlations and multivariate relationships). The modelled covariance matrix in the hybrid **B** formulation could be based on climatological ensemble estimates of the correlations and variances, following work in steps 1 and 2 above.

Localization is currently the main computational bottleneck in the hybrid **B** formulation, and work is needed to reduce this cost in order to make the EnVar formulation feasible with NEMOVAR. Recent developments with the `hdiag_nicas` software have opened up some interesting possibilities in this regard, by including an efficient localization algorithm based on a compactly-supported Gaussian-like function. The algorithm is made efficient by applying the localization function on a reduced-resolution grid, which is possible since localization scales are generally broad. The algorithm can also account for complex boundaries such as those found in a global ocean model, although the treatment is somewhat less-refined than what can be achieved by diffusion near complex boundaries. For localization, however, this is probably not important. The algorithm thus represents a potentially more efficient alternative to a localization algorithm based on the diffusion operator. In the hybrid formulation, however, the algorithms are complementary as the diffusion operator is more efficient for small correlation scales typically present in the modelled component.

The hybrid **B** formulation also benefits from a theoretically well-founded and practical algorithm (available in `hdiag_nicas`) for estimating localization scales and hybridization weights.

The hybrid **B** formulation also opens the way to using fully coupled background-error covariances, although this is recognized to be a longer term objective.

In parallel to the developments above, continued efforts should be made to improve the computational efficiency and scalability of the implicit diffusion-based correlation model. Important advances have been made in this area during this project, notably with the introduction of the Chebyshev iteration (CI) method to approximate the solution of the underlying implicit diffusion equation [Weaver *et al.*, 2016]. Nevertheless, further improvements are necessary to reduce the computational cost of the correlation model for applications involving high-resolution models or, as mentioned above, “large” correlation length-scales. Closely related to this work is the need to improve the efficiency and accuracy of approximate methods used to estimate normalization factors for the diffusion operator when using cycle-dependent correlation parameters.

There are different ways the cost of the correlation operator can be reduced while still remaining within the CI framework. In one approach that was studied during the project, it was shown that there was potential for speeding up the runtime of the correlation operator by allowing the implicit “time” steps (M) to be executed in parallel on each iteration of CI [Weaver *et al.*, 2017]. This can be done by reformulating the sequence of M -step implicit diffusion problems as a single problem involving a nonsymmetric matrix. This “time”-parallel version of the correlation operator could be implemented practically using a hybrid parallelization approach that combines Message Passage Interface tasks in the spatial domain with Open Multi-Processing threads spanning the pseudo-time dimension. Another approach is to employ selective use of single precision in the CI solver. Preconditioning is another possibility, with domain decomposition preconditioners, such as Additive Schwarz or Restricted Additive Schwarz [Cai and Sarkis, 1999], seemingly well suited for the problem at hand. Furthermore, for large-scale correlations, the diffusion computations could be done more economically on a grid with resolution comparable to the scale itself, which could be much coarser than the resolution of the native analysis grid.

Acknowledgements

The authors gratefully acknowledge the support of the ERA-CLIM2 oroject of the European Framework Programme 7 (Grant Agreement No. 607029), the AVENUE project of the RTRA STAE Foundation, and the French national programme LEFE/INSU.

A Appendix. Finite difference representation of the LCT

The estimation of the components of the LCT (\mathbf{H}) using centred finite differences is outlined in this appendix. Notation is similar to that used in the NEMO documentation [Madec, 2008]. Since the off-diagonal elements of the LCT involving the vertical coordinate s_3 have not been accounted for in NEMOVAR, we focus separately on the estimation of the diagonal component H_{33} , and of the horizontal components ($H_{11}, H_{12}, H_{21}, H_{22}$).

A.1 Estimating H_{33}

Consider a 1D scalar field α which we consider to be a function of the vertical coordinate s_3 . Let $s_3 = s_3(k)$ where k is the grid-point defined at the centre of the grid step, and $k + \frac{1}{2}$ the grid-point

