Ensemble of 4DVARs (En4DVar) data assimilation in a coastal ocean circulation model. Part I: Methodology and ensemble statistics

Ivo Pasmans[a,1], Alexander L. Kurapov[a,b]

[a] *College of Earth Ocean & Atmospheric Sciences, Oregon State University, Corvallis, OR, USA*
[b] *NOAA, Silver Spring, MD, USA*

Abstract

The ocean state off Oregon-Washington, U.S. West coast, is highly variable in time. Under these conditions the assumption made in traditional 4-dimensional variational data assimilation (4DVAR) that the prior model (background) error covariance is the same in every data assimilation (DA) window can be limiting. A DA system based on an ensemble of 4DVARs (En4DVar) has been developed in which the background error covariance is estimated from an ensemble of model runs and is thus time-varying. This part describes details of the En4DVar method and ensemble statistics verification tests. The control run and 39 ensemble members are forced by perturbed wind fields and corrected by DA in a series of 3-day windows. Wind perturbations are represented as a linear combination of empirical orthogonal functions (EOFs) for the larger scales and Debauchies wavelets for the smaller scales. The variance of the EOF expansion coefficients is based on estimates of the wind field error statistics derived using scatterometer observations and a Bayesian Hierarchical Model. It is found that the variance of the wind errors relative to the natural wind variability increases as

Email address: icpasman@uno.edu (Ivo Pasmans)
[1] Now at the University of New Orleans, New Orleans, LA, USA

the horizontal spatial scales decrease. DA corrections to the control run and ensemble members are calculated in parallel by the newly developed, cost-effective *cluster search minimization method*. For a realistic coastal ocean application, this method can generate a 30% wall time reduction compared to the restricted B-conjugate gradient (RBCG) method. Ensemble statistics are generally found to be consistent with background error statistics. In particular, ensemble spread is maintained without inflating. However, sea-surface height background errors can not be fully reproduced by the ensemble perturbations.

Keywords: 4DVAR, Coastal ocean, Data assimilation, Ensemble,
Numerical modelling, USA, Oregon

1. Introduction

Data assimilation (DA) is a procedure, e.g., used in meteorology and oceanography, in which the output of a numerical model is combined with observations to find the most-likely estimate for the true state of the system. DA algorithms require specification of the error statistics for the model and the observations. These statistics are often assumed to follow a multidimensional normal distribution with zero mean and a covariance that is static in time, i.e., the same from one assimilation cycle to the next. An example of such a DA system is the Oregon State University coastal ocean forecast system in an area offshore Oregon-Washington (OR-WA) at the U.S. West coast (Kurapov et al., 2011; Pasmans et al., 2019; Yu et al., 2012). This system applies the 4DVAR algorithm in a series of consecutive 3-day windows. Initial conditions at the beginning of each window are corrected to yield the nonlinear analysis that fits observations in this window. The simulation is then continued for another three days to provide the forecast. Summer dynamics in this region are dominated by the wind-forced upwelling and relaxation and the outflow of the Columbia River (Hickey et al., 2005, 2010; Huyer, 1983; Liu et al., 2009). In such a dynamic environment it is unlikely that the stationarity assumption on the error statistics holds. In particular, the temperature-salinity model error covariance will strongly depend on the presence of the river plume.

Over the past three decades, ensemble methods have been developed in meteorology to deal with non-stationarity in the error statistics. In these methods, the forecast ("background") error covariance is estimated from an ensemble of perturbed model runs. One of the earliest, and most popular, examples of such a method is the ensemble Kalman filter (Anderson, 2001; Bishop et al., 2001; Evensen, 1994; Lermusiaux and Robinson, 1999). More recently in meteorology, these ensemble Kalman filter systems have been combined with 4DVAR systems in which the background error covariance used by the 4DVAR system is estimated using the ensemble from the Kalman filter system (Buehner et al., 2009; Clayton et al., 2013; Zhang and Zhang, 2012). In this manuscript, we describe our approach to using an ensemble of 4DVARs (En4DVar) to provide a state-dependent background error covariance. This methodology will be tested with the OR-WA 4DVAR system. Generalization of the ensemble methodology to an ensembles of 4DVARs is nontrivial for three reasons. First, 4DVAR is computationally intensive. Calculation of a DA correction requires minimization of a cost function, or equivalently, solving a linear system with some symmetric and positive definite matrix

$\mathbf{A}$. This matrix is large and generally not available in explicit form. Instead only the algorithm for the product of $\mathbf{A}$ and a vector is at hand. An iterative method, e.g., the restricted B-conjugate gradient (RBCG) algorithm, is used to find an approximate solution of this system. Each iteration requires propagation of the tangent linear (TL) model over the analysis period, forward in time, and its adjoint counterpart (ADJ), backward in time. For practical systems, the 4DVAR cycle can take 10-100 times as much time as a single forward model run. Second, En4DVar only compounds this problem as it requires running the computationally intensive 4DVAR algorithm for each ensemble member. Third, the ensemble has to be initialized and evolved in such a manner that its covariance is representative of the background error covariance.

Over the last decade, much effort has been put into overcoming these challenges. Several solutions have been found. Instead of applying full 4DVAR to each ensemble member, it has become customary to calculate a low-rank approximation to $\mathbf{A}$, using e.g. Ritz pairs found by the Lanczos algorithm (Trefethen and Bau, 1997). In the Ensemble-Variational Integrated Localized Data Assimilation (EVIL) methodology (Aulign√© et al., 2016), minimization of the cost function is only carried out for the control, or deterministic, model run. From the Ritz vectors obtained as a by-product from this minimization, a low-rank approximation of $\mathbf{A}$ is constructed. The inverse of this approximation is then used to solve the linear system for the ensemble members. Desroziers and Berre (2012); Lorenc et al. (2017) followed similar approaches, but use the Ritz pairs solely to construct a preconditioner. Advances have also been made to speed up the 4DVAR minimization algorithm itself. Parallelisation can be applied to the TL and ADJ models by assigning the calculations for different parts of the domain to different processor cores, as well as to the minimization algorithm. The former is currently standard practice, while the latter is still an area of active research. One popular approach is to use an ensemble of concurrently produced nonlinear model runs to generate approximations of the TL and ADJ model. Examples of this approach are 4DEnVar (Amezcua et al., 2017; Desroziers et al., 2014; Gustafsson and Bojarova, 2014; Liu et al., 2008; Tian et al., 2017) and the Ensemble Kalman Smoother-4DVAR (EKS-4DVAR) (Mandel et al., 2016). 4DEnVar and EKS-4DVAR can be used to minimise the same cost function, but their implementation differs in two major ways. First, 4DEnVar uses all observations within a DA window to correct the initial condition, while EKS-4DVAR processes the observations in batches with each batch generating corrections to the model at, and prior to, the batch time. Second, both in En4DVar and EKS-4DVAR the propagation of perturbations from the background state to the next time step and into the observation space is carried out by a finite-difference scheme involving the nonlinear model and nonlinear sampling operators. The finite-difference scheme in EKS-4DVAR uses a smaller step size and thus better approximates the tangent linear model and linearised sampling operators used in classic 4DVAR. Background error localization in these methods is non-trivial. In absence of a TL model to propagate the localised background error covariance forward in time, localization is often assumed to be static in these methods. For limited-size ensembles with non-dense observation networks this can lead to a decrease in forecast performance compared to variational methods that do use TL and ADJ models (Poterjoy and Zhang, 2015; Poterjoy et al., 2016). A similar problem is encountered when attempting to implement a hybrid background covariance, a linear combination of an ensemble and climatological, static covariance, in these 4DEnVar systems. The use of a hybrid background covariance was found to improve the accuracy of the forecasts produced by traditional 4DVAR systems (Clayton et al., 2013; Kuhl et al., 2013). More specifically, it was found that the best fit to the assimilated observations is achieved when the climatological part makes up the major part of the background covariance (Clayton et al., 2013; Lorenc and Jardak, 2018). However, without TL and ADJ models to propagate the covariance back and forth in time, hybrid 4DEnVar failed to provide the same benefits (Lorenc et al., 2015). One different approach to parallelisation that does not suffer from these problems is taken by Rao and Sandu (2016) and Fisher and Gürol (2017). They make use of the TL and ADJ model and parallelize the 4DVAR minimization algorithm in time. This is done by dividing the analysis period into separate time intervals. The DA correction is found by minimizing a cost function that consists of the sum of the 4DVAR cost functions for each interval plus an additional term representing the constraint that the correction should be continuous going from one interval to another. Another recent approach that circumvents the problems with localisation and hybrid-covariances encountered in En4DVar is the Localized Ensemble-Based Tangent Linear Model (LETLM) in which the matrix for the tangent linear model is not constructed using a linearised version of the extensive nonlinear model, but retrieved from a simple regression against the ensemble members (Allen et al., 2017; Bishop et al., 2017; Frolov and Bishop, 2016; Frolov et al., 2018).

Yet another alternative approach, explored in this paper, is to try to accelerate the linear system solver by using several search directions in parallel. The current OR-WA system uses the restricted B-conjugate gradient method, RBCG (Gürol et al., 2014). In this method, an approximation to the linear system solution is sought in a low-dimensional Krylov space and the search space dimension is equal to the number of iterations. Several generic iterative solvers have already been developed in which the search space dimension grows faster than that. Consequently, less iterations are necessary to produce a good approximation to the solution. Among these are the Enlarged Krylov space method (Grigori et al., 2016) in which the search space is expanded by multiple directions per iteration simultaneously or the Augmented Krylov space methods where extra search directions are added to the system coming either from an earlier attempt to solve a similar system (Erhel and Guyomarc'h, 2000; Morgan, 1995), or from the eigenvectors of a preconditioner (Kharchenko and Yeremin, 1995), or from an attempt to solve the system with a different initial residual (Chapman and Saad, 1996). Additional search directions can also lie outside the Krylov subspace. E.g. Yaremchuk et al. (2017) uses model-based Empirical Orthogonal Functions (EOFs) to create search directions. Once a general search space is defined, the approximation can be defined as the vector in the search space that has the smallest distance to the exact solution in some appropriately chosen norm (Brezinski, 1999).

In the En4DVar system proposed in this this paper, the EVIL method (Auligné et al., 2016) is parallelized using a new variation of the enlarged and augumented Krylov space methods. Two principal and novel elements of the En4DVar will be described. The first is the cluster search method, used to enlarge the search space at each iteration at a price of running a relatively small number of TL-ADJ applications in parallel. The second is the use of a Bayesian Hierarchical Model to estimate the magnitude of the wind forcing perturbations for the ensemble members. Part II of this study (Pasmans et al., in preparation) will include (a) a comparison of the ensemble background error covariance produced by the En4DVar system and the static background error covariance based on the balance operator and (b) comparative tests of En4DVar and traditional 4DVAR implemented in the OR-WA coastal ocean forecast system. This paper is organized as follows: section 2 describes the experimental setup and the layout of the En4DVar system. Derivation of the cluster search method is presented in section 3. Wind perturbations for the ensemble members are discussed in section 4. In section 5 we check if the En4DVar statistics are representative of the background error statistics. Discussion and conclusions are presented in section 6.

2. The En4DVar System

Pasmans et al. (2019) describe the OR-WA coastal ocean forecast system in every detail, implemented as standard 4DVAR with a static background error covariance. A short summary is only provided here. The nonlinear model dynamics are described by the Regional Ocean Modeling System (ROMS, www.myroms.org) integrating three-dimensional, fully nonlinear, hydrostatic, Boussinesq equations featuring advanced numerics (Shchepetkin and McWilliams, 2003, 2005). The model domain is shown in Fig-

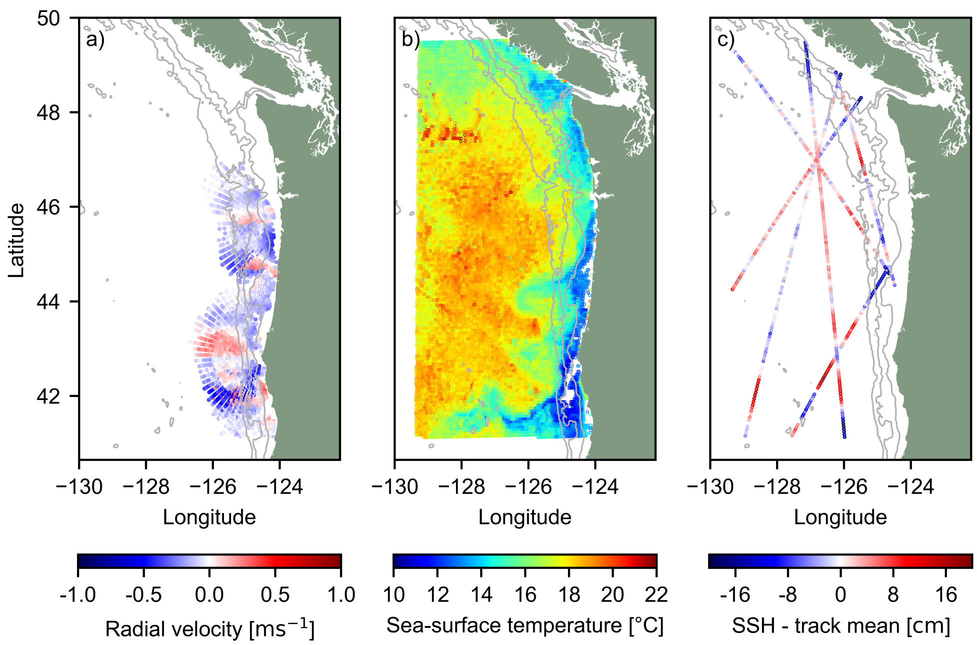

Figure 1: The model domain and observations assimilated in the window of 26-28 August, 2011: (a) high-frequency radar (HFR) daily-averaged radial velocity components, (b) sea-surface temperature (SST) and (c) sea-surface height (SSH, absolute dynamic topography minus the mean along each satellite track).

ure 1. The model resolution is approximately 2 km in the horizontal and 40 terrain-following layers in the vertical direction. The computational grid has 310 × 522 points. Non-tidal boundary conditions are taken from the global 1/12° Hycom-NCODA analyses (COAPS, 2015). Tidal forcing is added along the open boundaries (Egbert and Erofeeva, 2002, 2010). Atmospheric surface forcing is calculated based on the bulk flux algorithm (Fairall et al., 2003) using the 12-km resolution Northern American Mesoscale (NAM) model analysis fields (NOAA, 2011b). The fresh water discharge from the Columbia River, Fraser River, and 15 small rivers in Puget Sound is also included. Each hour the three-dimensional ocean state calculated by the model is saved to disk.

While standard ROMS includes TL and ADJ models, these are tightly integrated into the code such that implementing the En4DVar directly into ROMS was too challenging for us as the users. Instead, we utilize the stand-alone TL and ADJ AVRORA codes developed in-house (Kurapov et al., 2009, 2011; Yu et al., 2012) and integrate these with the nonlinear ROMS and other components of the En4DVar via Linux shell scripts, similarly to how it is done in the present OR-WA operational forecast system. The TL and ADJ runs are performed on a coarser, 4-km resolution model grid and their output is interpolated to and from the 2-km model grid.

All computations have been carried out on the COMET cluster with computer allocations made available through the XSEDE framework (Towns et al., 2014). Both the nonlinear ROMS and TL-ADJ AVRORA codes are run using message passing interface (MPI) parallelisation. The model grid is divided into horizontal tiles and computation in the interior of each tile is performed on a separate processor core. Owing to a relatively small grid size, a small number of $N_{cores} = 6$ tiles are used for each instance of the ADJ and TL model. More nodes are available to us, and later in this paper we discuss how these can be used to speed up the iterative minimization algorithm.

In this paper results from two experiments are compared, Ens and No DA. In experiment Ens, the En4DVar system is used to run $M = 40$ instances of the model forward in time. In the discussions below the run with index $m = 0$ is referred to as the control run, while instances $m = 1, 2, \ldots, M-1$ are referred to as ensemble members. Since the dynamics in this region are dominated by the wind forcing, we assume that the errors in the wind velocity are the dominant model error source. To include this error into the error statistics, the nonlinear forecasts for each ensemble member are run with perturbed wind velocities as detailed in section 4. It will be evaluated in section 5 whether the addition of these perturbations alleviates the need for the customary ensemble inflation (Anderson, 2001; Anderson and Anderson, 1999; Hamill et al., 2001). No wind perturbations are added to the control run. In our system, the analyses and forecasts from the unperturbed control run are considered to provide the best estimate of the ocean state. Unless explicitly stated otherwise, the control DA run is compared with the results from experiment No DA in section 5. Since no perturbations are added to the control run, the probability distributions of the errors in the control run will deviate from that in the ensemble members. Therefore the control run is not utilized in the calculation of $\mathbf{B}$. The ensemble members are all initialized from the same no-DA model output on 10 March 2011 and propagated forward in time without DA using the perturbed winds thus generating an ensemble of perturbed ocean states on 19 April 2011. Both Ens and No DA cases are then run and compared over the period from 19 April 2011 to 1 October 2011.

The set of observations for assimilation includes radial surface currents from high-frequency radars (HFR), alongtrack altimetry, and satellite sea-surface temperature (SST). Although the model includes tides, mainly to include their effect on river and ocean water mixing, our focus is on correcting subtidal variability. Surface tidal currents can be dominanted by non-stationary internal tides (e.g. Kurapov et al., 2003; Osborne et al., 2011) that are poorly predictable and poorly constrained by the available data. At the same time the daily-averaged HFR velocity data present a useful constraint on the 3-day ocean forecasts (Yu et al., 2012). Following our practice in the OR-WA system, we assimilate daily averaged HFR observations, matching those to the daily-averaged model outputs. Altimetry observations can suffer from large errors in the specification of the geoid. To suppress these and the tidal errors in the DA, we assimilate differences from the mean along-track SSH averaged over 24 hours. Details of the procedure can be found in Appendix A. More details on the data and their associated observational error variances are described in Pasmans et al. (2019). The only difference here is that the level 2 GOES-11 SST (Maturi et al., 2008) is used instead of polar-orbiting satellite products. The observational error covariance matrix $\mathbf{R}$ is diagonal.

As shown schematically in Figure 2, DA proceeds in a series of 3-day windows. At the beginning of each window, the initial conditions for the control run and each ensemble member are corrected. Then every model starts from the corrected initial conditions and is run forward using the non-linear ROMS for the length of the assimilation window (3 days) to obtain the analyses and continues for another 3 days to obtain the forecasts. The 3-day window length is long enough to let the non-tidal dynamics evolve and provide dynamically based space and time interpolation of the data. It is still short enough such that the correction to the initial conditions at the beginning of the DA window impacts forecasts.

To explain the DA method in more detail, we combine the temperature, salinity, sea surface height and horizontal velocity fields at the beginning

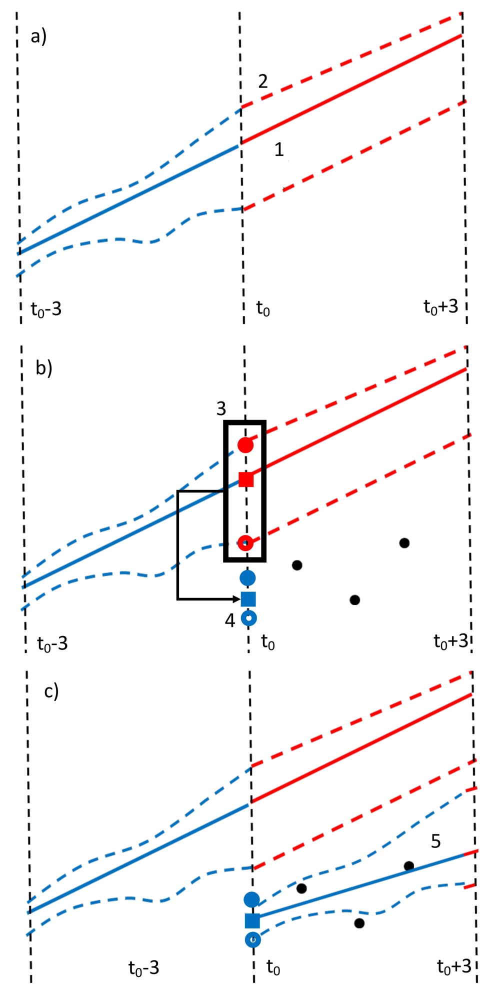

Figure 2: Overview of the En4DVar system. Panels a, b, and c show the progression of tasks. In (a), the control run (task 1, solid line) and ensemble members (with the envelope shown as dashed lines, task 2) are run for six days. The first three days are the analyses (blue), the last three days are the forecasts (red). In (b), task 3 is **B** calculation from the ensemble, and task 4 is the calculation of the DA corrections for the control run and the ensemble members using information from the observations (black circles). In (c), task 5 are the new six day model runs started from the corrected ocean states.

of the window and prior to DA into a vector of real numbers of length N: $\mathbf{x}_b^{(m)} \in \mathbb{R}^N$, $m = 0, \ldots, M-1$. The vector containing all observations within the window is denoted as $\mathbf{y} \in \mathbb{R}^D$. The innovation vector for each ensemble member is defined as $\mathbf{d}^{(m)} = \mathbf{y} - \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(m)}) + \epsilon^{(m)}$. Here $\mathcal{M}(\mathbf{x}_b^{(m)})$ is the nonlinear model trajectory started from the initial conditions $\mathbf{x}_b^{(m)}$ and $\mathbf{H}$ is the collection of data operators. A perturbation $\epsilon^{(m)}$ is added to the innovation vector for each ensemble member ($m = 1, \ldots, M-1$) to account for the uncertainty in the analysis introduced by the presence of observational errors (Burgers et al., 1998; Houtekamer and Mitchell, 1998). It is drawn from a normal distribution with zero mean and covariance $\mathbf{R}$. An overview of the symbols used is included as Appendix B.

