CHAPTER 17

Data Assimilation in Oceanography: Current Status and New Directions

Ibrahim Hoteit[1], Xiaodong Luo[2], Marc Bocquet[3], Armin Köhl[4], and Boujemaa Ait-El-Fquih[1]

[1]King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia; [2]International Research Institute of Stavanger (IRIS), Bergen, Norway; [3]CEREA, Joint Laboratory of École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France; [4]University of Hamburg, Hamburg, Germany

Characterizing and forecasting the state of the ocean is essential for various scientific, management, commercial, and recreational applications. This is, however, a challenging problem due to the large, multiscale and nonlinear nature of the ocean state dynamics and the limited amount of observations. Combining all available information from numerical models describing the ocean dynamics, observations, and prior information has proven to be the most viable approach to determine the best estimates of the ocean state, a process called data assimilation (DA). DA is becoming widespread in many ocean applications; stimulated by continuous advancement in modeling, observational, and computational capabilities. This chapter offers a comprehensive presentation of the theory and methods of ocean DA, outlining its current status and recent developments, and discussing new directions and open questions. Casting DA as a Bayesian state estimation problem, the chapter will gradually advance from the basic principles of DA to its most advanced methods. Three-dimensional DA methods, 3DVAR and Optimal Interpolation, are first derived, before incorporating time and present the most popular, Gaussian-based DA approaches: 4DVAR, Kalman filters and smoothers methods, which exploit past and/or future observations. Ensemble Kalman methods are next introduced in their stochastic and deterministic formulations as a stepping-stone toward the more advanced nonlinear/non-Gaussian DA methods, Particle and Gaussian Mixture filters. Other sophisticated hybrid extensions aimed at exploiting the advantages of both ensemble and variational methods are also presented. The chapter then concludes with a discussion on the importance of properly addressing the uncertainties in the models and the data, and available approaches to achieve this through parameters estimation, model errors quantification, and coupled DA.

Introduction

Following the tremendous progress in weather monitoring and forecasting, there is increased interest in developing global and regional systems for operational oceanography in order to provide estimates and forecasts of essential ocean variables. Outputs of such systems can be used to generate data products, applications, and services through national authorities and organizations, such as metocean service providers and environmental agencies. These can include nowcasts providing the most usefully accurate description of the present state of the ocean, forecasts of the future ocean conditions as far ahead as possible (typically one to two weeks), and reanalyses (hindcasts) assembling long-term datasets to describe the history of the studied region including time series showing trends and changes. Such products can provide crucial information for a wide

Hoteit, I., et al., 2018: Data assimilation in oceanography: Current status and new directions. In "*New Frontiers in Operational Oceanography*", E. Chassignet, A. Pascual, J. Tintoré, and J. Verron, Eds., GODAE OceanView, 465-512, doi:10.17125/gov2018.ch17.

variety of marine industrial and governmental activities and societal needs, including safety of life at sea, coastal extremes, pollution and contamination management, tourism, marine conservation, fisheries and aquaculture, exploration and drilling, desalination and plant cooling operations, shipping, harbor management and national security operations, etc. Operational oceanographic programs have been recently established in several countries of the European Union, as well as the United States of America and Japan.

Operational oceanography depends on the availability of ocean observations transmitted in (near-) real-time and time-dependent numerical models to project the information gathered by the observations into the future (or the past). The models can be constructed based on the observations (e.g., Hamilton et al., 2016; Dreano et al., 2015; Lguensat et al., 2017) by exploiting their statistical properties, but their outputs could be restricted to the measured quantities, which are limited in space and time. Such models have the advantage of being computationally very efficient, but their predictive capabilities are often limited to short temporal ranges and may not be efficient for predicting extremes. More established ocean models are developed based on the physical laws that govern the oceans general circulation (Navier-Stocks equations). Despite significant computational requirements, the dynamics of these models bring more information to the otherwise underdetermined ocean state estimation problem, which enhances the accuracy and dynamical consistency of the ocean state estimates.

Whether based on statistical properties or physical laws, ocean models are not perfect and can be subject to various sources of uncertainties. Observation-based models are, for instance, constructed based on statistical assumptions which may not be always relevant. Dynamical models, on the other hand, require atmospheric forcing that is not always available at the required spatial and temporal resolutions, and are themselves distorted by uncertainties. Numerical errors, missing physics, and poorly known parameterizations coefficients (e.g., diffusivity and viscosity) are also important sources of uncertainties in such models. These errors may accumulate over time and often deviate the model from the real ocean trajectory, even when the ocean state is perfectly known at the initial time. It is now recognized that the most efficient approach to obtain reliable ocean state estimates is to routinely constrain and adjust the model outputs with incoming ocean data through a data assimilation (DA) process (Ghil and Malanotte-Rizzoli, 1991; Wunsch, 1996; Bennett, 2005). In general, DA exploits the models as spatio-temporal interpolators of the data, and the data guide the models toward the true trajectory of the system. Effective operational oceanography relies on the ability to assimilate massive amounts of data gathered by monitoring systems in real time into advanced general circulation models (GCMs) on supercomputer facilities.

Although the theoretical framework of DA methods is well established on the Bayesian estimation theory (Law et al., 2015), the applications of these with state-of-the-art ocean general circulation models (OGCMs) is still strongly hampered by the large dimensional and nonlinear characteristics of these systems. As will be further discussed later, poor knowledge of the statistical properties of the models and data uncertainties also limits the efficiency of ocean DA systems. This chapter will provide a general overview of ocean data assimilation methods, presenting the state-of-the-art methods, their origins and relations, and discussing in particular major limitations and

future directions. It makes no attempt to be a comprehensive review of the extensive DA literature, which extends back to the late 1950s. This chapter will first describe the three-dimensional (3D) DA problem that only considers the current observations in the estimation of the ocean state before moving to more advanced four-dimensional (4D) DA methods, focusing on the most commonly used 4D variational (4DVAR) methods and ensemble Kalman filters. More sophisticated methods combining variational and ensemble methods and more advanced ones designed for non-Gaussian distributions and their potential use for enhancing ocean data assimilation systems will be also discussed. A summary of all the DA approaches discussed in this chapter and their main founding hypothesis is provided in Table 17.1.

DA Method	*Founding Hypothesis*	*Derivation*
3D variational assimilation (3DVar)	Gaussian model state and observation noise	Minimize a cost function that involves the model state and observation at a given time instance
4D variational assimilation (4DVar)	Gaussian model state, model errors and observation noise	Minimize a cost function that involves the model state and observation at a given time interval. Relations between model state variables are constrained by the dynamical model
Kalman filter (KF)	Linear dynamical model and observation operator; Gaussian model state, model errors and observation noise	Take the maximum a posteriori (MAP) estimate of the ocean state conditioned on previous observations
Extended Kalman filter (EKF); including reduced EKFs	Nonlinear dynamical model and/or observation operator; Gaussian model state, model errors and observation noise	Linearize nonlinear dynamical model and/or observation operator, and then take the MAP solution as in KF
Ensemble Kalman filters (EnKF), and ensemble optimal interpolation (EnOI)	As in EKF	Use an ensemble of model states to estimate the background statistics (prior mean and covariance), and corresponding KF update formulas to produce an analysis ensemble with targeted posterior mean and covariance
Particle filter (PF)	Nonlinear dynamical model and/or observation operator; non-Gaussian model state, model errors and observation noise	Use mixture of Dirac delta functions to approximate the prior and posterior distributions of the model state conditioned on previous observations
Gaussian mixture filter (GMF); including ensemble GMFs	As in PF	Use mixture of Gaussian distributions to approximate the prior and posterior distributions of the model state conditioned on previous observations, and also to approximate the distributions of the model errors and observation noise when necessary

Table 17.1. Founding hypotheses and derivations of data assimilation (DA) methods. Founding hypotheses describe assumption(s), e.g. linearity and/or Gaussianity, behind DA methods from a Bayesian perspective; whereas derivations summarize features utilized and/or actions taken to derive the DA methods. The table focuses on the filtering schemes, but the descriptions are also valid for the corresponding smoothing schemes.

Three-Dimensional Data Assimilation

The three-dimensional data assimilation (3DDA) problem refers to the space domain in which we look for the best estimate $\mathbf{x}^a$ (^{a} for analysis) of the ocean state $\mathbf{x}$ at some time given only the observation $\mathbf{y}$ of the state at that time. $\mathbf{y}$ may contain in situ measurements from cruises, profiles, gliders, and buoys, and satellite measurements of sea surface height and sea surface temperature. $\mathbf{x}$ is typically comprised of the prognostic model variables (needed to initialize the model) at every grid point of the domain, such as temperature, salinity, sea level, and velocities. We assume the observational model $\boldsymbol{H}$, possibly nonlinear, relating the ocean state to the observation is available.

$$\mathbf{y} = \boldsymbol{H}(\mathbf{x}). \tag{1}$$

Estimating $\mathbf{x}$ from $\mathbf{y}$ can be formulated as a weighted least-squares inverse problem in which we look for $\mathbf{x}$ that minimizes an objective function measuring the distance between the ocean state and the observations, of the form

$$\boldsymbol{J}_{3D}(\mathbf{x}) = (\mathbf{y} - \boldsymbol{H}(\mathbf{x}))^T \mathbf{W}^y (\mathbf{y} - \boldsymbol{H}(\mathbf{x})) = \|\mathbf{y} - \boldsymbol{H}(\mathbf{x})\|^2_{\mathbf{W}^y}, \tag{2}$$

where $\mathbf{W}^y$ is the data weight (definite positive) matrix introduced to specify the observations weights in the optimization (to assign, for example, less weights to uncertain measurements). In ocean applications, the number of observations p (i.e., dimension of $\mathbf{y}$) is typically much smaller than the number of state variables to be inferred n (i.e., dimension of $\mathbf{x}$). This makes the above problem underdetermined, and more information is needed to regularize it (Wunsch, 1996). This is commonly enforced by solving for the ocean state estimate that is not too far from a given prior state estimate $\mathbf{x}^b$ (^{b} for background), often taken as the most recent forecast (nowadays computed by the ocean model starting from the most recent state estimate). The objective function then becomes

$$\boldsymbol{J}_{3D}(\mathbf{x}) = \|\mathbf{y} - \boldsymbol{H}(\mathbf{x})\|^2_{\mathbf{W}^y} + \|\mathbf{x} - \mathbf{x}^b\|^2_{\mathbf{W}^b}, \tag{3}$$

where $\mathbf{W}^b$ is the background weight matrix.

This deterministic formulation of the ocean state estimation problem does not provide a framework for choosing the weight matrices or to quantify the uncertainties in the estimate. A more general approach to formulate the 3DDA problem is to encapsulate it within a Bayesian framework, which considers the ocean state $\mathbf{x}$ and the observation $\mathbf{y}$ as random variables, see for example Simon (2006) and Wikle and Berliner (2007). This naturally allows us to account for the uncertainties in the observation, which is often expressed as

$$\mathbf{y} = \boldsymbol{H}(\mathbf{x}) + \varepsilon, \tag{4}$$

where ε represents the observational errors, generally assumed unbiased, and to exploit a prior knowledge of the state and its uncertainty through their probability distributions. The solution of the estimation problem is then determined as the conditional probability distribution of the state given the observation $p_{\mathbf{x}|\mathbf{y}}$, which is computed via the Bayes' rule,

$$p_{\mathbf{x}|\mathbf{y}} = \frac{p_{\mathbf{y}|\mathbf{x}}\, p_{\mathbf{x}}}{p_{\mathbf{y}}}. \tag{5}$$

$p_{\mathbf{x}|\mathbf{y}}$ is also called *posterior* and represents an update of the *prior* distribution $p_{\mathbf{x}}$, while $p_{\mathbf{y}|\mathbf{x}}$ is the likelihood function of $\mathbf{y}$ given $\mathbf{x}$, and $p_{\mathbf{y}}$ is the marginal distribution of $\mathbf{y}$ representing a normalizing constant to ensure that the final solution is a probability distribution. An ocean state estimate $\mathbf{x}^a$ can then be obtained as the maximum a posteriori (MAP) estimate maximizing $p_{\mathbf{x}|\mathbf{y}}$, or the minimum-variance (MV) estimate (or posterior mean), which are equivalent when the posterior is Gaussian.

Assuming the *prior* and observation errors follow normal (Gaussian) distributions, then the posterior is given by (Talagrand, 2010)

$$p_{\mathbf{x}|\mathbf{y}} \propto \exp\left(-\frac{1}{2}\left[\|\mathbf{y} - H(\mathbf{x})\|^2_{\mathbf{R}^{-1}} + \|\mathbf{x} - \mathbf{x}^b\|^2_{\mathbf{B}^{-1}}\right]\right), \tag{6}$$

where $\mathbf{R}$ and $\mathbf{B}$ are respectively the observation and background error covariance matrices. Maximizing $p_{\mathbf{x}|\mathbf{y}}$ is equivalent to minimizing J_{3D}, with the weight matrices as the inverse of the covariance matrices, i.e. $\mathbf{W}^{\mathrm{y}} = \mathbf{R}^{-1}$ and $\mathbf{W}^b = \mathbf{B}^{-1}$.

When the observational operator is linear, and thus will be denoted by $\mathbf{H}$, the solution of the problem can be directly computed by setting the derivative of the convex objective function J_{3D} to zero, to obtain

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^b), \tag{7}$$

and its error covariance matrix

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}, \tag{8}$$

where $\mathbf{K}$ is the Gain matrix given by

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1}. \tag{9}$$

Note that in a non-Gaussian setting with a linear observational operator, the above solution remains the best linear unbiased estimator (BLUE; Talagrand, 2010), where "best" stands for minimum-variance.

Optimal Interpolation (OI; Bouttier and Courtier, 1999) is a popular algebraic simplification of the Gain matrix in the BLUE designed by viewing Eq. 7 as a list of scalar analysis equations, one per state variable of $\mathbf{x}$. Only observations located within a certain distance from the variable being analyzed are then used to compute the increment of that variable. This makes the OI scheme easy to parallelize and implement for efficient DA.

When the observational operator is nonlinear, the objective function is not convex and may exhibit several minima; but near the minimum one can linearize it (around the background) before computing the BLUE, exactly as above. One may also consider an iterative solution to the problem Eq. 7 by computing the linearization of the observational operator around the estimate of the last iteration (Simon, 2006).

A more straightforward approach is to apply an optimization algorithm to directly minimize the objective function J_{3D}, the most popular of which are the gradient-based optimization methods because of their fast convergence rate (Bouttier and Courtier, 1999). These methods use the gradient of the objective function to determine descent directions toward the minimum in an iterative procedure (Bouttier and Courtier, 1999). The gradient of J_{3D} with respect to $\mathbf{x}$ is

$$\nabla_{\mathbf{x}} J_{3D} = 2\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) - 2\mathbf{H}_{\mathbf{x}}^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x})), \quad \text{with } \mathbf{H}_{\mathbf{x}}^T = \nabla_{\mathbf{x}} H \tag{10}$$

which only requires the computation of the product of the inverse of the background and observation error covariances by a vector. As such, this offers the possibility to accommodate more sophisticated forms of the background error covariance matrix. This framework is known as the 3D variational (3DVAR) assimilation problem and is still heavily implemented in operational weather forecast centers.

The observational and background covariance matrices $\mathbf{R}$ and $\mathbf{B}$ are very important in determining the solution of the assimilation problem. These set the extent to which the background field (forecast) will be adjusted by the data by setting the weights of the background and data terms in the inversion. In practice, however, there is insufficient information to determine these matrices and ad hoc estimates are used instead. The observation errors are often assumed to be spatially uncorrelated, so that $\mathbf{R}$ is diagonal. Imposing correlation errors for data with important spatial coverage, such as satellite and radars, is important to avoid overweighting them in the assimilation. This could be conveniently implemented through an appropriate choice of a covariance model. A simpler way is to deliberately reduce the weights (i.e., overestimates the error variances) of the "clustered" observations, whose errors are expected to be correlated. Modeling $\mathbf{B}$ is more delicate as it needs to incorporate ocean balance properties and smoothness constraints, as the analysis increment completely lies within the subspace spanned by the directions of $\mathbf{B}$. The use of such constraints helps to dynamically spread the information in the observations, which should provide an analysis that could be more conveniently assimilated by the ocean model for forecasting. This was thoroughly discussed by Weaver et al. (2003, 2005) and Blayo et al. (2014). Use of ensemble of model outputs in the modeling of $\mathbf{B}$ has also became popular (e.g., Buehner, 2005).

Examples of ocean operational systems based on 3DDA methods include the US Naval Oceanographic Office NAVOCEANO system (Smedstad et al., 2003), the Meteorological Research Institute multivariate ocean variational estimation MOVE System (Usui et al., 2006), the European ECMWF system (Balmaseda et al., 2013), the Global Ocean Forecasting System (GOFS; Cummings and Smedstad, 2013), and the UK Met Office Forecasting Ocean Assimilation System (FOAM; Blockley et al., 2014).

Four-Dimensional Variational Assimilation

Four-dimensional variational assimilation (4DVAR) is a generalization of 3DVAR to the problem of estimating the state of a dynamical system using a set of observations that are available over a time interval. This not only includes information from future and past observations to estimate the

ocean state at a given time, but also enables exploitation of the dynamical information from the equations that govern the evolution of the ocean state in time (i.e., ocean model). In its most general form, the latter could be described by the Navier-Stokes equations (Temam, 1984), and we represent it here as a discrete-time dynamical operator M_k that integrates the ocean state $\mathbf{x}$ between two consecutive time steps t_{k-1} and t_k as

$$\mathbf{x}_k = M_k(\mathbf{x}_{k-1}) + \eta_k. \tag{11}$$

$\boldsymbol{\eta}_k$ is a stochastic term representing uncertainties in the model, referred to as model error, and is usually conveniently assumed stochastic following a Gaussian distribution of mean zero (unbiased) and covariance $\mathbf{Q}_k$.

4DVAR can be directly formulated from the least-square objective function (Eq. 3) by constraining the (sum of the) distances between the model state at the times of available data, according to some weight for each term. Here, we first derive 4DVAR as the MAP estimator of a Bayesian estimation problem before presenting the adjoint method to efficiently compute the gradient of the objective function and accordingly its optimum. We finish with a discussion on the main features and issues of this approach.

Bayesian formulation

The Bayesian estimation problem of the ocean state $\mathbf{x}_0, \ldots, \mathbf{x}_L$ over a time interval $[\mathrm{T}_0\ \mathrm{T}_L]$ given a set of available observations $\mathbf{y}_0, \ldots, \mathbf{y}_L$, related to the ocean state as in Eq. 1, involves the calculation of the conditional probability distribution, similar to Eq. 5.

$$p_{\mathbf{x}_0,\ldots,\mathbf{x}_l|\mathbf{y}_0,\ldots,\mathbf{y}_L} \propto p_{\mathbf{y}_0,\ldots,\mathbf{y}_L|\mathbf{x}_0,\ldots,\mathbf{x}_L}\ p_{\mathbf{x}_0,\ldots,\mathbf{x}_L}. \tag{12}$$

Assuming observation and model errors $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ are independent in time and mutually independent, and given the Markov Chain nature of the dynamical system (Eq. 11), standard conditional probability calculations lead to. See, for example, Simon (2006) and Law et al. (2015),

$$p_{\mathbf{x}_0,\ldots,\mathbf{x}_L|\mathbf{y}_0,\ldots,\mathbf{y}_L} \propto \Pi_{k=0}^{L} p_{\mathbf{y}_k|\mathbf{x}_k} \cdot p_{\mathbf{x}_0} \cdot \Pi_{k=1}^{L} p_{\mathbf{x}_k|\mathbf{x}_{k-1}}, \tag{13}$$

and under the assumption of Gaussian probability distributions we find analogous to Eq. 6

$$p_{\mathbf{x}_0,\ldots,\mathbf{x}_L|\mathbf{y}_0,\ldots,\mathbf{y}_L} \propto \exp\left(-\frac{1}{2}J_{4D}\right), \tag{14}$$

where

$$J_{4D}(\mathbf{x}_0,\ldots,\mathbf{x}_L) = \sum_{k=0}^{L} \|\mathbf{y}_k - H_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}} + \|\mathbf{x}_0 - \mathbf{x}_0^b\|^2_{\mathbf{B}_0^{-1}} + \sum_{k=0}^{L} \|\mathbf{x}_k - M_k(\mathbf{x}_{k-1})\|^2_{\mathbf{Q}_k^{-1}}. \tag{15}$$

The MAP estimator can thus be obtained by minimizing the 4D objective function $\boldsymbol{J}_{4D}$ as defined in Eq. 15, which is the same as optimizing

$$J_{4D}(\mathbf{x}_0, \eta_1 \ldots, \eta_L) = \sum_{k=0}^{L} \|\mathbf{y}_k - H_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}} + \|\mathbf{x}_0 - \mathbf{x}_0^b\|^2_{\mathbf{B}_0^{-1}} + \sum_{k=0}^{L} \|\eta_k\|^2_{\mathbf{Q}_k^{-1}} \qquad (16)$$

subject to the ocean dynamics as described by Eq. 11.