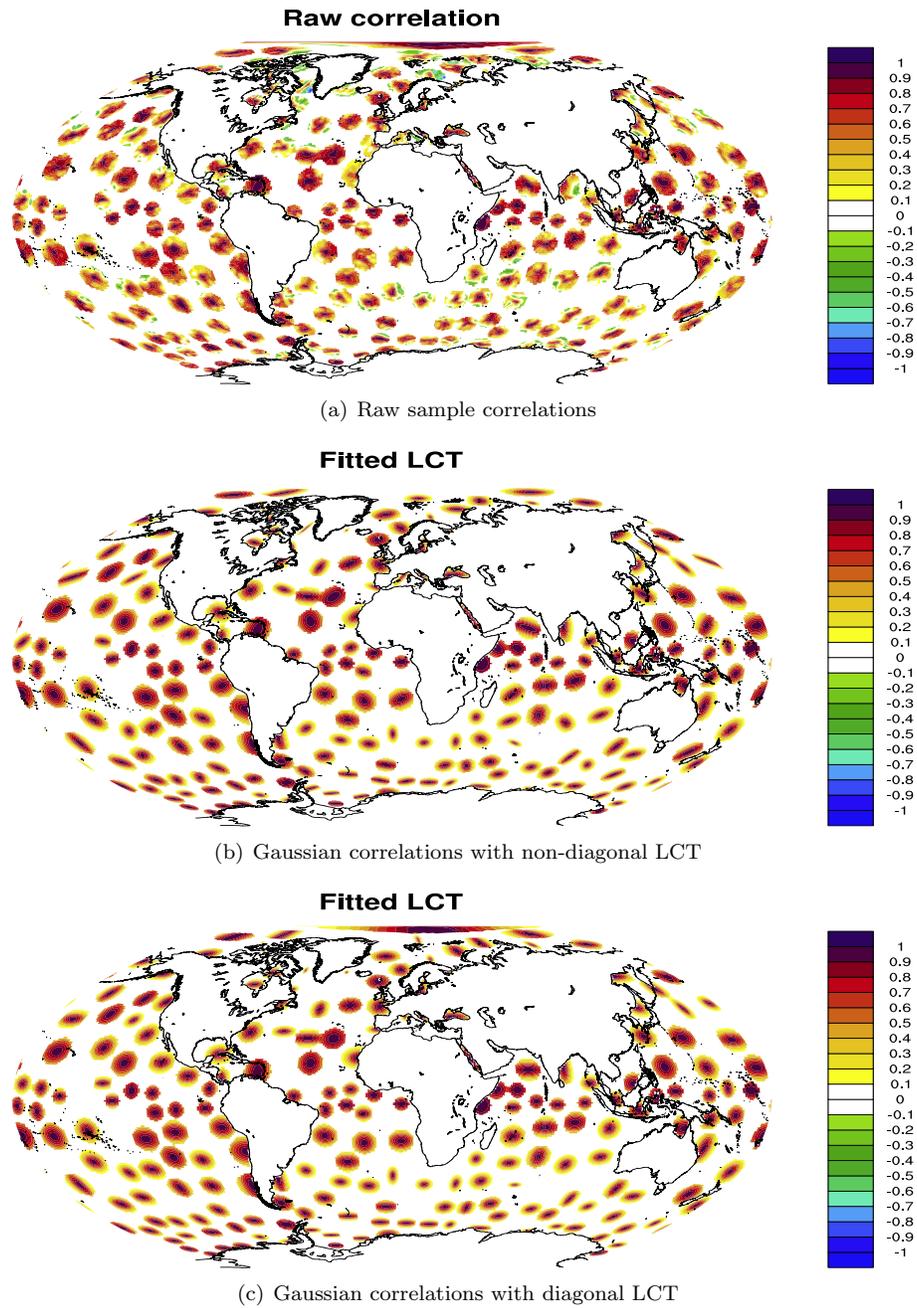


Figure 12: Examples of point correlations estimated using a 19-member ensemble on a single cycle of the ECMWF ORA system based on ORCA1. (a) Sample correlations estimated from the raw ensemble; (b) correlations estimated by fitting sample correlations to a locally homogeneous Gaussian function with non-diagonal LCT; and (c) as in (b) but using a diagonal LCT used in the Gaussian function. The computations have been done using the `hdiag_nicas` software (https://github.com/benjaminmenetrier/hdiag_nicas). Figure courtesy of B. Ménétrier.