The DA correction to the background state $\mathbf{x}_b^{(m)}$ is denoted as $\mathbf{x}^{(m)} \in \mathbb{R}^N$. It is found by minimizing the following cost function for each m (Courtier et al., 1994):

$$J(\mathbf{x}^{(m)}) = \frac{1}{2}\mathbf{x}^{(m)T}\mathbf{B}^{-1}\mathbf{x}^{(m)} + \frac{1}{2}(\mathbf{d}^{(m)} - \mathbf{H}\,\mathbf{M}^{(m)}\mathbf{x}^{(m)})^T\mathbf{R}^{-1}(\mathbf{d}^{(m)} - \mathbf{H}\,\mathbf{M}^{(m)}\mathbf{x}^{(m)}). \tag{1}$$

Here, $\mathbf{M}^{(m)}$ is the TL model, linearised with respect to $\mathcal{M}(\mathbf{x}_b^{(m)})$. $\mathbf{M}^{(m),T}$ is the ADJ model. $\mathbf{B}$ is the background error covariance obtained as the sample covariance of the ensemble members with localization as described in (Pasmans and Kurapov, 2017). To ensure that $\mathbf{B}$ represents dynamics on relatively slow, subtidal and subinertial temporal scales, each ensemble member is time-averaged over the 24h time interval centred at the beginning of the DA window, using the last 12h of the analysis and first 12h of the forecast from the previous window, before it is used in the calculation of $\mathbf{B}$.

The minimizer of (1) is sought as a solution of a linear, symmetric and positive-definite system of equations that can take different forms, e.g., depending on whether the solution is sought in a space of size N or D, and on how the system is preconditioned. RBCG proved to be an efficient solver in the data space of dimension D with good convergence (Gürol et al., 2014). It finds approximations of $\mathbf{x}^{(m)}$ that minimize the cost-function in (1) by solving the system

$$(\mathbf{I}+\mathbf{R}^{-1/2}\mathbf{H}\,\mathbf{M}^{(m)}\mathbf{B}\mathbf{M}^{(m),T}\,\mathbf{H}^{T}\mathbf{R}^{-1/2})\hat{\mathbf{x}}^{(m)} \stackrel{def}{=} \hat{\mathbf{A}}^{(m)}\hat{\mathbf{x}}^{(m)} = \hat{\mathbf{d}}^{(m)} \tag{2}$$

where $\mathbf{x}^{(m)} = \mathbf{B}\mathbf{M}^{(m),T}\mathbf{R}^{-1/2}\hat{\mathbf{x}}^{(m)}$ and $\hat{\mathbf{d}}^{(m)} = \mathbf{R}^{-1/2}\mathbf{d}^{(m)}$.

3. Cost Function Minimization

In this section we discuss several approaches to finding an approximation to $\hat{\mathbf{x}}^{(m)}$, the solution of (2), and propose the new, computationally efficient cluster search algorithm. We recognize that some mathematical details can be overwhelming to an educated reader who only wants to grasp the idea. For that reason we first provide a "high level" overview in the beginning of this session. This is followed by more formal sections 3.1-3.3 where all the necessary theoretical and algorithmic details are documented.

In RBCG, $\hat{\mathbf{x}}_i^{(m)}$, the i-th iteration approximation to $\hat{\mathbf{x}}^{(m)}$, is sought in the low-dimensional Krylov subspace $\mathcal{K}_i(\mathbf{d}^{(m)}, \hat{\mathbf{A}}^{(m)})$, where
$\mathcal{K}_i(\mathbf{z}, \mathbf{A}) = \mathrm{span}(\mathbf{z}, \mathbf{A}\mathbf{z}, \mathbf{A}^2\mathbf{z}, \ldots, \mathbf{A}^{i-1}\mathbf{z})$. This search space grows by one dimension per iteration. Let I' be the number of iterations necessary to bring $\hat{\mathbf{x}}_i^{(m)}$ within a certain prescribed error margin of $\hat{\mathbf{x}}^{(m)}$. Then obtaining $\hat{\mathbf{x}}_{I'}^{(m)}$ for each m can require a considerable amount of wall time as well as

computational resources. Indeed, every iteration requires the multiplication of $\hat{\mathbf{A}}^{(m)}$ with a vector. This demands that for each m a single implementation of the ADJ model over the analysis window is run, followed by application of $\mathbf{B}$, and the TL model. To carry this out, a total of $N_{cores} \times M$ processor cores need to be available in parallel (as described in more detail in section 3.1).

Faster convergence to the exact solution $\hat{\mathbf{x}}^{(m)}$, for each m, could be achieved by expanding the space in which $\hat{\mathbf{x}}_i^{(m)}$ is sought with vectors that lie outside $\mathcal{K}_i(\mathbf{d}^{(m)}, \hat{\mathbf{A}}^{(m)})$. Such vectors can be generated at no extra computational cost if we, similarly to Auligné et al. (2016), make the assumption that

$$\mathbf{M}^{(m)} \approx \mathbf{M}^{(0)} \stackrel{def}{=} \mathbf{M}, \tag{3}$$

and consequently $\hat{\mathbf{A}}^{(m)} \approx \hat{\mathbf{A}}^{(0)} \stackrel{def}{=} \hat{\mathbf{A}}$. In that case, the solution of (2) for different m can be combined into one system of equations. For each m, the search space will grow by M dimensions per iteration. This will allow to approximate $\hat{\mathbf{x}}^{(m)}$ with the same target accuracy in $I < I'$ iterations using the block diagonal CG method (see section 3.2). This approach will potentially exhibit faster convergence, but would still require the same $N_{cores} \times M$ cores per iteration as an ensemble of regular RBCGs. Given presently available resources, this method would be feasible for our relatively small OR-WA forecast system, but it can become prohibitively expensive for larger forecast systems requiring $N_{cores} = O(10^3)$ (e.g. Kurapov et al., 2017). For these systems, the new cluster search method is introduced (see section 3.3). It also depends on the assumption (3) and involves N_s new direction searches at every iteration, where $1 \leq N_s \ll M$. These new search directions are generated in parallel, requiring $N_{cores} \times N_s$ cores to be available simultaneously.

It serves as a compromise between RBCG and block diagonal CG.

3.1. RBCG

Using RBCG (Gürol et al., 2014), $\hat{\mathbf{x}}_i^{(m)}$ satisfies

$$\hat{\mathbf{x}}_i^{(m)} \in \mathcal{K}_i(\hat{\mathbf{d}}^{(m)}, \hat{\mathbf{A}}^{(m)}) : ||\hat{\mathbf{x}}_i^{(m)} - \hat{\mathbf{x}}^{(m)}||_E \leq ||\hat{\mathbf{w}} - \hat{\mathbf{x}}^{(m)}||_E \quad \forall \hat{\mathbf{w}} \in \mathcal{K}_i(\hat{\mathbf{d}}^{(m)}, \hat{\mathbf{A}}^{(m)}) \tag{4}$$

I.e., the i-th approximation to $\hat{\mathbf{x}}^{(m)}$ can be found as a linear combination of the vectors spanning the i-th Krylov space that minimizes the solution error in the E-norm $||\mathbf{w}||_E = (\mathbf{w}^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{A}}^{(m)} \mathbf{w})^{1/2}$ with $\hat{\mathbf{B}}^{(m)} = \hat{\mathbf{A}}^{(m)} - \mathbf{I} = \mathbf{R}^{-1/2} \mathbf{H} \, \mathbf{M}^{(m)} \mathbf{B} \mathbf{M}^{(m),T} \, \mathbf{H}^T \mathbf{R}^{-1/2}$. Then $\hat{\mathbf{x}}_i^{(m)}$ is uniquely determined as the E-projection of $\hat{\mathbf{x}}^{(m)}$ on $\mathcal{K}_i(\hat{\mathbf{d}}^{(m)}, \hat{\mathbf{A}}^{(m)})$:

$$\begin{aligned} \hat{\mathbf{x}}_i^{(m)} &= \hat{\mathbf{V}}_i (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{A}}^{(m)} \hat{\mathbf{V}}^T)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{A}}^{(m)} \hat{\mathbf{x}}^{(m)} \\ &= \hat{\mathbf{V}}_i (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{A}}^{(m)} \hat{\mathbf{V}})^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{d}}^{(m)} \end{aligned} \tag{5}$$

or alternatively,

$$\hat{\mathbf{x}}_i^{(m)} = \hat{\mathbf{V}}_i \mathbf{T}_i^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{d}}^{(m)} \tag{6}$$

where the column space of $\hat{\mathbf{V}}_i$ is equal to $\mathcal{K}_i(\hat{\mathbf{d}}^{(m)}, \hat{\mathbf{A}}^{(m)})$ and $\mathbf{T}_i = (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{A}}^{(m)} \hat{\mathbf{V}}_i$. Here $\hat{\mathbf{V}}_i$ and $\mathbf{T}_i$ depend on m via $\hat{\mathbf{A}}^{(m)}$ and $\hat{\mathbf{d}}^{(m)}$. This yields residuals $\hat{\mathbf{r}}_i^{(m)}$ that are by construction $\hat{\mathbf{B}}^{(m)}$-orthogonal to $\hat{\mathbf{V}}_i$. We refer to the column vectors of $\hat{\mathbf{V}}_i$ as the search directions. As $\hat{\mathbf{x}}_i^{(m)}$ is a projection, it is independent of the search directions chosen as long as they span the same space. In RBCG $i+1$-th search direction would be chosen to be E-orthogonal, i.e. conjugate, to $\hat{\mathbf{V}}_i$. Here the $i+1$-th search direction is chosen to be equal to $\mathbf{r}_i^{(m)}$ which is $\hat{\mathbf{B}}^{(m)}$-orthogonal to $\hat{\mathbf{V}}_i$. I.e.

$\hat{\mathbf{V}}_{i+1} = [\hat{\mathbf{V}}_i, \mathbf{r}_i^{(m)}]$. In this case, the E-orthonormalization is contained in $\mathbf{T}^{-1}(\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{V}}_i)^{-1}$ in (6). The pseudo-code for this method is included in Table C.8.

3.2. Full Parallelisation: Block Diagonal Conjugate Gradient Method

If (3) is assumed then $\hat{\mathbf{A}}^{(m)} = \hat{\mathbf{A}}$, $\hat{\mathbf{B}}^{(m)} = \hat{\mathbf{B}}$ and (2) for the different m can be combined into a single linear system

$$\hat{\mathbf{A}}\hat{\mathbf{X}} = \hat{\mathbf{D}}, \tag{7}$$

where $\hat{\mathbf{X}} = [\hat{\mathbf{x}}^{(0)}, \hat{\mathbf{x}}^{(1)}, \ldots, \hat{\mathbf{x}}^{(M-1)}]$ and $\hat{\mathbf{D}} = [\hat{\mathbf{d}}^{(0)}, \hat{\mathbf{d}}^{(1)}, \ldots, \hat{\mathbf{d}}^{(M-1)}] \in \mathbb{R}^{D\times M}$. Similar to (6), the i-th approximation $\hat{\mathbf{X}}_i$ can be found as

$$\hat{\mathbf{X}}_i = \hat{\mathbf{V}}_i(\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}\hat{\mathbf{A}}\hat{\mathbf{V}}_i)^{-1}\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}\hat{\mathbf{A}}\hat{\mathbf{X}} = \hat{\mathbf{V}}_i \mathbf{T}_i^{-1}(\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}\hat{\mathbf{V}}_i)^{-1}\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}\hat{\mathbf{D}} \tag{8}$$

where $\mathbf{T}_i = (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}\hat{\mathbf{V}}_i)^{-1}\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}\hat{\mathbf{A}}\hat{\mathbf{V}}_i$. $\hat{\mathbf{V}}_i = \hat{\mathbf{D}}$ if $i = 1$ and $\hat{\mathbf{V}}_i = [\hat{\mathbf{V}}_{i-1}, \hat{\mathbf{D}} - \hat{\mathbf{A}}\hat{\mathbf{X}}_i]$ if $i > 1$. $\mathbf{T}$, $\hat{\mathbf{V}}_i$ are independent of m. The column space of $\hat{\mathbf{V}}_i$ is now $\mathcal{K}_i(\hat{\mathbf{D}}, \hat{\mathbf{A}})$. The advantage here, compared to RBCG, is that the search space for each $\hat{\mathbf{x}}_i^{(m)}$, spanned by $\hat{\mathbf{V}}_i$, now has dimension $i \times M$ instead of i. The method results in matrices $\hat{\mathbf{V}}_i$ that are no longer $\hat{\mathbf{B}}$-orthogonal, but $\hat{\mathbf{B}}$-block orthogonal: if $\hat{\mathbf{v}}_p$ and $\hat{\mathbf{v}}_q$ are two columns of $\hat{\mathbf{V}}_i$ then $\hat{\mathbf{v}}_p^T \hat{\mathbf{B}}\hat{\mathbf{v}}_q = 0$ if $|p-q| \geq M$, but might be non-zero otherwise. The estimates $\hat{\mathbf{X}}_i$ retrieved in this way are the same as those found using the block diagonal CG method (O'Leary, 1980) with B-preconditioning. The pseudo-code for the block diagonal CG is given in Table C.9.

3.3. Partial Parallelisation: Cluster Search Method

In order to expand $\hat{\mathbf{V}}_{i-1}$ to $\hat{\mathbf{V}}_i$ in the block diagonal CG method, M applications of $\hat{\mathbf{A}}$ to a vector are necessary. This will require $N_{cores} \times M$

cores and can be prohibitively expensive for large systems. Here we introduce the cluster search method that requires $N_s \ll M$ concurrent applications of $\hat{\mathbf{A}}$ to create the expansion. In this case, we still look for a solution to (7) with $\hat{\mathbf{X}}_i$ still given by (8) but with $\hat{\mathbf{V}}_i$ constructed differently. In particular, we focus on $\hat{\mathbf{x}}^{(0)}$ as it is the control run that will be used to produce the operational forecasts and therefore minimization of the error in $\hat{\mathbf{x}}_i^{(0)}$ has top priority. To explain how $\hat{\mathbf{V}}_i$ is constructed, we momentarily assume that the eigendecomposition $\hat{\mathbf{A}} = \hat{\mathbf{U}}_0 \hat{\mathbf{\Lambda}}^2 \hat{\mathbf{U}}_0^T$ with $\hat{\mathbf{U}}_0^T \hat{\mathbf{U}}_0 = \mathbf{I}$ is available and require that: (i) $\hat{\mathbf{v}}_p^T \hat{\mathbf{B}} \hat{\mathbf{v}}_q = 0$ if $|p-q| \geq N_s$ similar to block diagonal CG and (ii) the residual for the control run, $\hat{\mathbf{r}}_i^{(0)}$, is in the column space of $\hat{\mathbf{V}}_{i+1}$ as is the case in RBCG for $m=0$.

Define $\hat{\mathbf{v}}' = \hat{\mathbf{U}}_0^T \hat{\mathbf{r}}_i^{(0)} = \hat{\mathbf{U}}_0^T \hat{\mathbf{A}} (\hat{\mathbf{x}}^{(0)} - \hat{\mathbf{x}}_i^{(0)})$, $\hat{\mathbf{v}}'' = \hat{\mathbf{U}}_0^T \hat{\mathbf{d}}^{(0)} = \hat{\mathbf{U}}_0^T \hat{\mathbf{A}} \hat{\mathbf{x}}^{(0)}$ and $\hat{\mathbf{U}}_i = \hat{\mathbf{U}}_0 - \hat{\mathbf{A}} \hat{\mathbf{V}}_i \mathbf{T}_i^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_0$. Then

$$\hat{\mathbf{U}}_0 \hat{\mathbf{v}}' = \hat{\mathbf{r}}_i^{(0)} = \hat{\mathbf{r}}_0^{(0)} - \hat{\mathbf{A}} \hat{\mathbf{x}}_i^{(0)} = \hat{\mathbf{r}}_0^{(0)} - \hat{\mathbf{A}} \hat{\mathbf{V}}_i \mathbf{T}_i^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \mathbf{r}_0^{(0)} = \hat{\mathbf{U}}_i \hat{\mathbf{v}}'' \quad (9)$$

To expand $\hat{\mathbf{V}}_i$ to $\hat{\mathbf{V}}_{i+1}$ we look for N_s new search vectors of the form $\mathbf{s}^{(n)} = \hat{\mathbf{U}}_i \hat{\mathbf{\Lambda}}^2 \mathbf{P}_n \hat{\mathbf{v}}''$ with $n = 1, 2, \ldots, N_s$ and $\mathbf{P}_n = \sum_{d \in D_n} \hat{\mathbf{e}}_d (\hat{\mathbf{e}}_d^T \hat{\mathbf{v}}'')$ with $\hat{\mathbf{e}}_d$ the unit vector in direction d. Here D_n is a subset of $\{1, 2, \ldots, D\}$ such that the union of $D_1, D_2, \ldots, D_{N_s}$ is $\{1, 2, \ldots, D\}$ and D_p and D_q are disjoint if $p \neq q$. Consequently, $\sum_{n=1}^{N_s} \mathbf{P}_n \hat{\mathbf{v}}'' = \hat{\mathbf{v}}''$. Combined with the equality $\hat{\mathbf{U}}_0 \hat{\mathbf{v}}' = \hat{\mathbf{U}}_i \hat{\mathbf{v}}''$ in (9) this then ensures that $\hat{\mathbf{r}}_i^{(0)}$ lies within $\hat{\mathbf{V}}_{i+1}$. Thus search vectors of the form $\mathbf{s}^{(n)}$ satisfy requirement (ii). Furthermore,

$$\begin{aligned} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_i &= \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_0 - \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{A}} \hat{\mathbf{V}}_i \mathbf{T}_i^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_0 \\ &= \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_0 - \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{A}} \hat{\mathbf{V}}_i (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{A}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_0 = \mathbf{0} \end{aligned} \quad (10)$$

This shows that $\hat{\mathbf{U}}_i$ is $\hat{\mathbf{B}}$-orthogonal to $\hat{\mathbf{V}}_i$ and since the N_s new search directions in $\hat{\mathbf{V}}_{i+1}$ are linear combinations of the column vectors of $\hat{\mathbf{U}}_i$, they satisfy requirement (i).

For the following we also need to be able to estimate $\hat{\mathbf{U}}_0^T\hat{\mathbf{A}}\hat{\mathbf{U}}_i$. For $i = 0$, $\hat{\mathbf{U}}_i = \hat{\mathbf{U}}_0$ and so $\hat{\mathbf{U}}_0^T\hat{\mathbf{A}}\hat{\mathbf{U}}_0 = \Lambda^2$, while for $i > 0$ exact expressions are not directly available. Instead we observe that the columns of $\hat{\mathbf{U}}_i$ are the residuals obtained after trying to find a solution to the linear system $\hat{\mathbf{A}}\mathbf{X}' = \hat{\mathbf{U}}_0$ in the search space $\mathbf{V}_i$. This system has the exact solution $\hat{\mathbf{U}}_0\Lambda^{-2}$. Here we make an ad-hoc assumption that these residuals are, in good approximation, multiples of the columns $\hat{\mathbf{U}}_0$, i.e. $\hat{\mathbf{U}}_i \approx \hat{\mathbf{U}}_0\Xi$ with Ξ diagonal. In this case, Ξ can be estimated as $\Xi^2 = \Xi^T\hat{\mathbf{U}}_0^T\hat{\mathbf{U}}_0\Xi \approx \hat{\mathbf{U}}_i^T\hat{\mathbf{U}}_i$.