The dimension of the ocean state can be very large in realistic applications, reaching up to 10^9–10^{10} in today's applications. And given that it should be determined for every time step, the required information exceeds by far the amount of available ocean data, even for coarse resolution models. One straightforward way to reduce the dimension of the 4DVAR optimization problem and mitigate its under-determined nature is to reduce the number of parameters by allowing only certain forms of model uncertainties. The extreme case is to assume the ocean model (Eq. 11) to be perfect, i.e., $\boldsymbol{\eta}_k = 0$, so the problem reduces to finding the initial condition $\mathbf{x}_0$ that best fits, within observation errors uncertainties, the model to the data by minimizing (as schematically illustrated in Fig. 17.1)

$$J_{4D}(\mathbf{x}_0) = \sum_{k=0}^{L} \|\mathbf{y}_k - H_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}} + \|\mathbf{x}_0 - \mathbf{x}_0^b\|^2_{\mathbf{B}_0^{-1}}, \quad \text{subject to } \mathbf{x}_k = M_k(\mathbf{x}_{k-1}). \qquad (17)$$

This is known as the strong constraint 4DVAR problem. Directly optimizing (Eq. 16) is known as the weak constraint 4DVAR problem.

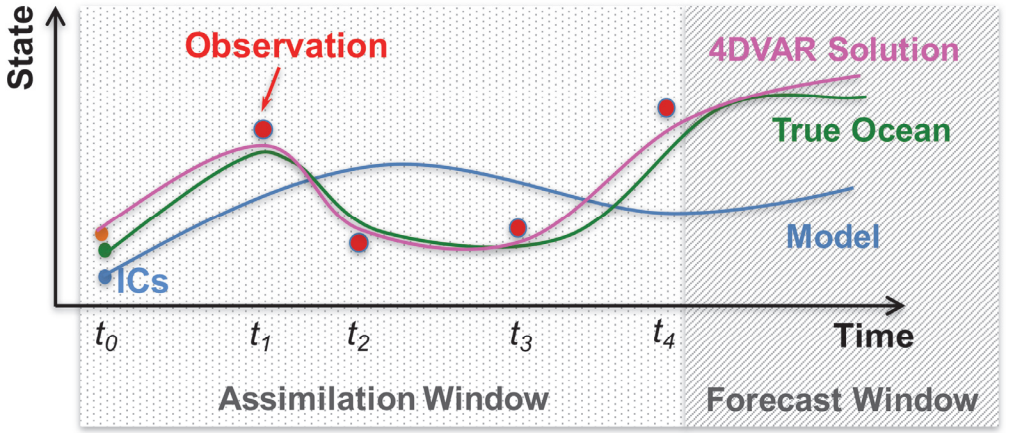

Figure 17.1. Schematic diagram of the 4DVAR assimilation procedure: fit the model to all available observations within an assimilation window to compute the analysis, from which integrate the ocean model for forecasting.

A large variety of different configurations exist between these extreme cases, for instance adjusting the ocean model parameters and inputs by including them as part of the estimation problem, i.e., as variables to be optimized in $\boldsymbol{J}_{4D}$. This may, for example, include the external forcing fields such as atmospheric and open boundaries conditions, the ocean topography, and/or internal parameters of ocean physics such as ocean mixing parameters, as has been successfully implemented in the Estimation of the Circulation and the Climate of the Ocean (ECCO) consortium (Wunsch and Heimbach, 2007; Köhl and Stammer, 2008) and the Regional Ocean Modeling System (ROMS; Moore et al., 2011). The objective function in this case would be comprised of the standard model-data misfit term along with prior and regularization terms to constrain the adjustments to the

optimized variables similar to the background term. This should be viewed as another approach to implementing a weak 4DVAR in which the model errors are non-additive but directly accounted for through appropriate dynamical parameterizations in the ocean model, which may help reduce the dimension of the optimization problem and impose dynamically balanced solutions for the adjusted variables and the estimated ocean state.

Solution of Four-Dimensional Variational (4DVAR) Assimilation

As in 3DVAR, gradient-based optimization algorithms are the standard methods to compute the 4DVAR solution. However, an important difference arises from the requirement that the solution needs to obey the model equations (Eq. 11). This leads to a so-called constrained optimization problem that is solved with the variational method. The variational principle, which is in functional space identical to setting the derivative of the objective function to zero, leads to the Euler-Lagrange equations. The latter are the adjoint equations to the tangent linear model equations, hence referred to as the adjoint method. The adjoint method provides the gradient by integrating the adjoint model backward in time, and is the most common approach to compute the gradient of the 4DVAR objective function $\boldsymbol{J}_{4D}$ (Le Dimet and Talagrand, 1986).

To understand why the backward integration of the adjoint gives the gradient, consider the strong constraint 4DVAR cost function. Using the chain rule for the derivatives of composite functions, one obtains

$$\nabla_{\mathbf{x}_0} J_{4D} = -2\sum_{k=0}^{L} \mathbf{M}_{k:0}^T \mathbf{H}_k^T \mathbf{R}_k^{-1}(\mathbf{y}_k - H_k(\mathbf{x}_k)) + 2\mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b), \tag{18}$$

where

$$\mathbf{M}_{k:0} = \mathbf{M}_k \cdots \mathbf{M}_1 \mathbf{M}_0, \quad \mathbf{M}_i = \nabla_{\mathbf{x}_i} M_i, \quad \text{and} \quad \mathbf{H}_i = \nabla_{\mathbf{x}_i} H_i\,. \tag{19}$$

This shows that the gradient of $\boldsymbol{J}_{4D}$ can be computed as $-2\tilde{\mathbf{x}}_0$ by integrating the adjoint model backward in time

$$\begin{cases} \tilde{\mathbf{x}}_{L+1} = 0, \\ \tilde{\mathbf{x}}_k = \mathbf{M}_k^T \tilde{\mathbf{x}}_{k+1} + \mathbf{H}_k^T \mathbf{R}_k^{-1}(\mathbf{y}_k - H_k(\mathbf{x}_k)) \quad \text{for} \quad k = L, \ldots, 0. \end{cases} \tag{20}$$

$\tilde{\mathbf{x}}$ is the so-called adjoint variable to $\mathbf{x}$. Comparing Eq. 20 with Eq. 18 shows that $\tilde{\mathbf{x}}_0$ is the gradient of Eq. 18. Moreover, it is obvious that $\tilde{\mathbf{x}}$ does not only provide the gradient at the initial time (and any given time), but will further provide gradients to model parameters (and inputs). For linear problems, the solution can be calculated directly from the set of adjoint and forward model equations. For nonlinear cases, the solution is computed iteratively, with each optimization iteration requiring one integration of the forward model starting from the parameter changes of the most recent iteration, based on the trajectory of which another integration of the adjoint model is performed backward in time to compute the gradient of the cost function.

This same adjoint machinery is also at the basis of the weak constraint 4DVAR problem, following the dual formulation (Courtier, 1997) or the Representer method (Bennett, 2005). Both approaches transfer the inversion into the data space, which allows to drastically reduce the

dimension of the weak 4DVAR optimization problem since the number of ocean data is commonly much smaller than the ocean state. The adjoint model is also used to compute the gradients of the 4DVAR objective function with respect to any model parameters, again using the chain rule (see, for example, Heimbach et al., 2002).

The weak 4DVAR methods provide very powerful tools to fit the ocean models to the available data; making the 4DVAR inversion problem highly underdetermined. Optimizing model errors at frequent model steps may enable efficient fit to the ocean observations but with a real risk of data overfitting and non-dynamical model errors adjustments. The role of the model errors covariance matrices $\mathbf{Q}_k$ becomes crucial, with unfortunately no established or efficient way to define these matrices (Wunsch, 1996; Hoteit et al., 2010).

Coding the adjoint model requires implementing the tangent linear model of the ocean model and its adjoint, and this can be a very demanding process. Automatic compilers have been developed to directly generate the adjoint code from the source code of the dynamical model (Giering and Kaminski, 1998). These may greatly facilitate the process of developing and maintaining the adjoint model to keep it up-to-date with forward model changes, but also impose some formats in the coding of the (forward) ocean model (Vlasenko et al., 2016). In addition to the technical challenge of generating an adjoint model, running the adjoint iteratively multiplies the cost of running a simulation by a factor of several hundreds. An additional difficulty arises in nonlinear models from the fact that the whole model trajectory needs to be known and stored at the time when the adjoint model is running. Checkpointing methods could be implemented into the adjoint code generation tools to efficiently reconstruct the trajectory (Heimbach et al., 2002).

Increasing efforts are being made to develop efficient methods that allow to either simplify the task of developing an adjoint code through reduced-order techniques, or completely by-passing the adjoint model through direct computation of the 4DVAR objective function gradients from forward model runs only. Reduced-order methods were developed around three related directions (Altaf et al., 2013a): (i) apply the optimization in a reduced space as a way to reduce the dimension of the optimization space to speed up the convergence rate (Robert et al., 2005; Hoteit and Köhl, 2006); (ii) develop a reduced-order model of the ocean model from which the adjoint model is derived (Vermeulen and Heemink, 2006; Fang et al., 2009); or (iii) directly develop a reduced-order adjoint model while still using the original forward ocean model for forward integrations (Altaf et al., 2013a; Yaremchuk et al. 2016). In this context, ensemble methods became popular as they were suggested to provide efficient tools to compute the gradients of the 4DVAR objective function, or to be used to parameterize the adjoint space in a hybrid assimilation framework (more on this in section below). Other adjoint-free optimization methods were tested, but these require dimension reduction before implementation to reduce their prohibitive computational burden (e.g., Hoteit, 2008).

The main difficulty in applying the adjoint method to ocean data assimilation problems is due to the nonlinear nature of the equations governing their dynamics. This problem is expected to become more severe as the resolution of ocean models continues to increase. In this case, the 4DVAR objective function becomes too irregular (non-convex), including multiple minima that

prevent a noticeable decrease in the objective function with gradient-based optimization techniques. This is associated with rapidly-growing response to perturbations ("intrinsic variability") that ultimately become unpredictable and are thus not controllable in the system. The adjoint gradient sensitivities then grow exponentially in time and become not useful in the optimization problem because for increasing assimilation windows the parameter range of validity of the linear approximation quickly becomes smaller than the uncertainty in the control parameters, which limits the length of the assimilation windows. In other words, the nonlinearity of the system invalidates the use of the gradient for descent (Pires et al., 1996; Köhl and Willebrand, 2002). Although short assimilation windows remain feasible, this may limit the benefit of the adjoint method, particularly since ocean observations are sparse and uncertain. Therefore, larger windows are desired to properly extract the large-scale parameter information via the dynamical constraint (Köhl and Willebrand, 2002). Large windows can be also useful to reduce dependence on the background covariance matrix and to provide enough time to infer enough sensitivities to, for instance, atmospheric forcing and/or boundaries conditions and other parameters if these were also to be adjusted in the 4DVAR system.

Since large scales are associated with longer predictability timescales, a way out is to separate the small from the large scales. This could be implemented by increasing viscosity and diffusivity terms in the backward adjoint run, which becomes close to the adjoint of a coarser, more linear model without local minima (Hoteit et al., 2005a). This approach works even with the original high-resolution forward model, because secondary minima become so dense over long periods of time that they appear as stochastic perturbations (Hoteit et al., 2005a). The limitation of this approach is that it may also start filtering out large-scale features over time because of the tight coupling between the different scales in the ocean. An alternative could be based on ensemble methods (Lea et al., 2000), but since the number of required ensembles grows as the gradients increase, this method quickly becomes unfeasible. It is still not clear to what extent the smoothing nature of the reduced-order and ensemble-based approaches, whether to derive an approximate adjoint or to directly compute the gradients, could mitigate this issue and help extend the assimilation windows.

A new approach was recently borrowed from the chaos theory to tackle the issues with nonlinearities, and it was applied for ocean parameters estimation of a climate model. Noticing that a coupling leads to synchronization of similar chaotic systems over long periods of time, a parameter estimation method based on the ability of a parameter-dependent system to synchronize with observations was developed in physics (Abarbanel, 2012). The coupling to the observations is included in the model as a relaxation term that, when strong enough, ultimately will turn the system into a non-chaotic system in which parameter estimation with the adjoint method may become again feasible. A caveat to this method is that the estimation takes place in a modified system and may no longer be optimal with respect to the original system. Moreover, because synchronization turns into a damping in the adjoint model, data will have a limited effect (in time) on the estimated parameters, particularly the initial conditions.

Recently, several 4DVAR regional ocean operational systems have been successfully developed and are currently routinely providing forecasts of ocean states, including the real-time forecasting

system of the Mid-Atlantic Bight (Zhang et al., 2010), the University of California Santa Cruz California Current forecasting system (Moore et al., 2011), the University of Hawaii forecasting system for the region surrounding the main Hawaiian Islands (Janeković et al., 2013), and the Navy Coastal Ocean forecasting of the Okinawa Trough (Smith et al., 2017b).

4DVAR is designed to compute the MAP as the final solution for the DA problem but not its covariance, which will be needed as a background for the next assimilation cycle. Although this could be conveniently estimated using the adjoint to compute a low-rank approximation of the Hessian matrix of the 4DVAR objective function, which is the inverse of the MAP error covariance matrix when the system is linear and the noise is Gaussian (Smith et al., 2015), strong nonlinearities mean that the Hessian will probably be computed around a local minima and may not reflect the global errors in estimation. The next section will present the Bayesian filtering approach that aims, in contrast, at directly computing the full conditional probability distribution of the ocean state given available observations.

Bayesian Filtering

The Bayesian estimation problem can be solved sequentially in time, as the observations become available. This is known as Bayesian filtering and it readily provides a suitable framework for operational oceanography where an ocean model is used for forecasting and the data are assimilated to update the model forecasts with the Bayes' rule every time they become available. Here, we are interested in estimating the ocean state at a given time t_k given all available observations up to t_k, which in a Bayesian setting involves the computation of $p_{\boldsymbol{x}_k|\boldsymbol{y}_0,\ldots,\boldsymbol{y}_k}$. Marginalizing Eq. 12, the filtering solution is then identical to the Bayesian estimator (Eq. 12) at the end of the assimilation window. This contrasts with a "smoother," which involves observations beyond time instant t_k, e.g., 4DVAR and ensemble Kalman smoothers (more on this in section below). This section will focus on the state estimation problem and introduce DA algorithms from a Bayesian filtering perspective. Most of these algorithms, as those presented below, aim at computing the MV estimator instead of the MAP estimator of 4DVAR.

The basis of Bayesian filtering is a state space model comprised of a dynamical model (Eq. 11) and an observation model (Eq. 1) that provides measurements of the ocean state in time. These provide $p_{\boldsymbol{x}_k|\,\boldsymbol{x}_{k-1}}$ and the likelihood $p_{\boldsymbol{y}_k|\,\boldsymbol{x}_k}$, respectively. Using standard conditional probability rules (e.g., Simon, 2006and Law et al., 2015), one can write

$$\begin{aligned} p_{\mathbf{x}_k|\mathbf{y}_0,\ldots,\mathbf{y}_k} &\propto p_{\mathbf{y}_k|\mathbf{y}_0,\ldots,\mathbf{y}_{k-1},\mathbf{x}_k} \cdot p_{\mathbf{x}_k|\mathbf{y}_0,\ldots,\mathbf{y}_{k-1}}, \\ &\propto p_{\mathbf{y}_k|\mathbf{x}_k}\; p_{\mathbf{x}_k|\mathbf{y}_0,\ldots,\mathbf{y}_{k-1}}. \end{aligned} \tag{21}$$

The posterior or analysis distribution is thus the product of the likelihood of the state given the new observation and the forecast (prior) distribution, which is the distribution of the state conditioned on all previous observations. This is called the update or analysis step. The forecast distribution can be computed from the analysis distribution at the previous time step by first computing the joint

distribution of ($\mathbf{x}_k$, $\mathbf{x}_{k-1}$) conditioned on $\mathbf{y}_0, \ldots, \mathbf{y}_{k-1}$, then integrating over $\mathbf{x}_{k-1}$ to obtain the desired marginal distribution as

$$p_{\mathbf{x}_k|\mathbf{y}_0,\ldots,\mathbf{y}_{k-1}} = \int p_{\mathbf{x}_k|\mathbf{x}_{k-1}}\, p_{\mathbf{x}_{k-1}|\mathbf{y}_0,\ldots,\mathbf{y}_{k-1}}\, d\mathbf{x}_{k-1}. \tag{22}$$

Therefore, if the analysis distribution is available at a given time, one can first compute the forecast distribution using Eq. 22, and then compute the analysis distribution at the next time using Eq. 21. One can then proceed recursively, starting from a prior distribution at the initial time and then alternate forecast and analysis steps to compute the analysis distribution at any given time. The filtering procedure is thus similar to a 3DDA procedure (as illustrated in Fig. 17.2), but operates on the state distribution rather than the state. This recursive framework provides a solution to the estimation problem and conceptually leads to the so-called optimal filter. In practice, however, difficulties often arise in computing the filter solution (distributions), largely due to the fact that evaluating the integrals in Eqs. 21-22 are numerically intractable in high-dimensional problems, such as the ocean. For such applications, one has to adopt certain approximations to derive some sub-optimal filters that provide accurate enough results at reasonable computational requirements. Reviewing different approaches for sub-optimal filtering is the focus of this section.

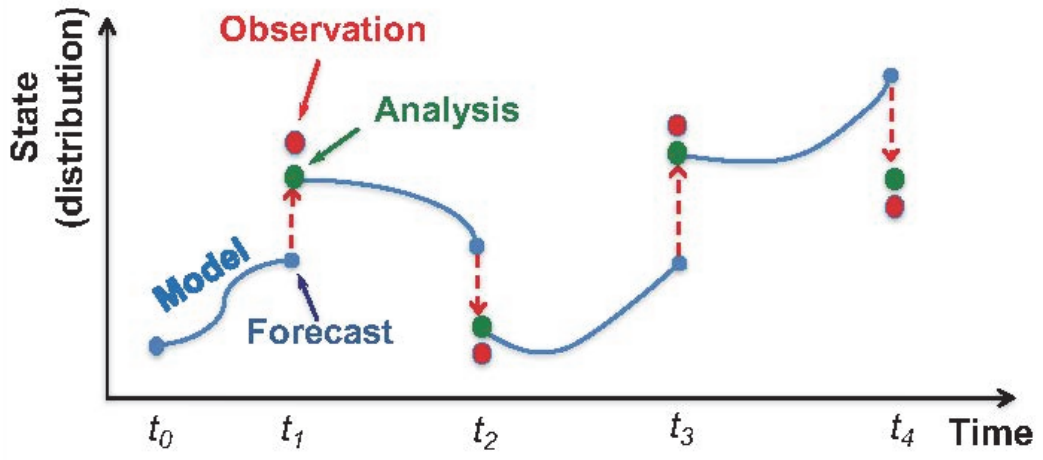

Figure 17.2. Schematic diagram of the sequential (3D and filtering) assimilation procedures, including 3DVAR, OI, KF, EnKFs, and PF. During the forecast step, the ocean model is integrated to the time of the next available observation starting from the most recent analysis. During the analysis step, the forecast is updated using the incoming observation to compute the analysis. Four assimilation cycles are shown.

Starting with the celebrated Kalman filter (KF; Kalman, 1960), which is designed for linear systems (dynamics and observations) and Gaussian errors, we present variants of the KF that enable its implementation for DA into large-scale nonlinear models. This includes the low-rank extended Kalman filters (LR-EKFs) and ensemble Kalman filters (EnKFs). The KF linear update step does not hold with the nonlinear ocean dynamics and for some of the ocean observations (e.g., acoustics). We will also present two nonlinear/non-Gaussian filters that are currently being investigated for potential use with realistic ocean data assimilation problems, the particle filter (PF) and the

Gaussian mixture filter (GMF). A schematic diagram illustrating the various filtering strategies is provided in Fig. 17.3.

