

OVERVIEW

Data assimilation in the geosciences: An overview of methods, issues, and perspectives

Alberto Carrassi¹  | Marc Bocquet² | Laurent Bertino¹ | Geir Evensen^{1,3}

¹Nansen Environmental and Remote Sensing Center, Bergen, Norway

²CEREA, Joint Laboratory École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France

³International Research Institute of Stavanger (IRIS), Bergen, Norway

Correspondence

Alberto Carrassi, Nansen Environmental and Remote Sensing Center, Bergen, Norway.
 Email: alberto.carrassi@nerc.no

Funding information

Embla of Nordforsk; Norwegian Research Council, Grant/Award Number: REDDA (#250711)

Edited by Eduardo Zorita, Domain Editor, and Mike Hulme, Editor-in-Chief

We commonly refer to state estimation theory in geosciences as data assimilation (DA). This term encompasses the entire sequence of operations that, starting from the observations of a system, and from additional statistical and dynamical information (such as a dynamical evolution model), provides an estimate of its state. DA is standard practice in numerical weather prediction, but its application is becoming widespread in many other areas of climate, atmosphere, ocean, and environment modeling; in all circumstances where one intends to estimate the state of a large dynamical system based on limited information. While the complexity of DA, and of the methods thereof, stands on its interdisciplinary nature across statistics, dynamical systems, and numerical optimization, when applied to geosciences, an additional difficulty arises by the continually increasing sophistication of the environmental models. Thus, in spite of DA being nowadays ubiquitous in geosciences, it has so far remained a topic mostly reserved to experts. We aim this overview article at geoscientists with a background in mathematical and physical modeling, who are interested in the rapid development of DA and its growing domains of application in environmental science, but so far have not delved into its conceptual and methodological complexities.

This article is categorized under:

Climate Models and Modeling > Knowledge Generation with Models

KEY WORDS

Bayesian methods, data assimilation, ensemble methods, environmental prediction

Contents

1	Introduction	2
2	State estimation: Formulation of the problem	2
3	A route to solutions: The gaussian approximation	7
4	Ensemble methods	12
5	Special topics	23
6	Where we are and where we go: A look at the perspectives	29
7	Conclusion	36
8	Appendices	45

1 | INTRODUCTION

The purpose of this article is to provide a comprehensive, state-of-the-art overview of the methods and challenges of data assimilation (DA) in the geosciences. We aim the article at geoscientists confronted with the problem of combining data with models and need to learn DA, but who are intimidated by the vast, and technical, literature. This work may guide them through a first journey into the topic while being at the same time as complete and precise as possible.

The finest mathematical details, at the crossing between different areas such as numerical methods, algebra, statistics, or dynamical systems, are essential to grasp the sense of the DA problem, to reveal its interdisciplinary nature and, for many, its beauty. We have nevertheless avoided all technicalities that were not, in our opinion, strictly necessary and have worked to make the narrative intelligible for geoscientists, or climate scientists, who do not usually possess a strong background in mathematics.

We provide the readers with an extensive bibliography, but also recognize that it is beyond the scope of this work to make it entirely exhaustive. In the last decade or so, DA has attracted much attention. The topic has recently reached the rank of a discipline per se, as testified by the appearance of books such as van Leeuwen, Cheng, and Reich (2015), Reich and Cotter (2015), Law, Stuart, and Zygalakis (2015), Asch, Bocquet, and Nodet (2016), or Fletcher (2017). These works have mainly addressed the mathematical dimension and formulation of the DA problem and complement seminal books on DA (e.g., Daley, 1993; Evensen, 2009b; Kalnay, 2002), that had spread the knowledge about DA and shaped it as an independent discipline. The present article places itself somehow in between these two classes of textbooks and aims at bridging the mathematical side of the DA methods with the practicalities and the physical intuitions and know how that have guided its dramatic development in climate science and geosciences in general.

The paper is structured to provide a first general formulation of the state estimation problem, from a Bayesian perspective, in Section 2. Having defined the issue, and illustrated the intrinsic difficulties to adopt the Bayesian approach in the geosciences, the two most popular families of DA algorithms, both based on a Gaussian approximation, namely the Kalman filter (KF) and the variational methods, are described in Section 3 and in the complementary Appendices A–D. Section 4 is entirely devoted to the ensemble methods, from their origin, through the most relevant successful variants and up to the new frontier of hybrid ensemble-variational methods. Four selected topics, characterizing the DA problems in the geosciences, are the subjects of Section 5, and are examples of the specific type of approximations and compromises, as well as the level of innovations DA has achieved. Finally Section 6 presents a prospect of the recent challenges and current directions of research and developments, with special attention to two, the particle filter (PF) and coupled DA (CDA).

2 | STATE ESTIMATION: FORMULATION OF THE PROBLEM

2.1 | Premises

The problem we intend to solve is the estimation of the state of a system, say the atmosphere, the ocean, or any component of the Earth system or its whole, at any arbitrary past, present, and future time. We possess two complementary, but both incomplete and inaccurate, sources of information: the *observations* and the *model*. DA provides the conceptual and methodological tools to tackle the problem by extracting synergies between model and observations and by exploiting their respective informational content. Given the nature of the modeling and observation infrastructure in the geosciences, DA is conveniently formalized as a discrete-model/discrete-observation estimation problem, as we will do in this overview. We remark, however, that, the *nudging method* (e.g., Hoke & Anthes, 1976; Lakshminarayanan & Lewis, 2013), one of the simplest and most straightforward approach to assimilate data in a dynamical model, is better formulated as continuous-model/continuous-data problem. Nudging will not be addressed in this overview, but it is worth mentioning that it has recently awoken new attention that has brought the introduction of new advanced formulations (e.g., Auroux & Blum, 2008; Pazo, Carrassi, & Lopez, 2016), and to the study of its connection with the synchronization problem (Duane, Tribbia, & Weiss, 2006). Interested readers can find a complete treatment of the continuous–continuous and discrete–continuous cases in many textbooks on estimation theory (see Bain & Crisan, 2009; Jazwinski, 1970).