In RBCG $\hat{\mathbf{x}}_i^{(0)}$ is defined as the vector in the search space $span(\hat{\mathbf{V}}_i)$ that minimizes the error (4) in the E-norm. The novel idea behind cluster search is to find a clustering $D_1, D_2, \ldots, D_{N_s}$ and the associated N_s new search vectors $\mathbf{s}^{(n)}$ such that the reduction of the expected error $||\hat{\mathbf{x}}^{(0)} - \hat{\mathbf{x}}_{i+1}^{(0)}||_E$ is larger than can be achieved using any other clustering. Using the properties of $\mathbf{P}_n$, the estimation $\hat{\mathbf{U}}_i \approx \hat{\mathbf{U}}_0\Xi_i$, and the orthonormality of $\hat{\mathbf{U}}_0$, we find that the E-norm of the expected error for $\hat{\mathbf{x}}_{i+1}^{(0)}$ can be estimated as

$$\begin{aligned}||\hat{\mathbf{x}}^{(0)} - \hat{\mathbf{x}}_i^{(0)} - \sum_{n=1}^{N_s}\alpha_n\hat{\mathbf{U}}_i\mathbf{P}_n\hat{\mathbf{v}}''||_E^2 \;\approx\; & \sum_{n=1}^{N_s}\sum_{d\in D_n}[(1-\lambda_d^{-2})v_d'^2 - 2\alpha_n v_d'\lambda_d^2(1-\lambda_d^{-2})v_d''\xi_d \\ & +\alpha_n^2\lambda_d^4(1-\lambda_d^{-2})v_d''^2\xi_d^2], \qquad (11)\end{aligned}$$

where λ_d and ξ_d are the d-th element on the diagonal of Λ and Ξ, correspondingly. To find the minimum of this function, we set the derivative of (11) as a function of α_n to zero and get

$$\hat{\alpha}_n = (\sum_{d\in D_n}(1-\lambda_d^{-2})\lambda_d^4 v_d''^2\xi_d^2\frac{v_d'}{v_d''\xi_d\lambda_d^2})(\sum_{d\in D_n}(1-\lambda_d^{-2})\lambda_d^4 v_d''^2\xi_d^2)^{-1} \qquad (12)$$

$$= \overline{\frac{v'}{v''\xi_d}\frac{1}{\lambda_d^2}}^n, \tag{13}$$

where $\overline{\cdot}^n$ denotes the weighted mean over the cluster D_n with weights $(1 - \lambda_d^{-2})\lambda_d^4 v_d''^2 \xi_d^2$. Inserting $\hat{\alpha}_n$ from (13) back into (11) gives that for our guesses of $\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(N_s)}$ the error squared obtains a minimum

$$\begin{aligned} &||\hat{\mathbf{x}}^{(0)} - \hat{\mathbf{x}}_i^{(0)} - \textstyle\sum_{n=1}^{N_s} \hat{\alpha}_n \hat{\mathbf{U}}_i \mathbf{P}_n \hat{\mathbf{v}}''||_E^2 \\ &\approx \textstyle\sum_{n=1}^{N_s} \sum_{d \in D_n} (1 - \lambda_d^{-2}) v_d''^2 \xi_d^2 \lambda_d^4 [\frac{v_d'^2}{v_d''^2 \xi_d^2 \lambda_d^4} - \hat{\alpha}_n^2] \\ &= \textstyle\sum_{n=1}^{N_s} W_n \, \mathrm{var}_n(\frac{v_d'}{v_d'' \xi_d \lambda_d^2}) \end{aligned} \tag{14}$$

with var_n the weighted variance over cluster n and $W_n = \sum_{d \in D_n} (1 - \lambda_d^{-2}) \lambda_d^4 v_d''^2 \xi_d^2$ the normalization coefficient for the n-th cluster. The K-means clustering algorithm (MacQueen, 1967) can now be used to find an approximate clustering that approximately minimizes (14). Once K-means produces a clustering, $\mathbf{s}^{(n)} = \hat{\mathbf{U}}_i \hat{\mathbf{\Lambda}}^2 \mathbf{P}_n \hat{\mathbf{v}}''$ are known and $\hat{\mathbf{V}}_{i+1} = [\hat{\mathbf{V}}_i, \mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(N_s)}]$.

In reality the eigenvalue decomposition of $\hat{\mathbf{A}}$ is not available. Instead it is used that if $\mathbf{R}$ and $\mathbf{B}$ are true estimates of the observational and background error covariance then $\hat{\mathbf{A}} = \langle \hat{\mathbf{d}} \hat{\mathbf{d}}^T \rangle$ (Desroziers et al., 2005). Here $\langle \cdot \rangle$ denote the expected value. Approximations to the eigenvectors and eigenvalues of $\hat{\mathbf{A}}$ are then found by calculating the eigenvalue decomposition of $\frac{1}{M}[\hat{\mathbf{d}}^{(0)}, \hat{\mathbf{d}}^{(1)}, \ldots, \hat{\mathbf{d}}^{(M-1)}][\hat{\mathbf{d}}^{(0)}, \hat{\mathbf{d}}^{(1)}, \ldots, \hat{\mathbf{d}}^{(M-1)}]^T = \frac{1}{M} \hat{\mathbf{D}} \hat{\mathbf{D}}^T \approx \langle \hat{\mathbf{d}} \hat{\mathbf{d}}^T \rangle$. The pseudocode for the cluster search method can be found in Table C.10. An overview of where the cluster search method enters the cost-function minimization algorithm is shown in Figure 3.

Notice that if $N_s = 1$ there is only one cluster and requirement (i) ensures that the new search direction is equal to $\mathbf{r}_i^{(0)}$. Consequently, the clustering method reverts to RBCG described in section 3.1 for $m = 0$. If $N_s = M$, each column vector of $\hat{\mathbf{U}}_i$ constitutes its own cluster and hence the new search directions are multiples of the column vectors of $\hat{\mathbf{U}}_i$. By construction the column vectors of $\hat{\mathbf{U}}_i$ are linear combinations of the i-th residuals from the different ensemble members. Consequently, $span(\hat{\mathbf{V}}_i)$ is equal in both the block diagonal CG and cluster search methods.

In section 5, we will compare convergence rates of RBCG, full parallelisation, and the cluster search methods in the realistic OR-WA system set-up. Before we can proceed with those, we next describe the wind perturbations that will be utilized in the tests of section 5.

4. Wind Perturbations

Conventionally, multiplicative ensemble inflation (Anderson and Anderson, 1999) is applied to the ensemble members to compensate for the fact that ensembles generally fail to account for all error sources. Multiplicative ensemble inflation implicitly assumes that the ensemble underestimates the relative growth of the background errors uniformly throughout the model. In the Oregon-Washington region surface currents, the strength of the coastal upwelling (Halpern, 1976; Huyer, 1983), and the location of the fresh water Columbia River plume (Hickey et al., 1998, 2005; Liu et al., 2009) all depend on the wind forcing. Therefore uncertainty in the wind forcing is assumed to be the dominant source of model error. In an attempt to better approximate the structure and strength of the background errors, we have opted for an approach different from multiplicative ensemble inflation. In this approach physically realistic wind perturbations are added to the ensemble members ($m = 1, \ldots, M-1$) and it is left up to the model physics to translate these wind forcing errors into background errors in the ocean state. Although no comparison with ensemble inflation will be made, we will verify later in this manuscript if adding the wind perturbations helps to avoid ensemble variance shrinking alleviating the need for the ensemble inflation (Hamill and Whitaker, 2005; Li et al., 2009; Whitaker and Hamill, 2002).

The perturbed wind fields for an ensemble members are generated as

$$\mathbf{w}(t) = \mathbf{w}_{NAM}(t) + \mathbf{w}_L(t) + \mathbf{w}_S(t) \tag{15}$$

with $\mathbf{w}_{NAM}(t) \in \mathbb{R}^{2N_w}$ the vector containing the meridional and zonal wind velocity components from the NAM model interpolated to the N_w ROMS model surface grid points. Fields $\mathbf{w}_L(t)$ and $\mathbf{w}_S(t)$ represent the large-scale and small-scale wind perturbations respectively.

For the large-scale perturbations, we use the empirical orthogonal function (EOF) decomposition of the series $\mathbf{w}'_{NAM}(t) = \mathbf{w}_{NAM}(t) - \langle \mathbf{w}_{NAM} \rangle$, where the winds are provided every 6 hr from 1 January 2011 00:00 to 31 December 2011 18:00 and $\langle \mathbf{w}_{NAM} \rangle$ is the mean wind field over this period. After the EOF decomposition the NAM wind field can be written as

$$\mathbf{w}_{NAM}(t) = \langle \mathbf{w}_{NAM} \rangle + \sum_{i=1}^{N_{EOF}} \beta_{NAM,i}(t)\mathbf{w}_{EOF,i} + \mathbf{w}_\perp(t), \tag{16}$$

where $\mathbf{w}_{EOF,i}$ is the EOF mode associated with the i-th largest singular value, $\mathbf{w}_\perp(t)^T\mathbf{w}_{EOF,i} = 0$ for $i = 1, 2, \ldots, N_{EOF}$ and $\beta_{NAM,i}(t)$ are the EOF expansion coefficients associated with different EOFs and different times. Here, we use 10 EOFs ($N_{EOF} = 10$) that explain 95% of the variance of $\mathbf{w}_{NAM}$ in time. Similarly to Hénaff et al. (2009) and Vervatis et al. (2016), we define the large-scale wind perturbation to be

$$\mathbf{w}_L(t) = \sum_{i=1}^{N_{EOF}} \beta_{L,i}(t)\mathbf{w}_{EOF,i}. \tag{17}$$

The expansion coefficients for the large-scale wind perturbations, $\beta_{L,i}$, with specified standard deviation $\hat{\sigma}_{EOF,i}$, are generated by an AR1-process

$$\begin{aligned} \beta_{L,i}(t) &= \hat{\sigma}_{EOF,i}\epsilon_{\beta,i}(t) && \text{for } t = 0 \\ \beta_{L,i}(t) &= c_\beta \beta_{L,i}(t-\Delta t) + \sqrt{1-c_\beta^2}\hat{\sigma}_{EOF,i}\epsilon_{\beta,i}(t) && \text{for } t \geq \Delta t. \end{aligned} \tag{18}$$

Here $\epsilon_{\beta,i}(t)$ is drawn from a standard normal distribution, $\Delta t = 6\,\mathrm{h}$ is the output time step of the NAM model and correlation coefficient $c_\beta = 0.4$ (Milliff et al., 2011). The two dominant wind EOFs scaled by the standard deviations of their expansion coefficients in the large-scale wind errors ($\hat{\sigma}_{EOF,i}$) are shown in Figure 4a,b.

The standard deviation of the large-scale expansion coefficients $\beta_{L,i}$ is estimated based on the differences between the 6-hourly NAM model wind output and the daily, 25 km ASCAT satellite wind product (Figa-Saldaa et al., 2002; NOAA, 2011a), and NDBC buoy (numbers 46089,46015, 46050, 46029, 46041) wind observations (NOAA, 2016). The estimate $\hat{\sigma}_{EOF,i}$ used is the mode of

$$\begin{aligned} &p(\sigma^2_{EOF,i}|\underline{w}_{obs}) = \\ &\int p(\underline{\mathbf{w}}_S, \underline{\beta}_L, \underline{\epsilon}_{obs}, \sigma^2_{EOF,1}, \ldots, \sigma^2_{EOF,N_{EOF}}, \sigma^2_S|\underline{w}_{obs})\, d\underline{\mathbf{w}}_S \\ &\times d\underline{\beta}_L d\underline{\epsilon}_{obs} \textstyle\prod_{j=1,j\neq i}^{N_{EOF}} d\sigma^2_{EOF,j} d\sigma^2_S, \end{aligned} \tag{19}$$

the conditional probability distribution for $\sigma^2_{EOF,i}$ given all the scatterometer and buoy wind observations in the model domain in 2011 mapped to the NAM model output times (vector $\underline{w}_{obs}$). For the buoy observations the time-mapping is done by selecting the buoy wind measurement closest to the NAM model output time, while the daily ASCAT observations are compared with the NAM model output time on the same day for which the RMSE between ASCAT observations and NAM model output is minimal. The conditional probability distribution in (19) is constructed using a Bayesian Hierarchical Model (BHM) similar to the one used in (Milliff et al., 2011; Wikle et al., 2001). The BHM consists of three stages:

$$\begin{aligned}
&\mathbf{w}_{obs}(t_j) = \mathbf{H}_{t_j}\mathbf{w}_{true}(t_j) + \epsilon_{obs}(t_j) \quad \text{(data stage)}\\
&\mathbf{w}_{true}(t_j) = \mathbf{w}_{NAM}(t_j) + \mathbf{w}_S(t_j) + \textstyle\sum_{i=1}^{N_{EOF}} \mathbf{w}_{EOF,i}\beta_{L,i}(t_j) \quad \text{(process stage)}\\
&\sigma^2_{obs},\, \sigma^2_S,\, \sigma^2_{EOF,i} \quad \text{(parameter stage)}
\end{aligned} \tag{20}$$

with the underbar denoting the concatenation of vectors taken at different NAM model output times t_j into one vector, e.g. $\underline{\mathbf{w}}_S = [\mathbf{w}_S(t_1); \mathbf{w}_S(t_2); \ldots; \mathbf{w}_S(t_{N_t})]$, $\mathbf{H}_{t_j}$ the linear operator that maps the wind velocities at time t_j from the model grid to the ASCAT and buoy observation locations, $\mathbf{w}_{true}(t_j) \in \mathbb{R}^{2N_w}$ the unknown true wind field at time t_j, $\epsilon_{obs}(t_j)$ the measurement error in the ASCAT/NDBC buoy wind observations, $\mathbf{w}_S(t_j) \in \mathbb{R}^{2N_w}$ the error in the small-scale wind field and $\beta_{L,i}(t_j)$ the expansion coefficient for the i-th EOF mode in the large-scale error in the wind field. Prior distributions of $\epsilon_{obs}(t_j)$, $\mathbf{w}_S(t_j)$, $\beta_{L,i}(t_j)$, σ^2_S and $\sigma^2_{EOF,i}$ are assumed to be:

$$\begin{aligned}
&\epsilon_{obs}(t_j) \sim N(\epsilon_{obs}(t_j); \mathbf{0}, \sigma^2_{obs}\mathbf{I})\\
&\mathbf{w}_S(t_j) \sim N(\mathbf{w}_S(t_j); \mathbf{0}, \sigma^2_S\mathbf{I})\\
&\beta_{L,i}(t_j) \sim N(\beta_{L,i}(t_j); 0, \sigma^2_{EOF,i})\\
&\sigma^2_S \sim IG(\sigma^2_S; a_S, b_S)\\
&\sigma^2_{EOF,i} \sim IG(\sigma^2_{EOF,i}; a_{EOF,i}, b_{EOF,i})
\end{aligned} \tag{21}$$

where $N(\mathbf{x};\mu,\mathbf{C})$ is a normal distribution with mean μ and covariance $\mathbf{C}$, $IG(x;a,b)$ the inverse gamma distribution with parameters a,b (see Appendix D for the details) and and $\sim$ denotes that a value or vector is randomly drawn from the given distribution. Any spatial structure in the small-scale errors $\mathbf{w}_S(t_j)$ is neglected. Based on ASCAT validation (Verspeek et al., 2013) σ_{obs} is set to $0.7\,\mathrm{m\ s^{-1}}$. We pick $a_{EOF,i} = \frac{1}{20}N_t$, $b_{EOF,i} = 0.1a_{EOF,i}\,var(\beta_{NAM,i})$ with N_t the number of days on which ASCAT observations are available and $var(\beta_{NAM,i})$ the variance of the coefficient β_i in (16). This gives an a priori distribution for $\sigma^2_{EOF,i}$ with mode $\frac{b_{EOF,i}}{1+a_{EOF,i}} \approx 0.1\,var(\beta_i)$. These values were chosen such that this mode corresponds to the $\hat{\sigma}^2_{EOF,i} = 0.09\,var(\beta_{NAM,i})$ estimate used by Hénaff et al. (2009) and Vervatis et al. (2016). Similarly, a_S and b_S are chosen to be $a_S = \frac{2}{20}N_tN_w$ and $b_S = \sigma^2_{obs}a_S$ given the a priori distribution of σ^2_S with a mode of approximately σ^2_{obs}.

The conditional probability distribution (19) derived using the BHM above is retrieved using a Gibbs sampler (see Appendix D) and is shown in Figure 5 for the nine dominant EOF modes. Also indicated in Figure 5 is the percentage of the variance in the NAM wind fields explained by each EOF. In addition, the 9% of this variance is shown by dashed lines, as this value was used in other studies to estimate $\hat{\sigma}_{EOF,i}$ (Hénaff et al., 2009; Vervatis et al., 2016). Figure 5 shows that the BHM estimate for $\hat{\sigma}_{EOF,i}$ is higher than the 9% estimate in all modes except mode one. The difference between the two estimates increases for increasing EOF number. For the higher EOFs (mode 4 and higher), which represent smaller spatial scales in the wind field (not shown), the 9% estimate severely underestimates the contribution of the mode to the error in the wind fields.

The study of scatterometer wind measurements over the Pacific Ocean shows that the power spectral density (PSD) of the wind field scales with $\kappa^{-\hat{\gamma}}$, where κ is the wave number and $\hat{\gamma} \approx 2$ (Chin et al., 1998). The PSD of the meridional NAM wind field determined using a Hamming window (Figure 6, solid blue line) decreases faster than this for $\kappa > 0.3$ rad km^{-1} owing to the limited (12-km) NAM resolution. As the NAM model cannot represent the small-scale wind field, probable small-scale wind fields are added to the ensemble members. Following Wikle et al. (2001), it is assumed that the small-scale wind field in (15) can be decomposed into Daubechies-2 wavelets (Cohen et al., 1993):

$$\mathbf{w}_S(t) = \gamma_0 \sum_{n=1}^{9} \sum_{i} \gamma_i^{(n)}(t) \psi_i^{(n)} \tag{22}$$

with $\gamma_i^{(n)}(t)$ coming from an AR1-process

$$\gamma_i^{(n)}(t) = c_\gamma \gamma_i^{(n)}(t - \Delta t) + \sqrt{1 - c_\gamma^2}\, \sigma_\gamma^{(n)} \epsilon_i^{(n)}(t) \tag{23}$$

Here n indicates the level of the wavelet, with the length scale of the wavelets doubling as the level increases with one and i running over all the wavelets that are available at level n. Similarly to the large-scale wind field, we use $c_\gamma = 0.4$. This wavelet approach yields small-scale wind perturbations that are local in space and are simultaneously constrained to a limited spectral band. By experimentation, the standard deviation of $\sigma_\gamma^{(n)}$ is chosen such that the dependence of the PSD S on the wave number κ of the total ensemble member wind field scales as $S(\kappa) \sim \kappa^{-2}$. This was achieved by setting $\sigma_\gamma^{(n)} = \exp[1.3(n-3)](0.5 - 0.5 \tanh[0.4(n-4)])$ and picking γ_0 such that the variance of the wind speed is $(0.55\,\mathrm{m\,s^{-1}})^2$. The PSDs for the wind fields of the different ensemble members on 8 August 2011 00:00 are shown together with the linear least-square log-log fit to the ensemble mean PSD for $\kappa > 0.1\,\mathrm{rad\,km}^{-1}$ in Figure 6. The fit confirms that the PSDs have indeed the desired $PSD \sim \kappa^{-2.0}$ relationship.

5. Results

5.1. Convergence

The effectiveness of the cluster search algorithm using a different number of clusters is compared with that of RBCG and block diagonal CG. Even though the DA correction $\mathbf{x}^{(m)}$ is only calculated after the last inner loop iteration $i = I$, the cost function (1) can be calculated for each inner loop iteration i if the substitution $\mathbf{x}_i^{(m)} = \mathbf{B}\mathbf{M}^T\mathbf{H}^T\mathbf{R}^{-1/2}\hat{\mathbf{x}}_i^{(m)}$ is made in (1):

$$J(\mathbf{x}_i^{(m)}) \;=\; \tfrac{1}{2}\hat{\mathbf{x}}_i^{(m),T}\hat{\mathbf{B}}\hat{\mathbf{x}}_i^{(m)} + \tfrac{1}{2}(\hat{\mathbf{d}}^{(m)} - \hat{\mathbf{B}}\hat{\mathbf{x}}_i^{(m)})^T(\hat{\mathbf{d}}^{(m)} - \hat{\mathbf{B}}\hat{\mathbf{x}}_i^{(m)}). \qquad (24)$$

Using (5) we find that $\hat{\mathbf{B}}\hat{\mathbf{x}}_i^{(m)} = \hat{\mathbf{B}}\hat{\mathbf{V}}_i\mathbf{T}_i^{-1}(\hat{\mathbf{V}}_i^T\hat{\mathbf{B}}\hat{\mathbf{V}}_i)^{-1}\hat{\mathbf{V}}_i^T\hat{\mathbf{B}}\hat{\mathbf{d}}^{(m)}$ which is readily available as $\hat{\mathbf{B}}\hat{\mathbf{V}}_i$ is stored. Using (24) the value of the cost function was calculated prior to each inner loop iteration using the cluster search method with different numbers of clusters ($1 \le N_s \le 40$). The cost function normalized by its value at the start of the minimization is shown in Figure 7a and b, for the windows starting on 31 May and 26 August 2011 respectively. Increase in the number of clusters N_s, and correspondingly the number of search directions at each iteration, consistently improves the rate at which the cost function decreases as the function of the inner loop iteration number. To provide a more quantitative assessment of the advantage of using several search directions in parallel, we compute the speed-up ratio

$$a(N_s) = \frac{I(1)}{I(N_s)}, \tag{25}$$

where $I(N_s)$ is the number of iterations needed to reach a specific reference level of the cost function $J = J_{\text{ref}}$ using the cluster method with N_s new search directions per iteration. $I(1)$ corresponds to RBCG. For J_{ref}, we choose the value in the case using $N_s = 4$ and $I = 12$ iterations as this will be adopted later as the standard setup in the long-term system evaluation in Part 2 of this study (Pasmans et al., in preparation). The speed-up ratios are shown in Figure 7c and indicate, e.g., that RBCG ($N_s = 1$) needs approximately 30% more iterations than cluster search with $N_s = 4$ to reach the same level of cost function reduction. A fit of a 2nd order polynomial to a (dashed black lines in Figure 7c) shows that the coefficient for the quadratic term is negative and significantly different from zero at a 95% significance level indicating that the additional benefit of adding more clusters diminishes as the number of clusters increases.