The Kalman Filter (KF)

In the KF (Kalman, 1960), the dynamical and observation operators in (Eqs. 11 and 1 are linear, and thus denoted respectively by **M** and **H** in this section; and the associated noise ε_k and η_k are Gaussian (i.e., the likelihood function $p_{\boldsymbol{y}_k|\,\boldsymbol{x}_k}$ and the transition distribution $p_{\boldsymbol{x}_k|\,\boldsymbol{x}_{k-1}}$ in (21) and (22) are Gaussian), independent in time and mutually independent. Starting from some initial Gaussian distribution, the forecast and posterior distributions, $p_{\boldsymbol{x}_k\,|\boldsymbol{y}_0,\ldots,\boldsymbol{y}_{k-1}}$ and $p_{\boldsymbol{x}_k\,|\boldsymbol{y}_0,\ldots,\boldsymbol{y}_k}$ remain Gaussian at subsequent time instants. The algorithm of the KF therefore reduces to recursively computing the means and covariance matrices of $p_{\boldsymbol{x}_k\,|\boldsymbol{y}_0,\ldots,\boldsymbol{y}_{k-1}}$ and $p_{\boldsymbol{x}_k\,|\boldsymbol{y}_0,\ldots,\boldsymbol{y}_k}$, which fully characterize their distributions. These represent the forecast and analysis MAPs (and MVs, which are identical in this case) and their errors covariance matrices, and are computed by recursive cycles of the following forecast and analysis steps.

Forecast step: Integrate the posterior mean, i.e., analysis state, $\mathbf{x}^a_{k-1}$ at time instant $t_{k\text{-}1}$ and the associated error covariance $\mathbf{P}^a_{k-1}$ forward with the dynamical model (Eq. 11) to compute the forecast state $\mathbf{x}^f_k$ and the associated error covariance $\mathbf{P}^f_k$ at the time of the next available observation t_k, as

$$\mathbf{x}^f_k = \mathbf{M}_k \mathbf{x}^a_{k-1}\,, \tag{23a}$$

$$\mathbf{P}^f_k = \mathbf{M}_k \mathbf{P}^a_{k-1} \mathbf{M}^T_k + \mathbf{Q}_k\,. \tag{23b}$$

Analysis step: Once the new observation $\mathbf{y}_k$ is available, update the forecast statistics $\mathbf{x}^f_k$ and $\mathbf{P}^f_k$ to their analysis counterparts, $\mathbf{x}^a_k$ and $\mathbf{P}^a_k$, using the BLUE (which is also the MAP here) in Eq. 7 with $\mathbf{B} = \mathbf{P}^f_k$,

$$\mathbf{x}^a_k = \mathbf{x}^f_k + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k \mathbf{x}^f_k)\,, \tag{24a}$$

$$\mathbf{P}^a_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{P}^f_k\,, \tag{24b}$$

$$\mathbf{K}_k = \mathbf{P}^f_k \mathbf{H}^T_k (\mathbf{H}_k \mathbf{P}^f_k \mathbf{H}^T_k + \mathbf{R}_k)^{-1}\,. \tag{24c}$$

Therefore, the only difference from a 3D assimilation setting is in the use of a time- (or flow-) dependent forecast (background) error covariance matrix that is updated in the analysis step as in Eq. 24b to account for the reduction in the estimation error (uncertainties) after the assimilation of an observation. The resulting analysis error covariance is then integrated by the model forward as in Eq. 23b to reflect an increase, or eventual decrease, in the initial analysis error during forecasting, depending on the ocean dynamics during that period (Pham et al., 1997), plus the contribution of the model errors.

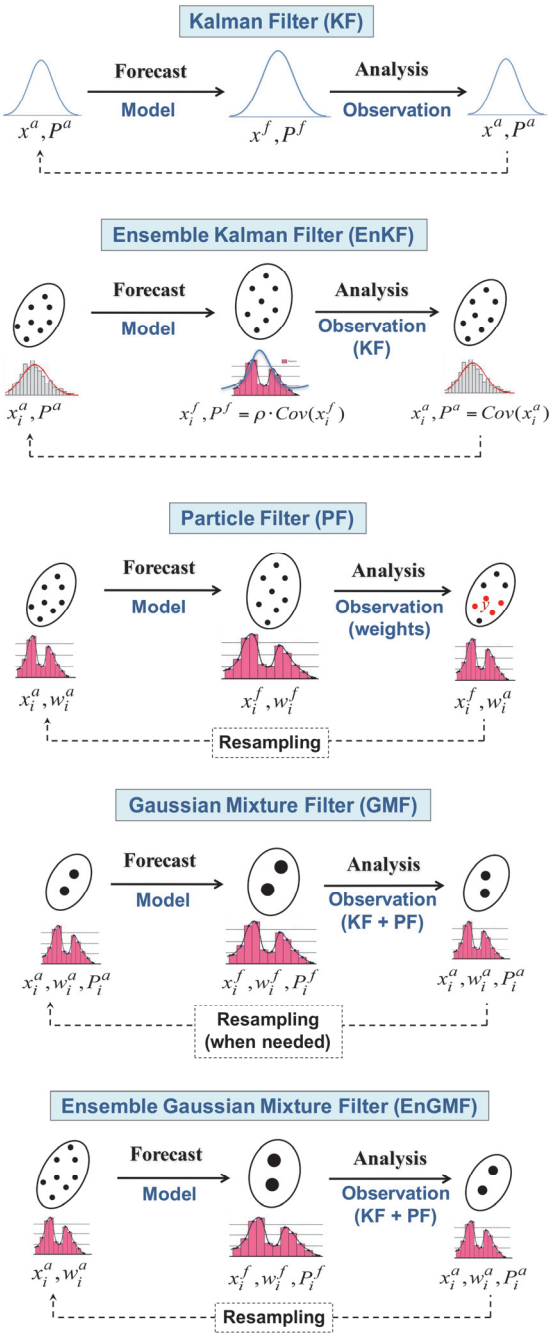

Figure 17.3. Schematic illustration of the different filtering strategies presented in Section 4. The Kalman filter (KF) only involves Gaussian distributions, characterized by their mean and covariances. Both the ensemble Kalman filter (EnKF) and the Particle filter (PF) integrate a set of ocean states sampled from the present distribution forward during the forecast step. While the EnKF assumes Gaussian background at the analysis step so that the forecast ensemble is updated with the incoming observation as in the KF, the PF applies a non-Guassian update to the samples weights only. The Gaussian mixture filter (GMF) maintains Gaussian mixture distributions at the forecast and analysis steps, in which the mixture weights are updated as in the PF and the mixture covariances as in the KF. As the EnKF, the ensemble GMF (EnGMF) integrates an ensemble of ocean states at the forecast step, but assumes Gaussian mixture background so that it applies a GMF analysis.

The application of the KF for ocean DA is hampered by (i) the nonlinear nature of the ocean dynamics (and eventually of some ocean observations), and (ii) the large dimension of the OGCMs (which is reaching up to $n \sim O(10^{10})$ in today's numerical ocean models). The first means that the KF cannot be directly implemented, and the second implies prohibitive computational burden (in term of storage and computation) in order to manipulate the KF error covariance matrices, of dimensions $n \times n$. Different simplified variants of the KF have been therefore proposed for ocean data assimilation, which can be split into two main categories, reduced extended Kalman filters and ensemble Kalman filters.

Reduced Extended Kalman Filters (REKFs)

To apply the KF to ocean assimilation problems, one can compute the forecast state as in Eq. 23a by just integrating the analysis state forward with the nonlinear model. Implementing the KF forecast error covariance calculation step (Eq. 23b) is, however, not as straightforward. A popular approach is to linearize the model (and eventually the observation operator) using, for instance, a first-order Taylor expansion, and then apply the KF to the linearized system. This leads to the popular, but no longer optimal, extended Kalman filter (EKF) (Jazwinski 1970), and eventually its higher-order variants depending on the retained order of the Taylor expansion (Anderson and Moore, 1979; Simon, 2006).

To avoid the prohibitive computational requirements of the EKF due to the large numerical dimension of realistic ocean models, different forms of reduced-state space or reduced-error space (i.e., low-rank error covariance matrix) approximations have been proposed (e.g., Fukumori and Malanotte-Rizzoli, 1995; Cane et al., 1996); Cohn and Todling, 1996; Verlaan and Heemink, 1997; Pham et al., 1997; Lermusiaux and Robinson, 1999; Farrell and Ioannou, 2001; Hoteit et al., 2002). A common feature of these REKFs is that they exploit information from a representative subspace of the full ocean state, or error subspace, and ignore information from the less influential complement subspace. This is supported by the dissipative and driven nature of the ocean dynamics, which concentrates energy at large scales, imposing a red spectrum of variability (Daley, 1991; Pham et al., 1997; Lermusiaux and Robinson, 1999). Consequently, EKF calculations are conducted on the retained subspace only, dramatically reducing the computational cost. The reduced state/error spaces, denoted by $\mathbf{L}$, can be set invariant in time, as in the reduced order EKF (ROEKF), or left to evolve with the model dynamics as in the singular evolutive extended Kalman (SEEK) filter, with the latter leading to more robust state estimates during periods of strong ocean variability (Hoteit and Pham, 2003). Both schemes operate with low-rank r EKF error covariance matrices, with r the dimension of the reduced space, while keeping the rest of the EKF algorithm mostly unchanged. The error covariance matrices are only evaluated through their low-rank counterparts, $\mathbf{L}$ and $\mathbf{U}$, where $\mathbf{P} = \mathbf{L}\mathbf{U}\mathbf{L}^T$ and $\mathbf{U}$ a $r \times r$ matrix representing the error variance in the reduced space, which avoids the storage of $\mathbf{P}$ and drastically reduces the EKF computational burden. More details can be found in Cane et al. (1996) and Pham et al. (1997). One caveat of this approach is in the treatment of the model error covariance matrix $\mathbf{Q}$ in Eq. 23b, as the rank of the forecast error covariance matrix $\mathbf{P}$ cannot indeed be preserved after adding $\mathbf{Q}$ unless the model error is projected on (and

therefore only treated in) the reduced space **L**, or simply neglected assuming perfect model (**Q** = 0) (Hoteit et al., 2007). This implies another approximation in the final EKF algorithm and would inevitably lead to an underestimation of the forecast error covariance.

REKFs have been applied to ocean DA in the global ocean as part of the Estimating the Circulation and Climate of the Global Ocean Data Assimilation Experiment (ECCO-GODAE) system (Kim et al., 2006), in the Pacific Ocean (Cane et al., 1996; Verron et al., 1999; Hoteit et al., 2002), and regionally in a nested implementation of the Ligurian Sea (Barth et al. 2007). They are also used operationally in, for instance, Monterey Bay (Haley et al., 2009), the Greek national POSEIDON-II system for the Mediterranean (Korres et al., 2010), and the European MERCATOR system (Lellouche et al., 2013). Given the complexity of the linearization step and its limitation with strongly nonlinear models (Evensen, 1994), as well as the difficulty of specifying and evolving a reduced subspace, these methods have dramatically lost popularity in recent years owing to the advances in ensemble Kalman filtering methods.

Ensemble Kalman Filters (EnKFs)

The main idea behind the EnKFs is to apply a Monte Carlo-like forecast step to integrate the KF analysis state and its error covariance forward through a set, or ensemble, of ocean states sampled from these two statistical moments (Evensen, 2003). The sampled analysis ensemble is then integrated forward with the (nonlinear) model to obtain the forecast ensemble, from which the forecast state and error covariance are taken as the sample mean and covariance of the ensemble. A KF analysis step is then applied to update the forecast ensemble every time a new observation is available. The ensemble formulation allows to avoid the manipulation of the KF error covariance matrices by performing the calculations on the ensemble members, which enables the implementation of the filter on large-scale ocean applications. Generally speaking, the Monte-Carlo forecast step requires N (= ensemble-size) ocean model integrations to compute the forecast ensemble, and the KF update step is applied in the low-rank ensemble subspace, typically of a dimension $N - 1$ (Pham, 2001). Another important advantage of the Monte Carlo forecast step is the possibility of implicitly accounting for the model errors through perturbations sampled from their distributions and then carried with the ensemble model runs (Evensen, 2003; Hoteit et al., 2007). This further allows avoiding the additive model error assumption, which is otherwise less general and difficult to account for in the REKFs (Pham et al., 1997; Hoteit et al., 2005b).

Because of their non-intrusive formulation and ease of implementation, remarkable robustness and effectiveness, and reasonable computational requirements, EnKF methods have become very popular in the geosciences. Many variants of the EnKF have been proposed in the literature, but a full review is beyond the scope of this chapter. They all operate as cycles of Monte Carlo forecast and KF update steps involving only the first two moments of the ocean state posterior, basically only differing in the sampling scheme of their analysis ensembles. Depending on whether the observations are perturbed before assimilation or not, the EnKFs are customarily classified as one of two types (Tippett et al., 2003): stochastic EnKFs (Burgers et al., 1998; Houtekamer et al., 2005; Hoteit et al., 2015) and deterministic EnKFs (Anderson, 2001; Bishop et al., 2001; Whitaker and

Hamill, 2002; Hoteit et al., 2012; Luo and Hoteit, 2014c). A stochastic EnKF essentially updates each forecast ensemble member with perturbed observations during the KF correction step. By contrast, a deterministic EnKF updates the ensemble mean and a specific (square-root) form of the sample (ensemble) error covariance matrix exactly as in the KF, without perturbing the observations. An analysis ensemble is then produced from the updated mean and covariance prior to the forecast step. The most popular deterministic EnKFs with publicly available codes are the singular evolutive interpolated KF–SEIK (Pham, 2001; Hoteit et al., 2002), the ensemble transform KF–ETKF (Bishop et al., 2001; Wang et al., 2004; Hunt et al., 2007), and the ensemble adjustment KF–EAKF (Anderson, 2001, 2009). With the continuous advances in computing capabilities, EnKF methods are becoming increasingly popular in the development of ocean operational systems, e.g. Toye et al. (2017). An EnKF is, for instance, already used operationally in the Norwegian North Atlantic and Arctic forecasting system, TOPAZ (Sakov et al., 2012). Below we focus on presenting the algorithms of the two "basic" forms of ensemble Kalman filtering; the original stochastic EnKF (Burgers et al. 1998; Houtekamer and Mitchell 1998) and two standard, closely-related deterministic EnKFs.

1) STOCHASTIC EnKF (SEnKF)

Assume an N-member analysis ensemble $\mathbf{X}^a_{k-1} = [\mathbf{x}^{a,i}_{k-1}, i = 1, 2 \ldots, N\}$ is available at the end of the $(k-1)^{th}$ assimilation cycle. The forecast ensemble at the next time t_k is obtained by integrating $\mathbf{x}^{a,i}_{k-1}$ with the dynamical model (11), i.e.

$$\mathbf{X}^f_k = \left\{ \mathbf{x}^{f,i}_k = M_k(\mathbf{x}^{a,i}_{k-1}) + \eta^i_k,\ i = 1, \ldots, N \right\}, \tag{25}$$

where η^i_k are sample dynamical noise drawn from the distribution of the model error term. The ensemble sample mean and covariance are taken as the forecast state and its error covariance matrix, respectively as

$$\mathbf{x}^f_k = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}^{f,i}_k, \qquad \mathbf{P}^f_k = \frac{1}{N-1}\sum_{i=1}^{N} \left(\mathbf{x}^{f,i}_k - \mathbf{x}^f_k\right)\left(\mathbf{x}^{f,i}_k - \mathbf{x}^f_k\right)^T. \tag{26}$$

In practice, $\mathbf{P}^f_k$ needs not be calculated. Instead, it is customary to approximate

$$\mathbf{P}^k_{xh} = \frac{1}{N-1}\sum_{i=1}^{N} \left(\mathbf{x}^{f,i}_k - \mathbf{x}^f_k\right)\left(\mathbf{y}^{f,i}_k - \mathbf{y}^f_k\right)^T, \tag{27}$$

$$\mathbf{P}^k_{hh} = \frac{1}{N-1}\sum_{i=1}^{N} \left(\mathbf{y}^{f,i}_k - \mathbf{y}^f_k\right)\left(\mathbf{y}^{f,i}_k - \mathbf{y}^f_k\right)^T,$$

where

$$\mathbf{y}^{f,i}_k = H_k(\mathbf{x}^{f,i}_k) \quad \text{and} \quad \mathbf{y}^f_k = \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}^{f,i}_k, \tag{28}$$

which enables to avoid the linearization of the nonlinear observation operator. The Kalman gain $\mathbf{K}_k$ is then approximated as

$$\mathbf{K}_k \approx \mathbf{P}^k_{xh}\left(\mathbf{P}^k_{hh} + \mathbf{R}_k\right)^{-1}. \tag{29}$$

When a new observation $\mathbf{y}_k$ is available, one computes the analysis ensemble from the forecast ensemble using the KF update step, as

$$\mathbf{x}_k^{a,i} = \mathbf{x}_k^{f,i} + \mathbf{K}_k \left(\mathbf{y}_k^{\varepsilon,i} - H_k(\mathbf{x}_k^{f,i}) \right), \quad \text{for} \quad i = 1, \cdots, N, \tag{30}$$

where $\mathbf{y}_k^{\varepsilon,i}$ are perturbed observations generated by adding to the observation $\mathbf{y}_k$ random perturbations sampled from the distribution of the observational error. Accordingly, the sample mean and covariance of the analysis ensemble $\mathbf{X}_k^a = \{\mathbf{x}_k^{a,i} : i = 1, \cdots, N\}$ are obtained in the spirit of Eq. 26, and the observations perturbations guarantee that they converge to the KF analysis and its error covariance with increasing ensemble size N. Integrating $\mathbf{X}_k^a$ forward to the time of the next available observation, one starts a new assimilation cycle, and so on.

The first two moments of the SEnKF analysis may only asymptotically match those of the KF (Evensen, 2003). In this sense, the SEnKF update step always introduces noise during the analysis (Nerger et al., 2005). The noise may become pronounced in typical oceanic data assimilation applications where the rank of the observational error covariance matrix $\mathbf{R}_k$ is much larger than the ensemble size, meaning that $\mathbf{R}_k$ will be greatly under-sampled (Altaf et al., 2014). Spurious correlations between the observation perturbations and the forecast perturbations could also lead to errors in the EnKF sample analysis error covariance matrices (Pham, 2001; Bowler et al., 2013). To mitigate this issue, one can either introduce a certain correction scheme as in Hoteit et al. (2015), or simply avoid perturbing the observations following a deterministic EnKF formulation, which will be discussed next. In the opposite case, that is when the ensemble size is larger than the number of observations, the SEnKF was shown to perform better than other EnKFs without perturbations in many situations (Anderson, 2010; Hoteit et al., 2012). Lawson and Hansen (2004) argued that the observation perturbations in the SEnKF tend to re-Gaussianize the ensemble distribution to explain the improved stability. Lei et al. (2010) also demonstrated that the SEnKF is generally more stable in certain circumstances, especially in the presence of wild outliers in the data. An important advantage of the SEnKF update step is that it readily provides an analysis ensemble for forecasting (that is randomly sampled from the assumingly Gaussian analysis distribution), avoiding the deterministic updating step that may distort some of the features of the forecast ensemble distribution as in the other ensemble KFs without perturbations (Lei et al., 2010; Hoteit et al., 2015). This allows more straightforward implementation of some auxiliary techniques, such as covariance localization and hybrid schemes, as will be further discussed below.

2) DETERMINISTIC EnKFS (DEnKFS)

The DEnKFs analysis ensemble is deterministically generated in order to perfectly match the KF estimate, and thus avoid the random perturbations of the SEnKF. There are infinite ways to match a mean and a covariance by an ensemble and accordingly various DEnKFs have been proposed, many of which are based on the square-root formulation of the KF, which was introduced as an approach to improve the stability of the KF by working on a certain square-root of the filter covariance matrix. Anderson (2001), Bishop et al. (2001), and Whitaker and Hamill (2002) exploited the readily square-root form of the ensemble-based covariances to propose deterministic EnKFs assimilating the data serially, one at a time, assuming uncorrelated observational errors (i.e.,

diagonal $\mathbf{R}_k$)[1]. This enables efficient (parallel) assimilation of very large number of observations (Houtekamer and Zhang, 2016). In contrast, the ensemble transform Kalman filter (ETKF; Bishop et al., 2001) and the singular evolutive interpolated Kalman (SEIK) filter (Pham, 2001; Hoteit et al., 2002) can directly handle any form of $\mathbf{R}_k$ by computing the analysis increment in the ensemble subspace. To assimilate large numbers of observations, one may apply local analysis steps using only neighbor observations (as in optimal interpolation), which needs to be used anyway to deal with the low-rank nature of ensemble sampled covariances, as will be further discussed below.