Throughout the text, the following notation convention is used: $\mathbf{x} \in \mathbb{R}^m$ means that \mathbf{x} is a m -dimensional vector whose components are real numbers; $f: \mathbb{R}^l \rightarrow \mathbb{R}^n$ signifies that the function f transforms an l -dimensional vector into an n -dimensional one; model values at discrete times are indicated as $\mathbf{x}(t_k) = \mathbf{x}_k$. We can now formalize the two ingredients entering the estimation problem.

2.1.1 | The dynamical model

Let us assume that a model of the natural processes of interest is available as a discrete stochastic-dynamical system,

$$\mathbf{x}_k = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}, \boldsymbol{\lambda}) + \boldsymbol{\eta}_k. \quad (1)$$

Here, $\mathbf{x}_k \in \mathbb{R}^m$ and $\boldsymbol{\lambda} \in \mathbb{R}^p$ are the model state and parameter vectors, respectively, $\mathcal{M}_{k:k-1} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is usually a nonlinear, possibly chaotic, function from time t_{k-1} to t_k , and $\boldsymbol{\eta}_k \in \mathbb{R}^m$ is the model error, represented here as a stochastic additive term, although it could be included into the parenthesis without loss of generality. The model parameters may include the external forcings or the boundary conditions.

The model error, $\boldsymbol{\eta}_k$, is intended to represent the error of a model prediction of the true unknown process at time t_k , initialized from the true state (perfect initial condition) at time t_{k-1} . It accounts for the cumulative effect, over the interval $t_k - t_{k-1}$, of errors in the parameters, $\boldsymbol{\lambda}$, errors in the numerical schemes used to integrate Equation (1) as well as the effect of the unresolved scales. The two latter arise from the spatiotemporal discretization from physical laws (e.g., the Navier Stokes equations) expressed as partial differential equations on a continuous media, into difference equations on a discrete grid or finite spectral modes. The appropriateness of the stochastic formulation of the model error is questionable in many cases (see Nicolis, 2003). Nevertheless it has the advantage of fitting very well to the probabilistic Bayesian approach to the DA problem, as we will clarify later. Alternative forms of model error treatment in DA have been proposed recently, including a deterministic one (see Carrassini & Vannitsem, 2016, and references therein), but they will not be developed in this overview.

The dynamical model can also include an explicit dependence on time (i.e., be nonautonomous), as can be the case if the system is subject to climate change driven by a time-dependent forcing, such as radiative forcing, anthropogenic changes in greenhouse gases, and aerosol concentrations. In such case, the system may not have an attractor nor an invariant measure on it (in practice, it does not possess statistical equilibrium), a situation that again would hamper the development of a consistent statistical framework for DA. Recent studies on the pullback or random attractor (Chekroun, Simonnet, & Ghil, 2011; Dijkstra, 2013) may open the path to suitable formulations of the estimation problem for nonautonomous systems.

2.1.2 | The observation model

Noisy observations of \mathbf{x}_k are available at discrete times and are represented as components of the observation vector $\mathbf{y}_k \in \mathbb{R}^d$. Assuming the noise is additive, they are related to the model state vector through

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \boldsymbol{\epsilon}_k. \quad (2)$$

Equation (2) defines the, generally nonlinear, observation operator, $\mathcal{H} : \mathbb{R}^m \rightarrow \mathbb{R}^d$, from model to observational space, which often involve spatial interpolations, convolutions, or spectral-to-physical space transformation in spectral models. Transformations based on physical laws for indirect measurements, such as radiative fluxes used to measure temperatures, can also be represented in this way (e.g., Kalnay, 2002). To simplify the notation, we have assumed the observation dimension is constant, so that $d_k = d$.

Similarly to model error, the observational error, $\boldsymbol{\epsilon}_k$, is also represented as a stochastic (i.e., random) additive term, and accounts for the instrumental error of the observing devices, deficiencies in the formulation of the observation operator itself, and the error of representation (or representativeness) (Janjić et al., 2017; Lorenc, 1986). The latter arises from the presence of unresolved scales and represents their effect on the scales explicitly resolved by the model. The error of representativeness is difficult to estimate due to lack of information at small scales, but ubiquitous in Earth science, because the description of a continuum fluid is made by an inevitably limited (albeit always growing) number of discrete grid points (or spectral bands); see Cohn (1997) for a discussion on this issue and the related “change of support” techniques in geostatistics (Chilès & Delfiner, 2012). Note that the additivity of noise is also a simplification since more general noisy observations $\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k, \boldsymbol{\epsilon}_k)$ could be considered.