Figure 8 compares differences in the initial condition corrections between RBCG ($N_s = 1$) and the cluster search with $N_s = 4$ on 26 August 2011. As the time available to perform DA is constrained in operational settings, the minimization in both these two cases is terminated after $I = 12$ inner loop iterations. The plots on the left show the DA correction calculated for SST, surface velocity and SSH fields with $N_s = 1$; and the plots on the right show the difference between the DA corrections in the cases with $N_s = 4$ and $N_s = 1$. While both methods yield similar large-scale corrections, they differ in details at the scale of geostrophic eddies.

For the ensemble members additional dependency on the search space comes from the fact that when cluster search is used the right hand side of (7) is replaced by its $\hat{\mathbf{B}}$-projection on the search space. I.e., on the right-hand side of (7), $\hat{\mathbf{d}}^{(m)}$ is effectively replaced by $\hat{\mathbf{V}}_i(\hat{\mathbf{V}}_i^T\hat{\mathbf{B}}\hat{\mathbf{V}}_i)^{-1}\hat{\mathbf{V}}_i^T\hat{\mathbf{B}}\hat{\mathbf{d}}^{(m)}$. For the control run this is not an issue as by construction $\hat{\mathbf{d}}^{(0)}$ lies in $span(\hat{\mathbf{V}}_1)$, but for the ensemble members this could result in the systematic elimination of a part of the errors contained in $\hat{\mathbf{d}}^{(m)}$. Such an elimination would result in DA corrections for the ensemble members that are too small and consequently an ensemble spread that will be too large. To test whether this is a valid concern, the normalized RMSE for each observation type, i.e. the RMS of the elements of $\hat{\mathbf{d}}^{(m)}$ associated with one type of observations, is calculated and compared with the RMSE after taking the $\hat{\mathbf{B}}$-projection of $\hat{\mathbf{d}}^{(m)}$ on $\hat{\mathbf{V}}_i$ with $i = 12$. If $\hat{\mathbf{d}}^{(m)}$ lies completely in $\hat{\mathbf{V}}_i$, as is the case for $m = 0$, the ratio of the latter over the former is one. The actual ratio in the experiment is calculated for each ensemble member and each window and the lower bound, upper bound and ensemble mean are shown in Figure 9. Figure 9 shows that, as expected, using the projection can result in the reduction of the RMSE (up to 40%). However, the figure also shows that taking the projection can increase the RMSE. This paradoxical behaviour emerges because the projection uses the $\hat{\mathbf{B}}$-inner product, while in the calculation of the RMSE involves the normal, Euclidean, inner product. Taking the mean of the ratios over all ensemble members shows that increases in the RMSEs created by the $\hat{\mathbf{B}}$-projection mostly, but not completely, offset the reductions in RMSEs and that the net result is a small decrease in the RMSE of 1.7% for SST, 3.6% for HFR and 2.8% for SSH observations. So, the projection effect might indeed result in overestimation of the error variances by the ensemble, but this effect is small.

5.2. Error Reduction

Table 1: RMS and the mean of the difference experiment minus observations as shown in Figure 10 and 11 for the different models and over the period 22 April to 28 September 2011.

	RMSE			Bias	
	SST [°C]	HFR [$\mathrm{cm\,s^{-1}}$]	SSH [cm]	SST [°C]	HFR [$\mathrm{cm\,s^{-1}}$]
No DA	1.17	18.3	6.4	-0.18	1.5
Control analysis	0.75	10.5	3.8	-0.04	0.2
Ensemble mean analysis	0.76	9.7	3.9	-0.05	-0.2
Control forecast	0.94	13.2	4.9	-0.07	0.4
Ensemble mean forecast	0.92	11.2	4.8	-0.13	-0.1

To test whether the system is effective correcting RMSE not only for the control run but also for the ensemble members, the RMSE between the data used in the assimilation and the nonlinear analyses and forecasts is calculated for the ensemble members as shown in Figure 10. Each line segment represents the RMSE in the analysis (left point) and in the forecast for the subsequent window (right point). Note that the forecast RMSE (right points) are calculated with respect to formally future observations. The En4DVar system is effective in reducing the RMSE: the analysis RMSE for the ensemble members exceeds that in experiment No DA (blue line) in less than 4% of the cases. Forecast RMSE for the ensemble members are smaller than No DA forecast RMSE in 73% of cases. The RMSEs for the ensemble members, however, are consistently larger than those for the control run (green line) as they are forced with perturbed wind fields and corrected with perturbed observations. However, the errors introduced by the perturbations cancel out in the ensemble mean. Indeed, the ensemble mean RMSE lies below that of the ensemble members. Note that in the ensemble Kalman filter there is no "control run" and the ensemble mean is used as the best estimate (Evensen, 1994). We could have used the same approach, but the additional communication between the computational nodes required to calculate the mean ocean state would have increased the wall time significantly. Table 1 shows that the RMSE of the control run is on par with that of the ensemble mean, with the exception of the RMSE in the HFR observations after 14 August 2011 (see Figure 10b). Hence, our choice to pick the control run over the ensemble mean to produce the forecasts will have only a limited negative impact on the forecast accuracy.

Figure 11 shows the observation-model bias in experiment No DA, the ensemble members, the control run, and the ensemble mean. As the along-track mean is removed from both the altimetry observations and their model equivalents prior to assimilation (see section 2), the along-track mean of both the assimilated altimetry observations and their model equivalents is by construction zero. Consequently, the bias along each track, and thus in general, is zero and is therefore not included in Figure 11. The bias in the HFR observation shows a spread around zero for both the ensemble members as well as the control run forecasts. The bias in the forecasts predictions for the SST observations, however, has a negative tendency with particularly large negative biases during the periods 13-16 May, 5-8 June, 21-24 July, 26-29 August, 25-28 September 2011. This results in an overall negative bias over the whole period as shown in Table 1. It is indicative of either insufficient surface heating in the model, too much mixing in the upper layer, or a positive bias in the satellite observations. Further verification against independent in-situ observations is described in Part II.

5.3. Ensemble Reliability

If the ensemble statistics are truly representative of the background error statistics, the ensemble is said to have high reliability. A rank diagram is a diagnostic that can be used to test this (Hamill, 2001). Figure 12 shows rank diagrams for the three different types of observations. The steps to construct these are as follows: (a) sample each ensemble member forecast at the observation locations and times and add a random observation error, (b) for each observation, count the number of ensemble forecasts that are lower than the measured value, (c) by definition, this number plus one is the rank of the observation, (d) count the frequency of each rank and divide by the total number of observations to determine the normalized frequency. If the ensemble is reliable, the rank diagram should be flat (Hamill, 2001). The 95%-confidence interval for the normalized frequency of a reliable ensemble is shown as dashed lines in Figure 12. The figure shows that the ensemble reliability is different for different fields. In the rank diagram for SST (Figure 12a) there is no distinctive peak. Instead the rank diagram has an upward slope. This can be due to the negative bias in the ensemble (see Figure 11a). For the HFR observations, mid-range ranks are relatively more abundant than the tails (see Figure 12b). This indicates that either the spread in the ensemble is larger than the standard deviation of the HFR background errors or that the the observational error magnitude is overestimated. The opposite is the case for SSH observations. Here the U-shape (Figure 12c) implies that the forecast ensemble underestimates the magnitude of the background errors.

Finally, estimates for the background error and observational covariances used in the DA system are compared with estimates obtained from the innovation statistics. The relations between innovation statistics and error variances are given by (Desroziers et al., 2005):

$$\left\langle (\mathbf{y} - \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(0)}))_d^2 \right\rangle = (\mathbf{HMBM}^T\mathbf{H}^T + \mathbf{R})_{dd} \tag{26}$$

$$\left\langle (\mathbf{y} - \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(0)}))_d (\mathbf{y} - \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(0)}) - \mathbf{HMx}^{(0)})_d \right\rangle = (\mathbf{R})_{dd} \tag{27}$$

$$\left\langle (\mathbf{HMx}^{(0)})_d (\mathbf{y} - \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(0)}))_d \right\rangle = (\mathbf{HMBM}^T\mathbf{H}^T)_{dd} \tag{28}$$

where $1 \le d \le D$ is the index of the observation and $\langle\cdot\rangle$ denotes the expectation value, $(\cdot)_d$ the d-th element of the vector, and $(\cdot)_{dd}$ the d-th element on the diagonal of a matrix. The expectation values on the left-hand side of (26)-(28) are approximated by averaging over all observations of the same type in each window. These estimates are shown as blue lines in Figure 13 (where the top, middle, and bottom plots are for (26), (27), and (28) correspondingly). An approximation to the right-hand side of (27) is obtained by averaging $(\mathbf{R})_{dd}$ over all the observations of the same type. For the right-hand side of (28) an approximation is obtained by doing the same for

$$\begin{aligned} (\mathbf{B}_{ens})_{dd} &\stackrel{def}{=} (\mathbf{HMBM}^T\mathbf{H}^T)_{dd} = \tfrac{1}{M-2}\textstyle\sum_{m=1}^{M-1} (\mathbf{H}\mathcal{M}(\mathbf{x}_b^{(m)}) - \overline{\mathbf{H}\mathcal{M}(\mathbf{x}_b)})_d^2 \\ \overline{\mathbf{H}\mathcal{M}(\mathbf{x}_b)} &= \tfrac{1}{M-1}\textstyle\sum_{m=1}^{M-1} \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(m)}) \end{aligned} \tag{29}$$

Error standard deviations based on these estimates are displayed as dashed black lines in Figure 13.

Equations (26)-(28) only hold if $\mathbf{R}$ and $\mathbf{B}$ correctly represent the true observational and background error covariances. Figure 13 shows to what extend this is the case in our system. Figure 13g,h,i show that the ensemble error standard deviation (black line) grows over time for all three types of observations and hence that the wind perturbations are sufficient to prevent the ensemble spread from collapsing even without ensemble inflation. For SST the error standard deviation estimates from the innovation statistics are in agreement with the specified standard deviations (see Figure 13a,d,g). Error standard deviation estimates for HFR observations are consistent up to 1 July 2011 too. After 1 July 2011, however, the total error standard deviation estimate is too large (Figure 13b). The standard deviations for the observational errors agree (Figure 13e), so the overestimation is due to the fact that after 1 July the ensemble background error standard deviation estimate (black line Figure 13h) is larger than the standard deviation error estimate based on the forecast-observation differences (blue line in Figure 13h). This finding is consistent with the shape of the rank histogram in Figure 12b indicating overdispersion in the ensemble. Further investigation (not shown here) indicates that the difference between the estimates for background error standard deviation can be attributed nearly entirely to the sparse HFR observations taken far offshore (depth $> 2\,\mathrm{km}$). Closer to the shore (depth $< 1\,\mathrm{km}$), where numerous, closely-spaced, HFR observations are available for DA to reduce the background error, the ensemble estimates for the observational and background error standard deviations and those based on (26)-(28) show good agreement (Figure 13b, light blue/grey). Initially, the total SSH error standard deviation estimate from $\mathbf{R}$ and $\mathbf{B}_{ens}$ (black line in Figure 13c) is smaller than the total SSH error standard deviation from the innovations (blue line in Figure 13c). One could put forward the hypothesis that this is due to the fact that the standard deviation from $\mathbf{B}_{ens}$ (black line in Figure 13i) is smaller than the observational error estimate used in the DA (black line in Figure 13f) resulting in small SSH DA corrections. This would be a satisfactory explanation near the shore (depth $< 1\,\text{km}$) where ensemble estimates for $\hat{\mathbf{B}}$ and $\hat{\mathbf{B}} + \mathbf{R}$ remain nearly constant over time (Figure 13i, light blue/grey). However, in general the standard deviation from $\mathbf{B}_{ens}$ keeps increasing over time, growing beyond the specified observational error standard deviation of $2\,\text{cm}$, and (27) is not satisfied: the innovation statistics estimate (blue line in Figure 13f) for the observational error standard deviation continues to lie above the specified standard deviation (black line in Figure 13f). Either we have underestimated the observational error standard deviation while specifying $\mathbf{R}$ or the structure of the background errors is such that the system cannot remove them effectively.

6. Conclusions and Discussion

The development of ensemble-based 4DVAR systems has been one of the main focus areas in numerical weather prediction. Similarly, there is a rationale to applying ensemble-based 4DVAR systems for oceanic prediction. Utility of a static $\mathbf{B}$ can be limiting in shelf applications. The OR-WA forecast system, used in this study as a test ground for En4DVar, is a good example where model error statistics are influenced by high temporal and spatial ocean state variability. Before En4DVar can be applied successfully to the OR-WA or any other coastal system, many technical details must be worked out as outlined in this manuscript. The newly developed En4DVar systems need to go through statistical tests for self-consistency using actual observations, which help us understand the system behaviour and potential biases in the data.

Critical to a successful implementation of En4DVar for large prediction systems will be the development of time-efficient cost function minimization algorithms that take advantage of the massive parallel computer architectures. The cluster search method developed and tested in this study explores N_s search directions in parallel at each inner loop iteration. It was found that using a relatively small number of parallel direction computations, $N_s = 4$, can reduce the wall clock time by 30% compared to RBCG to achieve the same level of cost function reduction. It can be interesting to see in future studies whether combination of this method with saddle point algorithms (Rao and Sandu, 2016; Fisher and Gürol, 2017) can deliver an even better 4DVAR performance given the same limited number of cores.

Our system did not employ ensemble inflation (e.g. Anderson, 2001; Anderson and Anderson, 1999; Hamill et al., 2001), but generated background errors by perturbing the wind fields in the ensemble members in a realistic way. Although no comparison was made with ensemble inflation, and thus it cannot be concluded that wind perturbations are superior to ensemble inflation, no collapse of the ensemble spread was observed and therefore wind perturbations alleviated the need for ensemble inflation. The common assumption (Hénaff et al., 2009; Palmer et al., 2009; Vervatis et al., 2016) that the variance of the wind errors is proportional to the natural time-variability was found to be unrealistic. This possibly due to the atmospheric model being less able to represent small scales, or due to the inability to represent all possible small-scale error modes correctly with a very limited set of EOFs. Instead we found using a BHM that the wind errors increase in proportion to the natural variability as the spatial scale of the wind error decreases. Based on this, we agree with the findings of Whitaker and Hamill (2012) that additive inflation is more suitable to representing model errors, like the errors in wind forcing, than multiplicative inflation in which error variances are assumed to be proportional to the temporal variance in the signal.

Even though the En4DVar system was effective in reducing forecast errors compared to the case No DA, the rank diagram analysis suggests that the ensemble fails to represent the background error statistics perfectly: the ensemble overestimates the spread in the surface velocity background errors, while it underestimates the spread in the SSH background errors (see Figure 12). Although the rank diagrams in Figure 12 are not uniform, the bias and the maximum/minimum frequency-ratio of the diagrams is not exceptionally large compared to the rank diagram analyses in ensemble Kalman filter DA studies (e.g. Cookson-Hills et al., 2017; Fujita et al., 2007; Leeuwenburgh, 2007; Meng and Zhang, 2008).

This concludes the introduction of the new En4DVar system for the OR-WA system and evaluation of the error. In Part II of this study Pasmans et al. (in preparation) we will discuss if En4DVar yields better quality predictions than the traditional 4DVAR with balance operator background covariance currently used in the operational OR-WA system.

7. Acknowledgements

This work was supported by the National Oceanic and Atmospheric Administration (NOAA) Coastal Ocean Modelling Testbed (COMT) [grant number NA13NOS0120139], the NOAA Quantitative Observing System Assessment Program (QOSAP), Integrated Ocean Observing System / Northwest Association of Networked Ocean Observing Systems (IOOS/NANOOS) [grant number NA16NOS0120019], the National Aeronautics and Space Administration (NASA) SWOT Science Definition Team project [grant number NNX13AD89G]. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) [allocation TG-OCE160001], which is supported by the National Science Foundation [grant number ACI-1548562]. The views, opinions, and findings contained in this paper are those of the authors and should not be construed as an official NOAA or U.S. government position, policy, or decision. We would like to thank the four anonymous reviewers for their very motivating and insightful comments.

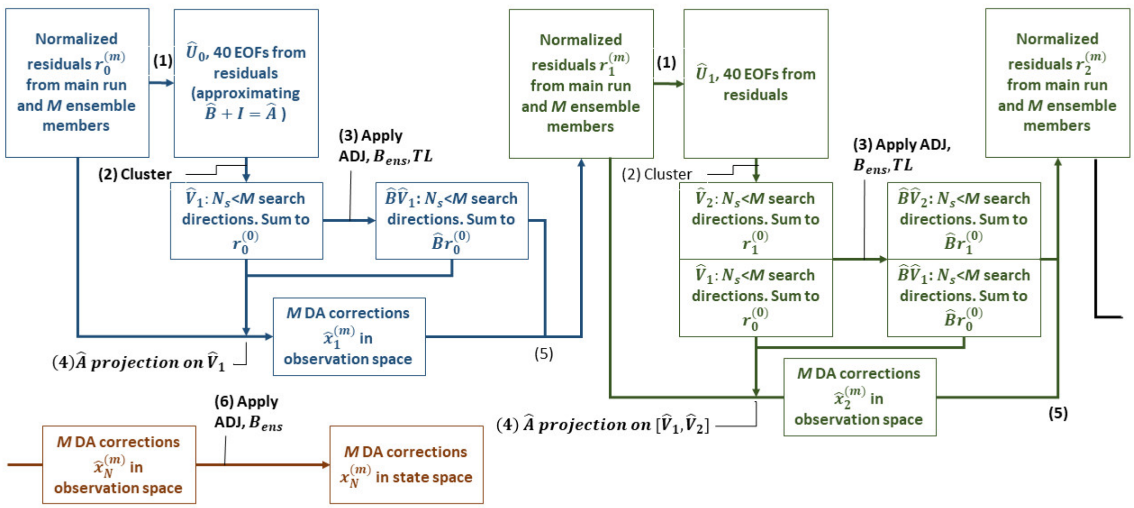

Figure 3: Overview of two (blue and green) inner loop iterations of the 4DVAR algorithm using the cluster search method. (1) For the control run and each ensemble member the residuals are normalized by the observation error standard deviation, (2) the EOF modes and singular values of the residuals are determined as they provide an estimate of $\hat{\mathbf{B}}$, (3) the EOF modes are combined using the cluster search algorithm into M' search directions such that the search direction sum to the residual of the control run, (4) the ADJ, $\mathbf{B}$, and TL are applied to each search direction, (5) for each ensemble member a solution in the shared search space is sought and new residuals are calculated (6) in the end the ADJ and $\mathbf{B}$ are applied to each solution, generating a correction in the model space.