defined at the interface with the adjacent grid step. The derivative of α at adjacent points in the s_3 direction is defined at the mid-point between them, and can be represented by the difference operator

$$\frac{n_k}{e_k} \delta_k[\alpha] = \frac{n_k}{e_k} \left(m_{k+\frac{1}{2}} \alpha_{k+\frac{1}{2}} - m_{k-\frac{1}{2}} \alpha_{k-\frac{1}{2}} \right),$$

where e_k is the scale factor (metric coefficient) defined at grid-point k , $m_{k\pm\frac{1}{2}}$ is a land/ocean mask (equals 0 for land; 1 for ocean), and n_k is a mask to enforce either Neumann or Dirichlet boundary conditions (equals 0 for Neumann; 2 for Dirichlet). The average of α at adjacent points in the s_k direction is defined at the mid-point between them and represented by the operator

$$\bar{\alpha}^k = \left(\frac{1}{m_{k+\frac{1}{2}} + m_{k-\frac{1}{2}}} \right) \left(m_{k+\frac{1}{2}} \alpha_{k+\frac{1}{2}} + m_{k-\frac{1}{2}} \alpha_{k-\frac{1}{2}} \right).$$

Let the perturbation state ε in Eq. (11) be defined at integer grid-points k where, for clarity of notation, we have dropped the hat and prime symbols on ε . The product of vertical derivatives at each grid-point is needed to compute \mathbf{g}_{33} in (12):

$$\{g_{33}\}_{k+\frac{1}{2}} \equiv \left\{ \left(\frac{\partial \varepsilon}{\partial s_3} \right) \left(\frac{\partial \varepsilon}{\partial s_3} \right) \right\}_{k+\frac{1}{2}} = \left(\frac{1}{e_{k+\frac{1}{2}}} \delta_{k+\frac{1}{2}}[\varepsilon] \right)^2.$$

To estimate \mathbf{h}_{33} (Eq. (13)), we compute the ensemble average of \mathbf{g}_{33} . This gives the values of the derivative product at the mid-point of the cell interface. These values are then averaged back to the centre of the cell to define H_{33} ,

$$\{H_{33}\}_k = \bar{h}_{33}^{-k},$$

and then inverted to give H_{33}^{-1} . The diffusion tensor coefficient κ_{33} is defined at the mid-point of the cell where it multiplies the vertical derivative of the field being diffused. κ_{33} is obtained from H_{33}^{-1} by averaging H_{33}^{-1} back to the mid-point and dividing it by $2M - d - 2$ (see Eq. (8)) where $d = 1$ or 3 depending on whether the 1D or 3D diffusion-based correlation model is used. For the vertical component, the double averaging is, in fact, unnecessary since \mathbf{h}_{33} is already defined at the mid-point. This is done to be consistent with the horizontal case described below, which does require averaging to collocate the horizontal tensor elements. The double averaging done in the vertical introduces additional filtering in the computation of κ_{33} .

A.2 Estimating $(H_{11}, H_{12}, H_{21}, H_{22})$

Consider a 2D scalar field $\alpha = \alpha(s_1, s_2)$ where the coordinates $s_1 = s_1(i, j)$ and $s_2 = s_2(i, j)$. The grid-points (i, j) are defined at the centre of a grid cell, and the points $(i+\frac{1}{2}, j)$ and $(i, j+\frac{1}{2})$ are defined at the east and north interfaces, respectively, between adjacent grid cells. The derivative of α at adjacent points in the s_1 or s_2 direction will be defined at the mid-point between them, and can be represented by the difference operators

$$\begin{aligned} \frac{n_{i,j}}{e_{i,j}} \delta_i[\alpha] &= \frac{n_{i,j}}{e_{i,j}} \left(m_{i+\frac{1}{2},j} \alpha_{i+\frac{1}{2},j} - m_{i-\frac{1}{2},j} \alpha_{i-\frac{1}{2},j} \right), \\ \frac{n_{i,j}}{e_{i,j}} \delta_j[\alpha] &= \frac{n_{i,j}}{e_{i,j}} \left(m_{i,j+\frac{1}{2}} \alpha_{i,j+\frac{1}{2}} - m_{i,j-\frac{1}{2}} \alpha_{i,j-\frac{1}{2}} \right) \end{aligned}$$