Confronting model with data is inherent to the scientific methods since Galileo's era. Nevertheless, the DA problem in Earth science has some characteristic criticalities that makes it unique. In geosciences, we usually have $d \ll m$, that is, the amount of available data is insufficient to fully describe the system and one cannot strongly rely on a data-driven approach: the model is paramount. It is the model that fills the spatial and temporal gaps in the observational network: it propagates information from observed-to-unobserved areas and from the observing times to any other causally related. This capability has been pivotal in DA for numerical weather prediction (NWP) with notable examples being the ocean areas and the Southern Hemisphere, where the lack of routine observations platforms has been compensated by the dynamical model (see Daley, 1993; Kalnay, 2002). The other peculiarity is the use of massive dataset ($d \approx \mathcal{O}(10^7)$) and huge models states ($m \approx \mathcal{O}(10^9)$). Thus, DA methods are designed to achieve the best possible use of a never sufficient (albeit constantly growing) amount of data, and to attain an efficient data-model fusion, in a short period of time (typically 3–6 hr for NWP). This poses a formidable computational challenge, and makes DA an example of big-data problems.

2.2 | Bayesian formulation of the state estimation problem

With the two complementary pieces of information in hand, model and data, we can move forward and formalize their fusion. Because of the assumed random nature of both the model and observational error, they can be described in terms of probability density functions (pdfs), and the Bayesian approach offers a natural framework to understand the DA problem. Our main focus here is on state estimation, that is, the issue of estimating \mathbf{x} , based on the model and the observation. Nevertheless, in many physical applications, one is often interested in the joint estimate of the system's state and of its parameters, λ (see Evensen, 2009a, and references therein). We shall allude to parameter estimation later, but its extensive exposition is beyond the scope of this overview.

In the Bayesian formulation, the output of the estimation process is the posterior distribution $p(\mathbf{x}|\mathbf{y})$ of the unknown process \mathbf{x} conditioned on the data \mathbf{y} , which can be obtained using Bayes' rule

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \quad (3)$$

In Equation (3), the three components of the Bayesian inference appear: $p(\mathbf{x})$ is the prior pdf that gathers all the knowledge before assimilating the new observations, $p(\mathbf{y}|\mathbf{x})$ is the likelihood of the data conditioned on the state \mathbf{x} (i.e., what would be the observation if the true state were known?), and $p(\mathbf{y})$ is the marginal distribution of the observation, $p(\mathbf{y}) = \int d\mathbf{x} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$, that is, the distribution of \mathbf{y} whichever the value of the state. The distribution $p(\mathbf{y})$ is independent of \mathbf{x} and is treated as a normalization coefficient. The postulate of a prior is a distinctive feature of the Bayesian approach, which allows the introduction of arbitrary information about the system before data are included. Its choice is subjective and one can in principle use any distribution that suits a study's specific purposes, either based on climatology (i.e., from historical knowledge about the system), on theoretical physical principles or even on subjective expert's opinion. However, in many practical cases, the search for a good informative prior is not straightforward, although its choice may strongly affect the results, sometimes adversely.

So far, we have left the times of \mathbf{x} and \mathbf{y} undefined. Both the model and the observational error sequences, $\{\eta_k: k = 1, \dots, K\}$ and $\{\epsilon_k: k = 1, \dots, K\}$ are assumed to be independent in time, mutually independent, and distributed according to the pdfs p_η and p_ϵ , respectively. These pdfs are related to the prior and likelihood terms of Bayes' rule as follows:

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = p_\eta[\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})], \quad (4)$$

$$p(\mathbf{y}_k|\mathbf{x}_k) = p_\epsilon[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)], \quad (5)$$

where the model dependency on the parameters, λ , has been dropped to simplify the notation. Let us define the sequences of system states and observations within the interval $[t_0, t_K]$ as $\mathbf{x}_{K:0} = \{\mathbf{x}_K, \mathbf{x}_{K-1}, \dots, \mathbf{x}_0\}$ and $\mathbf{y}_{K:1} = \{\mathbf{y}_K, \mathbf{y}_{K-1}, \dots, \mathbf{y}_1\}$, respectively. Since the observational errors are assumed to be independent in time we can split the products of the probabilities

$$p(\mathbf{y}_{K:1}|\mathbf{x}_{K:0}) = \prod_{k=1}^K p(\mathbf{y}_k|\mathbf{x}_k) = \prod_{k=1}^K p_\epsilon[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)], \quad (6)$$

meaning that the mutual likelihood of all the observations in the interval $t_K - t_0$ is the product of the individual likelihoods at each time. We will further assume that the process is Markovian, which means that the state \mathbf{x} at time t_k , conditioned on all past states, only depends on the most recent state at time t_{k-1} and split the prior pdf accordingly

$$p(\mathbf{x}_{K:0}) = p(\mathbf{x}_0) \prod_{k=1}^K p(\mathbf{x}_k|\mathbf{x}_{k-1}) = p(\mathbf{x}_0) \prod_{k=1}^K p_\eta[\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})]. \quad (7)$$

By combining Equation (6) and (7) using Bayes' rule, Equation (3), we get the posterior distribution as a product

$$p(\mathbf{x}_{K:0}|\mathbf{y}_{K:1}) \propto p(\mathbf{x}_0) \prod_{k=1}^K p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1}) = p(\mathbf{x}_0) \prod_{k=1}^K p_\epsilon[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)]p_\eta[\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})]. \quad (8)$$

Equation (8) is of central importance: it states that a new update can be obtained as soon as new data is available; it is called a hidden Markov chain in statistics. It is worth to mention that although the hypothesis of uncorrelated-in-time model and observational errors provides a notable mathematical advantage and it has been key to obtain Equation (8), that is, in fact almost never a valid one in realistic geosciences applications. Model error between two successive updates will be in most cases very naturally time correlated. Similarly, observational error will also be time correlated when, for instance, measurements are taken by the same instruments (e.g., from a satellite at two successive passages) and are processed using a physical models.