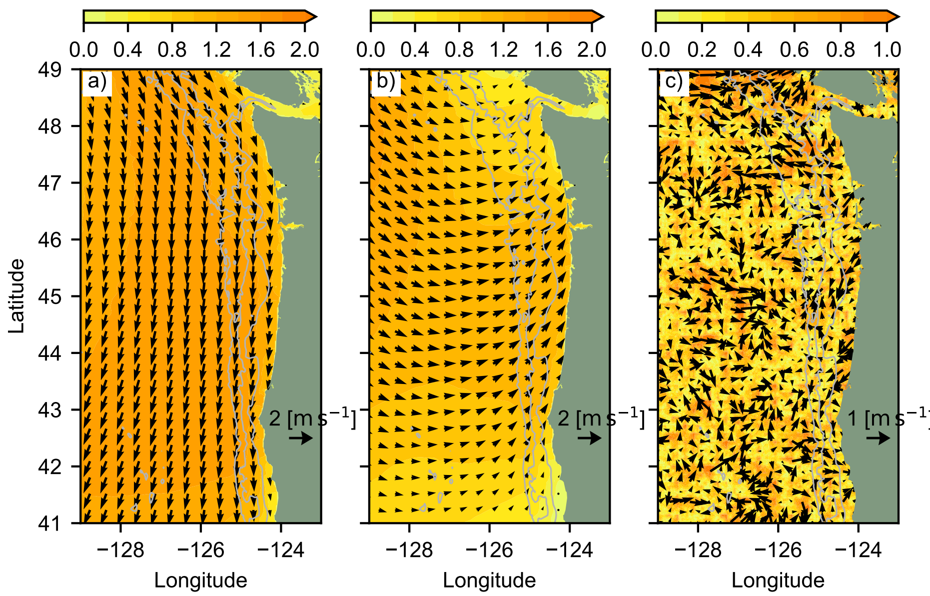

Figure 4: (a) $\hat{\sigma}_{EOF,1}\mathbf{w}_{EOF,1}$, first EOF of the wind field scaled by the standard deviation of the wind perturbations for this mode, (b) $\hat{\sigma}_{EOF,2}\mathbf{w}_{EOF,2}$ the second EOF of the wind field scaled by the standard deviation of the wind perturbations for this mode, (c) $\mathbf{w}_S$, example of a small-scale wind error field for one time. Colour scale shows the wind error speed of the different wind fields in $\mathrm{m\,s^{-1}}$.

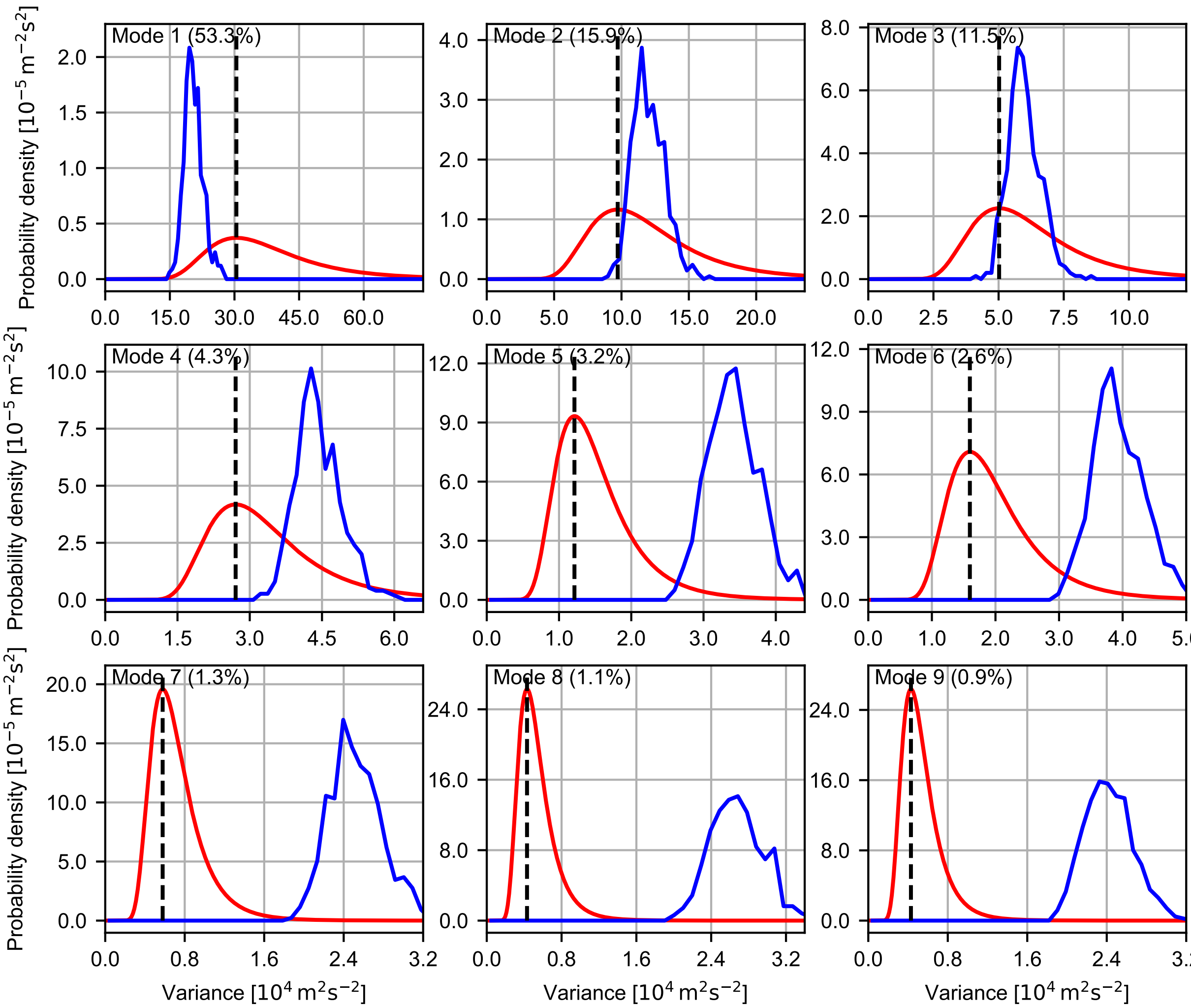

Figure 5: A priori (red) and posteriori (blue) probability density distributions for the variance of the EOF coefficients $\sigma^2_{EOF,i}$. The variance for which the blue distribution attains its maximum is used as estimate $\hat{\sigma}^2_{EOF,i}$. For reference, the values based on estimates used previously in literature, $\hat{\sigma}_{EOF,i}$ is 30% of the standard deviation in $\beta_{NAM,i}(t)$, is marked by the vertical, black dashed line.

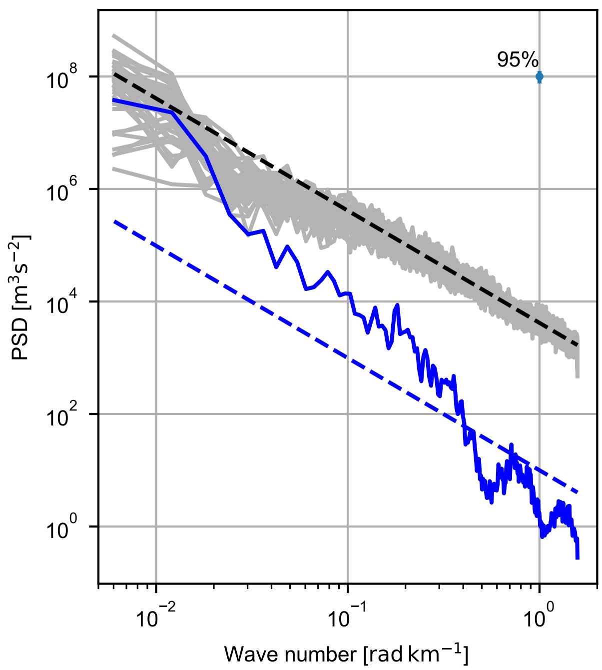

Figure 6: Power spectral density of the meridional wind field as a function of the meridional wave number from the interpolated NAM wind field (blue) and the perturbed wind fields of the ensemble members (grey) on 8 August 2011 00:00. Also shown are the fits to the mean of the ensemble power spectral densities (dashed black line) and the NAM power spectral density (dashed blue line) and the 95% confidence interval for the PSD estimates.

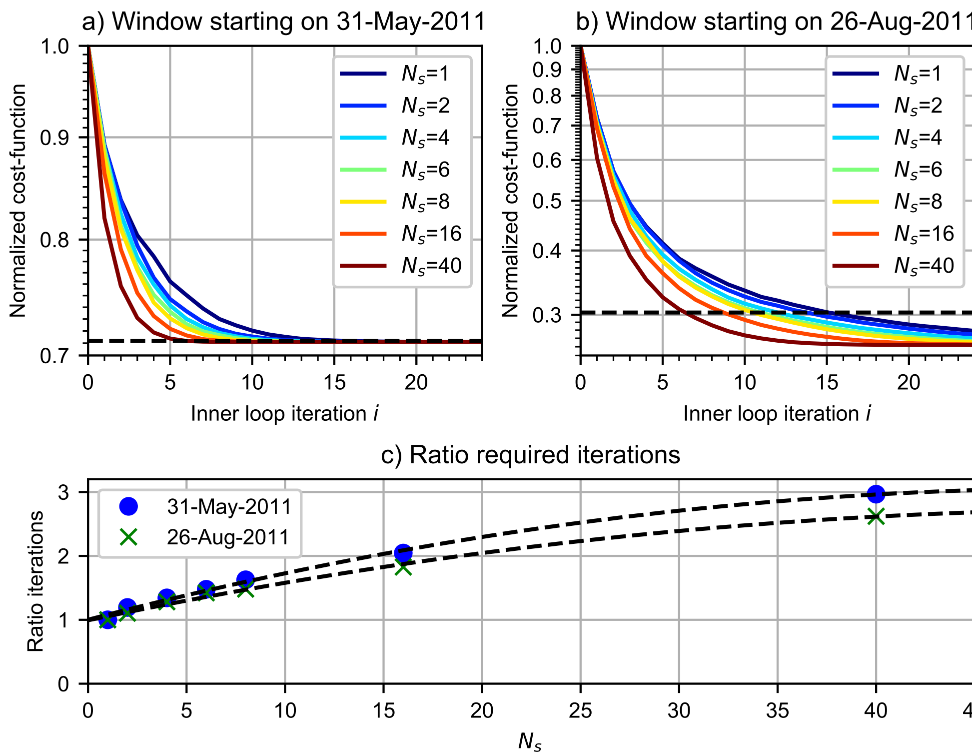

Figure 7: The cost function (5) versus the number of inner loop iterations using different levels of parallelisation N_s for the window starting (a) on 31 May 2011 and (b) on 26 August 2011. For reference, the value of the cost function after 12 inner loop iterations with $N_s = 4$ is marked by a dashed black line. The cost function values are normalized by the value before the start of 4DVAR. Panel (c) shows the speed-up ratio $a(N_s)$ (25).

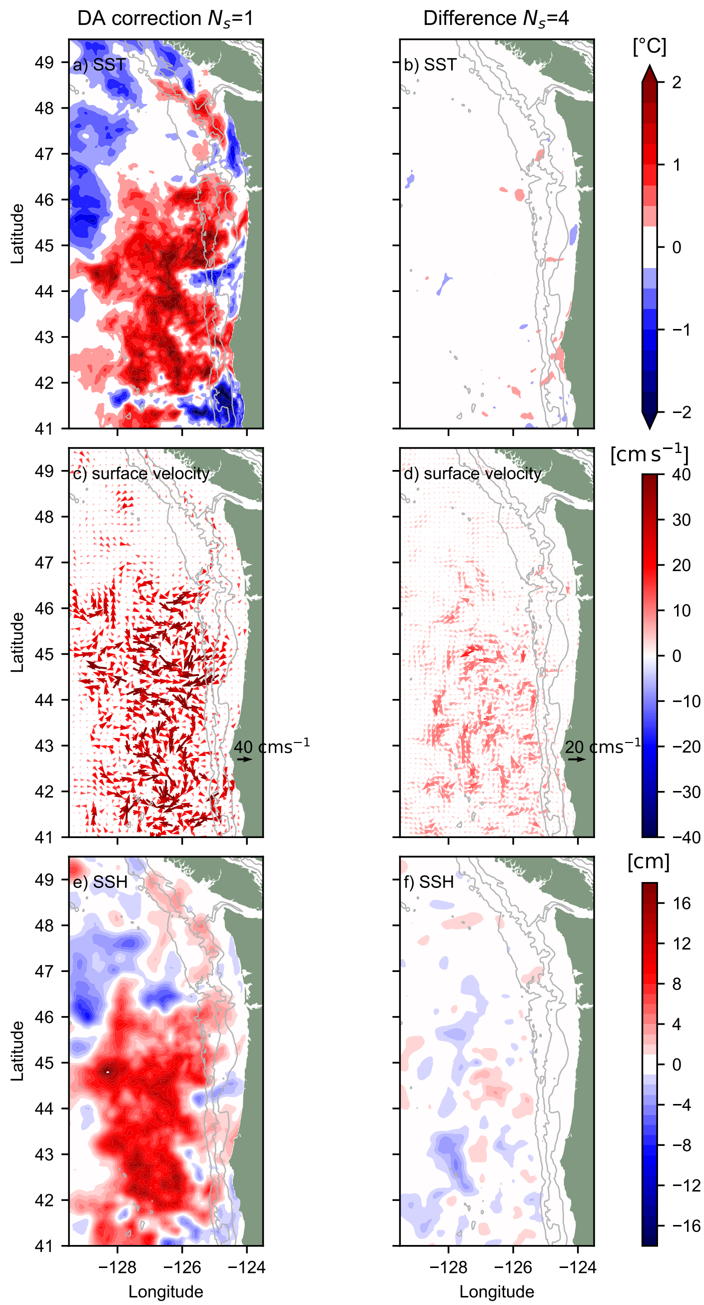

Figure 8: (Left) DA correction using $i = 12$ and $N_s = 1$ (RBCG method). (Right) The difference between the DA correction with $N_s = 4$ and $N_s = 1$. (Top) SST, (middle) surface velocity field (bottom) and SSH.

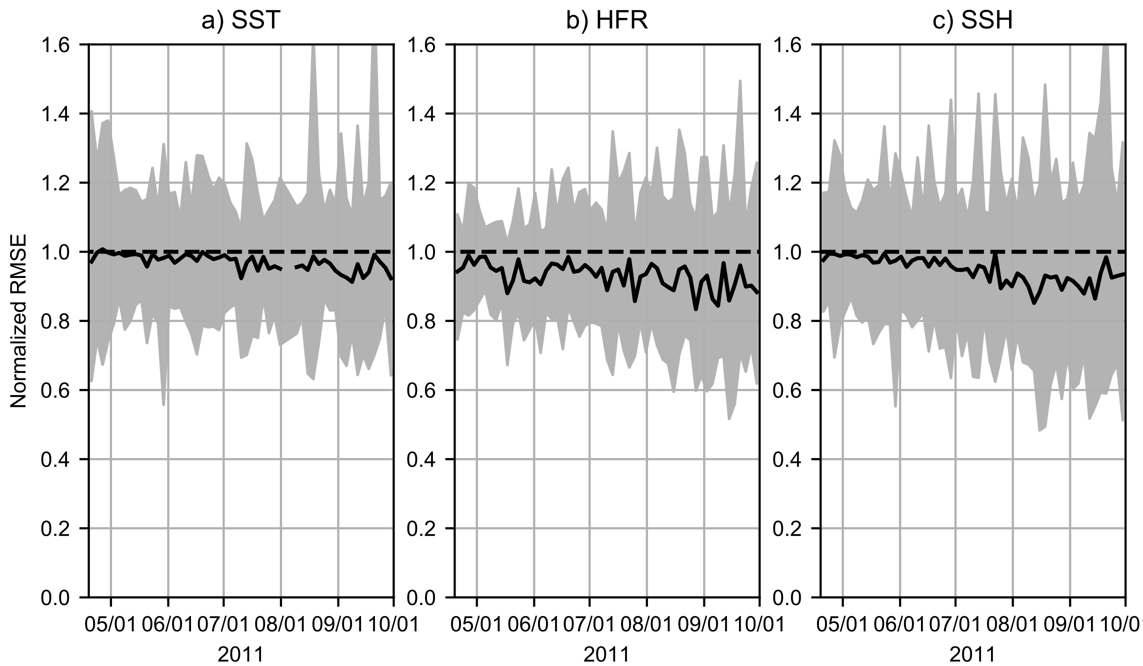

Figure 9: The normalized forecast error RMSE, as contained in vector $\hat{\mathbf{d}}$, after $\hat{\mathbf{B}}$-projection expressed as ratio versus the RMS of the total forecast error for (a) SST, (b) HFR and (c) SSH observations for each DA window. Grey area shows the range of this ratio over all ensemble members, while the black solid line marks the ensemble mean of the ratio. Dashed black line marks the value of 1 that would be obtained if the innovation vector $\hat{\mathbf{d}}$ of the ensemble lies completely in the search space.

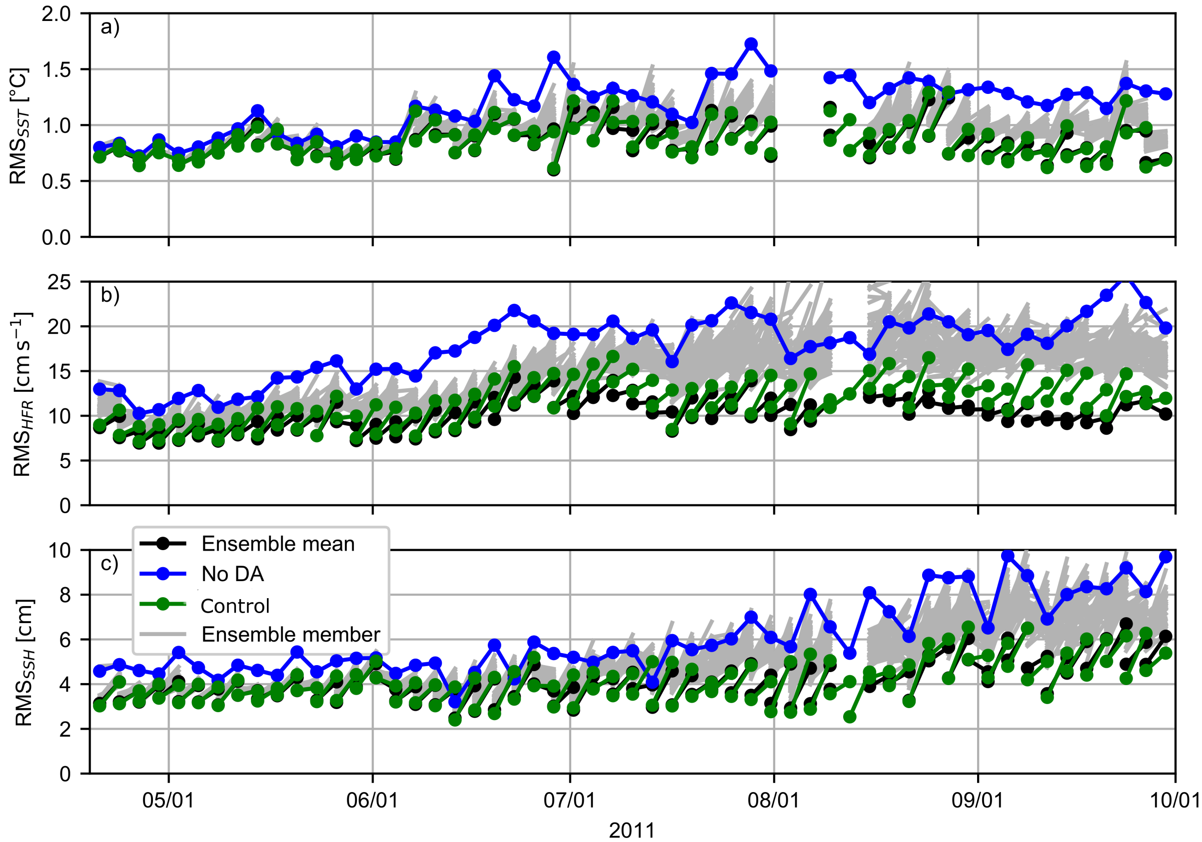

Figure 10: RMSE per 3-day window for (a) SST, (b) HFR daily-averaged velocity and (c) SSH observations from the model without DA (blue), the Control run (green), the different ensemble members (grey) and the ensemble average (black). The left side of each line piece marks the RMSE in the analysis, the right side the RMSE in consecutive the forecast.