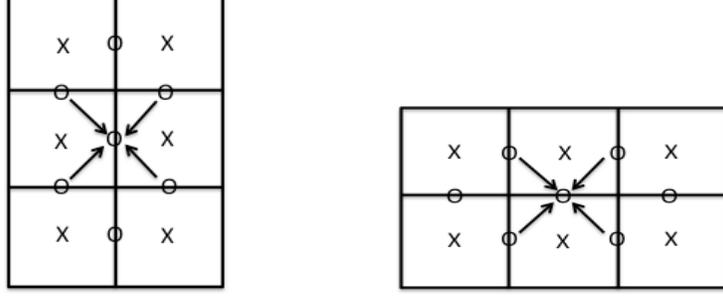


Figure 13: With centred finite differences, the horizontal derivatives of a scalar variable defined at the location X are located at the half grid-points defined by O. When $m \neq n$ the derivative with respect to s_m is located at a different point from the derivative with respect to s_n . In this case, the computation of the derivative product $(\partial\varepsilon/\partial s_m)(\partial\varepsilon/\partial s_n)$ requires averaging or interpolating one of the derivatives to the location of the other derivative as illustrated.

where $e_{i,j}$ is the scale factor (metric coefficient) defined at grid-point (i,j) , $m_{i\pm\frac{1}{2},j}$ and $m_{i,j\pm\frac{1}{2}}$ are land-ocean masks (equal to 0 or 1), and $n_{i,j}$ is a mask to enforce Neumann or Dirichlet boundary conditions (equal to 0 or 2, respectively). Similarly, the average of α at adjacent points in the s_1 or s_2 direction are defined at the mid-point between them, and can be represented by the operators

$$\bar{\alpha}^i = \left(\frac{1}{m_{i+\frac{1}{2},j} + m_{i-\frac{1}{2},j}} \right) (m_{i+\frac{1}{2},j}\alpha_{i+\frac{1}{2},j} + m_{i-\frac{1}{2},j}\alpha_{i-\frac{1}{2},j}),$$

$$\bar{\alpha}^j = \left(\frac{1}{m_{i,j+\frac{1}{2}} + m_{i,j-\frac{1}{2}}} \right) (m_{i,j+\frac{1}{2}}\alpha_{i,j+\frac{1}{2}} + m_{i,j-\frac{1}{2}}\alpha_{i,j-\frac{1}{2}}).$$

Let the normalized perturbation state ε be defined at integer grid-points (i,j) . The products involving the horizontal derivative components are estimated using the difference formulae

$$\{g_{11}\}_{i+\frac{1}{2},j} \equiv \left\{ \left(\frac{\partial\varepsilon}{\partial s_1} \right) \left(\frac{\partial\varepsilon}{\partial s_1} \right) \right\}_{i+\frac{1}{2},j} = \left(\frac{n_{i+\frac{1}{2},j}\delta_{i+\frac{1}{2}}[\varepsilon]}{e_{i+\frac{1}{2},j}} \right)^2, \quad (41)$$

$$\{g_{22}\}_{i,j+\frac{1}{2}} \equiv \left\{ \left(\frac{\partial\varepsilon}{\partial s_2} \right) \left(\frac{\partial\varepsilon}{\partial s_2} \right) \right\}_{i,j+\frac{1}{2}} = \left(\frac{n_{i,j+\frac{1}{2}}\delta_{j+\frac{1}{2}}[\varepsilon]}{e_{i,j+\frac{1}{2}}} \right)^2, \quad (42)$$

$$\{g_{12}\}_{i+\frac{1}{2},j} \equiv \left\{ \left(\frac{\partial\varepsilon}{\partial s_1} \right) \left(\frac{\partial\varepsilon}{\partial s_2} \right) \right\}_{i+\frac{1}{2},j} = \left(\frac{n_{i+\frac{1}{2},j}\delta_{i+\frac{1}{2}}[\varepsilon]}{e_{i+\frac{1}{2},j}} \right) \left(\frac{\overline{n_{i,j}\delta_{j+\frac{1}{2}}[\varepsilon]}^{i+\frac{1}{2}}}{e_{i,j}} \right), \quad (43)$$

$$\{g_{21}\}_{i,j+\frac{1}{2}} \equiv \left\{ \left(\frac{\partial\varepsilon}{\partial s_2} \right) \left(\frac{\partial\varepsilon}{\partial s_1} \right) \right\}_{i,j+\frac{1}{2}} = \left(\frac{n_{i,j+\frac{1}{2}}\delta_{j+\frac{1}{2}}[\varepsilon]}{e_{i,j+\frac{1}{2}}} \right) \left(\frac{\overline{n_{i,j}\delta_{i+\frac{1}{2}}[\varepsilon]}^{j+\frac{1}{2}}}{e_{i,j}} \right). \quad (44)$$

Equations (41) and (42) require, as in the 1D (vertical) case, simply squaring the derivative estimates at the mid-point where they are defined. Equations (43) and (44), on the other hand, require first

averaging one of the derivatives to the location of the other derivative before they can be multiplied. This is illustrated in Figure 13. The need to average the derivatives introduces inaccuracies in the estimate of g_{12} and g_{21} .