Here, we present the ETKF for illustration and discuss its similarities with SEIK (Nerger et al., 2012). Following the same forecast step as the SEnKF, a forecast ensemble $\mathbf{X}_k^f = \{\mathbf{x}_k^{f,i}, i = 1, \cdots, N\}$ is available at time t_k as in Eq. 25, with the forecast state as the sample mean $\mathbf{x}_k^f$ and its error covariance as $\mathbf{P}_k^f$ given by Eq. 26. Instead of directly working on $\mathbf{P}_k^f$, one first constructs a square-root $\mathbf{S}_k^f$ of $\mathbf{P}_k^f$ from the forecast ensemble perturbations,

$$\mathbf{S}_k^f = \frac{1}{\sqrt{N-1}} \left[\mathbf{x}_k^{f,1} - \mathbf{x}_k^f, \cdots, \mathbf{x}_k^{f,N} - \mathbf{x}_k^f \right]; \tag{31}$$

and its equivalent in the observation space,

$$\mathbf{x}_k^{a,i} = \mathbf{x}_k^a + \sqrt{N-1}\,(\mathbf{S}_k^a)_i, \quad \text{for } i = 1, \cdots, N. \tag{32}$$

ETKF updates the ensemble forecast and error covariance exactly as in the KF, using Eqs. 24a and 24c,

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k \left(\mathbf{y}_k - \mathbf{y}_k^f \right), \tag{33a}$$

$$\mathbf{K}_k = \mathbf{S}_k^f (\mathbf{S}_k^h)^T \left(\mathbf{S}_k^h \left(\mathbf{S}_k^h \right)^T + \mathbf{R}_k \right)^{-1}, \tag{33b}$$

whereas the covariance update formula (Eq. 24b) can be computed as

$$\mathbf{P}_k^a = \mathbf{S}_k^f \mathbf{V}_k \left(\mathbf{S}_k^f \right)^T, \tag{34a}$$

$$\mathbf{V}_k = \left((\mathbf{S}_k^h)^T \mathbf{R}_k^{-1} \mathbf{S}_k^h + \mathbf{I}_N \right)^{-1}, \tag{34b}$$

where $\mathbf{I}_N$ is the N-dimensional identity matrix. A spectral decomposition is then applied, so that

$$(\mathbf{S}_k^h)^T \mathbf{R}_k^{-1} \mathbf{S}_k^h = \mathbf{E}_k \mathbf{D}_k \mathbf{E}_k^T,$$

where $\mathbf{E}_k$ consists of eigenvectors of $(\hat{\mathbf{S}}_k^h)^T \mathbf{R}_k^{-1} \hat{\mathbf{S}}_k^h$, and $\mathbf{D}_k$ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues. One can then express $\mathbf{V}_k$ as

$$\mathbf{V}_k = \mathbf{T}_k (\mathbf{T}_k)^T, \quad \text{with} \quad \mathbf{T}_k = \mathbf{E}_k (\mathbf{D}_k + \mathbf{I}_N)^{-1/2}. \tag{35}$$

so that

$$\mathbf{P}_k^a = \mathbf{S}_k^a (\mathbf{S}_k^a)^T, \quad \text{where} \quad \mathbf{S}_k^a = \mathbf{S}_k^f \mathbf{T}_k. \tag{36}$$

[1] For correlated observational errors, one may transform the observations by the inverse of the square-root of the observation error covariance matrix $\mathbf{R}_k$ to obtain a new set of uncorrelated observations that could be serially assimilated.

Accordingly, the analysis ensemble $\mathbf{X}_k^a = \{\mathbf{x}_k^{a,i}, i = 1, \cdots, N\}$ can then be generated using

$$\mathbf{x}_k^{a,i} = \mathbf{x}_k^a + \sqrt{N-1}\,(\mathbf{S}_k^a)_i, \quad \text{for } i = 1, \cdots, N, \tag{37}$$

where $(\mathbf{S}_k^a)$ denotes the i^{th} column of $\mathbf{S}_k^a$, so that the sample covariance of $\mathbf{X}_k^a$ is exactly $\mathbf{P}_k^a$. This, however, does not guarantee that its sample mean is $\mathbf{x}_k^a$ in Eq. 33a unless

$$\sum_{i=1}^{N} (\mathbf{S}_k^a)_i = 0\,. \tag{38}$$

To avoid this bias, Wang et al. (2004) latter followed SEIK formulation and proposed to take $\mathbf{S}_k^a = \mathbf{S}_k^f \mathbf{T}_k \mathbf{Z}_k$, where

$$\mathbf{Z}_k \mathbf{1} = \mathbf{0} \quad \text{and} \quad \mathbf{Z}_k \mathbf{Z}_k^T = \mathbf{I}_N\,, \tag{39}$$

with $\mathbf{1}$ being a vector whose elements are all 1s. Another feature of the SEIK filter is the use of a "random" matrix $\mathbf{Z}_k$, aiming at "redistributing" the error variance among the ensemble members to help in the mitigation of ensemble degeneracy (Sakov and Oke, 2008).

Many other variants of the KF are implemented in a similar way to the DEnKFs, i.e. ensemble forward propagation for forecasting and Kalman-based update with the observations, with the only differences in the formulation of the analysis step to update the forecast ensemble to the analysis ensemble. We cite here, for example, the unscented Kalman filter Julier et al. (2000); Julier and Uhlmann (2004); Luo and Moroz (2009), the divided difference filter (Ito and Xiong, 2000; Luo et al., 2012). However, these generally require an ensemble size that is larger than the dimension of ocean state, which makes them prohibitive for large-scale applications. In contrast, the EnKF formulation is more efficient and has found numerous applications in many ocean data assimilation problems.

3) ENSEMBLE OPTIMAL INTERPOLATION (EnOI)

Integrating large ensembles with an OGCM is computationally demanding. Following the optimal interpolation formulation of the DA problem, which uses a static pre-selected background covariance in the update step, ensemble optimal interpolation (EnOI) methods were proposed (Evensen, 2003; Hoteit et al., 2002; Oke et al., 2007). EnOI is a very cost-effective alternative to an EnKF, in which the static background covariance is estimated as the sample covariance matrix of an adequately pre-selected ensemble, generally describing the error growing modes or representing the variability of the studied ocean. Ocean large-scale dynamics evolve slowly within relatively short time windows, which justifies keeping the ensemble members static in time or over certain periods (Hoteit et al., 2002). In doing so, the EnKF reduces to an OI scheme in which only the analysis mean is integrated forward with the model for forecasting. The update step could be implemented based on a stochastic (e.g., SEnKF) update scheme (Counillon and Bertino, 2009) or a deterministic (e.g. SEIK) update scheme (Hoteit et al., 2002). Dropping the model integration of the ensemble members from the EnKF algorithm not only allows to drastically reduce the computational burden, but also to avoid the degeneracy of its members (more on this below). The method was found to be quite competitive compared to an EnKF at a fraction of the computational cost (Hoteit et al., 2002; Oke et al., 2007; Sakov and Sandery, 2015; Toye et al., 2017). However,

its performance may be limited during periods of rapidly evolving dynamics, which are generally not well captured by a static background covariance (Hoteit et al., 2002; Hoteit and Pham, 2004). To account for the seasonal and intra-seasonal variability of the ocean flow, Xie and Zhu (2010) proposed to implement EnOI with an ensemble selected at every assimilation cycle from monthly climatology ocean states with a three-month moving window centered at the assimilation time. Currently, EnOI is used operationally in the Australian Bluelink system (Oke et al., 2008).

4) AUXILIARY TECHNIQUES TO ENHANCE THE PERFORMANCE OF ENKFS

Realistic EnKF ocean data assimilation problems are typically implemented with small ensembles of the order of 100 members or less (Aanonsen et al., 2009; Hoteit et al., 2015; Houtekamer and Zhang, 2016) to restrict the computational cost. The downside of using small ensembles is, however, twofold: (i) rank-deficient (low rank or degrees-of-freedom) forecast/background ensemble compared to the ocean state and observations dimensions (Hamill et al., 2009; Houtekamer and Zhang, 2016), and (ii) important Monte Carlo sampling errors (Anderson, 2012; Hamill et al., 2001; Luo et al., 2018). Together with the ubiquitous nonlinearity of the ocean dynamics, the implementation of EnKFs with small ensembles for OGCM DA requires explicit compensation for the effects of a finite ensemble. For instance, Bocquet et al. (2015) derived a prior probability density function conditional on the background ensemble to account for the sampling errors due to a small ensemble. Iterative methods were also introduced in the framework of the EnKFs to deal with strong nonlinearities. Many other auxiliary techniques have been proposed in the literature, including the most popular ones presented here: *covariance inflation*, *localization*, and *hybrid covariance*.

(i) Covariance inflation: As suggested by its name, covariance inflation "inflates" the covariance of an EnKF forecast, or analysis, ensemble by some positive factor at each assimilation cycle. The rationale behind covariance inflation can be explained from different points of views. It is, for instance, often justified based on the observation that the EnKF ensemble covariances are systematically underestimated due to the effect of finite ensembles (Whitaker and Hamill, 2002) and/or to account for neglected model errors (Pham et al., 1997; Hoteit and Pham, 2004). This helps mitigateg the ensemble collapse due to a lack of spread. In such cases, covariance inflation may be directly applied to the background ensemble through either an additive (Houtekamer and Mitchell, 2005; Lee et al., 2017; Yang et al., 2015) or multiplicative (Anderson and Anderson, 1999; Anderson, 2007a, 2009; Miyoshi, 2011) factor, or to the analysis ensemble through a certain relaxation term (Whitaker and Hamill, 2012; Zhang et al., 2004). An alternative point of view is to relate covariance inflation to robust filtering (Luo and Hoteit, 2011) in the context of an ensemble implementation of the H_∞ filter (Simon, 2006). This leads to different forms of inflation, including the conventional additive, multiplicative, or relaxation methods, and also encompasses the less conventional inflation methods such as those modifying the eigenvalues of the estimation error covariance in the ensemble space (Altaf et al., 2013b; Ott et al., 2004; Bai et al., 2016).

In practice, the value of the inflation factor is often set by trial and error, but adaptive inflation methods, spatially and in time, have gained popularity recently (e.g., Hoteit et al., 2002; Anderson, 2007a; Anderson, 2009; Li et al., 2009a; Bocquet, 2011; Luo and Hoteit, 2013; Miyoshi, 2011; and

Lee et al., 2017). In Anderson (2009), the inflation factor is treated as a random variable, and is then updated at each assimilation cycle. Similar ideas have been applied by Li et al. (2009a), Miyoshi (2011), and Gharamti (2018). Other approaches to "estimating" the value of the inflation factor have been also proposed, including the use of the forecast error statistics to guide the choice of the inflation factor to stabilize the EnKF and prevent filter divergence (Hoteit et al., 2005b; Luo and Hoteit, 2013, 2014c; Lee et al., 2017).

(ii) Localization: A small ensemble not only introduces spurious correlations between physically uncorrelated model variables in the ensemble covariance, but also provides limited degrees-of-freedom (rank of the ensemble) to fit the observations (Whitaker and Hamill, 2002). A straightforward way shown to be very efficient in many applications for dealing with such problems is to taper the long-range correlations in the EnKFs ensemble covariance matrices (Hamill et al., 2001; Houtekamer and Mitchell, 1998), a technique known as "localization." Most localization schemes are based on the distances between the physical locations of model variables and/or observations. For instance, Houtekamer and Mitchell (1998) introduced a local analysis scheme that updates an ocean state variable using only the observations located in its neighborhood. Hamill et al. (2001) adopted a covariance localization scheme in which one replaces the background covariance matrix with a Schur product between the background covariance matrix and a tapering matrix, whereas each element of the tapering matrix is computed using the Gaspari-Cohn function that depends on the physical distance between a pair of model variable and an observation (Gaspari and Cohn, 1999). For large-scale problems, directly manipulating the background covariance may become quite demanding in terms of computer memory. To alleviate this problem, other localization schemes have been proposed in which the Schur product is conducted between a certain tapering matrix and other quantities, such as the cross-covariance matrix between the forecast state and observation ensembles, the covariance matrix of the forecast observations ensemble (Houtekamer and Mitchell, 2001), and the Kalman gain matrix (Anderson, 2007b; Zhang and Oliver, 2010), or even the ensemble of forecast perturbations (Sakov and Bertino, 2011). Alternatively, to reduce the size of the involved matrices, localization has been also implemented in the observation space (Fertig et al., 2007) and in the context of a local EnKF (Bishop and Hodyss, 2007; Hunt et al., 2007; Ott et al., 2004).

The distance-based localization methods discussed above require that both the model variables and the observations have associated physical locations. In certain applications, however, some ocean variables/observations may not be associated with a physical location, e.g. non-local or spatial-temporal (or 4D) observations such as acoustics. In these situations, it becomes difficult to "localize" such observations with a distance-based localization method (Bocquet, 2016; Fertig et al., 2007; Luo et al., 2018). Therefore, adaptive localization methods have been proposed, some of them not based on physical distances. For instance, Bishop and Hodyss (2007) conducted a Schur product between the background error covariance matrix and a tapering matrix, whereas the latter was conducted by raising each element of a sample correlation matrix of model variables to a certain power. Anderson (2007b) and Zhang and Oliver (2010) used multiple background ensembles to compute a set of Kalman gain matrices, and then constructed the tapering matrices based on the

sample statistics of the Kalman gain matrices. Anderson (2012, 2016) and De La Chevrotière and Harlim (2017) also proposed localization methods correcting for sampling errors of the correlation coefficients between the pairs of model variables and observations. Sample correlation coefficients were also used in adaptive localization schemes (Evensen, 2009; Rasmussen et al., 2015a). More recently, Luo et al. (2018) elaborated on how to "localize" non-local and/or 4D observations through detections of causal relations between model variables and simulated observations.

(iii) Hybrid covariance: The use of small ensembles generally means that a significant part of the state (error) space is not represented by the ensemble. This implies that the ensemble subspace will not offer enough degrees-of-freedom to fit a large number of observations, and produces unrealistic confidence in the filter forecast (Song et al., 2010). The hybrid EnKF-OI/3DVAR method (Hamill and Snyder, 2000) is another approach that one could consider to enhance the performance of the EnKF without significantly increasing its computational cost. In this method, and at every assimilation cycle, the filter forecast covariance is estimated as a linear combination of a flow-dependent ensemble covariance sampled by an EnKF and a (pre-selected) static background covariance, i.e.,

$$\mathbf{P}_k^{Hybrid} = \alpha \mathbf{P}_k^{EnKF} + \beta \mathbf{B}^{Static}, \tag{40}$$

as a way to compensate for the complement of the ensemble's subspace α and β are tuning parameters that are usually set by trial and error, but could also be optimized adaptively as in Gharamti et al. (2014a). This technique has been successfully applied in several ocean applications (see e.g. Counillon and Bertino, 2009; and Tsiaras et al., 2017) and was shown to be quite efficient at improving the EnKFs robustness and performances.

Other forms of hybrid methods have been also proposed, for example, the semi-evolutive SEIK filter (Hoteit et al., 2001) in which only a selected part of the ensemble is updated by the model, and the adaptive EnKF, which selects new members from a static ensemble to enrich the EnKF ensemble based on the analysis error (Song et al., 2010).

Computing the ensemble SEnKF update based on a hybrid covariance is rather straightforward; obtaining it from a square-root EnKF is not (Bocquet et al., 2015; Auligné et al., 2016). Currently, many DA systems, including most operational ones, make use of hybrid covariances.

(iv) Iterative EnKFs: To handle nonlinear observations, one may adopt an iterative optimization scheme for the update step, similar to the variational DA methods (Courtier et al., 1994). In the context of ensemble DA, one may interpret iterative methods through a Bayesian perspective (Emerick and Reynolds, 2012). Alternatively, one can recast ensemble DA as a stochastic optimization problem (Oliver et al., 1996), and solve the problem using different optimization algorithms. Various iterative methods were introduced in the context of the EnKF (Zupanski, 2005; Gu and Oliver, 2007; Lorentzen and Nævdal, 2011; Sakov et al., 2012; Bocquet and Sakov, 2012; Luo and Hoteit, 2014c; Gharamti et al., 2015a), the ensemble smoother (EnS) (Emerick and Reynolds, 2012; Chen and Oliver, 2013; Luo et al., 2015), and the iterative ensemble Kalman smoother (IEnKS) (Bocquet and Sakov, 2014).

Non-Gaussian Filtering

The different Kalman filter options presented above are all based, in some way or another, on Gaussian distributions for the background/forecast and the noise (and linear observation operator). Given the nonlinear nature of the ocean dynamics, the forecast distribution will not be Gaussian even when the analysis distribution at the previous step is Gaussian. As such, all Kalman-type filters are sub-optimal in the context of nonlinear Bayesian filtering. Relaxing the assumption of Gaussian distributions is an active area of research in ocean DA. This field is very well developed in the mathematics and electric engineering communities, and mathematically sound non-Gaussian Bayesian filters have been already developed, the most famous of which is the PF. The excessively large dimension of the ocean models means a prohibitive number of realizations to sample the ocean state distribution, which precludes any brute force implementation of these techniques for ocean DA. Given this hard constraint on the number of samples that could be considered in a realistic ocean application, our goal here should be more practical to derive approximate nonlinear/non-Gaussian Bayesian filtering schemes that are robust and efficient, in terms of computational cost and performance, at least competitive with EnKFs, for potential application on realistic ocean data assimilation problems.

Nonlinear/non-Gaussian Bayesian filtering recently became an active area of research in the ocean community, which has mainly focused on two types of filters, namely the PF (Gordon et al., 1993) and the GMF (Sorenson and Alspach, 1971). Both approaches resort to some (truncated) statistical mixture models to describe the forecast and analysis distributions of the Bayesian filter, so that an approximate numerical solution can be computed. The two filtering strategies are summarized below with appropriate references.

1) PARTICLE FILTERING (PF)

The PF uses mixture models of Dirac delta densities (or a random set of ocean states) to approximate/discretize the prior (forecast) and posterior (analysis) state distributions. More specifically, suppose that at the $(k-1)^{th}$ assimilation cycle, the posterior is approximated by

$$p_{\mathbf{x}_{k-1}|\mathbf{y}_0,\ldots,\mathbf{y}_{k-1}} \approx \sum_{i=1}^{N} w_{k-1}^{i}\,\delta\left(\mathbf{x}_{k-1}-\mathbf{x}_{k-1}^{a,i}\right), \tag{41}$$

where δ denotes the Dirac delta function, $\mathbf{x}_{k-1}^{a,i}$ are the particles at the analysis step (similar to the ensemble members in the EnKF), w_{k-1}^{i} are the associated weights, and N is the total number of particles. The parameters of this mixture are then updated recursively based on the Bayesian filter steps (Gordon et al., 1993; Doucet et al., 2001; Van Leeuwen, 2009; Bocquet et al., 2010) as follows.

Forecast step: As in the EnKF, the analysis particles $\mathbf{x}_{k-1}^{a,i}$ are integrated forward with the dynamical model to obtain the forecast particles $\mathbf{x}_{k}^{f,i}$ at the next time t_k. The associated weights w_{k-1}^{i} remain unchanged.

Analysis step: The incoming observation $\boldsymbol{y}_k$ is used to update the weights only, while the particles themselves are kept unchanged, i.e. $\mathbf{x}_k^{a,i} = \mathbf{x}_k^{f,i}$. Roughly speaking, the particles will see their weights increase if they are close to $\boldsymbol{y}_k$ and decrease otherwise, according to

$$w_k^i = \frac{1}{c_k} w_{k-1}^i \, p(\mathbf{y}_k | \mathbf{x}_k^{f,i}), \tag{42}$$

where c_k is a constant that normalizes the weights of the posterior distribution, but does not need to be computed in practice. In the case of Gaussian observational error, the likelihood $p(y_k|\mathbf{x}_k^{f,i})$ becomes the Gaussian distribution of mean $\boldsymbol{H}_k(\mathbf{x}_k^{f,i})$ and covariance $\mathbf{R}_k$.