Depending on which time period is needed for state estimation, it is possible to define three estimation problems (Wiener, 1949):

1. **Prediction:** estimate $p(\mathbf{x}_l|\mathbf{y}_{k:1})$ with $l > k$.
2. **Filtering:** estimate $p(\mathbf{x}_k|\mathbf{y}_{k:1})$.
3. **Smoothing:** estimate $p(\mathbf{x}_{K:0}|\mathbf{y}_{K:1})$, or selected marginals of this pdf, such as $p(\mathbf{x}_l|\mathbf{y}_{K:1})$, with $0 \leq l < K$.

A schematic illustration of the three problems is given in Figure 1.

The prediction problem (Figure 1 top panel) is formally addressed by integrating Equation 4, that is, by solving the Chapman–Kolmogorov equation for the propagation of a pdf by the model

$$p(\mathbf{x}_l|\mathbf{y}_{k:1}) = \int d\mathbf{x}_k p_{\epsilon}[\mathbf{x}_l - \mathcal{M}_{l:k}(\mathbf{x}_k)] p(\mathbf{x}_k|\mathbf{y}_{k:1}), \quad (9)$$

given the conditional pdf at time t_k , $p(\mathbf{x}_k|\mathbf{y}_{k:1})$.

The filtering problem (Figure 1 mid panel) is the most common in geophysical applications, and is characterized by sequential processing, in which measurements are utilized as they become available (Bengtsson, Ghil, & Källén, 1981; Jazwinski, 1970). An analysis step, in which the conditional pdf $p(\mathbf{x}_k|\mathbf{y}_{k:1})$ is updated using the latest observation, \mathbf{y}_k , alternates with a forecast step which propagates this pdf forward until the time of a new observation. The analysis is based on the application of Bayes' Equation (3), which becomes

$$p(\mathbf{x}_k|\mathbf{y}_{k:1}) = \frac{p_{\epsilon}[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)] p(\mathbf{x}_k|\mathbf{y}_{k-1:1})}{\int d\mathbf{x}_k p_{\epsilon}[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)] p(\mathbf{x}_k|\mathbf{y}_{k-1:1})}, \quad (10)$$

while in the prediction step one integrates the Chapman–Kolmogorov Equation (9) from $k-1$ to k . The process is then repeated, sequentially, with the outcome of the Chapman–Kolmogorov equation providing the prior distribution for the next analysis step.

Estimating the conditional pdf of the state at any time t_k , $0 \leq k \leq K$, based on all observations (past, present, and future), is known as the smoothing problem (Figure 1 bottom panel). This is relevant when one is interested in a retrospective analysis after the observations have been collected, and it is very useful in generating reanalysis as well as in the estimation of the parameters λ . Reanalysis programs offer an homogeneous reconstructed estimate of the system based on observations covering an extended period of time (up to a century), using state-of-the-art model and DA methods of the time when the reanalysis is produced. Reanalyses are paramount in climate diagnosis and several NWP, ocean, and climate centers issue their own reanalysis products (e.g., among many others, Dee et al., 2011; Xie, Bertino, Counillon, Lisæter, & Sakov, 2017).

The joint smoothing pdf is given in Equation (8), and is based on the recursive use of Bayes' formula for the update at each t_k , and the Chapman–Kolmogorov equation for the pdf propagation from t_{k-1} to t_k . To see this, the smoothing distribution at time t_k can be written using the subsequent state at time t_{k+1}

$$p(\mathbf{x}_k|\mathbf{y}_{K:1}) = \int d\mathbf{x}_{k+1} p(\mathbf{x}_k|\mathbf{x}_{k+1}, \mathbf{y}_{K:1}) p(\mathbf{x}_{k+1}|\mathbf{y}_{K:1}). \quad (11)$$

Furthermore, we note the following

$$p(\mathbf{x}_k|\mathbf{x}_{k+1}, \mathbf{y}_{K:1}) = p(\mathbf{x}_k|\mathbf{x}_{k+1}, \mathbf{y}_{k:1}) = \frac{p(\mathbf{x}_{k+1}|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{k:1})}{p(\mathbf{x}_{k+1}|\mathbf{y}_{k:1})}, \quad (12)$$

where we used again the fact that observations $\{\mathbf{y}_{k+1}, \dots, \mathbf{y}_K\}$ are independent of \mathbf{x}_k when \mathbf{x}_{k+1} is known (see Section 2.2), and $p(\mathbf{x}_k|\mathbf{y}_{k:1})$ is the filter solution (the analysis) at time t_k . Together, Equation (11) and (12) suggest a forward-backward recursive algorithm, in which a forward filtering is followed by a backward smoothing. Starting from $p(\mathbf{x}_0)$, the forward phase consists of a sequential filter from t_1 to t_K , which allows one to estimate, and to store, the sequence of filter pdfs $p(\mathbf{x}_k|\mathbf{y}_{k:1})$. In the backward phase, from t_{K-1} to t_1 , one evaluates Equation (12) using the previously stored filter solutions, and finally estimates the smoothing pdf recursively via Equation (11) using the smoothing pdf, $p(\mathbf{x}_{k+1}|\mathbf{y}_{K:1})$, from the previous iteration. We will see in Section 3.1.2 that the above recursion possesses an analytical solution when the dynamical and observational models are linear and all error pdfs are Gaussian.