To estimate \mathbf{h}_{mn} ($m, n = 1, 2$) (Eq. (13)), we compute the ensemble average of the corresponding elements of \mathbf{g}_{mn} . Notice that components of \mathbf{h}_{11} and \mathbf{h}_{12} are not defined at the same mid-points as the components of \mathbf{h}_{22} and \mathbf{h}_{21} . In order to compute the inverse Hessian tensor, the elements of the \mathbf{h}_{mn} must be colocated. The procedure adopted here is to average the components to the centre point (i, j) of the cell where ε is defined:

$$\begin{aligned}\{H_{11}\}_{i,j} &= \overline{h_{11}}^i, \\ \{H_{22}\}_{i,j} &= \overline{h_{22}}^j, \\ \{H_{12}\}_{i,j} &= \{H_{21}\}_{i,j} = \frac{1}{2} \left(\overline{h_{12}}^i + \overline{h_{21}}^j \right),\end{aligned}$$

where the last expression enforces symmetry. The resulting 2×2 matrix can then be easily inverted at each grid point. After inverting the tensor, the elements are divided by $2M - d - 2$ (Eq. (8)), where $d = 2$ or 3 depending on whether the 2D or 3D diffusion-based correlation model is used, and then averaged to the appropriate cell interface points where the diffusion tensor elements are defined.

References

- Balmaseda MA, Mogensen K, Weaver AT. 2013. Evaluation of the ECMWF Ocean Reanalysis ORAS4. *Q. J. R. Meteorol. Soc.* **139**: 1132–1161.
- Bannister RN. 2017. A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. R. Meteorol. Soc.* **143**: 607–633.
- Belo Pereira M, Berre L. 2006. The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Weather Rev.* **134**: 2466–2489.
- Cai XC, Sarkis A. 1999. Restricted additive Schwarz preconditioner for general sparse linear systems. *SIAM J. Sci. Comput.* **21**: 792–797.
- Chorti A, Hristopulos DT. 2008. Nonparametric identification of anisotropic (elliptic) correlations in spatially distributed data sets. *IEEE Trans. Signal Process* **56**: 43–62.
- Daget N, Weaver AT, Balmaseda MA. 2009. Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Q. J. R. Meteorol. Soc.* **135**: 1071–1094.
- Daley R. 1991. *Atmospheric Data Analysis*. Cambridge Atmospheric and Space Sciences Series, Cambridge University Press: Cambridge, UK.
- Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* **131**: 1433–1452.
- Desroziers G, Camino JT, Berre L. 2014. 4denvar: Link with 4D state formulation of variational assimilation and different possible implementations. *Q. J. R. Meteorol. Soc.* **138**: 2097–2110.
- Gürol S, Weaver AT, Moore AM, Piacentini A, Arango HG, Gratton S. 2014. **B**-preconditioned minimization algorithms for variational data assimilation with the dual formulation. *Q. J. R. Meteorol. Soc.* **140**: 539–556.