After an analysis (or forecast) step, the MV estimate of the ocean state is then obtained as the weighted-average of the particles, i.e. $\sum_{i=1}^{N} w_k^i \, \mathbf{x}_k^{a(f),i}$.

The theory of the PF is well established and the convergence of the PF state distribution toward the Bayesian filter distribution has been proven given an infinite number of particles (Doucet et al., 2001). In practice, however, one is restricted to a finite number of particles and the PF will suffer from the degeneracy of its particles, a phenomenon in which most of the weights concentrate on very few particles after only a few assimilation cycles. The effective number of the particles then decreases (Snyder et al., 2008) and the filter often collapses. This happens because the particles drift away from the true state, with the observations exerting no feedback on the particles. To overcome this, a resampling technique is needed. The basic idea of resampling is to draw new particles according to their estimated weights, and then assign them uniform weights (Gordon et al., 1993).

Many forms of PFs have been suggested, mainly differing in their resampling strategies. There is a rich literature on this topic and readers are referred to, for example, Arulampalam et al. (2002), Doucet et al. (2001), Doucet and Johansen (2011), Duan et al. (2010) and the references therein, for more information. However, even with resampling, the PF still requires a large number of particles to achieve an accurate solution (Doucet et al., 2000b). This makes the brute force implementation of the PF for DA with computationally demanding, realistic OGCMs a challenging problem (Anderson, 2003; Snyder et al., 2008; Van Leeuwen, 2009). Several strategies are currently being investigated to enable the implementation of the PF for large-scale ocean DA problems, many of which try to somehow exploit the future observations for efficient sampling of a limited number of particles (e.g., Van Leeuwen, 2010; Chorin et al., 2010); split the ocean state into smaller vectors to reduce the dimension of the problem, a form of localization (e.g., Ait-El-Fquih and Hoteit, 2016; Penny and Miyoshi, 2016; Poterjoy, 2016); or apply some transformation to move the particles toward high-probability regions so that a single particle does not dominate the total weight (e.g., Luo and Hoteit, 2014a; Reich, 2013; El-Sheikh et al., 2014). A more recent school of thought is investigating combinations of the EnKF and the PF (Frei and Künsch, 2013; Shen and Tang, 2015; Ait-El-Fquih and Hoteit, 2017), exploiting the robustness of the EnKF and (asymptotic) optimality of the PF.

2) GAUSSIAN MIXTURE FILTERING (GMF)

The idea here is to use mixture models of Gaussian densities to approximate the state distributions, i.e., of the form

$$\sum_{i=1}^{N} w^i \Phi(\mathbf{x} : \mathbf{x}^i, \mathbf{P}^i), \tag{43}$$

where w^i are the weights of each Gaussian component Φ in the mixture of mean $\mathbf{x}^i$ and covariance $\mathbf{P}^i$. The $\mathbf{x}^i$s are also called particles in analogy to the PF. This is based on Alspach and Sorenson (1972) who demonstrated that when the likelihood is Gaussian and the observation operator is linear, a Bayesian update of a Gaussian mixture (GM) prior leads to a GM posterior for which the parameters of the Gaussian components are updated as in the KF, and their weights are updated as in the PF. Two GMF strategies can be then distinguished for ocean DA, depending on whether the forecast step is initiated from GM or Dirac mixture (DM) posteriors. If initiated from a GM, applying a local linearization around the centers of the Gaussian components of the mixture allows to carry the forecast step as a set of EKFs operating in parallel. More precisely, the GM posterior is integrated by an ocean model into a GM prior to the next assimilation step with the same weights, and centers and covariances computed from their prior counterparts using the EKF (Sorenson and Alspach, 1971; Chen and Liu, 2000; Bengtsson et al., 2003; Hoteit et al., 2008). In doing so, the prior and posterior distributions always remain as GMs, with the weights updated as in the PF, and the parameters of the Gaussian components as in the EKF, or any other nonlinear Kalman filter, such as an EnKF, (Hoteit et al., 2008; Luo et al., 2010; Hoteit et al., 2012; Sondergaard and Lermusiaux, 2013). This is generally referred to as Gaussian mixture filtering (GMF). If initiated from a DM, the forecast step is identical to that of the PF/EnKF, but the GM posterior needs to be resampled into a DM before forecasting (after every update step). We refer to this type of filtering the ensemble GMF (EnGMF; Anderson and Anderson, 1999; Liu et al., 2015, 2016). Resampling is then part of the EnGMF, but could be conducted only when needed (Hoteit et al., 2008, 2012). Even though the GMFs are less prone to the degeneracy of their particles, because the PF update of its weights is normalized by the covariance of the innovation vector (or observation prediction error) instead of the observation error covariance as in the standard PF, particles might still collapse and resampling a new GM with uniform weights may improve the filter robustness and performance (Hoteit et al., 2008). Another approach to mitigate the weight's collapse is to somehow follow an approach combining EnKF and PF, as suggested by Bengtsson et al. (2003).

GMFs are basically implemented as an ensemble of nonlinear KFs running in parallel, and as such can be computationally very demanding. Hoteit et al. (2008) suggested using a uniform low-rank covariance for all GM components to reduce the computational burden. Hoteit et al. (2012) later investigated a GMF in which each component of the forecast/analysis GM is represented by a different ensemble as a way to use a different covariance matrix for each Gaussian component. Sondergaard and Lermusiaux (2013) constructed the GM (from the forecast particles) in a reduced state space using an expectation-maximization (EM) algorithm. This enabled them to consider a variable number of Gaussian components in the GM, which was determined based on the Bayesian

information criterion. The implementation of an EnGMF can be more straightforward, as it can be conveniently done via an EnKF code since they share the same forecast step, and a Kalman update step to each particle (or ensemble member). It is important to realize, however, that the EnGMF does not need to perturb the observations and may eventually use a different covariance matrix to update each particle depending on how the GM prior is estimated from the forecast particles (Liu et al., 2016). Of course one needs to add a resampling step for the EnGMF; a deterministic resampling step by analogy to the DEnKF was found beneficial when dealing with a small number of particles (Liu et al., 2015), but defining an optimal strategy for resampling from a GM remains an open problem.

Smoothing

Filters condition the ocean state with past (including current) observations, and are thus naturally designed for operational/forecasting settings. In contrast, smoothers condition the ocean state with past and future observations, exactly as 4DVAR. As such, smoothing estimates are expected to be more accurate than filtering estimates because of the information gain from future observations, and are thus more relevant to performing re-analyses. Smoothing is also useful to estimate model parameters whose influence on the state may be spread in time. Unlike 4DVAR, which is designed to compute the MAP estimator only, smoothers are usually implemented through additional update steps of the filters' estimates, with the future data, involving the associated distributions.

The ensemble smoothing update presented in this section is based on a linear analysis, which is an approximation for nonlinear models. More advanced but significantly more costly schemes that could go beyond the Gaussian approximation of the EnKF are available, such as the iterative ensemble Kalman smoother (Bocquet and Sakov, 2014) with multi-pass iterative updates. All these smoothers should be equivalent with linear model and Gaussian statistics (Cosme et al., 2012), and sometimes even with a nonlinear model (Raanes, 2016).

We are, therefore, interested here in estimating the ocean state $\mathbf{x}_k$ (in the past) at t_k, given all the observations up to T_L ($\geq t_k$). Let $\mathbf{E}_k^f$ be the matrix of the forecast ensemble members at t_k, column-wise, and consider the collection of these matrices within the assimilation window $[T_0\ T_L]$. This collection represents an ensemble of trajectories of the model, which can be seen as a prior within $[T_0\ T_L]$. Let $\mathbf{E}_k^a$ be the matrix of the EnKF-updated ensemble members at $t_k \in [T_0\ T_L]$ after assimilating the observation vector $\boldsymbol{y}_k$. Whatever the EnKF's choice, the ensemble member updates are linear combinations of the prior members, so that one can actually write,

$$\mathbf{E}_k^a = \mathbf{E}_k^f \mathbf{T}_k, \tag{44}$$

where $\mathbf{T}_k$ is an update transform matrix that depends on the observations, the prior and the EnKF flavor (Evensen, 2003). In the case of the stochastic and the deterministic EnKFs, these could be obtained from Eqs. 29 and 36, respectively.

The advantage of expressing the ensemble members, instead of the perturbations, update as in Eq. 44 is to include the analysis as part of the transformation. Then, the ensemble update at t_l ($t_k \leq t_l \leq T_L$) given this observation vector at t_k is

$$\mathbf{E}_l^a = \mathbf{M}_{k:l}\mathbf{E}_k^a = \mathbf{M}_{k:l}(\mathbf{E}_k^f\mathbf{T}_k) = (\mathbf{M}_{k:l}\mathbf{E}_k^f)\mathbf{T}_k = \mathbf{E}_l^f\mathbf{T}_k, \tag{45}$$

where $\mathbf{M}_{k:l}$ is the tangent linear model from t_k to t_l. As a consequence, to perform an update of the ensemble at any time within the assimilation window, one just needs to apply the transform $\mathbf{T}_k$ on the right to the ensemble. This generalizes to the case where $T_0 \leq t_l \leq t_k \leq T_L$ by setting $\mathbf{M}_{k:l} = \mathbf{M}_{k:l}^{-1}$.

From this translational invariance of the update, which was put forward by Evensen (2003) and Hunt et al. (2004), one can build the following ensemble Kalman smoothing (EnKS) algorithm. First, run the EnKF throughout the assimilation window up to T_L and store the ensemble $\mathbf{E}_l^a$ and the ensemble transform matrices $\mathbf{T}_l$ at any time t_l where a smoothing analysis is needed. Second, to compute the smoothing updated ensemble $\mathbf{E}_k^s$ at t_k within the assimilation window, one simply needs to apply the translational invariance principle and use the transform matrices from t_l to T_L:

$$\mathbf{E}_k^s = \mathbf{E}_k^a \prod_{l=k}^{L} \mathbf{T}_l. \tag{46}$$

Hence, this smoother consists of two passes: an EnKF forward run, followed by a retrospective update of the ensembles (that need updating). This makes this EnKS a very elegant procedure, but with the drawback of requiring important storage.

Domain localization of the EnKS can be implemented as in the EnKF. A common mistake is to enforce inflation, often required by an EnKF, in the forward EnKF pass as well as in the backward smoothing pass. Instead, inflation should only be applied in the forward EnKF step, since sampling errors counteracted by inflation have already been accounted for in the forward pass. Cosme et al. (2010) and Nerger et al. (2014) offer examples of application of the EnKS in ocean data assimilation.

Hybrid Ensemble-Variational Methods

As discussed in the previous sections, there are benefits and drawbacks in using an EnKF or a 4DVAR. The EnKF involves a flow-dependent representation of the errors, via the ensemble. In applications involving high-dimensional models, this representation is nevertheless rank-deficient and relies on localization. The computational cost of the ensemble propagation is demanding, but this can be mitigated by trivial parallelization. On the downside, the EnKF is generally efficient for moderate model nonlinearity because of its second-order moments approximation of the error statistics. 4DVAR, as a nonlinear variational method, can handle nonlinearities. This, however, requires the adjoint of the observation and propagation models, which is a strong technical drawback as it is time-consuming to derive and maintain the adjoint code. Moreover, the traditional 4DVAR does not propagate the flow-dependent error statistics, but only the state analysis. This is however mitigated by the possibility of using a full-rank background error covariance matrix, which is not possible in a standard EnKF. Finally, 4DVAR does not lend itself easily to parallelization. A

full discussion can be found in Lorenc (2003), Kalnay et al. (2007), and Chapter 7 of Asch et al. (2016).

There have been various attempts to merge these two types of methods in order to combine their strengths while avoiding some of their drawbacks, which we refer to as hybrid or hybrid ensemble-variational methods.

(i) Ensemble of Data Assimilations: To implement flow-dependence of the error representation in a DA system, one could consider an ensemble of such systems. The outcomes of the analyses from the DA ensemble could then be combined to form a flow-dependent background matrix, as in the EnKF. The implementation of such an approach is straightforward, requiring minimal changes to an existing operational DA system. It is called ensemble of data assimilations, and has mostly been used in National Weather Prediction centers that operate 4DVAR assimilation and, in particular, Météo-France (Raynaud et al., 2009; Berre et al., 2015) and the European Centre for Medium-Range Weather Forecasts (ECMWF; Bonavita et al., 2011, 2012). In such a setting, each 4DVAR analysis, indexed by i, uses a different first guess $\mathbf{x}_0^i$ in the optimization procedure, and perturbed observation vectors $\{\mathbf{y}_k^i\}_{k=0,\ldots,L}$, as in the SEnKF, to maintain statistical consistency. The background covariance is typically a hybrid one as in Eq. 40, comprised of the static covariances of the traditional 4DVAR and incorporating the sample covariances from the prior ensemble. The resulting ensemble of analyses is then propagated to the next analysis. The sample covariances may need to be regularized by localization. In this context, covariance localization has been implemented via the so-called $\boldsymbol{\alpha}$ control variable trick (Lorenc, 2003; Buehner, 2005; Wang et al., 2007), or using wavelet truncations (Berre et al., 2015).

(ii) Iterative Ensemble Kalman Filter and Smoother: Most ensemble variational and hybrid methods have been designed empirically. By contrast, the iterative ensemble Kalman smoother (IEnKS) has been derived from Bayes' rule (Bocquet and Sakov, 2013, 2014) with well-identified approximations. The analysis step of the IEnKS consists in a nonlinear minimization of the 4DVAR cost function over an assimilation window $[T_0, T_L]$, with the goal of estimating the initial state at T_0, assuming a perfect model. It uses a previously forecast ensemble at T_0 to generate second-order background statistics in exactly the same way as the ETKF. Different from 4DVAR, the optimization is carried out in the space spanned by the ensemble rather than the full state-space. This restricts the search of the analysis state, but enables the use of efficient minimization techniques such as Gauss-Newton, Levenberg-Marquardt, and trust-region methods. Analogous to the strong 4DVAR cost function in Eq. 17, the IEnKS cost function is typically of the form

$$J_{IEnKS}(\mathbf{w}) = \|\mathbf{w}\|^2 + \sum_{k=L-S+1}^{L} \left\| \mathbf{y}_k - H_k \left(M_{0:k}(\mathbf{x}_0^b + \mathbf{S}_0^b \mathbf{w}) \right) \right\|_{\mathbf{R}_k^{-1}}^2, \tag{47}$$

where $\mathbf{w}$ is the vector of unknown coordinates in the ensemble space, $\mathbf{x}_0^b$ and $\mathbf{S}_0^b$ are respectively the prior ensemble mean and the matrix of the prior ensemble anomalies at T_0, and $\boldsymbol{M}_{0:k}$ is the model resolvent integrating the ocean state from T_0 to t_k. The sensitivities of the observations to $\mathbf{w}$ required for the calculation of the gradient of the cost function can be computed by finite-difference methods, or using finite-spread representation as in the EnKF. This avoids the need for the tangent-linear and

adjoint models. Once a minimum is found, an approximate Hessian can be obtained at the corresponding minimizer in the ensemble space. Just like with the ETKF, this allows to generate an updated ensemble of anomalies. In the forecast step, the newly generated ensemble is integrated with the model from t_L to t_{L+S}, to be used as the prior of the next assimilation cycle. The only approximations in the scheme are the Gaussian modeling of the prior and posterior ensemble, and, to a lesser extent, the finite-size of the ensemble.

The "shift" parameter S can obviously be chosen to be L or, more generally, any value between 1 and $L + 1$, if all observations are to be assimilated. This degree of freedom, barely exploited in 4DVAR, is interesting because it allows flexibility in the transfer of information from one assimilation window to the next through the ensemble. The case $S = 1$, $L = 0$ with a linear observation model, coincides with the ETKF (Hunt et al., 2007). The case $S = 1$, $L = 0$ and a general (nonlinear) observation model coincides with the maximum likelihood ensemble filter (MLEF) (Zupanski, 2005). The case $S = 1$, $L = 1$ is known as the iterative ensemble KF (IEnKF, Sakov et al., 2012; Bocquet and Sakov, 2012). The case $S = L$ with a single iteration in the analysis and further mild restrictions corresponds to the 4D-ETKF (Hunt et al., 2004).

The name smoother comes from Bell (1994), who first proposed an iterative (full-rank) Kalman smoother. The IEnKS can be used as a smoother as well as a filter. As opposed to the EnKS, and with a nonlinear model, each iteration should improve the filtering solution, i.e., the state estimate at the present time. With low-order chaotic models, the IEnKS has been shown to outperform any scalable known method such as the EnKF, the EnKS, and 4DVAR, for both filtering and smoothing.

As any ensemble method, the IEnKF and IEnKS require localization. This is not as straightforward to implement as in the EnKF since it should be applied within the full assimilation window. Because the dynamics of the error do not generally commute with localization (Fairbairn et al., 2014; Bocquet and Sakov, 2014), suboptimalities can appear for long windows. A solution—the so-called dynamically covariant localization—has been proposed in Bocquet (2016) and yields good results with simple models. If the adjoint model is available, then this issue could be dealt with in a more efficient way (Bocquet, 2016).

The IEnKS is a mathematically justified Hybrid Ensemble-Variational method that can also be also useful in understanding and rationalizing the so-called Four-Dimensional Ensemble Variational (4DEnVAR) methods.

(iii) Four-Dimensional Ensemble Variational (4DEnVAR): This class of methods was developed by the National Weather Prediction centers within a 4DVAR framework. The primary goal was to avoid maintaining the adjoint of the forecast model. Like the IEnKS and similar to the reduced 4DVAR put forward early in oceanography by Robert et al. (2005) and Hoteit and Köhl (2006), the analysis is performed within the subspace spanned by the ensemble (Liu et al., 2008). Observation perturbations are usually generated stochastically using, for instance, a stochastic EnKF (Liu et al., 2009; Buehner et al., 2010a). Hence, in addition to avoiding the need for an adjoint model, flow-dependent error estimation is introduced. The 4DEnVAR implementations usually come with a hybrid background. Just like the IEnKS, localization is necessary and theoretically more challenging than with an EnKF (Desroziers et al., 2016). Many 4DEnVAR variants have been

suggested, depending on the availability of the adjoint models and how the perturbations are generated (Buehner et al., 2010a,b; Zhang and Zhang, 2012; Clayton et al., 2013; Poterjoy and Zhang, 2015). Several 4DEnVar weather systems are now operational or on the verge of being so (Buehner et al., 2013; Gustafsson et al., 2014; Desroziers et al., 2014; Lorenc et al., 2015; Kleist and Ide, 2015; Buehner et al., 2015; Bannister, 2017). A recent further sophistication is to construct an ensemble of data assimilations of 4DEnVAR in order to generate the perturbations (Bowler et al., 2017; Arbogast et al., 2017).

Discussion and Future Developments

The theoretical framework of ocean DA methods is now well established around the Bayesian estimation theory, and many robust methods have been developed for efficient assimilation of available ocean data into state-of-the-art general circulation ocean models. These provide various tools to predict the past and/or future state of the ocean, conditioned on available data, and also to quantify its uncertainties. Evaluating the uncertainties is important for decision making and proper weighting of the most recent estimate against incoming observations for computing the next estimate. At present, 4DVAR and EnKFs are the state-of-the-art techniques for ocean data assimilation, and most ocean centers are currently developing their operational systems around these approaches. Both have been extensively studied and their characteristics are now well understood. They were further shown to provide viable solutions in many ocean DA applications.

4DVAR and EnKF have their own advantages and weaknesses, and this lead to the development of a new class of Hybrid Ensemble-Variational methods, aiming at combining the strength of these two approaches. The hybrid techniques are currently being actively investigated by the atmospheric community, but nothing should preclude their applications for ocean DA. Other advances in the developments of auxiliary techniques that were proven important for enhancing the performances of 4DVAR and EnKF are also expected to improve the accuracy of the ocean DA products, which may include working on more sophisticated parameterizations of (ensemble) covariance matrices, enhanced resampling techniques, enforcing dynamical balances and constraints, etc.