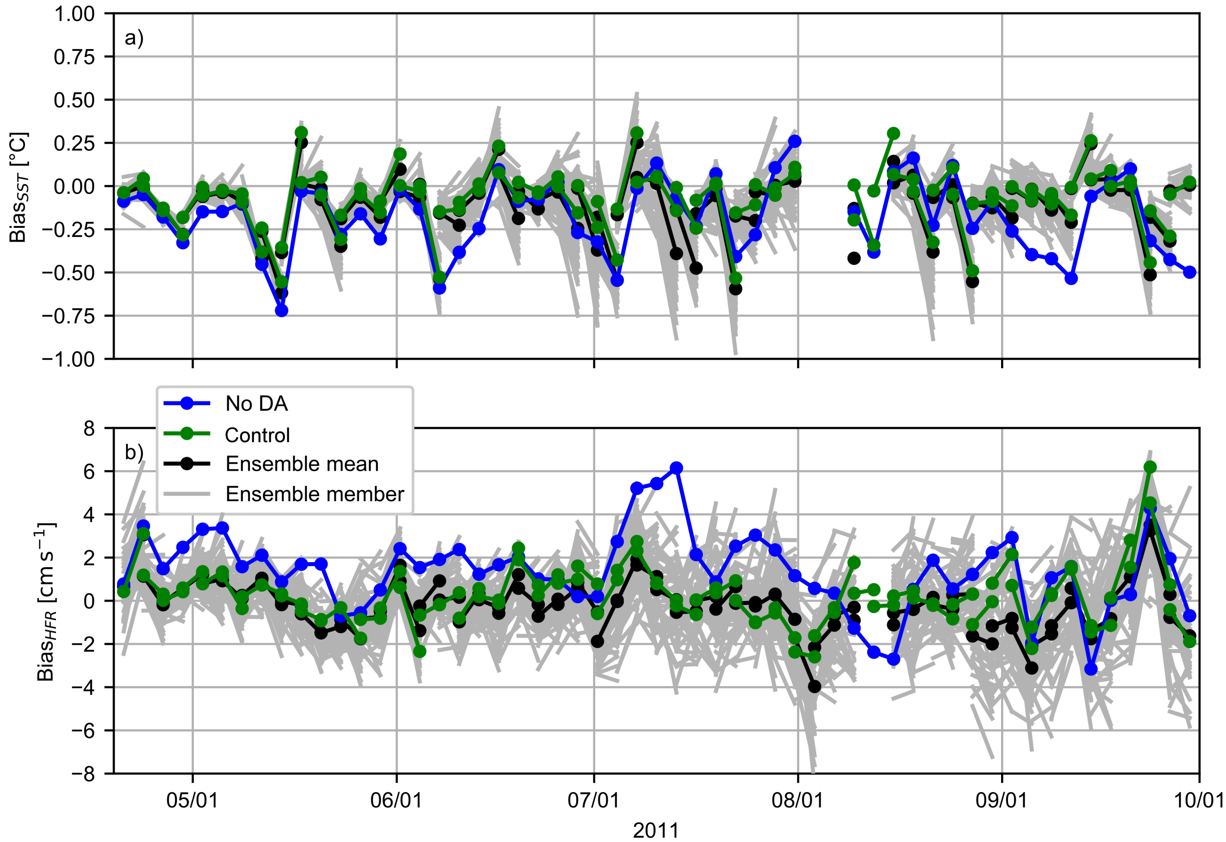

Figure 11: The 3-day model bias (model-observations) per window for (a) SST and (b) HFR daily-averaged velocity observations from the model without DA (blue), the control run (green), the different ensemble members (grey) and the ensemble average (black). The left side of each line piece marks the bias in the analysis, the right side the bias in the forecast.

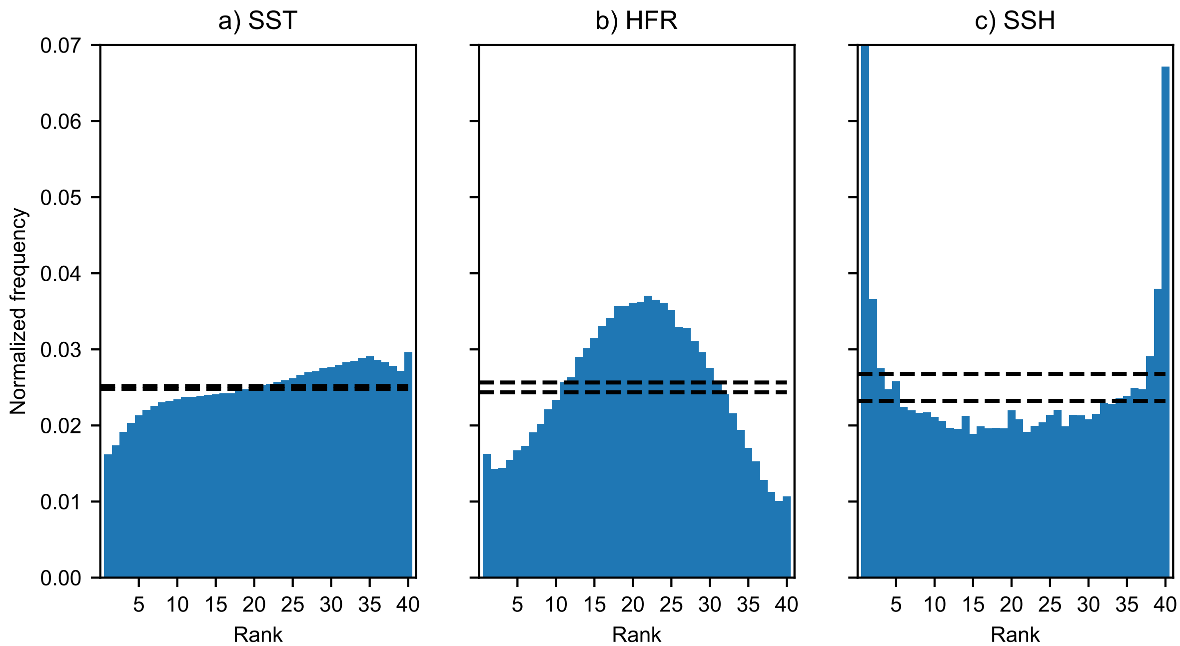

Figure 12: Rank of the perturbed (a) SST, (b) HFR daily-averaged surface velocity and (c) SSH observations within the ensemble.

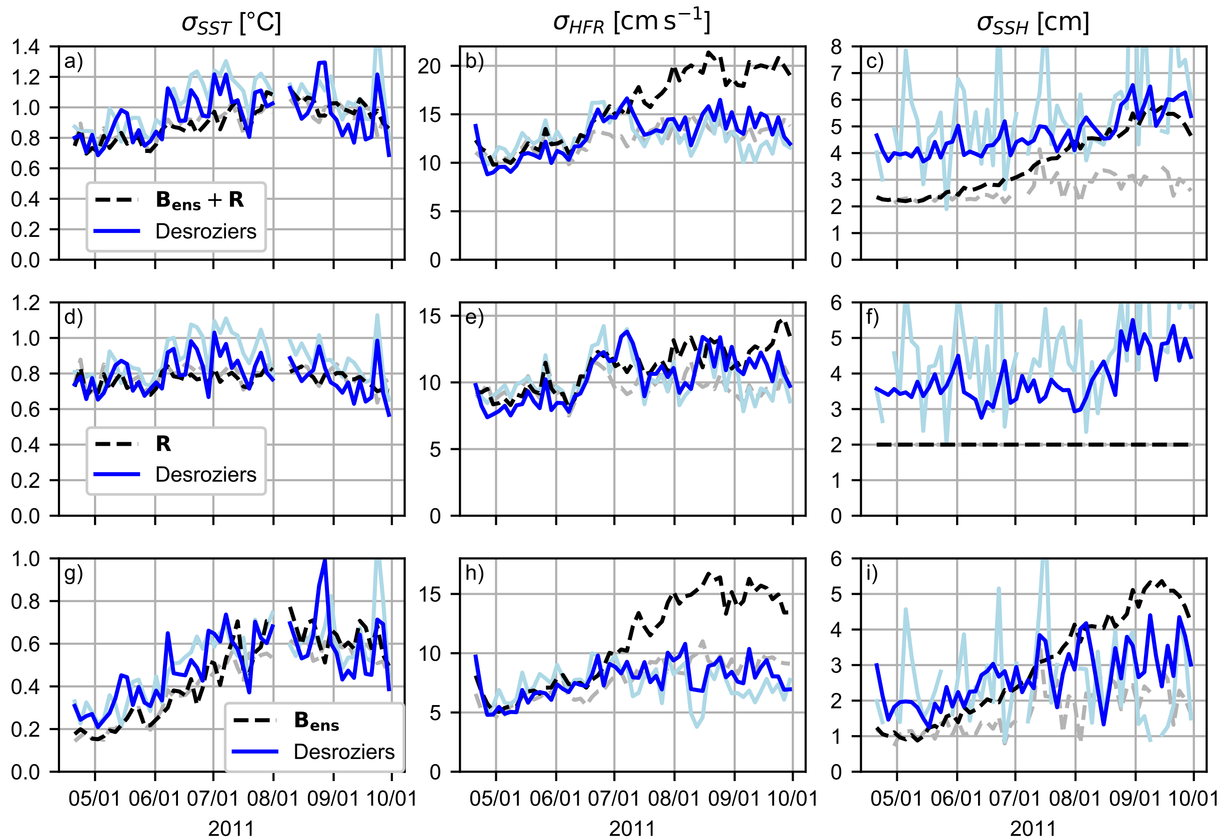

Figure 13: Estimates for total error variance (first row), observational error variance (centre row) and background error variance (bottom row) based on SST observations (left column), HFR observations (centre column) and SSH observations (right column). Blue lines are estimates based on (26)-(28), while the black lines indicate estimates based on the specified $\mathbf{R}$ and the ensemble variance. Same for the light blue (grey) lines respectively, but now using only observations taken in water of less than 1 km depth.

Appendix A. SSH observations

Let R be the track number of a set of SSH observations. A single pass of a SSH satellite through the domain takes at most several minutes. Therefore all SSH observations during this pass are assumed to have been made at the same time t. Here t is chosen to be the mean of the observation times during the pass. Let $\mathbf{x}_k(R,t)$ be the location of the k-th observation of the SSH along-track R at time t. Define

$$\begin{aligned} \zeta_k'(R,t) &= SSHA(\mathbf{x}_k(R,t),t) + MDT(\mathbf{x}_k(R,t)) \\ &+ \textstyle\sum_{l=1}^{8} T^{-1} \int_{t_0}^{t_1} A_l(\mathbf{x}_k(R,t)) \cos[\omega_l \tau - \phi_l(\mathbf{x}_k(R,t))]\, d\tau \end{aligned} \tag{A.1}$$

Here $SSHA(\mathbf{x}_k(R,t),t)$ and $MDT(\mathbf{x}_k(R,t))$ are respectively the detided sea-surface height anomaly and mean dynamic topography at location $\mathbf{x}_k(R,t)$ and time t as provided by the SSH satellite data provider, ω_l, $A_l(\mathbf{x}_k(R,t)$ and $\phi_l(\mathbf{x}_k(R,t))$ are the angular frequency, amplitude and the phase of the l-th tidal component at location $\mathbf{x}_k(R,t)$. A_l and ϕ_l are estimated for the M2, S2, N2, K2, K1, O1, P1, Q1 tide by regression from the No DA model run using T_TIDE (Pawlowicz et al., 2002). t_0 is the maximum of $t - 12\text{h}$ and the begin of the current DA window, t_1 is the minimum of $t + 12\,\text{h}$ and the end of the current DA window and $T = t_1 - t_0$. Then the SSH observation provided to the DA system at time t and location $\mathbf{x}_k(R,t)$ is:

$$\zeta_k(R,t) = \zeta_k'(R,t) - K^{-1} \sum_j \zeta_j'(R,t) \tag{A.2}$$

with K the total number of SSH observations in track R at time t.

The innovation corresponding to this observation is then calculated as

$$\begin{aligned} d &= \zeta_k(R,t) - [\zeta_{model}(\mathbf{x}_k(R,t),t) - T^{-1} \textstyle\int_{t_0}^{t_1} \zeta_{model}(\mathbf{x}_k(R,t),\tau)\, d\tau] \\ &+ K^{-1} \textstyle\sum_j [\zeta_{model}(\mathbf{x}_j(R,t),t) - T^{-1} \int_{t_0}^{t_1} \zeta_{model}(\mathbf{x}_j(R,t),\tau)\, d\tau] \end{aligned} \tag{A.3}$$

By applying this procedure, we attempt to correct a non-tidal SSH slope along each track, but not the average level.

Appendix B. List of Symbols

Symbol	Equivalent	Meaning
$\|\| \cdot \|\|_E$	$\|\|\mathbf{y}\|\|_E = \sqrt{\mathbf{y}^T \hat{\mathbf{A}} \mathbf{y}}$	Energy norm associated with the symmetric, strictly positive definite matrix $\hat{\mathbf{A}}$.
$A_l(\mathbf{x})$		Amplitude of the l-th tidal component at location $\mathbf{x}$.
$\hat{\mathbf{A}}$	$\hat{\mathbf{B}} + \mathbf{I}$	Covariance between the total error in the normalized observations (see $\hat{\mathbf{d}}$).
$a_{EOF,i}$		Shape parameter in the inverse gamma distribution for $\sigma^2_{EOF,i}$.
$a(N_s)$		Ratio of the number of RBCG inner loop iterations necessary to reach the same amount of cost-function reduction as in the cluster search method with N_s clusters.
a_S		Shape parameter in the inverse gamma distribution for σ^2_S.
$\mathbf{B}$		Background error covariance. Covariance between the background errors in the initial condition.
$\hat{\mathbf{B}}$	$\mathbf{R}^{-1/2}\mathbf{HMBM}^T\mathbf{H}^T\mathbf{R}^{-1/2}$	Covariance between the background errors in the normalized observations (see $\hat{\mathbf{d}}^{(m)}$).
$b_{EOF,i}$		Scale parameter in the inverse gamma distribution for $\sigma^2_{EOF,i}$.
b_S		Scale parameter in the inverse gamma distribution for $\sigma^2_{S,i}$.
c_β	0.4	AR1-process parameter used in generating large-scale wind error EOF mode expansion coefficient $\beta_{L,i}$.

Table B.2: List of symbols used in the text.

Symbol	Equivalent	Meaning
c_β	0.4	AR1-process parameter used in generating small-scale wind error Debauchies wavelet coefficient $\gamma_i^{(n)}$.
D		Number of observations; dimension of the observation space.
D_n		Set containing the indices of the singular values that are assigned to the n-th cluster.
$\hat{\mathbf{d}}^{(m)}$	$\mathbf{R}^{-1/2}\mathbf{y} - \mathbf{R}^{-1/2}\mathbf{H}\mathcal{M}(\mathbf{x}_b^{(m)})$	Innovations from the m-th ensemble run normalized by the observational error standard deviations.
$\hat{\mathbf{D}}$	$[\hat{\mathbf{d}}^{(1)}, \hat{\mathbf{d}}^{(2)}, \ldots, \hat{\mathbf{d}}^{(M-1)}]$	Matrix containing normalized ensemble innovation vectors as columns.
$\mathbf{H}$		Linear sampling operator generating model predictions that can be compared with the observations.
$\mathbf{I}$		Identity matrix.
$J(\mathbf{x}^{(m)})$	$\frac{1}{2}\mathbf{x}^{(m)T}\mathbf{B}^{-1}\mathbf{x}^{(m)} + \frac{1}{2}\|\|\hat{\mathbf{d}} - \mathbf{R}^{-1/2}\mathbf{H}\mathbf{M}\hat{\mathbf{x}}^{(m)}\|\|^2$	4DVAR cost-funtion for the initial condition correction $\mathbf{x}^{(m)}$.
$\mathcal{K}_i(\mathbf{d}, \hat{\mathbf{A}})$	$span(\mathbf{d}, \hat{\mathbf{A}}\mathbf{d}, \ldots, \hat{\mathbf{A}}^{i-1}\mathbf{d})$	i-th Krylov space generated by $\hat{\mathbf{A}}$ starting from vector $\hat{\mathbf{d}}$.
M		Number of parallel nonlinear model runs, i.e. control run plus $M-1$ ensemble members.

Table B.3: List of symbols used in the text. (continued)

Symbol	Equivalent	Meaning
$\mathcal{M}(\mathbf{x}_b)$		Nonlinear ROMS model propagating initial condition $\mathbf{x}_b$ to all model output times.
$\mathbf{M}$		Tangent linear model linearised around the background from the control run.
$\mathbf{M}^{(m)}$		Tangent linear model linearised around the background from the m-th ensemble member.
$\mathbf{M}^T$		Adjoint model.
N		Dimension of the model space.
N_s		Number of clusters used in the cluster search method.
$\mathbf{P}_n$		Projection operator setting the elements of a vector of which the indices are not contained in D_n to zero.
$\mathbf{R}$		Diagonal observational error covariance matrix.
$\mathbf{s}^{(n)}$		n-th search vector added in the current inner loop iteration with $1 \leq n \leq N_s$. Varies per inner loop iteration.
$\mathbf{T}$	$(\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{A}} \hat{\mathbf{V}}_i$	$\hat{\mathbf{B}}$-projection of $\hat{\mathbf{A}}\hat{\mathbf{V}}$ onto the search space.
$\hat{\mathbf{U}}_0$		Eigenvectors of the matrix $\hat{\mathbf{D}}\hat{\mathbf{D}}^T$.
$\hat{\mathbf{U}}_i$	$\hat{\mathbf{U}}_0 - \hat{\mathbf{A}} \hat{\mathbf{V}}_i \mathbf{T}_i^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_0$	Difference between the eigenvectors (in $\hat{\mathbf{U}}_0$) and the approximation of those eigenvectors in the search space.

Table B.4: List of symbols used in the text (continued).

Symbol	Equivalent	Meaning
$\hat{\mathbf{v}}'$	$\hat{\mathbf{U}}_0^T\hat{\mathbf{r}}_0^{(0)}$	Expansion of the initial 4DVAR residual ($\hat{\mathbf{d}}^{(0)}$) with respect to the column vectors of $\hat{\mathbf{U}}_0$.
$\hat{\mathbf{v}}''$		Expansion of the i-th residual $\hat{\mathbf{r}}_i^{(0)}$ with respect to the column vectors of $\hat{\mathbf{U}}_i$.
$\hat{\mathbf{V}}_i$		Matrix with all search vectors used in the i-th inner loop iteration as columns.
$\mathbf{w}_{EOF,i}$		i-th empirical orthogonal functions obtained from the time series of NAM model wind fields.
$\mathbf{w}_L(t)$		Large-scale spatial wind error field at time t.
$\mathbf{w}_{NAM}(t)$		Vector containing the North American Mesoscale (NAM) model wind field at time t interpolated onto the ROMS model grid.
$\mathbf{w}_{obs}(t)$		Vector containing the ASCAT scatterometer and NOAA buoy wind observations at time t.
$\mathbf{w}_S(t)$		Small-scale spatial wind error field at time t.
$\mathbf{x}^{(m)}$	$\mathbf{x}^{(m)} = \mathbf{B}\mathbf{M}^T\mathbf{R}^{-1/2}\hat{\mathbf{x}}^{(m)}$	4DVAR correction to the initial conditions in the model space. I.e. correction that minimizes cost-function J.
$\hat{\mathbf{x}}^{(m)}$		4DVAR correction to the initial conditions before mapping to the model space.
$\hat{\mathbf{x}}_i^{(m)}$		Approximation to $\hat{\mathbf{x}}^{(m)}$ obtained after the i-th inner loop iteration.

Table B.5: List of symbols used in the text (continued).

Symbol	Equivalent	Meaning
$\mathbf{x}_{ana}^{(m)}$	$\mathbf{x}_b^{(m)} + \hat{\mathbf{x}}^{(m)}$	Ocean model state vector at the beginning of the DA window after the 4DVAR correction $\mathbf{x}^{(m)}$ has been applied.
$\mathbf{x}_b^{(m)}$		Initial condition for the DA window prior to application of the 4DVAR correction for ensemble member m.
$\hat{\mathbf{X}}_i$	$[\hat{\mathbf{x}}_i^{(0)}, \hat{\mathbf{x}}_i^{(1)}, \ldots, \hat{\mathbf{x}}_i^{(M-1)}]$	Matrix having the approximation to $\hat{\mathbf{x}}^{(m)}$ obtained after the i-th inner loop iteration as $m+1$-th column.
$\mathbf{y}$		Vector having the observed values (SST,SSH,HFR) as elements.
α_n		Coefficient in front of $\mathbf{s}^{(n)}$ in $\hat{\mathbf{x}}_{i+1}^{(0)} - \hat{\mathbf{x}}_i^{(0)}$. Changes in each inner loop iteration.
$\hat{\alpha}_n$		Estimator of α_n that miminizes the error in estimate $\hat{\mathbf{x}}_{i+1}^{(0)}$ in the energy norm, i.e. value that minimizes $\lvert\lvert\hat{\mathbf{x}}^{(0)} - \hat{\mathbf{x}}_i^{(0)} - \sum_{n=1}^{N_s} \alpha_n \mathbf{s}^{(n)}\rvert\rvert_E$
$\beta_{L,i}(t)$		EOF expansion coefficient for the i-th EOF mode in the large-scale wind error at time t.
$\beta_{NAM,i}$		EOF expansion coefficient for the i-th EOF mode in the NAM model wind fields at time t.
$\gamma_i^{(n)}(t)$		Coefficient for the i-th Debauchie wavelet at the n-th level in the small-scale wind error field at time t.
$\epsilon_{obs}(t)$		Observational error in the wind velocity $\mathbf{w}_{obs}(t)$ observations.