- Guttorp P, Gneiting T. 2006. Miscellanea studies in the history of probability and statistics XLIX: on the Matérn correlation family. *Biometrika*. **93**: 989–995.
- Hristopulos DT. 2002. New anisotropic covariance models and estimation of anisotropic parameters based on the covariance tensor identity. *Stoch. Environ. Res. Risk Assess* **16**: 43–62.
- Kepert JD. 2009. Covariance localization and balance in an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.* **135**: 607–633.
- Laloyaux P, Balmaseda MA, Dee D, Mogensen K, Janssen P. 2016. A coupled data assimilation system for climate reanalysis. *Q. J. R. Meteorol. Soc.* **142**: 65–78.
- Leutbecher M. 2009. Diagnosis of ensemble forecasting systems. In: *Annual Seminar: Diagnosing of forecasting and data assimilation systems*. ECMWF, Reading, UK, pp. 235–266.
- Madec G. 2008. *NEMO Ocean Engine*. Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27, ISSN No 1288-1619: France.
- Martin MJ, Balmaseda M, Bertino L, Brasseur P, Brassington G, Cummings J, Fujii Y, Lea DJ, Lellouche JM, Mogensen K, Oke PR, Smith GC, Testut CE, Waagbø, Waters J, Weaver AT. 2015. Status and future of data assimilation in operational oceanography. *J. Oper. Oceanography* **8**: s28–s48.
- Ménétrier B, Auligné T. 2015. Optimized localization and hybridization to filter ensemble-based covariances. *Mon. Weather Rev.* **143**: 3931–3947.
- Ménétrier B, Montmerle T, Michel Y, Berre L. 2015a. Linear filtering of sample covariances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization. *Mon. Weather Rev.* **143**: 1622–1643.
- Ménétrier B, Montmerle T, Michel Y, Berre L. 2015b. Linear filtering of sample covariances for ensemble-based data assimilation. Part II: Application to a convective-scale NWP model. *Mon. Weather Rev.* **143**: 1644–1664.
- Michel Y. 2013. Estimating deformations of random processes for correlation modelling: methodology and the one-dimensional case. *Q. J. R. Meteorol. Soc.* **139**: 771–783.
- Michel Y, Ménétrier B, Montmerle T. 2016. Objective filtering of the local correlation tensor. *Q. J. R. Meteorol. Soc.* **142**: 2314–2323.
- Mirouze I, Blockley EW, Lea DJ, Martin MJ, Bell MJ. 2016. A multiple length scale correlation operator for ocean data assimilation. *Tellus A*. **68**: 29744.
- Mirouze I, Weaver AT. 2010. Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Q. J. R. Meteorol. Soc.* **136**: 1421–1443.
- Mogensen K, Balmaseda MA, Weaver AT. 2012. The NEMOVAR ocean assimilation system and its implementation for ECMWF’s ocean analysis for System 4. Technical Memorandum 668, ECMWF, Reading, UK. Also registered as a CERFACS Technical Report No. TR-CMGC-12-30.
- Pannekoucke O, Massart S. 2008. Estimation of the local diffusion tensor and normalization for heterogeneous correlation modelling using a diffusion equation. *Q. J. R. Meteorol. Soc.* **134**: 1425–1438.
- Ricci S, Weaver AT, Vialard J, Rogel P. 2005. Incorporating temperature-salinity constraints in the background error covariance of variational ocean data assimilation. *Mon. Weather Rev.* **133**: 317–338.
- Sato Y, De Pondeva MSFV, Purser RJ, Parrish DF. 2009. Ensemble-based background error covariance implementations using spatial recursive filters in NCEP’s grid-point statistical interpolation system. Office Note 459, National Centers for Environmental Prediction, Camp Springs, USA.
- Waters J, Lea DJ, Martin MJ, Mirouze I, Weaver AT, While J. 2015. Implementing a variational

- data assimilation system in an operational 1/4 degree global ocean model. *Q. J. R. Meteorol. Soc.* **141**: 333–349.
- Weaver AT, Courtier P. 2001. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.* **127**: 1815–1846.
- Weaver AT, Deltel C, Machu E, Ricci S, Daget N. 2005. A multivariate balance operator for variational ocean data assimilation. *Q. J. R. Meteorol. Soc.* **131**: 3605–3625.
- Weaver AT, Gürol S, Tshimanga J, Chrust M, Piacentini A. 2017. “Time”-parallel diffusion-based correlation operators. Technical Memorandum 808, ECMWF, Reading, UK. Also registered as a CERFACS Technical Report No. TR-PA-17-119.
- Weaver AT, Mirouze I. 2013. On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation. *Q. J. R. Meteorol. Soc.* **139**: 242–260.
- Weaver AT, Tshimanga J, Piacentini A. 2016. Correlation operators based on an implicitly formulated diffusion equation solved with the Chebyshev iteration. *Q. J. R. Meteorol. Soc.* **142**: 455–471.
- Weaver AT, Vialard J, Anderson DLT. 2003. Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part I: Formulation, internal diagnostics, and consistency checks. *Mon. Weather Rev.* **131**: 1360–1378.
- Yaremchuk M, Nechaev D. 2013. Covariance localization with the diffusion-based correlations models. *Mon. Weather Rev.* **141**: 848–860.
- Zuo H, Balmaseda MA, de Boisseson E, Hirahara S, Chrust M, de Rosnay P. 2017. A generic ensemble generation scheme for data assimilation and ocean analysis. Technical Memorandum 795, ECMWF, Reading, UK.