In addition, recent advances in the development of non-Gaussian/nonlinear DA methods for large-scale DA problems hold the promise of edging us closer toward the optimal Bayesian estimate. These techniques were indeed demonstrated to be superior to EnKFs and 4DVAR in many idealized DA problems. However, many challenges still require more investigations into, for example, how best to parameterize the involved probability distributions or to resample a reasonable-size ensemble/sample from the estimated distributions. These should be further complemented by further studies to better understand the relevance of more advanced assimilation for the estimation of the various ocean spatial and temporal scales (Subramanian et al., 2012), particularly in relation to the associated computational cost and the spatial and temporal coverage of the assimilated observations.

All of these developments focused on introducing new DA techniques or enhancing existing ones. Another very important aspect of any assimilation system is the treatment of the underlying

uncertainties in the system, which basically characterize the final assimilation solution. These, however, have not received the attention they deserve by the ocean data assimilation community, with the exception of some uncoordinated attempts to address some of the challenges, and the problem remains open. Uncertainties can be inherited from the incomplete ocean models dynamics, measurements sensors, and various parameters and inputs. These could also be introduced by poor knowledge of the statistics of the modeled noise, which are required by the assimilation algorithms. There is a variety of approaches to quantify, reduce, and account for the uncertainties in an assimilation system that are yet to be fully exploited by the ocean community. We end this chapter with a general discussion and a summary of available promising tools to efficiently handle the uncertainties in an ocean DA system.

(i) State-Parameters Estimation: Various parameterizations are used in the general circulation ocean models (e.g., mixing parameters) and, as such, the ocean state depends on a set of (physical) parameters that may not be perfectly known. Estimating these parameters along with the state should reduce the uncertainties in the system and eventually lead to an improved final assimilation solution.

Estimating the parameters of an ocean model is technically an inverse problem that can be addressed by a deterministic (least-squares or variational-like) approach aiming at determining the parameters that best fit the model to available data, or as a Bayesian inverse problem computing the posterior distribution of the parameters given a prior and the likelihood (e.g., Aster et al., 2005; Tarantola, 2005). The first approach is basically an optimization problem, very similar to 4DVAR (Wunsch, 1996; Elbern et al., 2007). Markov Chain Monte Carlo (MCMC) is the reference numerical method to compute the posterior solution of the Bayesian inverse problem, but it can require a prohibitive number of ocean model runs to converge (Metropolis et al., 1953; Malinverno, 2002; Sraj et al., 2016). Surrogate models have been used to enable the application of such methods for large-scale problems, see for example Li et al. (2015), Sraj et al. (2016) and Sripitana et al. (2017) for ocean applications. An advantage of DA, even though its methods are often suboptimal, is that its framework may allow for joint state-parameters estimations. The 4DVAR framework readily enables for ocean parameters estimation (Wunsch and Heimbach, 2007; Liu et al., 2012) within a strong or weak constraint formulation, but it is an offline approach. The recursive nature of filters and smoothers allows for simultaneous online estimation of the state and parameters, and accordingly straightforward updates with new observations.

The technical framework for online estimation of the state and the parameters of a dynamical system based on observations is well established (Harvey and Phillips, 1979; Pagan, 1980; Hamilton, 1986; Aksoy et al., 2006; Rasmussen et al., 2015b; Ait-El-Fquih et al., 2016) and is already being heavily exploited in hydrology and subsurface flow applications (e.g., Oliver and Chen, 2011; Moradkhani et al., 2005; Gharamti et al., 2015b). It has been recently applied in the context of a storm surge ocean model by Sripitana et al. (2018). Within the Bayesian estimation framework, and similarly to the state, the parameters are treated as random variables with a given prior distribution. The state-parameters estimation problem then consists of computing the distribution of the augmented (or joint) state-parameters vector, say $[\mathbf{x}, \theta]$, conditioned on available

observations. The assimilation methods presented in this chapter can be then directly used to address this problem by applying them to $[\mathbf{x}, \theta]$ instead of $\mathbf{x}$, assuming constant dynamics for the time-evolution of the parameters. This is the joint state-parameters estimation approach. Another method that became popular more recently is the so-called dual approach, which consists of separately updating the state and the parameters using two interactive filters, one acting on the parameters and the other on the state given the parameters. This approach is commonly referred to as Rao-blackwelisation in the PF community where it has been introduced to reduce the variance of the joint state and parameters estimation (Doucet et al., 2000a). It was applied early on in an ensemble context in hydrology (Moradkhani et al., 2005) to mitigate for noticeable inconsistencies issues in the joint approach (Chen and Zhang, 2006), but only recently was used in a Bayesian consistent framework (Ait-El-Fquih et al., 2016). The dual estimation approach has been proven to be especially beneficial in strongly nonlinear applications (Gharamti et al., 2014b; Ait-El-Fquih et al., 2016).

Another state-parameters estimation problem that is important to address considers the statistical parameters instead of (or together with) the physical parameters. This mainly involves the parameters of the dynamical model errors and observations noise distributions (Särkkä and Hartikainen, 2013; Ardeshiri et al., 2015), and could be particularly useful in the context of stochastic parameterizations. It should help sampling relevant stochastic terms based on the available data and the dynamics in hand.

The most common approach to simultaneously estimate these so-called hyper-parameters along with the state consists of maximizing the likelihood of the whole set of available observations given these parameters (Hamilton, 1986). However, because of the complexity of the likelihood function, an analytical evaluation of the maximum likelihood estimate of the hyper-parameters is generally not feasible. A number of approximate numerical solutions have therefore been proposed, mostly relying on gradient-based iterative optimization methods (Harvey and Phillips, 1979; Pagan, 1980). To overcome the common limitations of gradient-based methods, the popular iterative Expectation-Maximization (EM) algorithm has been first introduced to linear systems (Shumway and Stoffer, 1982), and later extended to nonlinear systems involving PF- and EnKF-like assimilation algorithms (e.g., Cappé et al., 2005; Stroud and Bengtsson, 2007; Frei and Künsch, 2012; Ueno et al., 2010; and Dreano et al., 2017). The EM iterations alternate between an expectation (E) step, which constructs an expectation-type cost function of the log-likelihood evaluated at the parameters current estimate, and a maximization (M) step, which computes new parameters maximizing the expected log-likelihood created in the E step. The resulting parameters' estimates are then used to determine the distribution of the hyper-parameters in the following E step.

More recently a Bayesian approach has been proposed, considering the hyper-parameters as random variables with a prior distribution, commonly a Wishart distribution. This leads to a posterior that is also a Wishart distribution and for which sufficient statistics can be obtained in closed forms (Robert, 2007). The Bayesian approach is an online procedure that computes a new estimate at each observation time step, while the maximum likelihood approach is viewed as an off-line (or smoother) approach. Furthermore, the maximum likelihood approach computes only a point

estimate of the hyper-parameters, while the Bayesian approach provides a full distribution from which any (point) estimate can be obtained.

(ii) Accounting for Model Errors: It is important to properly quantify the statistics of the forecast (background) in the assimilation (Dee, 1995; Li et al., 2009b). In the description of the state-space model (dynamical and observation models) of the DA problem, model errors were traditionally represented as white noise, Gaussian of mean zero and a given covariance. This assumption is at the basis of the KF- and 4DVAR- like assimilation algorithms, even though the contribution of this additive error term, through its covariance, is often treated crudely or even neglected. The model bias, or mean of model errors, can be conveniently treated as a state-parameters estimation problem (Zupanski and Zupanski, 2006) or closely related techniques (Dee and da Silva, 1998; Chepurin et al., 2005). This can be further extended to the case of time-dependent bias, for which a computationally efficient scheme expressing the errors in terms of very few degrees-of-freedom has been proposed by Danforth and Kalnay (2008). The dimension of the model error covariance can be prohibitively large; accurately estimating its coefficients would require a large amount of data that is simply not available (Mehra, 1970).

In 4DVAR, the model error is accounted for as a control term in the objective cost function (Trémolet, 2007). Strong constraint 4DVAR completely ignores it to simplify its optimization problem, projecting all errors on the initial conditions or further on other model parameters and inputs in slightly modified variants. Weak constraint 4DVAR methods, such as the representer method, provide elegant ways to directly tackle the problem but the end result heavily depends on the specification of appropriate model error covariance matrices to properly weight these terms in the cost function and spread the information to non-observed locations (Di Lorenzo et al., 2007). Imposing dynamical constraints is also important, as for the background covariance (Ngodock and Carrier, 2014).

EnKFs (smoothers) algorithms can directly handle the model error covariance matrix as in the KF (smoother), but this is not as straightforward for the large dimensional ocean DA problem. In this context, it is more efficient to exploit the Monte Carlo framework of the ensemble methods and account for the model errors through perturbations sampled from given statistics (covariance) (Mitchell and Houtekamer, 2000; Hamill and Whitaker, 2005; Hoteit et al., 2007), as in a fully nonlinear DA scheme. Likewise, this could be also applied in the case of non-additive noise, e.g. as a stochastic parameterization scheme (Buizza et al., 1999; Wu et al., 2008). These methods, or a combination of them, may provide an efficient approach, computationally and dynamically, to account for ocean models errors in an ensemble (nonlinear) setting. Additive or not, the parameters and statistics of such terms could be quantified using the state-parameters and hyper-parameters methods discussed in the previous section (Ardeshiri et al., 2015; Dreano et al., 2017). One may also opt for an offline approach to quantify the statistics of the model errors or of their parameterizations based on the available data (Daley, 1992) or a set of assimilated increments (analysis minus model forecast), which became popular recently (van Leeuwen, 2015). One may further consider identifying missing physics in the model following a similar approach (van Leeuwen, 2015).

(iii) Coupled Data Assimilation: Building more accurate ocean models with better predictive skills can be achieved through parameters estimation and quantification of modeling errors, but also by implementing more complete dynamics. Ocean models are designed to be forced by prescribed atmospheric fields, a framework that does not allow to fully account for the ocean feedback to the atmosphere. This is also true for the atmospheric models, which require ocean surface fields on the ocean-atmosphere boundary layer. Following the continuous progress in computing resources and the desire to solve the most complete dynamics and extend the predictability of the atmospheric and oceanic forecasting systems, fully coupled ocean-atmosphere models have been under development for many years (e.g., Delworth et al., 1993; Stockdale et al., 1998). Another important benefit from models coupling is that it provides a dynamical framework to exploit the observations of the state of one system in the assimilation of the other model, e.g. use atmospheric observations to update the ocean model state, and vice-versa. This is referred to as the coupled ocean-atmosphere data assimilation (CDA) problem.

CDA allows to exchange information between the models through the coupling dynamics and the assimilated observations. This should further provide state estimates that are more consistent with the dynamics of both systems. Nowadays, most operational centers agree that CDA is the goal for analysis and prediction of the climate system, particularly on subseasonal-to-seasonal and longer timescales (Penny and Hamill, 2017). The most straightforward CDA method would be to follow an augmented state approach in which the states and observations of the coupled models are concatenated into one state vector and one observation vector, basically considering the coupled system as a single state-space model. One could then directly apply any of the DA methods presented in this chapter for simultaneous assimilation of all available observations.

Because of the complexity of coding and maintaining the adjoint of a coupled ocean-atmosphere system, the non-intrusive and portable nature of the EnKF algorithms have made them more popular for CDA (Tardif et al., 2015; Sluka et al., 2016). These methods may also offer more flexibility in implementation, such as using different filters and ensemble sizes for each model (Luo and Hoteit, 2014b). However, the multi-scale nature, in space and time, of the dynamics of the coupled system requires revisiting the calculations of the ensembles cross-correlations (Luo and Hoteit, 2014b; Lu et al., 2015) and accordingly adapting the necessary auxiliary techniques, such as the ensemble localization and inflation (Frolov et al., 2016). Another important complexity in CDA is related to the turbulent dynamics of the atmospheric component together with the longer than synoptic timescales that make coupled problems interesting, but difficult to handle with a linear ensemble assimilation technique. Nonlinear/non-Gaussian assimilation methods may thus prove to be useful for such systems. However, the algorithmic and computational complexity of the coupled system, involving two or more different general circulation models, remains an important factor in limiting the development of CDA systems. To simplify the CDA problem, the notion of "weak" CDA has been introduced (e.g. Lu et al., 2015), in which each model of the coupled system assimilates its own observations. The coupling is achieved only via the forecast step, as opposed to performing coupled analysis steps, which is referred to as "strong" CDA. Weak CDA relies solely on the coupling dynamics to spread the observations information between the models and, thus, may miss

opportunities to exploit some useful information from the observations of the other models during assimilation. Nevertheless, the development of CDA systems is underway and the characteristics of cross-covariances between the errors in the atmosphere and ocean model forecasts are being explored (Smith et al., 2017a), demonstrating the additional potential of the "strong" CDA.

Despite these challenges, the promises of delivering more skillful assimilative and predictive models, and particularly long-term (e.g., subseasonal-to-seasonal and longer) forecasts, will make CDA an active area of research in the years to come. Ocean models are also often coupled with many other components of the earth system, such as wave models to better describe wave-ocean interactions and properly resolve the surface roughness that is important for the atmosphere, various transport models for tracking, and biogeochemical models to simulate the ocean ecosystem variability. This presents endless possibilities for developing multi-assimilative models that combine different models, observations, and assimilation methods.

Acknowledgements

Xiaodong Luo was partly supported by the project "Advancing permafrost carbon climate feedback-improvements and evaluations of the Norwegian Earth System Model with observations" funded by the Research Council of Norway (ID 250740), and the project (ID 230303) funded by the Research Council of Norway and the industry partners – ConocoPhillips Skandinavia AS, Aker BP ASA, Eni Norge AS, Maersk Oil Norway AS, DONG Energy A/S, Denmark, Statoil Petroleum AS, ENGIE E&P NORGE AS, Lundin Norway AS, Halliburton AS, Schlumberger Norge AS, Wintershall Norge AS – through The National IOR Centre of Norway. CEREA is a member of Institute Pierre-Simon Laplace (IPSL). The authors would like to thank Naila Raboudi for her help and comments.

References

Aanonsen, S., G. Nævdal, D. Oliver, A. Reynolds, and B. Valls, 2009: The ensemble Kalman filter in reservoir engineering: a review. SPE Journal, 14, 393–412, SPE-117274-PA.

Abarbanel, H., 2012: Analysis of observed chaotic data. Springer Science & Business Media.

Ait-El-Fquih, B., M. E. Gharamti, and I. Hoteitl, 2016: A Bayesian consistent dual ensemble Kalman filter for state-parameter estimation in subsurface hydrology. Hydrology and Earth System Sciences, 20, 3289–3307.

Ait-El-Fquih, B., and I. Hoteit, 2016: A variational Bayesian multiple particle filtering scheme for large-dimensional systems. IEEE Transactions on Signal Processing, 64 (20), 5409–22.

Ait-El-Fquih, B., and I. Hoteit, 2018: An efficient state-parameter filtering scheme combining ensemble Kalman and particle filters. Mon. Wea. Rev., https://doi.org/10.1175/MWR-D-16-0485.1.

Aksoy, A., F. Zhang, and J. Nielsen-Gammon, 2006: Ensemble-based simultaneous state and parameter estimation with MM5. Geophysical Research Letters, 33, L12 801.

Alspach, D., and H. Sorenson, 1972: Nonlinear Bayesian estimation using Gaussian sum approximations. IEEE transactions on automatic control, 17 (4), 439–448.

Altaf, M. U., M. E. Gharamti, A. W. Heemink, and I. Hoteit, 2013a: A reduced adjoint approach to variational data assimilation. Computer Methods in Applied Mechanics and Engineering, 254, 1–13.

Altaf, U., T. Buttler, T. Mayo, C. Dawson, A. Heemink, and I. Hoteit, 2014: A comparison of ensemble Kalman filters for short range storm surge assimilation. Mon. Wea. Rev., 142, 2899–2914.

Altaf, U. M., T. Butler, X. Luo, C. Dawson, T. Mayo, and H. Hoteit, 2013b: Improving short range ensemble Kalman storm surge forecasting using robust adaptive inflation. Mon. Wea. Rev., 141, 2705–2720.

Anderson, B. D. O., and J. B. Moore, 1979: Optimal filtering. Englewood Cliffs, 21, 22–95.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. Mon. Wea. Rev., 129, 2884–2903.

Anderson, J. L., 2003: A local least squares framework for ensemble filtering. Mon. Wea. Rev., 131 (4), 634–642.
Anderson, J. L., 2007a: An adaptive covariance inflation error correction algorithm for ensemble filters. Tellus, 59A (2), 210–224.
Anderson, J. L., 2007b: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. Physica D: Nonlinear Phenomena, 230 (1), 99–111.
Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. Tellus, 61A, 72–83.
Anderson, J. L., 2010: A non-Gaussian ensemble filter update for data assimilation. Mon. Wea. Rev., 138, 4186–4198.
Anderson, J. L., 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. Mon. Wea. Rev., 140 (7), 2359–2371.
Anderson, J. L., 2016: Reducing correlation sampling error in ensemble Kalman filter data assimilation. Mon. Wea. Rev., 144 (3), 913–925.
Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Mon. Wea. Rev., 127, 2741–2758.
Arbogast, E., G. Desroziers, and L. Berre, 2017: A parallel implementation of a 4DEnVar ensemble. Q. J. R. Meteor. Soc., doi:10.1002/qj.3061.
Ardeshiri, T., E. Özkan, U. Orguner, and F. Gustafsson, 2015: Approximate Bayesian smoothing with unknown process and measurement noise covariances. IEEE Signal Processing Letters, 22 (12), 2450–2454.
Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp, 2002: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on Signal Processing, 50, 174–188.
Asch, M., M. Bocquet, and M. Nodet, 2016: Data assimilation: Methods, algorithms, and applications. SIAM.
Aster, R. C., B. Borchers, and C. H. Thurber, 2005: Parameter Estimation and Inverse Problems. Elsevier, New York, 301 pp.
Auligné, T., B. Ménétrier, A. C. Lorenc, and M. Buehner, 2016: Ensemble–variational integrated localized data assimilation. Mon. Wea. Rev., 144 (10), 3677–3696.
Bai, Y., Z. Zhang, Y. Zhang, and L. Wang, 2016: Inflating transform matrices to mitigate assimilation errors with robust filtering based ensemble Kalman filters. Atmospheric Science Letters, 17, 470–478.
Balmaseda, M. A., K. Mogensen, and A. T. Weaver, 2013: Evaluation of the ECMWF ocean reanalysis system ORAS4. Q. J. R. Meteor. Soc., 139 (674), 1132–1161.
Bannister, R. N., 2017: A review of operational methods of variational and ensemble-variational data assimilation. Q. J. R. Meteor. Soc., 143 (703), 607–633.
Barth, A., A. Alvera-Azcárate, J.-M. Beckers, M. Rixen, and L. Vandenbulcke, 2007: Multigrid state vector for data assimilation in a two-way nested model of the ligurian sea. J. Mar. Sys., 65 (1), 41–59.
Bell, B. M., 1994: The iterated Kalman smoother as a Gauss–Newton method. SIAM Journal on Optimization, 4 (3), 626–636.
Bengtsson, T., C. Snyder, and D. Nychka, 2003: Toward a nonlinear ensemble filter for high-dimensional systems. Journal of Geophysical Research: Atmospheres, 108 (D24).
Bennett, A. F., 2005: Inverse modeling of the ocean and atmosphere. Cambridge University Press.
Berre, L., H. Varella, and G. Desroziers, 2015: Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4DVar at Météo-France. Q. J. R. Meteor. Soc., 141 (692), 2803–2812.
Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with ensemble transform Kalman filter. Part I: Theoretical aspects. Mon. Wea. Rev., 129, 420–436.
Bishop, C. H., and D. Hodyss, 2007: Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. Q. J. R. Meteor. Soc., 133, 2029–2044.
Blayo, E., M. Bocquet, E. Cosme, and L. F. Cugliandolo, 2014: Advanced data assimilation for geosciences. Oxford University Press, lecture Notes of the Les Houches School of Physics: Special Issue, 608 pp.
Blockley, E. W., M. J. Martin, A. J. McLaren, A. G. Ryan, J. Waters, D. J. Lea, I. Mirouze, K. A. Peterson, A. Sellar, and D. Storkey, 2014. Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts. Geosci. Model. Dev., 7, 2613–2638.
Bocquet, M., 2011: Ensemble Kalman filtering without the intrinsic need for inflation. Nonlin. Proc. Geophys., 18, 735–750.
Bocquet, M., 2016: Localization and the iterative ensemble Kalman smoother. Q. J. R. Meteor. Soc., 142, 1075–1089.
Bocquet, M., C. Pires, and L. Wu, 2010: Beyond Gaussian statistical modeling in geophysical data assimilation. Mon. Wea. Rev., 138 (8), 2997–3023.