We conclude the section with some comparative considerations between the filter and smoother. The filtering solution at an arbitrary t_k ($0 \leq k \leq K$) is obtained by sequential updating until t_k and it thus accounts for all observations before t_k . In contrast, the smoothing solution at the same time t_k also accounts for future observations until t_K , and it is thus generally more accurate than the filtering one. At the final time t_K , both solutions have incorporated the same amount of data, so that in the absence of approximations, they will coincide (e.g., Kalnay & Yang, 2010).

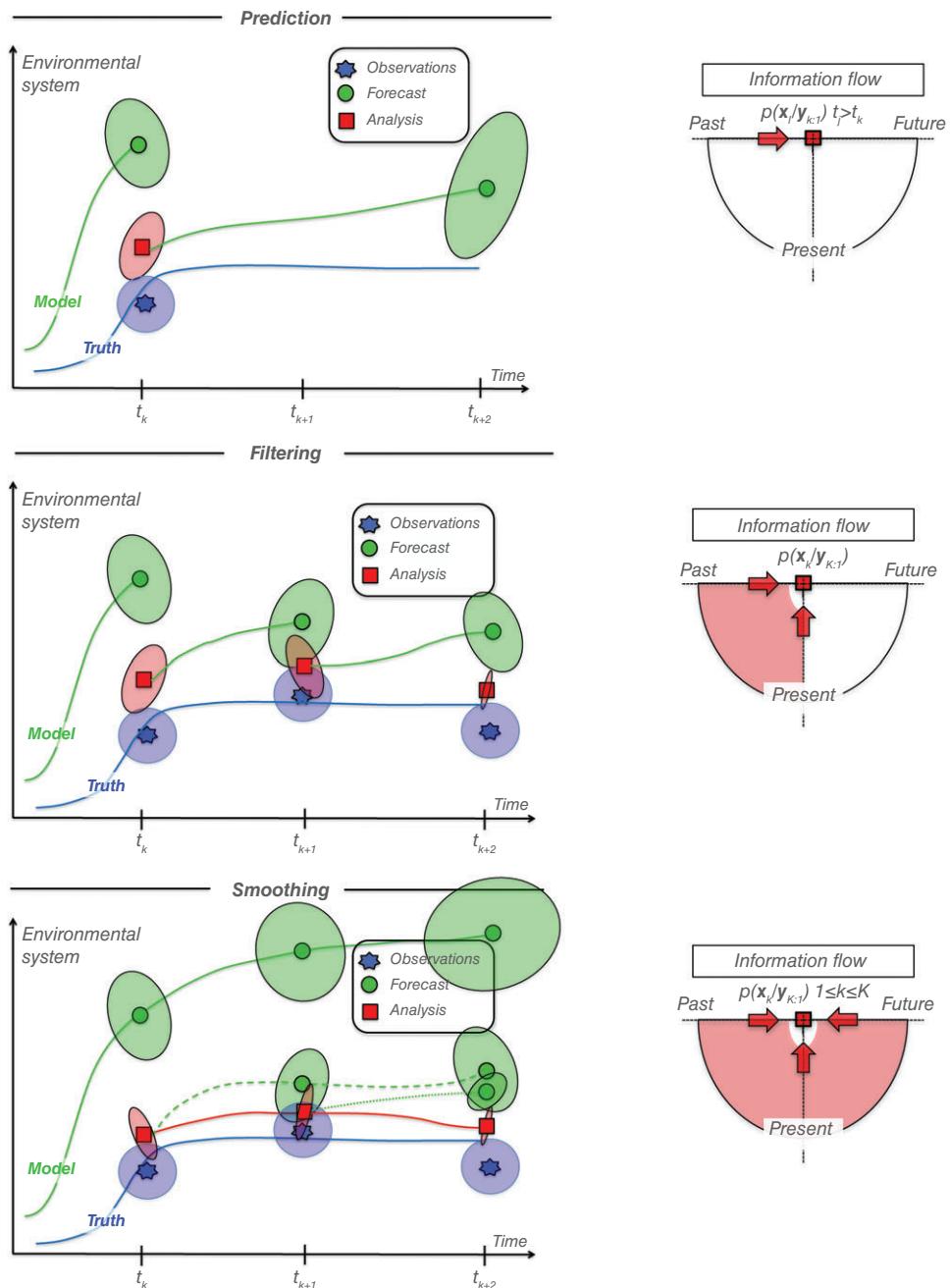


FIGURE 1 Illustration of the three estimation problems: **Prediction** (top), **filtering** (middle), and **smoothing** (bottom). The true unknown signal is represented by the blue line. Observation (blue), forecast (green), and analysis (red) pdfs are displayed as ellipsoids of proportional size, that is, the smaller the size the smaller the estimated uncertainty, so the larger the confidence; observational error is assumed constant. The associated blue stars for the observations, green squares for the forecast, and red square for the analysis have to be intended as point estimators based on the corresponding pdfs; one, not unique, choice can be the mean of the pdfs (c.f., Section 3, for a discussion on the choice of the estimator). **Prediction** (top panel): An analysis is produced at t_k using the forecast and the observation at t_k : The analysis uncertainty is smaller than the forecast uncertainty. From the analysis at t_k , a prediction is issued until t_{k+2} . The prediction error grows in time (as exemplar of a chaotic behavior typical in geophysical systems; c.f., Section 5.1) and the forecast uncertainty at t_{k+2} (the green ellipsoid) is larger than the analysis uncertainty at t_k (red ellipsoid). Information is propagated only forward from t_k as depicted in the information flow diagram (top right). **Filter** (mid panels): A prediction is issued from the analysis at t_k until the next observations at t_{k+1} ; the forecast uncertainty at t_{k+1} is larger than that in the analysis at t_k . At time t_{k+1} , a new analysis is performed by combining the forecast and the observations at t_{k+1} . The analysis uncertainty (red ellipsoid) is smaller than both the forecast and the observational uncertainties (green and blue ellipsoid/circle). From the analysis at t_{k+1} , the process is repeated and a new forecast until t_{k+2} is issued, and a new analysis is performed using the observation at t_{k+2} . The information flow diagram (mid right) depicts how the information is carried from both the past (as in the prediction problem) and from the present using current data. **Smoothening** (bottom panels): All observations between t_k and t_{k+2} contribute simultaneously to the analysis, which is now the entire trajectory within the smoothing interval $[t_k, t_{k+2}]$. At the final time, t_{k+2} , the smoother and filter have assimilated the same amount of observations, so their analyses at t_{k+2} , and their associated estimated uncertainties, are approximately the same (compare the red ellipsoids at t_{k+2} for smoother and filter), but the smoother is more accurate at any other time in the window. The smoother solutions at t_k and t_{k+1} provide initial conditions for predictions until t_{k+2} (dotted and solid green lines, respectively). At final time, t_{k+2} , there are three forecasts initialized respectively by the analyses at t_{k-1} (not shown), t_k and t_{k+1} , and the associated uncertainties (green ellipsoids) are inverse proportional to the length of the prediction, with the forecast initialized at t_{k+1} being the most accurate.

3 | A ROUTE TO SOLUTIONS: THE GAUSSIAN APPROXIMATION