Table B.6: List of symbols used in the text (continued).

Symbol	Equivalent	Meaning
ζ		Sea-surface height
$\hat{\mathbf{\Lambda}}$		Diagonal matrix with eigenvalues of $\hat{\mathbf{A}}$ on the diagonal.
λ_d		d-th eigenvalue of matrix $\hat{\mathbf{A}}$.
Ξ		Diagonal maxtrix having ξ_d on its diagonal.
ξ_d		Ratio of the norm of the d-th column of $\hat{\mathbf{U}}_0$ over the norm of the d-th column of $\hat{\mathbf{U}}_i$.
$\sigma^2_{EOF,i}$		Variance in $\beta_{L,i}(t)$, equal to the variance of the i-th EOF mode in the time-series for the large-scale wind errors.
$\hat{\sigma}^2_{EOF,i}$		Estimator for $\sigma^2_{EOF,i}$.
σ^2_{obs}		Variance of the observational error in the ASCAT and NOAA buoy wind observations.
σ^2_S		Variance of the small-scale wind errors, i.e. the elements of $\mathbf{w}_S(t)$.
$\phi_l(\mathbf{x})$		Phase of tidal component l at location $\mathbf{x}$.
ω_l		Angular frequency of the l-th tidal components.

Table B.7: List of symbols used in the text (continued).

Appendix C. Pseudocode Minimization Algorithms

Table C.8: Pseudocode for RBCG using (6) for ensemble member m.

$\hat{\mathbf{V}}_0 = [\,]$; $\hat{\mathbf{W}}_0 = [\,]$; $\hat{\mathbf{r}} = \hat{\mathbf{d}}^{(m)}$
for $i = [1 : I]$
$\quad\hat{\mathbf{V}}_i = [\hat{\mathbf{V}}_{i-1}, \hat{\mathbf{r}}]$
$\quad$Use AVRORA TL-ADJ to calculate $\hat{\mathbf{W}}_i = [\hat{\mathbf{W}}_{i-1}, \hat{\mathbf{B}}\hat{\mathbf{r}}^{(m)}]$
$\quad\hat{\mathbf{A}}\hat{\mathbf{V}}_i = \hat{\mathbf{W}}_i + \hat{\mathbf{V}}_i$
$\quad\mathbf{T} = (\hat{\mathbf{V}}_i^T \hat{\mathbf{W}}_i)^{-1} \hat{\mathbf{W}}_i^T (\hat{\mathbf{A}}\hat{\mathbf{V}}_i)$
$\quad\mathbf{p} = \mathbf{T}^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{W}}_i)^{-1} \hat{\mathbf{W}}^T \hat{\mathbf{d}}^{(m)}$
$\quad\hat{\mathbf{x}}^{(m)} = \hat{\mathbf{V}}_i \mathbf{p}$ and $\hat{\mathbf{B}}\hat{\mathbf{x}}^{(m)} = \hat{\mathbf{W}}_i \mathbf{p}$
$\quad\hat{\mathbf{r}}^{(m)} = \hat{\mathbf{d}}^{(m)} - \hat{\mathbf{B}}\hat{\mathbf{x}}^{(m)} - \hat{\mathbf{x}}^{(m)}$
end for
Use AVRORA ADJ to calculate correction $\mathbf{x}^{(m)} = \mathbf{B}\mathbf{M}^T \mathbf{R}^{-1/2} \hat{\mathbf{x}}^{(m)}$

Table C.9: Pseudocode for block diagonal CG with B-preconditioning using (6).

Initialize $\hat{\mathbf{V}}_0 = [\,]$ and $\hat{\mathbf{W}}_0 = [\,]$
Set $\mathbf{D} = [\hat{\mathbf{d}}^{(0)}, \hat{\mathbf{d}}^{(1)}, \ldots, \hat{\mathbf{d}}^{(M-1)}]$ and $\mathbf{S} = \hat{\mathbf{D}}$

for $i = [1 : I]$

  $\hat{\mathbf{V}}_i = [\hat{\mathbf{V}}_{i-1}, \mathbf{S}]$

  parfor $m = [1 : M]$

    Take $\hat{\mathbf{r}}$ to be the m-th column vector of $\mathbf{S}$

    Use AVRORA TL-ADJ to calculate $\mathbf{w}^{(m-1)} = \hat{\mathbf{B}}\mathbf{r}$

  end parfor

  $\hat{\mathbf{W}}_i = [\hat{\mathbf{W}}_{i-1},\ \mathbf{w}^{(0)}, \mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(M-1)}]$

  $\hat{\mathbf{A}}\hat{\mathbf{V}}_i = \hat{\mathbf{W}}_i + \hat{\mathbf{V}}_i$

  $\mathbf{T} = (\hat{\mathbf{V}}_i^T \hat{\mathbf{W}}_i)^{-1} \hat{\mathbf{W}}_i^T (\hat{\mathbf{A}}\hat{\mathbf{V}}_i)$

  $\mathbf{P} = \mathbf{T}^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{W}}_i)^{-1} \hat{\mathbf{W}}^T \hat{\mathbf{D}}$

  $\hat{\mathbf{X}} = \hat{\mathbf{V}}_i \mathbf{P}$ and $\hat{\mathbf{B}}\hat{\mathbf{X}} = \hat{\mathbf{W}}_i \mathbf{P}$

  $\mathbf{S} = \hat{\mathbf{D}} - \hat{\mathbf{B}}\hat{\mathbf{X}} - \hat{\mathbf{X}}$

end for

parfor $m = [0 : M-1]$

  Let $\hat{\mathbf{x}}^{(m)}$ be the $m+1$-th column of $\hat{\mathbf{X}}$

  Use AVRORA ADJ to calculate correction $\mathbf{x}^{(m)} = \mathbf{B}\mathbf{M}^T \mathbf{R}^{-1/2} \hat{\mathbf{x}}^{(m)}$

end parfor

Table C.10: Pseudocode for the cluster search method. $\mathbf{e}_j$ is a unit vector in direction j.

Initialize $\hat{\mathbf{V}}_0 = [\,]$ and $\hat{\mathbf{W}}_0 = [\,]$

Set $\mathbf{D} = [\hat{\mathbf{d}}^{(0)}, \hat{\mathbf{d}}^{(1)}, \ldots, \hat{\mathbf{d}}^{(M)}]$ and $\mathbf{r} = \hat{\mathbf{d}}^{(0)}$

Use SVD to decompose $\frac{1}{\sqrt{M}}\mathbf{D} = \hat{\mathbf{U}}\boldsymbol{\Lambda}\mathbf{Z}^T$. Discard $\mathbf{Z}$

Set $\hat{\mathbf{U}}' = \hat{\mathbf{U}}$

Calculate $\hat{\mathbf{v}}'' = \hat{\mathbf{U}}^T\hat{\mathbf{d}}^{(0)}$

---

for $i = [1 : I]$

  for $m = [1, 2, \ldots, M]$

    $\xi_m = (\hat{\mathbf{U}}\mathbf{e}_m)^T(\hat{\mathbf{U}}'\mathbf{e}_m)$

  end for

  $\hat{\mathbf{v}}' = \hat{\mathbf{U}}^T\mathbf{r}$

  Get $D_1, D_2, \ldots, D_{N_s} = \text{func\_cluster}(\Lambda, \mathbf{v}', \mathbf{v}'', \xi)$

  for $n = [1 : N_s]$

    $\mathbf{s}^{(n)} = \sum_{j \in D_n}(\hat{\mathbf{U}}'\mathbf{e}_j)(\mathbf{e}_j^T\mathbf{v}'')$

  end for

  $\hat{\mathbf{V}}_i = [\hat{\mathbf{V}}_{i-1}, \mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(N_s)}]$

  parfor $n = [1 : N_s]$

    Use AVRORA TL-ADJ to calculate $\mathbf{w}^{(n)} = \hat{\mathbf{B}}\mathbf{s}^{(n)}$

  end parfor

  $\hat{\mathbf{W}}_i = [\hat{\mathbf{W}}_{i-1},\ \mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \ldots, \mathbf{w}^{(N_s)}]$

  $\hat{\mathbf{A}}\hat{\mathbf{V}}_i = \hat{\mathbf{W}}_i + \hat{\mathbf{V}}_i$

  $\mathbf{T} = (\hat{\mathbf{V}}_i^T\hat{\mathbf{W}}_i)^{-1}\hat{\mathbf{W}}_i^T(\hat{\mathbf{A}}\hat{\mathbf{V}}_i)$

  $\mathbf{P} = \mathbf{T}^{-1}(\hat{\mathbf{V}}_i^T\hat{\mathbf{W}}_i)^{-1}\hat{\mathbf{W}}^T\hat{\mathbf{D}}$ and $\mathbf{Q} = \mathbf{T}^{-1}(\hat{\mathbf{V}}_i^T\hat{\mathbf{W}}_i)^{-1}\hat{\mathbf{W}}^T\hat{\mathbf{U}}$

  $\hat{\mathbf{X}} = \hat{\mathbf{V}}_i\mathbf{P}$ and $\hat{\mathbf{B}}\hat{\mathbf{X}} = \hat{\mathbf{W}}_i\mathbf{P}$

  $\mathbf{U}' = \hat{\mathbf{U}} - \hat{\mathbf{W}}_i\mathbf{Q} - \hat{\mathbf{V}}_i\mathbf{Q}$

  $\mathbf{r} = \mathbf{d}^{(0)} - (\hat{\mathbf{B}}\hat{\mathbf{X}} + \hat{\mathbf{X}})\mathbf{e}_1$

end for

Table C.10: Pseudocode for the cluster search method. $\mathbf{e}_j$ is a unit vector in direction j (continued).

parfor $m = [0 : M-1]$

Let $\hat{\mathbf{x}}^{(m)}$ be the $m+1$-th column of $\hat{\mathbf{X}}$

Use AVRORA to calculate correction $\mathbf{x}^{(m)} = \mathbf{B}\mathbf{M}^T\mathbf{R}^{-1/2}\hat{\mathbf{x}}^{(m)}$

end parfor

---

function $[D_1, D_2, \ldots, D_{N_s}] =$ **func_cluster**$(\Lambda, \hat{\mathbf{v}}', \hat{\mathbf{v}}'', \xi)$

Calculate weight $w_m = {v''}_m^2 \xi_m^2 \lambda_m^4 (1 - \lambda_m^{-2})$ with $\lambda_m = (\Lambda)_{mm}$, $\xi_m = \mathbf{e}_m^T \xi$, and $v''_m = \mathbf{e}_m^T \mathbf{v}''$.

Set $z_m = \frac{v'_m}{v''_m \xi_m \lambda_m^2}$ with $v'_m = \mathbf{e}_m^T \mathbf{v}'$

for $m = [1 : N_s]$

$\overline{z}_m = \min(z) + \frac{1}{N_s}(m - \frac{1}{2})(\max(z) - \min(z))$

end for

for $j = [1 : 1000]$

(Re)initialize $D_1, D_2, \ldots, D_{N_s} = [\,]$

for $m = [1 : M]$

Find l that minimizes $|\overline{z}_l - z_m|$

$D_l = [D_l, m]$

end for

for $m = [1 : N_s]$

$\overline{z}_m = (\sum_{l \in D_m} w_m z_m)(\sum_{l \in D_m} w_m)^{-1}$

end for

end function

Appendix D. Estimation Conditional Distribution $\sigma_{EOF,i}$

The integral on the right-hand side of the conditional probability distribution $\sigma^2_{EOF,i}$ (19)

$$\begin{aligned} &p(\sigma^2_{EOF,i}|\underline{w}_{obs}) = \\ &\int p(\underline{\mathbf{w}}_S, \underline{\beta}_L, \underline{\epsilon}_{obs}, \sigma^2_{EOF,1}, \ldots, \sigma^2_{EOF,N_{EOF}}, \sigma^2_S|\underline{w}_{obs})\, d\underline{\mathbf{w}}_S \\ &\times d\underline{\beta}_L d\underline{\epsilon}_{obs} \textstyle\prod_{j=1,j\neq i}^{N_{EOF}} d\sigma^2_{EOF,j} d\sigma^2_S \end{aligned} \tag{D.1}$$

is approximated by drawing 500 samples of distribution (19) with $\mathbf{s} = (\underline{\mathbf{w}}_S, \underline{\beta}_L, \sigma^2_{\beta_1}, \ldots, \sigma^2_{\beta_{N_{EOF}}}, \sigma^2_S)$ from distribution (19) $p(\mathbf{s})$ followed by the creation of a normalized histogram of $\sigma^2_{EOF,i}$ from these samples. The sampling is carried out using a Gibbs sampler (Casella and George, 1992) and consists of sequentially drawing components of $\mathbf{s}$ under the condition that the other components remain constant. I.e., a new sample $\mathbf{s}' = (\underline{\mathbf{w}}'_S, \underline{\beta}'_L, \sigma'^2_{EOF,1}, \ldots, \sigma'^2_{EOF,N_{EOF}}, \sigma'^2_S)$ is constructed from the previous sample $\mathbf{s} = (\underline{\mathbf{w}}_S, \underline{\beta}_L, \sigma^2_{EOF,1}, \ldots, \sigma^2_{EOF,N_{EOF}}, \sigma^2_S)$ by sequentially drawing

1. for each t_j

$$\begin{aligned} \beta'_L(t_j) &\sim p(\beta'_L(t_j)|\underline{\mathbf{w}}_{obs}, \underline{\mathbf{w}}_S, \Sigma^2_{EOF}, \sigma^2_S) \\ &\sim p(\mathbf{w}_{obs}(t_j)|\beta'_L(t_j), \mathbf{w}_S(t_j), \Sigma^2_{EOF}, \sigma^2_S) \\ &\times p(\beta'_L(t_j)|\mathbf{w}_S(t_j), \Sigma^2_{EOF}, \sigma^2_S) \\ &\sim N(\mathbf{w}_{obs}(t_j) - \mathbf{H}_{t_j}\mathbf{w}_{NAM}(t_j) - \mathbf{H}_{t_j}\mathbf{W}\beta'_L(t_j) \\ &\quad -\mathbf{H}_{t_j}\mathbf{w}_S(t_j); \mathbf{0}, \sigma^2_{obs}\mathbf{I}) N(\beta'_L(t_j); \mathbf{0}, \Sigma^2_{EOF}) \\ &\sim N(\beta'_L(t_j); \sigma^{-2}_{obs}\mathbf{C}\mathbf{W}^T\mathbf{H}^T_{t_j}[\mathbf{w}_{obs}(t_j) - \mathbf{H}_{t_j}\mathbf{w}_{NAM}(t_j) \\ &\quad -\mathbf{H}_{t_j}\mathbf{w}_S(t_j)], \mathbf{C}) \\ &\quad \text{with} \quad \mathbf{C}^{-1} = \sigma^{-2}_{obs}\mathbf{W}^T\mathbf{H}^T_{t_j}\mathbf{H}_{t_j}\mathbf{W} + \Sigma^{-2}_{EOF} \end{aligned} \tag{D.2}$$

2. for each t_j

$$
\begin{aligned}
\mathbf{w}'_S(t_j) &\sim p(\mathbf{w}'_S(t_j)|\underline{\mathbf{w}}_{obs}, \underline{\mathbf{w}}'_{EOF}, \Sigma^2_{EOF}, \sigma^2_S) \\
&\sim p(\mathbf{w}_{obs}(t_j)|\mathbf{w}'_S(t_j), \beta'_L(t_j), \Sigma^2_{EOF}, \sigma^2_S) \\
&\times\ p(\mathbf{w}'_S(t_j)|\beta'_L(t_j), \Sigma^2_{EOF}, \sigma^2_S) \\
&\sim N(\mathbf{w}_{obs}(t_j) - \mathbf{H}_{t_j}\mathbf{w}_{NAM}(t_j) - \mathbf{H}_{t_j}\mathbf{W}\beta'_L(t_j) \\
&\quad -\mathbf{H}_{t_j}\mathbf{w}'_S(t_j); \mathbf{0}, \sigma^2_{obs}\mathbf{I})N(\mathbf{w}'_S(t_j); \mathbf{0}, \sigma^2_S\mathbf{I}) \\
&\sim N(\mathbf{w}'_S(t_j); \sigma^{-2}_{obs}\mathbf{C}\mathbf{H}^T_{t_j}[\mathbf{w}_{obs}(t_j) - \mathbf{H}_{t_j}\mathbf{w}_{NAM}(t_j) \\
&\quad -\mathbf{H}_{t_j}\mathbf{W}\beta'_L(t_j))], \mathbf{C}) \\
&\quad \text{with} \quad \mathbf{C}^{-1} = \sigma^{-2}_{obs}\mathbf{H}^T_{t_j}\mathbf{H}_{t_j} + \sigma^{-2}_S\mathbf{I}
\end{aligned}
\tag{D.3}
$$

3. for each $i = 1, 2, \ldots, N_{EOF}$:

$$
\begin{aligned}
\sigma'^2_{EOF,i} &\sim p(\sigma'^2_{EOF,i}|\underline{\mathbf{w}}_{obs}, \underline{\beta}'_L, \underline{\mathbf{w}}'_S, \sigma^2_S) \\
&\sim p(\underline{\beta}'_{L,i}|\sigma'^2_{EOF,i})p(\sigma'^2_{EOF,i}) \\
&\sim N(\underline{\beta}'_{L,i}|\mathbf{0}, \sigma'^2_{EOF,i})IG(\sigma'^2_{EOF,i}|a_{EOF,i}, b_{EOF,i}) \\
&\sim IG(\sigma'^2_{EOF,i}|a_{EOF,i} + \tfrac{1}{2}N_t, b_{EOF,i} + \tfrac{1}{2}\textstyle\sum_j \beta^2_{EOF,i}(t_j))
\end{aligned}
\tag{D.4}
$$

4.

$$
\begin{aligned}
\sigma'^2_S &\sim p(\sigma'^2_S|\underline{\mathbf{w}}_{obs}, \underline{\beta}'_L, \underline{\mathbf{w}}'_S, \Sigma'^2_{EOF}) \\
&\sim p(\underline{\mathbf{w}}'_S|\sigma'^2_S)p(\sigma'^2_S) \\
&\sim N(\underline{\mathbf{w}}'_S|\mathbf{0}, \sigma'^2_S\mathbf{I})IG(\sigma'^2_S|a_S, b_S) \\
&\sim IG(\sigma'^2_S|a_S + N_tN_w, b_S + \tfrac{1}{2}\textstyle\sum_j ||\mathbf{w}^2_S(t_j)||^2)
\end{aligned}
\tag{D.5}
$$

with $\mathbf{w}_{obs}(t_j)$ the ASCAT and buoy wind observations at time t_j, $\mathbf{w}_{NAM}(t_j)$ the NAM model wind field at time t_j interpolated onto the ROMS model grid, $\mathbf{H}_{t_j}$ the operator that interpolates the wind field to the observation locations at time t_j, $\epsilon_{obs}(t_j)$ the measurement error in the ASCAT/NDBC buoy wind observations, $\mathbf{w}_S(t_j) \in \mathbb{R}^{2N_w}$ the error in the small-scale wind field and $\beta_{EOF,i}(t_j)$ the contribution of the i-th EOF to the large-scale error in the wind field, $\beta_{EOF,i}$ the expansion coefficient of the large-scale wind errors, $\mathbf{\Sigma}$ the diagonal matrix having $\sigma_{EOF,1}, \ldots, \sigma_{EOF,N_{EOF}}$ on its diagonal and $\mathbf{W}$ the matrix having $\mathbf{w}_{EOF,i}$ as its i-th column. The   denotes that a value is randomly dran from a distribution, $N(\mathbf{x}; \mu, \mathbf{C}) = (2\pi)^{-\frac{1}{2}D} \det(\mathbf{C})^{-\frac{1}{2}} \exp\big(-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{C}^{-1}(\mathbf{x}-\mu)\big)$ is a normal distribution with mean μ and covariance $\mathbf{C}$ and $IG(x; a, b) = \Gamma(a)^{-1} b^a x^{-a-1} \exp(-\frac{b}{x})$ the inverse gamma distribution with scale parameters a, b. In the second lines of (D.2)-(D.5) Bayes' theorem has been used. In order to insure that the samples generated are uncorrelated 10000 samples are generated with the Gibbs sampler, but only every 20th sample is retained.

References

Allen, D.R., Bishop, C.H., Frolov, S., Hoppel, K.W., Kuhl, D.D., Nedoluha, G.E., 2017. Hybrid 4DVAR with a local ensemble tangent linear model: Application to the shallow-water model. Mon. Wea. Rev. 145, 97–116. doi:10.1175/MWR-D-16-0184.1.

Amezcua, J., Goodliff, M., Leeuwen, P.J.V., 2017. A weak-constraint 4DEnsembleVar. Part I: Formulation and simple model experiments. Tellus A 69, 1271564. doi:10.1080/16000870.2016.1271564.