Bocquet, M., P. N. Raanes, and A. Hannart, 2015: Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation. Nonlin. Processes Geophys., 22, 645.

Bocquet, M., and P. Sakov, 2012: Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems. Nonlin. Processes Geophys., 19 (3), 383–399.

Bocquet, M., and P. Sakov, 2013: Joint state and parameter estimation with an iterative ensemble Kalman smoother. Nonlin. Processes Geophys., 20 (5), 803–818.

Bocquet, M., and P. Sakov, 2014: An iterative ensemble Kalman smoother. Q. J. R. Meteor. Soc., 140 (682), 1521–1535.

Bonavita, M., L. Isaksen, and E. Hólm, 2012: On the use of EDA background error variances in the ECMWF 4DVar. Q. J. R. Meteor. Soc., 138 (667), 1540–1559.

Bonavita, M., L. Raynaud, and L. Isaksen, 2011: Estimating background-error variances with the ECMWF ensemble of data assimilations system: Some effects of ensemble size and day-to-day variability. Q. J. R. Meteor. Soc., 137 (655), 423–434.

Bouttier, F., and P. Courtier, 1999: Date assimilation concepts and methods. meteorological training lecture notes, ECMWF, shinfield park. Reading.

Bowler, N. E., J. Flowerdew, and S. R. Pring, 2013: Tests of different flavors of EnKF on a simple model. Q. J. R. Meteor. Soc., 139, 1505–1519.

Bowler, N. E., and Coauthors, 2017: Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations. Q. J. R. Meteor. Soc., 143 (704), 1280–1302.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. Q. J. R. Meteor. Soc., 131 (607), 1013–1043.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part-I: Description and single-observation experiments. Mon. Wea. Rev., 138 (5), 1550–1566.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010b: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. Mon. Wea. Rev., 138 (5), 1567–1586.

Buehner, M., J. Morneau, and C. Charette, 2013: Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. Nonlin. Processes Geophys., 20 (5), 669–682.

Buehner, M., and Coauthors, 2015: Implementation of deterministic weather forecasting systems based on ensemble–variational data assimilation at environment Canada. Part I: The global system. Mon. Wea. Rev., 143 (7), 2532–2559.

Buizza, R., M. Milleer, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Q. J. R. Meteorol. Soc., 125 (560), 2887–2908.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: On the analysis scheme in the ensemble Kalman filter. Mon. Wea. Rev., 126, 1719–1724.

Cane, M. A., A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert, and A. J. Busalacchi, 1996: Mapping tropical pacific sea level: Data assimilation via a reduced state space Kalman filter. Journal of Geophysical Research: Oceans, 101 (C10), 22 599–22 617.

Cappé, O., E. Moulines, and T. Rydén, 2005: Inference in Hidden Markov Models. Springer-Verlag.

Chassignet, E. P., H. E. Hurlburt, O. M. Smedstad, G. R. Halliwell, A. J. Hogan, P. J. Wallcraft, R. Baraille, and R. Bleck, 2007: The HYCOM (Hybrid Coordinate Ocean Model) data assimilative system. J. Mar. Sys., 65 (1), 60–83.

Chen, R., and J. Liu, 2000: Mixture Kalman filters. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62 (3), 493–508.

Chen, Y., and D. Oliver, 2013: Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. Computational Geosciences, 17, 689–703.

Chen, Y., and D. Zhang, 2006: Data assimilation for transient flow in geologic formations via ensemble Kalman filter. Advances in Water Resources, 29, 1107–1122.

Chepurin, G. A., J. A. Carton, and D. P. Dee, 2005: Forecast model bias correction in ocean data assimilation. Mon. Wea. Rev., 133, 1328–1342.

Chorin, A., M. Morzfeld, and X. Tu, 2010: Implicit particle filters for data assimilation. Commun. Appl. Math. Comput. Sci., 5, 221–240.

Clayton, A. M., A. C. Lorenc, and D. M. Barker, 2013: Operational implementation of a hybrid ensemble 4DVar global data assimilation system at the met office. Q. J. R. Meteor. Soc., 139 (675), 1445–1461.

Cohn, S., and R. Todling, 1996: Approximate data assimilation schemes for stable and unstable dynamics. J. Meteor. Soc. Japan, 74, 63–75.

Cosme, E., J.-M. Brankart, J. Verron, P. Brasseur, and M. Krysta, 2010: Implementation of a reduced rank square-root smoother for high resolution ocean data assimilation. Ocean Modelling, 33 (1), 87–100.

Cosme, E., J. Verron, P. Brasseur, J. Blum, and D. Auroux, 2012: Smoothing problems in a Bayesian framework and their linear Gaussian solutions. Mon. Wea. Rev., 140 (2), 683–695.

Counillon, F., and L. Bertino, 2009: Ensemble optimal interpolation: multivariate properties in the Gulf of Mexico. Tellus, 61A, 296 – 308.

Courtier, P., 1997: Dual formulation of four-dimensional variational assimilation. Q. J. R. Meteor. Soc., 123 (544), 2449–2461.

Courtier, P., J.-N. The´paut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4DVar using an incremental approach. Q. J. R. Meteor. Soc., 120, 1367–1387.

Cummings, J.A., O.M. Smedstad. 2013. Variational data assimilation for the global ocean. In: Park S., Xu L. (eds) Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol II). Springer, Berlin, Heidelberg.

Daley, R., 1991: Atmospheric data analysis. Cambridge Atmospheric and Space Science Series, Cambridge University Press, 6966, 25.

Daley, R., 1992: Estimating model-error covariances for application to atmospheric data assimilation. Mon. Wea. Rev., 120, 1735–1746.

Danforth, C. M., and E. Kalnay, 2008: Using singular value decomposition to parameterize state-dependent model errors. J. Atmos. Sci., 65, 1467–1478.

De La Chevrotie`re, M., and J. Harlim, 2017: A data-driven method for improving the correlation estimation in serial ensemble Kalman filters. Mon. Wea. Rev., 145, 985–1001.

Dee, D. P., 1995: Online estimation of error covariance parameters for atmospheric data assimilation. Mon. Wea. Rev., 123, 1128–1145.

Dee, D. P., and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. Q. J. R. Meteorol. Soc., 124, 269–295.

Delworth, T., S. Manabe, and R. J. Stouffer, 1993: Interdecadal variations of the thermohaline circulation in a coupled ocean-atmosphere model. Journal of Climate, 6, 1993–2011.

Desroziers, G., E. Arbogast, and L. Berre, 2016: Improving spatial localization in 4DEnVar. Q. J. R. Meteor. Soc., 142 (701), 3171–3185.

Desroziers, G., J.-T. Camino, and L. Berre, 2014: 4DEnVar: Link with 4D state formulation of variational assimilation and different possible implementations. Q. J. R. Meteor. Soc., 140 (684), 2097–2110.

Di Lorenzo, E., A. M. Moore, H. G. Arango, B. D. Cornuelle, A. J. Miller, B. Powell, B. S. Chua, and A. F. Bennett, 2007: Weak and strong constraint data assimilation in the inverse regional ocean modeling system (ROMS): Development and application for a baroclinic coastal upwelling system. Ocean Modelling, 16, 160–187.

Doucet, A., N. De Freitas, and N. Gordon, Eds., 2001: Sequential Monte Carlo methods in practice. Springer Verlag.

Doucet, A., N. De Freitas, K. P. Murphy, and S. J. Russell, 2000a: Rao-blackwellised particle filtering for dynamic Bayesian networks. Proceedings of the 16th world Conference on UAI, Stanford, California, USA, 176–83.

Doucet, A., S. Godsill, and C. Andrieu, 2000b: On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and Computing, 10 (3), 197–208.

Doucet, A., and A. M. Johansen, 2011: A tutorial on particle filtering and smoothing: fifteen years later. Oxford Handbook of Nonlinear Filtering.

Dreano, D., B. Mallick, and I. Hoteit, 2015: Filtering remotely sensed chlorophyll concentrations in the Red Sea using a space–time covariance model and a Kalman filter. Spatial Statistics, 13, 1–20.

Dreano, D., P. Tandeo, M. Pulido, B. Ait-El-Fquih, T. Chonavel, and I. Hoteit, 2017: Estimating model-error covariances in nonlinear state-space models using Kalman smoothing and the expectation-maximization algorithm. Q. J. R. Meteorol. Soc., 143, 1877–1885.

Duan, L., C. L. Farmer, and I. M. Moroz, 2010: Regularized particle filter with Langevin resampling step. AIP Conference Proceedings, 1281 (1), 1080–1083.

El-Sheikh, A., I. Hoteit, and M. Wheeler, 2014: Nested sampling particle filter for nonlinear data assimilation. Q. J. R. Meteorol. Soc., 140, 1640–1653.

Elbern, H., A. Strunk, H. Schmidt, and O. Talagrand, 2007: Emission rate and chemical state estimation by four-dimensional variational inversion. Atmos. Chem. Phys., 7, 3749–3769.

Emerick, A. A., and A. C. Reynolds, 2012: Ensemble smoother with multiple data assimilation. Computers & Geosciences, 55, 3–15.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans, 99 (C5), 10 143– 10 162.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean Dynamics, 53, 343–367.

Evensen, G., 2009: The ensemble Kalman filter for combined state and parameter estimation. IEEE Control Syst., 29 (3), 83–104.

Fairbairn, D., S. Pring, A. Lorenc, and I. Roulstone, 2014: A comparison of 4DVar with ensemble data assimilation methods. Q. J. R. Meteor. Soc., 140 (678), 281–294.

Fang, F., C. C. Pain, I. M. Navon, M. D. Piggott, G. J. Gorman, P. E. Farrell, P. A. Allison, and A. J. H. Goddard, 2009: A POD reduced-order 4DVar adaptive mesh ocean modelling approach. International Journal for Numerical Methods in Fluids, 60, 709–732.

Farrell, B. F., and P. J. Ioannou, 2001: State estimation using a reduced-order Kalman filter. Journal of the Atmospheric Sciences, 58 (23), 3666–3680.

Fertig, E. J., B. R. Hunt, E. Ott, and I. Szunyogh, 2007: Assimilating non-local observations with a local ensemble Kalman filter. Tellus A, 59, 719–730.

Frei, M., and H. Ku¨nsch, 2012: Sequential state and observation noise covariance estimation using combined ensemble Kalman and particle filters. Mon. Wea. Rev., 140, 1476–95.

Frei, M., and H. R. Ku¨nsch, 2013: Bridging the ensemble Kalman and particle filters. Biometrika, 100, 781–800.

Frolov, S., C. H. Bishop, T. Holt, J. Cummings, and D. Kuhl, 2016: Facilitating strongly coupled ocean-atmosphere data assimilation with an interface solver. Mon. Wea. Rev., 144, 3–20.

Fukumori, I., and P. Malanotte-Rizzoli, 1995: An approximate kaiman filter for ocean data assimilation: An example with an idealized gulf stream model. J. Geophys. Res.: Oceans, 100 (C4), 6777–6793.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. Q. J. R. Meteor. Soc., 125, 723–757.

Gharamti, M., J. Valstar, and I. Hoteit, 2014a: An adaptive hybrid EnKF -OI scheme for efficient state-parameter estimation of reactive contaminant transport models. Advances in Water Resources, 71, 1–15.

Gharamti, M. E., 2018: Enhanced adaptive inflation algorithm for ensemble filters. Mon. Wea. Rev., 146, 623–640.

Gharamti, M. E., B. Ait-El-Fquih, and I. Hoteit, 2015a: An iterative ensemble Kalman filter with one-step-ahead smoothing for state-parameters estimation of contaminant transport models. Journal of Hydrology, 527, 442–457.

Gharamti, M. E., B. Ait-El-Fquih, and I. Hoteit, 2015b: An iterative ensemble Kalman filter with one-step-ahead smoothing for state-parameters estimation of contaminant transport models. Journal of Hydrology, 527, 442–457.

Gharamti, M. E., A. Kadoura, S. Sun, J. Valstar, and I. Hoteit, 2014b: Constraining a compositional flow model with flow-chemical data using an ensemble Kalman filter. Water Resources Research, 50, 2444–2467.

Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. Advances in Geophysics, 33, 141–266.

Giering, R., and T. Kaminski, 1998: Recipes for adjoint code construction. ACM Transactions on Mathematical Software (TOMS), 24 (4), 437–474.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear and non-Gaussian Bayesian state estimation. IEE Proceedings F in Radar and Signal Processing, 140, 107–113.

Gu, Y., and D. Oliver, 2007: An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. SPE Journal, 12, 438–446.

Gustafsson, N., J. Bojarova, and O. Vignes, 2014: A hybrid variational ensemble data assimilation for the high resolution limited area model (HIRLAM). Nonlin. Processes Geophys., 21 (1), 303–323.

Haley, P. J., and Coauthors, 2009: Forecasting and reanalysis in the Monterey Bay/California current region for the autonomous ocean sampling network-ii experiment. Deep Sea Research Part II: Topical Studies in Oceanography, 56 (3), 127–148.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. Mon. Wea. Rev., 128 (8), 2905–2919.

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

Hamill, T. M., J. S. Whitaker, J. L. Anderson, and C. Snyder, 2009: Comments on “Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems”. J. Atmos. Sci., 66, 3498–3500.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon. Wea. Rev., 129, 2776–2790.

Hamilton, F., T. Berry, and T. Sauer, 2016: Ensemble Kalman filtering without a model. Physical Review X, 6 (1), 011 021.

Hamilton, J. D., 1986: State-space models. R.F. Engle & D. McFadden, H. of Economics, Ed., Vol. 4, 1st ed., Elsevier, chap. 50, 3039–80.

Harvey, A. C., and G. D. A. Phillips, 1979: Maximum likelihood estimation of regression models with autoregressive-moving average disturbance. Biometrika, 66, 49–58.
Heimbach, P., C. Hill, and R. Giering, 2002: Automatic generation of efficient adjoint code for a parallel Navier-Stokes solver. Computational Science ICCS 2002, 1019–1028.
Hoteit, I., D.-T. Pham, M. E. Gharamti, and X. Luo, 2015: Mitigating observation perturbation sampling errors in the stochastic EnKF. Mon. Wea. Rev., 143 (7).
Hoteit, I., X. Luo, and D. T. Pham, 2012: Particle Kalman filtering: A nonlinear Bayesian framework for ensemble Kalman filters. Mon. Wea. Rev., 140, 528–542.
Hoteit, I., B. Cornuelle, and P. Heimbach, 2010: An eddy permitting variational data assimilation system for estimating the of the tropical pacific. J. Geophys. Res., 115, C03001, doi:10.1029/2009JC005 347.
Hoteit, I., 2008: A reduced-order simulated annealing approach for four-dimensional variational data assimilation in meteorology and oceanography. Int. J. Numer. Meth. Fluids, 58 (11), 1181–1199.
Hoteit, I., D. T. Pham, G. Triantafyllou, and G. Korres, 2008: A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography. Mon. Wea. Rev., 136, 317–334.
Hoteit, I., G. Triantafyllou, and G. Korres, 2007: Using low-rank ensemble Kalman filters for data assimilation with high dimensional imperfect models. J. Num. Ana. Ind. Appl. Math., 2 (1-2), 67–78.
Hoteit, I., and A. Köhl, 2006: Efficiency of reduced-order, time-dependent adjoint data assimilation approaches. Journal of Oceanography, 62 (4), 539–550.
Hoteit, I., B. Cornuelle, A. Köhl, and D. Stammer, 2005a: Treating strong adjoint sensitivities in tropical eddy-permitting variational data assimilation. Q. J. R. Meteorol. Soc., 131, 3659–3682.
Hoteit, I., G. Korres, and G. Triantafyllou, 2005b: Comparison of extended and ensemble based Kalman filters with low and high-resolution primitive equations ocean models. Nonlin. Processes Geophys., 12, 755–765.
Hoteit, I., and D.-T. Pham, 2004: An adaptively reduced-order extended Kalman filter for data assimilation in the tropical pacific. J. Mar. Sys., 45 (3), 173–188.
Hoteit, I., and D. T. Pham, 2003: Evolution of the reduced state space and data assimilation schemes based on the Kalman filter. Journal of the Meteorological Society of Japan, 81, 21–39.
Hoteit, I., D. T. Pham, and J. Blum, 2002: A simplified reduced order Kalman filtering and application to altimetric data assimilation in tropical pacific. J. Mar. Sys., 36, 101–127.
Hoteit, I., D. T. Pham, and J. Blum, 2001: A semi-evolutive partially local filer for data assimilation. Marine Pollution Bulletin, 43, 164–174.
Houtekamer, P., and F. Zhang, 2016: Review of the ensemble Kalman filter for atmospheric data assimilation. Mon. Wea. Rev., 144, 4489–4532.
Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. Mon. Wea. Rev., 126, 796–811.
Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble filter for atmospheric data assimilation. Mon. Wea. Rev., 129, 123–137.
Houtekamer, P. L., and H. L. Mitchell, 2005: Ensemble Kalman filtering. Q. J. R. Meteor. Soc., 131, 3269–3289.
Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. Mon. Wea. Rev., 133 (3), 604–620.
Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. Physica D, 230 (1), 112–126.
Hunt, B. R., and Coauthors, 2004: Four-dimensional ensemble Kalman filtering. Tellus A, 56 (4), 273–277.
Ito, K., and K. Xiong, 2000: Gaussian filters for nonlinear filtering problems. IEEE Transactions on Automatic Control, 45, 910–927.
Janekovic, I., B. Powell, D. Matthews, M. McManus, and J. Sevadjian, 2013: 4DVar data assimilation in a nested, coastal ocean model: A hawaiian case study. Journal of Geophysical Research: Oceans, 118 (10), 5022–5035.
Jazwinski, A. H., 1970: Stochastic Processes and Filtering Theory. Academic Press.
Julier, S., J. Uhlmann, and H. Durrant-Whyte, 2000: A new method for the nonlinear transformation of means and covariances in filters and estimators. IEEE Transactions on Automatic Control, 45, 477–482.
Julier, S., and J. K. Uhlmann, 2004: Unscented filtering and nonlinear estimation. Proc. IEEE, 92, 401–422.
Kalman, R., 1960: A new approach to linear filtering and prediction problems. J. Basic Eng., 82 (1), 35–45.
Kalnay, E., H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabera-Poy, 2007: 4DVar or ensemble Kalman filter? Tellus A, 59 (5), 758–773.
Kim, S.-B., I. Fukumori, and T. Lee, 2006: The closure of the ocean mixed layer temperature budget using level-coordinate model fields. Journal of Atmospheric and Oceanic Technology, 23 (6), 840–853.

Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the ncep gfs. Part I: System description and 3D-hybrid results. Mon. Wea. Rev., 143 (2), 433–451.

Köhl, A., and D. Stammer, 2008: Variability of the meridional overturning in the North Atlantic from the 50 years GECCO state estimation. J. Phys. Oceanogr., 38, 1913–1930.

Köhl, A., and J. Willebrand, 2002: An adjoint method for the assimilation of statistical characteristics into eddy-resolving ocean models. Tellus, 54, 406–425.

Korres, G., K. Nitti, L. Perivoliotis, K. Tsiaras, A. Papadopoulos, G. Triantafyllou, I. Hoteit, and K. Abdullah, 2010: Forecasting the Aegean Sea hydrodynamics within the Poseidon-ii operational system. Journal of Operational Oceanography, 3 (1), 37–49.

Law, K., A. Stuart, and K. Zygalakis, 2015: Data assimilation: a mathematical introduction, Vol. 62. Springer.

Lawson, W. G., and J. A. Hansen, 2004: Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. Mon. Wea. Rev., 132 (8), 1966–1981.

Le Dimet, F. X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. Tellus A, 38A (2), 97–110.

Lea, D., M. Allen, and T. Haine, 2000: Sensitivity analysis of the climate of a chaotic system. Tellus A, 52, 523–532.

Lee, Y., A. J. Majda, and D. Qi, 2017: Preventing catastrophic filter divergence using adaptive additive inflation for baroclinic turbulence. Mon. Wea. Rev., 145, 669–682.