With the pdfs for smoother, filter, and prediction in hand, it still remains to be decided which estimator suits best a specific problem. Those pdfs, in fact, describe fully the probability of all possible states of the system but hardly make good targets for an estimation algorithm. Two “natural” options, the mean of the distribution or its mode, characterize two traditional approaches to state estimation. Methods designed targeting the mean are named minimum squared error estimators because, irrespective of the properties of the pdf (i.e., whether or not it is symmetric and unimodal), the mean is always the minimum squared error estimate (Jazwinski, 1970). The mode is the peak of the distribution, the most probable state, and methods targeting it are referred to as maximum a posteriori estimators. Other possible options are viable, such as the median, but they are generally more difficult to compute do not have the same wide range of applicability as the mean or the mode.

Despite its appealing and clear logic, the huge dimensions of typical models and datasets used in environmental science hamper the use of a fully Bayesian approach. The problem dimension renders it extremely difficult, or practically impossible, to define and to evolve the pdfs. To overcome this computational, albeit fundamental, issue it is usually assumed that the uncertainties of each piece of information (observations, model, and prior) are Gaussian distributed. This hypothesis leads to a substantial simplification: the pdfs can be completely described by their first and second moments: the mean and the covariance matrix.

The Gaussian approximation is at the core of most DA procedures successfully used in the geosciences, and this overview will largely discuss methods that rely upon it to different extents. This section presents two main approaches to the estimation problem in which DA in the geosciences is rooted: the minimum variance Kalman-like (Section 3.1) and the maximum a posteriori-variational methods (Section 3.2). Section 3 is complemented by four Appendices (A–D), where interested readers can find additional details on the methods.

3.1 | The linear and Gaussian case—The Kalman filter and Kalman smoother

The dynamical and observational models, Equations (1) and (2), are both assumed to be linear

$$\mathbf{x}_k = \mathbf{M}_{k:k-1} \mathbf{x}_{k-1} + \boldsymbol{\eta}_k, \quad \boldsymbol{\eta}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k), \quad (13)$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \boldsymbol{\epsilon}_k, \quad \boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k), \quad (14)$$

with $\mathbf{M}_{k:k-1}$ and \mathbf{H}_k being $m \times m$ and $d \times m$ matrices, respectively. The observational and model noises are assumed to be mutually uncorrelated and in time, unbiased, and Gaussian distributed with covariance matrices $\mathbf{R}_k \in \mathbb{R}^{d \times d}$ and $\mathbf{Q}_k \in \mathbb{R}^{m \times m}$, respectively.

3.1.1 | Kalman filter

Under the aforementioned hypotheses of all sources of errors being Gaussian and both the observational and dynamical models being linear, the KF is the exact (optimal) solution to estimate recursively the conditional mean and the associated uncertainty, the conditional covariance, of the filtering pdf $p(\mathbf{x}_k | \mathbf{y}_{k:0})$. The process is sequential, meaning that observations are assimilated in chronological order, and the KF alternates a forecast step, when the mean and covariance of the prediction conditional pdf, $p(\mathbf{x}_k | \mathbf{y}_{k-1:0})$, are evolved, with an analysis step in which the mean and covariance of the filtering conditional pdf, $p(\mathbf{x}_k | \mathbf{y}_{k:0})$, are updated. By virtue of the Gaussian assumption, the mean and covariance suffice to fully describe the pdfs and the KF provides an exact set of closed equations. The KF formulae are given below without their derivation, which goes beyond the scope of this overview, but the reader can find numerous alternative ways in classical books or reviews on filtering (Bain & Crisan, 2009; Bengtsson et al., 1981; Cohn, 1997; Jazwinski, 1970; Talagrand, 1997; Wikle & Berliner, 2007).