Anderson, J.L., 2001. An ensemble adjustment Kalman filter for data assimilation. Mon. Weather Rev. 129, 2884–2903. doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

Anderson, J.L., Anderson, S.L., 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Mon. Wea. Rev. 127, 2741–2758. doi:10.1175/1520-0493(1999)127¡2741:AMCIOT¿2.0.CO;2.

Aulign\u00e9, T., M\u00e9n\u00e9trier, B., Lorenc, A.C., Buehner, M., 2016. Ensemble-variational integrated localized data assimilation. Mon. Weather Rev. 144, 3677–3696. doi:10.1175/MWR-D-15-0252.1.

Bishop, C.H., Etherton, B.J., Majumdar, S.J., 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. Mon. Weather Rev. 129, 420–436. doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Bishop, C.H., Frolov, S., Allen, D.R., Kuhl, D.D., Hoppel, K., 2017. The Local Ensemble Tangent Linear Model: an enabler for coupled model 4D-Var. Q. J. R. Meteorol. Soc. 143, 1009–1020. doi:10.1002/qj.2986.

Brezinski, C., 1999. Multiparameter descent methods. Linear Algebra Its Appl. 296, 113–141. doi:10.1016/S0024-3795(99)00112-3.

Buehner, M., Houtekamer, P.L., Charette, C., Mitchell, H.L., He, B., 2009. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. Mon. Weather Rev. 138, 1567–1586. doi:10.1175/2009MWR3158.1.

Burgers, G., Jan van Leeuwen, P., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126, 1719–1724. doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Casella, G., George, E.I., 1992. Explaining the Gibbs Sampler. Am. Sat. 46, 167–174. doi:10.2307/2685208.

Chapman, A., Saad, Y., 1996. Deflated and augmented Krylov subspace techniques. Numer. Linear Algebra Appl. 4, 43–66. doi:10.1002/(SICI)1099-1506(199701/02)4:1<3C43::AID-NLA99>3E3.0.CO;2-Z.

Chin, T.M., Milliff, R.F., Large, W.G., 1998. Basin-scale, high-wavenumber sea surface wind fields from a multiresolution analysis of scatterometer data. J. Atmos. Oceanic Technol. 15, 741–763. doi:10.1175/1520-0426(1998)015<0741:BSHWSS>2.0.CO;2.

Clayton, A.M., Lorenc, A.C., Barker, D.M., 2013. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. Q. J. R. Meteor. Soc. 139, 1445–1461. doi:10.1002/qj.2054.

COAPS, 2015. HYCOM + NCODA global 1/12° analysis, Version GLBa0.08, january 2011 to december 2011. Center for Ocean-Atmospheric Prediction Studies, https://hycom.org/data/glba0pt08.

Cohen, A., Daubechies, I., Vial, P., 1993. Wavelets on the interval and fast wavelet transforms. Appl. Comput. Harmon. A. 1, 54–81. doi:10.1006/acha.1993.1005.

Cookson-Hills, P., Kirshbaum, D.J., Surcel, M., Doyle, J.G., Fillion, L., Jacques, D., Baek, S.J., 2017. Verification of 24-h quantitative precipitation forecasts over the Pacific Northwest from a high-resolution ensemble Kalman filter system. Wea. Forecasting 32, 1185–1208. doi:10.1175/WAF-D-16-0180.1.

Courtier, P., Thépaut, J.N., Hollingsworth, A., 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. Q. J. R. Meteor. Soc. 120, 1367–1387. doi:10.1002/qj.49712051912.

Desroziers, G., Berre, L., 2012. Accelerating and parallelizing minimizations in ensemble and deterministic variational assimilations. Q. J. R. Meteor. Soc. 138, 1599–1610. doi:10.1002/qj.1886.

Desroziers, G., Berre, L., Chapnik, B., Poli, P., 2005. Diagnosis of observation, background and analysis-error statistics in observation space. Q. J. R. Meteor. Soc. 131, 3385–3396. doi:10.1256/qj.05.108.

Desroziers, G., Camino, J., Berre, L., 2014. 4DEnVar: Link with 4D state formulation of variational assimilation and different possible implementations. Q. J. R. Meteor. Soc. 140, 2097–2110. doi:10.1002/qj.2325.

Egbert, G.D., Erofeeva, S.Y., 2002. Efficient inverse modeling of barotropic ocean tides. J. Atmos. Oceanic Technol. 19, 183–204. doi:10.1175/1520-0426(2002)019<0183:EIMOBO>2.0.CO;2.

Egbert, G.D., Erofeeva, S.Y., 2010. TPXO atlas, Version 7.2. OSU TOPEX/Poseidon Global Inverse Solution. `http://volkov.oce.orst.edu/tides/global.html`.

Erhel, J., Guyomarc'h, F., 2000. An augmented conjugate gradient method for solving consecutive symmetric positive definite linear systems. SIAM J. Matrix Anal. & Appl. 21, 1279–1299. doi:10.1137/S0895479897330194.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99. doi:10.1029/94JC00572.

Fairall, C.W., Bradley, E.F., Hare, J.E., Grachev, A.A., Edson, J.B., 2003. Bulk parameterization of airsea fluxes: Updates and verification for the COARE algorithm. J. Climate 16, 571–591. doi:10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2.

Figa-Saldaa, J., Wilson, J.J.W., Attema, E., Gelsthorpe, R., Drinkwater, M.R., Stoffelen, A., 2002. The advanced scatterometer (ASCAT) on the meteorological operational (MetOp) platform: A follow on for European wind scatterometers. Can. J. Remote Sens 28, 404–412. doi:10.5589/m02-035.

Fisher, M., Gürol, S., 2017. Parallelization in the time dimension of four-dimensional variational data assimilation. Q. J. R. Meteor. Soc. 143, 1136–1147. doi:10.1002/qj.2997.

Frolov, S., Allen, D.R., Bishop, C.H., Langland, R., Hoppel, K.W., Kuhl, D.D., 2018. First application of the Local Ensemble Tangent Linear Model (LETLM) to a realistic model of the global atmosphere. Mon. Wea. Rev. 146, 2247–2270. doi:10.1175/MWR-D-17-0315.1.

Frolov, S., Bishop, C.H., 2016. Localized ensemble-based tangent linear models and their use in propagating hybrid error covariance models. Mon. Wea. Rev. 144, 1383–1405. doi:10.1175/MWR-D-15-0130.1.

Fujita, T., Stensrud, D.J., Dowell, D.C., 2007. Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. Mon. Weather Rev. 135, 1846–1868. doi:10.1175/MWR3391.1.

Grigori, L., Moufawad, S., Nataf, F., 2016. Enlarged Krylov subspace conjugate gradient methods for reducing communication. SIAM J. Matrix Anal. & Appl. 37, 744–773. doi:10.1137/140989492.

Gürol, S., Weaver, A.T., Moore, A.M., Piacentini, A., Arango, H.G., Gratton, S., 2014. B-preconditioned minimization algorithms for variational data assimilation with the dual formulation. Q.J.R. Meteorol. Soc. 140, 539–556. doi:10.1002/qj.2150.

Gustafsson, N., Bojarova, J., 2014. Four-dimensional ensemble variational (4D-En-Var) data assimilation for the HIgh Resolution Limited Area Model (HIRLAM). Nonlinear Process. Geophys. 21, 745–762. doi:10.5194/npg-21-745-2014.

Halpern, D., 1976. Structure of a coastal upwelling event observed off Oregon during July 1973. Deep-Sea Res. Oceanogr. Abstr. 23, 495–508. doi:10.1016/0011-7471(76)90861-5.

Hamill, T.M., 2001. Interpretation of rank histograms for verifying ensemble forecasts. Mon. Weather Rev. 129, 550–560. doi:10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

Hamill, T.M., Whitaker, J.S., 2005. Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. Mon. Weather Rev. 133, 3132–3147. doi:10.1175/MWR3020.1.

Hamill, T.M., Whitaker, J.S., Snyder, C., 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon. Weather Rev. 129, 2776–2790. doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2. 00494.

Hénaff, M.L., Mey, P.D., Marsaleix, P., 2009. Assessment of observational networks with the representer matrix spectra method-application to a 3D coastal model of the Bay of Biscay. Ocean Dyn. 59, 3–20. doi:10.1007/s10236-008-0144-7.

Hickey, B.M., Geier, S., Kachel, N., MacFadyen, A., 2005. A bi-directional river plume: The Columbia in summer. Cont. Shelf Res. 25, 1631–1656. doi:10.1016/j.csr.2005.04.010.

Hickey, B.M., Kudela, R.M., Nash, J.D., Bruland, K.W., Peterson, W.T., MacCready, P., Lessard, E.J., Jay, D.A., Banas, N.S., Baptista, A.M., Dever, E.P., Kosro, P.M., Kilcher, L.K., HornerDevine, A.R., Zaron, E.D., McCabe, R.M., Peterson, J.O., Orton, P.M., Pan, J., Lohan, M.C., 2010. River influences on shelf ecosystems: Introduction and synthesis. J. Geophys. Res. Oceans 115. doi:10.1029/2009JC005452.

Hickey, B.M., Pietrafesa, L.J., Jay, D.A., Boicourt, W.C., 1998. The Columbia River plume study: Subtidal variability in the velocity and salinity fields. J. Geophys. Res. Oceans 103, 10339–10368.

Houtekamer, P.L., Mitchell, H.L., 1998. Data assimilation using an ensemble Kalman filter technique. Mon. Weather Rev. 126, 796–811. doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.O;2.

Huyer, A., 1983. Coastal upwelling in the California Current System. Prog. Oceanogr. 12, 259–284. doi:10.1016/0079-6611(83)90010-1.

Kharchenko, S.A., Yeremin, A.Y., 1995. Eigenvalue translation based preconditioners for the GMRES(k) method. Numer. Linear Algebra Appl. 2, 51–77. doi:10.1002/nla.1680020105.

Kuhl, D.D., Rosmond, T.E., Bishop, C.H., McLay, J., Baker, N.L., 2013. Comparison of hybrid ensemble/4DVar and 4DVar within the NAVDAS-AR data assimilation framework. Mon. Weather Rev. 141, 2740–2758. doi:10.1175/MWR-D-12-00182.1.

Kurapov, A.L., Egbert, G.D., Allen, J.S., Miller, R.N., 2009. Representer-based analyses in the coastal upwelling system. Dyn. Atmos. Oceans 48, 198–218. doi:10.1016/j.dynatmoce.2008.09.002.

Kurapov, A.L., Egbert, G.D., Allen, J.S., Miller, R.N., Erofeeva, S.Y., Kosro, P.M., 2003. The M2 internal tide off Oregon: Inferences from data assimilation. J. Phys. Oceanogr. 33, 1733–1757. doi:10.1175/1520-0485(2003)033<1733:TMITOO>2.0.CO;2.

Kurapov, A.L., Foley, D., Strub, P.T., Egbert, G.D., Allen, J.S., 2011. Variational assimilation of satellite observations in a coastal ocean model off Oregon. J. Geophys. Res. Oceans 116, C05006. doi:10.1029/2010JC006909.

Kurapov, A.L., Pelland, N.A., Rudnick, D.L., 2017. Seasonal and interannual variability in along-slope oceanic properties off the US West Coast: Inferences from a high-resolution regional model. J. Geophys. Res. Oceans 122, 5237–5259. doi:10.1002/2017JC012721.

Leeuwenburgh, O., 2007. Validation of an EnKF system for OGCM initialization assimilating temperature, salinity, and surface height measurements. Mon. Weather Rev. 135, 125–139. doi:10.1175/MWR3272.1.

Lermusiaux, P.F.J., Robinson, A.R., 1999. Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. Mon. Weather Rev. 127, 1385–1407. doi:10.1175/1520-0493(1999)127<1385:DAVESS>2.0.CO;2.

Li, H., Kalnay, E., Miyoshi, T., Danforth, C.M., 2009. Accounting for model errors in ensemble data assimilation. Mon. Weather Rev. 137, 3407–3419. doi:10.1175/2009MWR2766.1.

Liu, C., Xiao, Q., Wang, B., 2008. An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. Mon. Weather Rev. 136, 3363–3373. doi:10.1175/2008MWR2312.1.

Liu, Y., MacCready, P., Hickey, B.M., 2009. Columbia River plume patterns in summer 2004 as revealed by a hindcast coastal ocean circulation model. Geophys. Res. Lett. 36, L02601. doi:10.1029/2008GL036447.

Lorenc, A.C., Bowler, N.E., Clayton, A.M., Pring, S.R., Fairbairn, D., 2015. Comparison of Hybrid-4DEnVar and Hybrid-4DVar data assimilation methods for global NWP. Mon. Weather Rev. 143, 212–229. doi:10.1175/MWR-D-14-00195.1.

Lorenc, A.C., Jardak, M., 2018. A comparison of hybrid variational data assimilation methods for global NWP. Q. J. R. Meteorol. Soc. , 2748–2760doi:10.1002/qj.3401.

Lorenc, A.C., Jardak, M., Payne, T., Bowler, N.E., Wlasak, M.A., 2017. Computing an ensemble of variational data assimilations using its mean and perturbations. Q. J. R. Meteor. Soc. 143, 798–805. doi:10.1002/qj.2965.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, in: Le Cam, L.M., Neyman, J. (Eds.), Proceedings of the Berkeley symposium on mathematics and statistics and probability, 5th, Oakland, CA, USA.. pp. 281–297.

Mandel, J., Bergou, E., Gürol, S., Gratton, S., Kasanický, I., 2016. Hybrid Levenberg-Marquardt and weak-constraint ensemble Kalman smoother method. Nonlinear Process. Geophys. 23, 59–73. doi:10.5194/npg-23-59-2016.

Maturi, E., Harris, A., Mittaz, J., Merchant, C., Potash, B., Meng, W., Sapper, J., 2008. NOAA's sea surface temperature products from operational geostationary satellites. Bull. Amer. Meteor. Soc. 89, 1877–1888. doi:10.1175/2008BAMS2528.1.

Meng, Z., Zhang, F., 2008. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with

3DVAR in a month-long experiment. Mon. Weather Rev. 136, 3671–3682. doi:10.1175/2008MWR2270.1.

Milliff, R.F., Bonazzi, A., Wikle, C.K., Pinardi, N., Berliner, L.M., 2011. Ocean ensemble forecasting. Part I: Ensemble Mediterranean winds from a Bayesian hierarchical model. Q. J. R. Meteor. Soc. 137, 858–878. doi:10.1002/qj.767.

Morgan, R., 1995. A restarted GMRES method augmented with eigenvectors. SIAM J. Matrix Anal. & Appl. 16, 1154–1171. doi:10.1137/S0895479893253975.

NOAA, 2011a. ASCAT METOPA-A, April 2011 to September 2011. Center for Satellite Application and Research (STAR). `https://manati.star.nesdis.noaa.gov/datasets/ASCATData.php/`.

NOAA, 2011b. North American Mesoscale Forecast System (NAM) 12 km analysis, january 2011 to december 2011. National Centers for Environmental Information. `https://www.ncei.noaa.gov/thredds/catalog/namanl/catalog.html`.

NOAA, 2016. Pacific North buoys. National Data Buoy Center. `http://www.ndbc.noaa.gov/`.

O'Leary, D.P., 1980. The block conjugate gradient algorithm and related methods. Linear Algebra Appl. 29, 293–322. doi:10.1016/0024-3795(80)90247-5.

Osborne, J.J., Kurapov, A.L., Egbert, G.D., Kosro, P.M., 2011. Spatial and temporal variability of the M2 internal tide generation and propagation on the Oregon shelf. J. Phys. Oceanogr. 41, 2037–2062. doi:10.1175/JPO-D-11-02.1.

Palmer, T., Buizza, R., F., D.R., Jung, T., Leutbecher, M., Shutts, G., Steinheimer, M., Weisheimer, A., 2009. Stochastic Parametrization and Model Uncertainty. Technical Memorandum No. 598. ECMWF, Reading.

Pasmans, I., Kurapov, A., Barth, J., Ignatov, A., Kosro, P., Shearman, R., 2019. Why gliders appreciate good company: Glider assimilation in the Oregon-Washington coastal ocean 4DVAR system with and without surface observations. J. Geophys. Res. Oceans 124, 750–772. doi:10.1029/2018JC014230.

Pasmans, I., Kurapov, A.L., 2017. A Monte Carlo background covariance localization method for an ensemble-variational assimilation system. Mon. Weather Rev. 145, 4543–4557. doi:10.1175/MWR-D-16-0424.1.

Pasmans, I., Kurapov, A.L., Barth, J., Kosro, P.M., Shearman, R.K., in preparation. Ensemble 4dvar (En4DVar) data assimilation in a coastal ocean circulation model. Part II: Implementation offshore Oregon-Washington, USA. Ocean Model. .

Pawlowicz, R., Beardsley, B., Lentz, S., 2002. Classical tidal harmonic analysis including error estimates in MATLAB using T_Tide. Comput. & Geosci. 28, 929–937. doi:10.1016/S0098-3004(02)00013-4.

Poterjoy, J., Zhang, F., 2015. Systematic comparison of four-dimensional data assimilation methods with and without the tangent linear model using hybrid background error covariance: E4DVar versus 4DEnVar. Mon. Weather Rev. 143, 1601–1621. doi:10.1175/MWR-D-14-00224.1.

Poterjoy, J., Zhang, F., Poterjoy, J., Zhang, F., 2016. Comparison of hybrid four-dimensional data assimilation methods with and without the tangent linear and adjoint models for predicting the life cycle of hurricane Karl (2010). doi:10.1175/MWR-D-15-0116.1.

Rao, V., Sandu, A., 2016. A time–parallel approach to strong-constraint four–dimensional variational data assimilation. J. Comput. Phys. 313, 583–593. doi:10.1016/j.jcp.2016.02.040.

Shchepetkin, A.F., McWilliams, J.C., 2003. A method for computing horizontal pressure-gradient force in an oceanic model with a nonaligned vertical coordinate. J. Geophys. Res. Oceans 108, 3090. doi:10.1029/2001JC001047.

Shchepetkin, A.F., McWilliams, J.C., 2005. The Regional Oceanic Modeling System (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. Ocean Model. 9, 347–404. doi:10.1016/j.ocemod.2004.08.002.

Tian, X., Zhang, H., Feng, X., Xie, Y., 2017. Nonlinear least square En4DVar to 4DEnVar methods for data assimilation: Formulation, analysis, and preliminary evaluation. Mon. Weather Rev. 146, 77–93. doi:10.1175/MWR-D-17-0050.1.

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G.D., Roskies, R., Scott, J.R., Wilkins-Diehr, N., 2014. XSEDE: Accelerating scientific discovery. Comp. Sci. & Eng. 16, 62–74. doi:10.1109/MCSE.2014.80.

Trefethen, L.N., Bau, D., 1997. Numerical linear algebra. SIAM, Philadelphia, PA.

Verspeek, J., Verhoef, A., Stoffelen, A., 2013. ASCAT-B NWP ocean calibration and validation. Technical Report SAF/OSI/CDOP2/KNMI/TEC/RP/199. KNMI, De Bilt.

Vervatis, V., Testut, C.E., De Mey, P., Ayoub, N., Chanut, J., Quattrocchi, G., 2016. Data assimilative twin-experiment in a high-resolution Bay of Biscay configuration: 4DEnOI based on stochastic modeling of the wind forcing. Ocean Model. 100, 1–19. doi:10.1016/j.ocemod.2016.01.003.

Whitaker, J.S., Hamill, T.M., 2002. Ensemble data assimilation without perturbed observations. Mon. Weather Rev. 130, 1913–1924. doi:10.1175/1520-0493(2002)130 $<$ 1913:EDAWPO $>$ 2.0.CO;2.

Whitaker, J.S., Hamill, T.M., 2012. Evaluating methods to account for system errors in ensemble data assimilation. Mon. Weather Rev. 140, 3078–3089. doi:10.1175/MWR-D-11-00276.1.

Wikle, C.K., Milliff, R.F., Nychka, D., Berliner, L.M., 2001. Spatiotemporal hierarchical Bayesian modeling tropical ocean surface winds. J. Am. Stat. Assoc. 96, 382–397. doi:10.1198/016214501753168109.

Yaremchuk, M., Martin, P., Beattie, C., 2017. A hybrid approach to generating search subspaces in dynamically constrained 4-dimensional data assimilation. Ocean Model. 117, 41–51. doi:10.1016/j.ocemod.2017.08.003.

Yu, P., Kurapov, A.L., Egbert, G.D., Allen, J.S., Kosro, P.M., 2012. Variational assimilation of HF radar surface currents in a coastal ocean model off Oregon. Ocean Model. 49-50, 86–104. doi:10.1016/j.ocemod.2012.03.001.

Zhang, M., Zhang, F., 2012. E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. Mon. Wea. Rev. 140, 587–600.