Lei, J., P. Bickel, and C. Snyder, 2010: Comparison of ensemble Kalman filters under non-Gaussianity. Mon. Wea. Rev., 138, 1293–1306.

Lellouche J.-M., O. Le Galloudec, M. Drévillon, C. Régnier, E. Greiner, G. Garric, N. Ferry, C. Desportes, C.-E. Testut, C. Bricaud, R. Bourdallé-Badie, B. Tranchant, M. Benkiran, Y. Drillet, A. Daudin, and C. De Nicola, 2013. Evaluation of global monitoring and forecasting systems at Mercator Océan. Ocean Science, 9, 57–81,

Lermusiaux, P., and A. Robinson, 1999: Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. Mon. Wea. Rev., 127, 1385–1407.

Lguensat, R., P. Tandeo, P. Ailliot, M. Pulido, and R. Fablet, 2017: The analog data assimilation. Mon. Wea. Rev., 145 (10), 4093–4107.

Li, G., M. Iskandarani, M. Le Henaff, J. Winokur, O. Le Maitre, and O. Knio, 2015: Quantifying initial and wind forcing uncertainties in the Gulf of Mexico. Comput. Geosci., 20, 1133–1153.

Li, H., E. Kalnay, and T. Miyoshi, 2009a: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. Q. J. R. Meteor. Soc., 135, 523–533.

Li, H., E. Kalnay, T. Miyoshi, and C. M. Danforth, 2009b: Accounting for model errors in ensemble data assimilation. Mon. Wea. Rev., 137, 3407–3419.

Liu, B., B. Ait-El-Fquih, and I. Hoteit, 2016: Efficient kernel-based ensemble Gaussian mixture filtering. Mon. Wea. Rev., 144 (2), 781–800.

Liu, C., A. Köhl, and D. Stammer, 2012: Adjoint-based estimation of eddy-induced tracer mixing parameters in the global ocean. Journal of Physical Oceanography, 42 (7), 1186–1206.

Liu, C., Q. Xiao, and B. Wang, 2008: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. Mon. Wea. Rev., 136 (9), 3363–3373.

Liu, C., Q. Xiao, and B. Wang, 2009: An ensemble-based four-dimensional variational data assimilation scheme. Part II: Observing system simulation experiments with advanced research WRF (ARW). Mon. Wea. Rev., 137 (5), 1687–1704.

Liu, L., H. Ji, and Z. Fan, 2015: Improved iterated-corrector-PHD with Gaussian mixture implementation. Signal Processing, 114, 89–99.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP a comparison with 4DVar. Q. J. R. Meteor. Soc., 129 (595), 3183–3203.

Lorenc, A. C., N. E. Bowler, A. M. Clayton, S. R. Pring, and D. Fairbairn, 2015: Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. Mon. Wea. Rev., 143 (1), 212–229.

Lorentzen, R., and G. Nævdal, 2011: An iterative ensemble Kalman filter. IEEE Transactions on Automatic Control, 56, 1990 –1995.

Lu, F., Z. Liu, S. Zhang, and Y. Liu, 2015: Strongly coupled data assimilation using leading averaged coupled covariance (lacc). Part I: Simple model study. Mon. Wea. Rev., 143, 3823–3837.

Luo, A., L. M. Moroz, and I. Hoteit, 2010: Scaled unscented transform Gaussian sum filter: Theory and application. Physica-D, 239, 684–701.

Luo, X., T. Bhakta, and G. Nædal, 2018: Correlation-based adaptive localization with applications to ensemble-based 4D seismic history matching. SPE Journal, in press, SPE-185936-PA.

Luo, X., and I. Hoteit, 2011: Robust ensemble filtering and its relation to covariance inflation in the ensemble Kalman filter. Mon. Wea. Rev., 139, 3938–3953.
Luo, X., and I. Hoteit, 2013: Covariance inflation in the ensemble Kalman filter: a residual nudging perspective and some implications. Mon. Wea. Rev., 141, 3360–3368.
Luo, X., and I. Hoteit, 2014a: Efficient particle filtering through residual nudging. Q. J. R. Meteor. Soc., 140, 557–572.
Luo, X., and I. Hoteit, 2014b: Ensemble Kalman filtering with a divided state-space strategy for coupled data assimilation problems. Mon. Wea. Rev., 142, 4542–4558.
Luo, X., and I. Hoteit, 2014c: Ensemble Kalman filtering with residual nudging: an extension to the state estimation problems with nonlinear observations. Mon. Wea. Rev., 142, 3696–3712.
Luo, X., I. Hoteit, and I. M. Moroz, 2012: On a nonlinear Kalman filter with simplified divided difference approximation. Physica D, 241, 671–680.
Luo, X., and I. M. Moroz, 2009: Ensemble Kalman filter with the unscented transform. Physica D, 238, 549–562.
Luo, X., A. Stordal, R. Lorentzen, and G. Nævdal, 2015: Iterative ensemble smoother as an approximate solution to a regularized minimum-average-cost problem: theory and applications. SPE Journal, 20, 962–982.
Malinverno, A., 2002: Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. Geophysical Journal International, 151, 675–688.
Mehra, R., 1970: On the identification of variances and adaptive Kalman filtering. IEEE Transactions on Automatic Control, 15, 175–184.
Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953: Equation of state calculations by fast computing machines. J. Chem. Phys., 21, 1087–1092.
Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. Mon. Wea. Rev., 128, 416–433.
Miyoshi, T., 2011: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. Mon. Wea. Rev., 139, 1519–1535.
Moore, A. M., and Coauthors, 2011: The regional ocean modeling system (roms) four-dimensional variational data assimilation systems: Part II–performance and application to the California Current system. Progress in Oceanography, 91 (1), 50–73.
Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. R. Houser, 2005: Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources, 28 (2), 135–147.
Nerger, L., W. Hiller, and J. Schröter, 2005: A comparison of error subspace Kalman filters. Tellus A, 57 (5), 715–735.
Nerger, L., T. Janjic, J. Schrter, and H. W., 2012: A unification of ensemble square root Kalman filters. Mon. Wea. Rev., 140 (7), 2335–2345.
Nerger, L., S. Schulte, and A. Bunse-Gerstner, 2014: On the influence of model nonlinearity and localization on ensemble Kalman smoothing. Q. J. R. Meteor. Soc., 140 (684), 2249–2259.
Ngodock, H., and M. Carrier, 2014: A 4D-Var system for the Navy coastal ocean model. Part II: Strong and weak constraint assimilation experiments with real observations in Monterey Bay. Mon. Wea. Rev., 142 (6), 2108–2117.
Oke, P. R., G. B. Brassington, and A. Griffin, and D. A. Schiller, 2008: The Bluelink ocean data assimilation system (bodas). Ocean Modelling, 21 (1), 46–70.
Oke, P. R., P. Sakov, and S. P. Corney, 2007: Impacts of localisation in the EnKF and EnOI: experiments with a small model. Ocean Dynamics, 57 (1), 32–45.
Oliver, D., and Y. Chen, 2011: Recent progress on reservoir history matching: A review. Computational Geosciences, 15, 185–221.
Oliver, D. S., N. He, and A. C. Reynolds, 1996: Conditioning permeability fields to pressure data. ECMOR V-5th European Conference on the Mathematics of Oil Recovery.
Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. Tellus, 56A, 415–428.
Pagan, A., 1980: Some identification and estimation results for regression models with stochastically varying coefficients. Journal of Econometrics, 13, 341–63.
Penny, S. G., and T. M. Hamill, 2017: Coupled data assimilation for integrated earth system analysis and prediction. Bulletin of the American Meteorological Society, 97 (7), ES169–ES172.
Penny, S. G., and T. Miyoshi, 2016: A local particle filter for high-dimensional geophysical systems. Nonlin. Processes Geophys, 23, 391–405.
Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. Mon. Wea. Rev., 129, 1194–1207.

Pham, D. T., J. Verron, and C. Roubaud, 1997: Singular evolutive Kalman filter with EOF initialization for data assimilation in oceanography. J. Mar. Syst., 16, 323–340.

Pires, C., R. Vautard, and O. Talagrand, 1996: On extending the limits of variational assimilation in nonlinear chaotic systems. Tellus, 48, 96–121.

Poterjoy, J., 2016: A localized particle filter for high-dimensional nonlinear systems. Mon. Wea. Rev., 144, 59–76.

Poterjoy, J., and F. Zhang, 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. Wea. Rev., 143 (5), 1601–1621.

Raanes, P. N., 2016: On the ensemble Rauch-Tung-Striebel smoother and its equivalence to the ensemble Kalman smoother. Q. J. R. Meteor. Soc., 142 (696), 1259–1264.

Rasmussen, J., H. Madsen, K. H. Jensen, and J. C. Refsgaard, 2015a: Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance. Hydrology and Earth System Sciences, 19, 2999–3013.

Rasmussen, J., H. Madsen, K. H. Jensen, and J. C. Refsgaard, 2015b: Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: Evaluating the effect of ensemble size and localization on filter performance. Hydrology Earth System Science, 19, 2999–3013.

Raynaud, L., L. Berre, and G. Desroziers, 2009: Objective filtering of ensemble-based background-error variances. Q. J. R. Meteor. Soc., 135 (642), 1177–1199.

Reich, S., 2013: A nonparametric ensemble transform method for Bayesian inference. SIAM Journal on Scientific Computing, 35, A2013–A2024.

Robert, C., 2007: The Bayesian choice: From decision-theoretic foundations to computational implementation. Springer Science & Business Media, New York.

Robert, C., S. Durbiano, E. Blayo, J. Verron, J. Blum, and F.-X. Le Dimet, 2005: A reduced-order strategy for 4DVar data assimilation. J. Mar. Sys., 57 (1), 70–82.

Sakov, P., and L. Bertino, 2011: Relation between two common localisation methods for the EnKF. Computational Geosciences, 15 (2), 225–237.

Sakov, P., and P. R. Oke, 2008: Implications of the form of the ensemble transformation in the ensemble square root filters. Mon. Wea. Rev., 136 (3), 1042–1053.

Sakov, P., D. S. Oliver, and L. Bertino, 2012: An iterative EnKF for strongly nonlinear systems. Mon. Wea. Rev., 140 (6), 1988–2004.

Sakov P., F. Counillon, L. Bertino, K.A. Lisæter, P. Oke, A. Korablev. 2012. TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. Ocean Science, 8, 633–656.

Sakov, P., and P. A. Sandery, 2015: Comparison of EnOI and EnKF regional ocean reanalysis systems. Ocean Modelling, 89, 45–60.

Särkkä, S., and J. Hartikainen, 2013: Non-linear noise adaptive Kalman filtering via variational Bayes. Proceedings of the IEEE 2013 International Workshop on Machine Learning for Signal Processing.

Shen, Z., and Y. Tang, 2015: A modified ensemble Kalman particle filter for non-Gaussian systems with nonlinear measurement functions. Journal of Advances in Modeling Earth Systems, 7, 50–66.

Shumway, R. H., and D. S. Stoffer, 1982: An approach to time series smoothing and forecasting using the EM algorithm. Journal of Time Series Analysis, 3 (4), 253–264.

Simon, D., 2006: Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches. Wiley-Interscience, 552 pp.

Sluka, T. C., S. G. Penny, E. Kalnay, and T. Miyoshi, 2016: Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation. Geophysical Research Letters, 43, 752–759.

Smedstad, O. M., H. E. Hurlburt, E. J. Metzger, R. C. Rhodes, J. F. Shriver, A. J. Wallcraft, and A. B. Kara, 2003: An operational eddy resolving 1/16 global ocean nowcast/forecast system. J. Mar. Sys., 40, 341–361.

Smith, K. D., A. M. Moore, and H. G. Arango, 2015: Estimates of ocean forecast error covariance derived from hessian singular vectors. Ocean Modelling, 89, 104–121.

Smith, P. J., A. S. Lawless, and N. K. Nichols, 2017a: Estimating forecast error covariances for strongly coupled atmosphere–ocean 4D-Var data assimilation. Mon. Wea. Rev., 145 (10), 4011–4035.

Smith, S., H. Ngodock, M. Carrier, J. Shriver, P. Muscarella, and I. Souopgui, 2017b: Validation and operational implementation of the Navy Coastal Ocean Model Four-Dimensional Variational Data Assimilation system (NCOM 4DVar) in the Okinawa Trough. Data Assimilation for Atmospheric, Oceanic and Hydro- logic Applications (Vol. III), Springer, 405–427.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. Mon. Wea. Rev., 136, 4629–4640.

Sondergaard, T., and P. F. Lermusiaux, 2013: Data assimilation with Gaussian mixture models using the dynamically orthogonal field equations. Part I: Theory and scheme. Mon. Wea. Rev., 141 (6), 1737–1760.
Song, H., I. Hoteit, B. Cornuelle, and A. C. Subramanian, 2010: An adaptive approach to mitigate background covariance limitations in the ensemble Kalman filter. Mon. Wea. Rev., 138 (7), 2825–2845.
Sorenson, H. W., and D. L. Alspach, 1971: Recursive Bayesian estimation using Gaussian sums. Automatica, 7, 465 – 479.
Sraj, I., S. Zedler, C. Jackson, O. Knio, and I. Hoteit, 2016: Polynomial chaos-based Bayesian inference of K-profile parameterization in a general circulation model of the Tropical Pacific. Mon. Wea. Rev., 144 (12), 4621–4640.
Sripitana, A., T. Mayo, I. Sraj, O. Knio, C. Dawson, O. Le Maitre, and I. Hoteit, 2017: Assessing an ensemble Kalman filter inference of manning's n coefficient of a storm surge model against a polynomial chaos-based MCMC. Ocean Dynamics, 67, 1067–1094.
Stockdale, T. N., D. L. T. Anderson, J. O. S. Alves, and M. A. Balmaseda, 1998: Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. Nature, 392, 370–373.
Stroud, J. R., and T. Bengtsson, 2007: Sequential state and variance estimation within the ensemble Kalman filter. Mon. Wea. Rev., 135, 3194–3208.
Subramanian, A. C., I. Hoteit, I. Cornuelle, A. J. Miller, and H. Song, 2012: Linear versus nonlinear filtering with scale-selective corrections for balanced dynamics in a simple atmospheric model. Journal of the Atmospheric Sciences, 69, 3405–3419.
Talagrand, O., 2010: Variational assimilation. Data assimilation, Springer, 41–67.
Tarantola, A., 2005: Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics (SIAM), 339 pp.
Tardif, R., G. J. Hakim, and C. Snyder, 2015: Coupled atmosphere–ocean data assimilation experiments with a low-order model and CMIP5 model data. Climate Dynamics, 45, 1415–1427.
Temam, R., 1984: Navier-Stokes equations: Theory and Numerical Analysis, Vol. 343. AMS Chelsea Publishing.
Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square-root filters. Mon. Wea. Rev., 131, 1485 – 1490.
Toye, H., P. Zhan, G. Gopalakrishnan, A. Kartadikaria, H. Huang, O. Knio, and I. Hoteit, 2017: Ensemble data assimilation in the Red Sea: sensitivity to ensemble selection and atmospheric forcing. Ocean Dynamics, 67, 915–933.
Trémolet, Y., 2007: Model-error estimation in 4D-Var. Q. J. R. Meteorol. Soc., 133, 1267–1280.
Tsiaras, K. P., I. Hoteit, S. Kalaroni, G. Petihakis, and G. Triantafyllou, 2017: A hybrid ensemble-OI Kalman filter for efficient data assimilation into a 3-D biogeochemical model of the Mediterranean. Ocean Dynamics, 67, 673–690.
Ueno, G., T. Higuchi, T. Kagimoto, and N. Hirose, 2010: Maximum likelihood estimation of error covariances in ensemble-based filters and its application to a coupled atmosphere–ocean model. Q. J. R. Meteorol. Soc., 136, 1316–1343.
Usui, N., S. Ishizaki, Y. Fujii, H. Tsujino, T. Yasuda, and M. Kamachi, 2006: Meteorological research institute multivariate ocean variational estimation (MOVE) system: Some early results. Advances in Space Research, 37 (4), 806–822.
Van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. Mon. Wea. Rev., 137, 4089–4114.
Van Leeuwen, P. J., 2010: Nonlinear data assimilation in geosciences: An extremely efficient particle filter. Q. J. R. Meteor. Soc., 136 (653), 1991–1999.
van Leeuwen, P. J., 2015: Representation errors and retrievals in linear and nonlinear data assimilation. Q. J. R. Meteorol. Soc., 141, 1612–1623.
Verlaan, M., and A. W. Heemink, 1997: Tidal flow forecasting using reduced rank square-root filters. Stochastic hydrology and Hydraulics, 11 (5), 349–368.
Vermeulen, P. T., and A. W. Heemink, 2006: Model-reduced variational data assimilation. Mon. Wea. Rev., 134, 2888–2899.
Verron, J., L. Gourdeau, D. T. Pham, R. Murtugudde, and A. Busalacchi, 1999: An extended Kalman filter to assimilate satellite altimeter data into a nonlinear numerical model of the tropical Pacific Ocean: Method and validation. Journal of Geophysical Research: Oceans, 104 (C3), 5441–5458.
Vlasenko, A. V., A. Köhl, and D. Stammer, 2016: The efficiency of geophysical adjoint codes generated by automatic differentiation tools. Computer Physics Communications, 199, 22–28.
Wang, X., C. H. Bishop, and S. J. Julier, 2004: Which is better, an ensemble of positive-negative pairs or a centered simplex ensemble. Mon. Wea. Rev., 132, 1590–1605.
Wang, X., C. Snyder, and T. M. Hamill, 2007: On the theoretical equivalence of differently proposed ensemble–3DVAR hybrid analysis schemes. Mon. Wea. Rev., 135 (1), 222–227.

Weaver, A., J. Vialard, and D. L. T. Anderson, 2003: Three-and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part I: Formulation, internal diagnostics, and consistency checks. Mon. Wea. Rev., 131 (7), 1360–1378.

Weaver, A. T., C. Deltel, É. Machu, S. Ricci, and N. Daget, 2005: A multivariate balance operator for variational ocean data assimilation. Q. J. R. Meteor. Soc., 131 (613), 3605–3625.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. Mon. Wea. Rev., 130 (7), 1913–1924.

Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. Mon. Wea. Rev., 140, 3078–3089.

Wikle, C. K., and L. M. Berliner, 2007: A Bayesian tutorial for data assimilation. Physica D: Nonlinear Phenomena, 230 (1), 1–16.

Wu, L., V. Mallet, M. Bocquet, and B. Sportisse, 2008: A comparison study of data assimilation algorithms for ozone forecasts. Journal of Geophysical Research: Atmospheres, 113 (D20), doi:10.1029/2008JD009991.

Wunsch, C., 1996: The ocean circulation inverse problem. Cambridge University Press, Cambridge, UK.

Wunsch, C., and P. Heimbach, 2007: Practical global oceanic state estimation. Physica D, 230, 197–208.

Xie, J., and J. Zhu, 2010: Ensemble optimal interpolation schemes for assimilating Argo profiles into a hybrid coordinate ocean model. Ocean Modelling, 33 (3), 283–298.

Yang, S.-C., E. Kalnay, and T. Enomoto, 2015: Ensemble singular vectors and their use as additive inflation in EnKF. Tellus A: Dynamic Meteorology and Oceanography, 67, 26 536.

Yaremchuk, M., P. Martin, A. Koch, and C. Beattie, 2016: Comparison of the adjoint and adjoint-free 4DVar assimilation of the hydrographic and velocity observations in the Adriatic Sea. Ocean Modelling, 97, 129–140.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. Mon. Wea. Rev., 132, 1238–1253.

Zhang, M., and F. Zhang, 2012: E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. Mon. Wea. Rev., 140 (2), 587–600.

Zhang, W. G., J. L. Wilkin, and H. G. Arango, 2010: Towards an integrated observation and modeling system in the New York Bight using variational methods. Part I: 4D-Var data assimilation. Ocean Modelling, 35 (3), 119–133.

Zhang, Y., and D. S. Oliver, 2010: Improving the ensemble estimate of the Kalman gain by bootstrap sampling. Mathematical Geosciences, 42, 327–345.

Zupanski, D., and M. Zupanski, 2006: Model error estimation employing an ensemble data assimilation approach. Mon. Wea. Rev., 134, 1337–1354.

Zupanski, M., 2005: Maximum likelihood ensemble filter: theoretical aspects. Mon. Wea. Rev., 133, 1710–1726.