Let us indicate the forecast/analysis mean state and covariance as \mathbf{x}_k^f/a and \mathbf{P}_k^f/a , respectively. The KF equations then read

$$\text{Forecast Step} \quad \mathbf{x}_k^f = \mathbf{M}_{k:k-1} \mathbf{x}_{k-1}^a, \quad (15)$$

$$\mathbf{P}_k^f = \mathbf{M}_{k:k-1} \mathbf{P}_{k-1}^a \mathbf{M}_{k:k-1}^T + \mathbf{Q}_k. \quad (16)$$

$$\text{Analysis step} \quad \mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1}, \quad (17)$$

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^f), \quad (18)$$

$$\mathbf{P}_k^a = (\mathbf{I}_k - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f. \quad (19)$$

Given as inputs the matrices \mathbf{Q}_k , \mathbf{R}_k , \mathbf{H}_k , and \mathbf{M}_k , for $k \geq 1$, and the initial condition for the mean, $\mathbf{x}_0^a = \mathbf{x}_0$, and initial error covariance, $\mathbf{P}_0^a = \mathbf{P}_0$, Equations (15)–(19) estimate sequentially the state, $\mathbf{x}_k^{f/a}$, and the associated error covariance, $\mathbf{P}_k^{f/a}$, at any time $k > 1$. The matrix $\mathbf{K}_k \in \mathbb{R}^{m \times d}$ is the Kalman gain and it contains the regression coefficients of the optimal linear combination between the prior (given here by the forecast \mathbf{x}_k^f) and the observations. The resulting state estimate, the analysis \mathbf{x}_k^a , has

minimum error variance and is unbiased. In fact, the KF Equations (15)–(19) can be obtained without the need of the Gaussian hypothesis and the KF solution is also said the best linear unbiased estimator (BLUE) (see Asch et al., 2016, their Section 3.4.3). Similarly, the KF equations can also be seen and derived using a maximum entropy approach (see Giffin & Urniezius, 2014; Mitter & Newton, 2005).

3.1.2 | Kalman smoother

We discussed in Section 2.2 that a recursive estimate of the pdf of the system state at any arbitrary time in the observing window, conditioned on all observations in the window, that is, the pdf $p(\mathbf{x}_k|\mathbf{y}_{K:1})$, can be obtained using a forward and backward recursions, in which a forward-in-time filter is followed by a backward-in-time smoother. Under the hypotheses of linear dynamical and observational models and Gaussian errors, this bidirectional recursion can be solved analytically and it is referred to as the Kalman smoother (KS, Jazwinski, 1970). Like for the KF, we describe the KS formulae in the following, but we do not provide their derivation which can be found, for instance, in Jazwinski (1970).

Assume that a forward in time KF has already been implemented using Equations (15–19) for $k = 1, \dots, K$, and the forecast and analysis means and covariances, $\mathbf{x}_k^{f/a}$ and $\mathbf{P}_k^{f/a}$, have been computed and stored. With the filtering solutions at disposal, we can run the KS recursion backward in time, for $k = K-1, \dots, 1$, to compute the smoothing mean and covariance, \mathbf{x}_k^{sm} and \mathbf{P}_k^{sm} , according to

$$\mathbf{S}_k = \mathbf{P}_k^a \mathbf{M}_{k+1:k}^T (\mathbf{M}_{k+1:k} \mathbf{P}_k^a \mathbf{M}_{k+1:k}^T + \mathbf{Q}_{k+1})^{-1} = \mathbf{P}_k^a \mathbf{M}_{k+1:k}^T \mathbf{P}_{k+1}^{-f}, \quad (20)$$

$$\text{KS Mean} \quad \mathbf{x}_k^{\text{sm}} = \mathbf{x}_k^a + \mathbf{S}_k (\mathbf{x}_{k+1}^{\text{sm}} - \mathbf{x}_{k+1}^f). \quad (21)$$

$$\text{KS Covariance} \quad \mathbf{P}_k^{\text{sm}} = \mathbf{P}_k^a + \mathbf{S}_k (\mathbf{P}_{k+1}^{\text{sm}} - \mathbf{P}_{k+1}^f) \mathbf{S}_k^T, \quad (22)$$

with initial conditions, at time t_K , $\mathbf{x}_K^{\text{sm}} = \mathbf{x}_K^a$ and $\mathbf{P}_K^{\text{sm}} = \mathbf{P}_K^a$. As anticipated in Section 2.2, the KF and KS solutions are equivalent at final time t_K , but not at earlier times. The KS formulation given above is also known as Rauch–Tung–Striebel smoother; more details on the smoothing techniques can be found in Cosme, Verron, Brasseur, Blum, and Auoux (2012).

Appendix A discusses in more details some key properties of the KF and KS.

3.2 | Variational approach

Variational methods are designed to estimate the model trajectory that “best fits” all the observations within a prescribed observing window $[t_0, t_K]$. Given that the model trajectory is adjusted globally to fit all observations simultaneously, the state estimate at an arbitrary time t_k , $t_0 \leq t_k \leq t_K$, is influenced by all observations within the window. The problem is formalized as the one of minimizing an appropriate scalar *cost function* that quantifies the model-to-data misfit (Sasaki, 1970; Thompson, 1969).

Assuming data are available at each time t_k in the window (this assumption can easily be relaxed without any major changes in the exposition that follows), the variational problem reads

$$\mathbf{x}_{K:0}^a = \operatorname{argmin}(\mathcal{J}(\mathbf{x}_{K:0})) \quad k = 1, \dots, K, \quad (23)$$

where the notation *argmin* is used to signify that $\mathbf{x}_{K:0}^a$ is the solution for which the cost function, $\mathcal{J}(\mathbf{x}_{K:0})$, attains its minimum. Note that $\mathcal{J}(\mathbf{x}_{K:0})$ is a function of the entire model trajectory within the window and the variational analysis solution, the trajectory $\mathbf{x}_{K:0}^a$, provides the closest possible fit to the data, while being consistent with the model dynamics over the entire window. Note furthermore that, in formulating the variational problem in Equation (23), we did not require the model being linear and we have not yet made any hypothesis on the statistical nature of the errors. The most common forms of the variational problem in the geosciences are discussed and derived in Section 3.2.1. For example, when both observational and model errors are assumed uncorrelated in time, unbiased, and Gaussian distributed with covariances \mathbf{R}_k and \mathbf{Q}_k , respectively, these errors are both penalized in the cost function as follows:

$$\mathcal{J}(\mathbf{x}_{K:0}) = \frac{1}{2} \sum_{k=0}^K \|\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\|_{\mathbf{R}_k^{-1}}^2 + \frac{1}{2} \sum_{k=1}^K \|\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\|_{\mathbf{Q}_k^{-1}}^2 + \frac{1}{2} \|\mathbf{x}_0 - \mathbf{x}^b\|_{\mathbf{B}^{-1}}^2, \quad (24)$$

where the notation, $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}$, is used to indicate the weighted Euclidean norm. In Equation (24), we furthermore assumed to have a prior, commonly referred to as *background* in variational methods literature, for the initial state at the start of the window, in the form of a Gaussian pdf with mean, $\mathbf{x}^b \in \mathbb{R}^m$, and covariance, $\mathbf{B} \in \mathbb{R}^{m \times m}$, respectively. We will recall and derive Equation (24) in Section 3.2.1.

The calculus of variations can be used to find the extremum of $\mathcal{J}(\mathbf{x}_{K:0})$ and leads to the corresponding Euler–Lagrange equations (see Bennett, 1992; Le Dimet & Talagrand, 1986). These arise by searching for the solution of a constrained minimization problem in which the solution is required to follow the model dynamics by appending the model equations such that

$$\mathcal{L}(\mathbf{x}_{K:0}, \boldsymbol{\Gamma}) = \mathcal{J}(\mathbf{x}_{K:0}) + \sum_{k=1}^K \boldsymbol{\Gamma}_k^T (\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}) - \boldsymbol{\psi}_k), \quad (25)$$

with $\boldsymbol{\Gamma}_k^T \in \mathbb{R}^m$ the Lagrange multiplier vector at time t_k . The model error term can also be part of the control variable and it is indicated here as $\boldsymbol{\psi}_k$ to differentiate it from the stochastic model error, $\boldsymbol{\eta}_k$, of Equation (1). Although rigorous and formally appealing, solving the Euler–Lagrange equations for problems with the size and complexity of typical geophysical systems is practically impossible, unless drastic approximations are introduced. When the dynamics are linear and the amount of observations is not very large, the Euler–Lagrange equations can be solved with the *method of representers* which has been very successful in the early days of DA for the ocean (Bennett, 1992), and later extended to nonlinear dynamics by Uboldi and Kamachi (2000). A nice property of the representer method is that it reduces the dimension of the problem to the number of measurements (smaller than the whole model state at all times or the model state dimension; see Bennett (1992)).

Nevertheless, the representers method is still far from being applicable for realistic high-dimensional problems such as NWP. An attractive alternative is represented by the *descent methods*, which use the gradient vector of the cost function in an iterative minimization procedure (Talagrand & Courtier, 1987, see also Appendix B). This latter approach, in its many variants, is adopted in most of the operational NWP centers that employ variational assimilation (Fisher & Andersson, 2001). At the cost function minimum, its gradient must vanish and its Hessian matrix be positive definite (this condition is the extension to vector functions of the classical condition for a minimum of a scalar function for which its first derivative, the gradient, must vanish and its second derivative, here the Hessian, must be positive, here positive definite):

$$\nabla_{\mathbf{x}_{K:0}} \mathcal{J}(\mathbf{x}_{K:0}) = \mathbf{0}, \quad \nabla_{\mathbf{x}_{K:0}}^2 \mathcal{J}(\mathbf{x}_{K:0}) > 0, \quad (26)$$

where $\mathbf{0} \in \mathbb{R}^{m \times K}$ is a matrix with all entries equal to zero, and the second inequality stands for the positive definiteness of the Hessian matrix. The *control variable* with respect to which the minimization is performed is the entire model trajectory over the observing window. The variational problem defined by Equation (23) is usually referred to as *weak constraint* given that the model dynamics is affected by errors and it thus constitutes a weak constraint during the minimization (Sasaki, 1970).

An important particular case is given by the *strong constraint* variational assimilation in which the model is assumed to be perfect (Le Dimet & Talagrand, 1986; Lewis & Derber, 1985; Penenko & Obraztsov, 1976) (i.e., $\boldsymbol{\psi}_k = 0$). In this case, the dynamical model is (assumed) purely deterministic and its solutions are uniquely determined by specifying the initial conditions. This transforms the problem from a constrained into an unconstrained one and the cost function into a function of the initial condition alone, $\mathcal{J}(\mathbf{x}_{K:0}) \Rightarrow \mathcal{J}(\mathbf{x}_0)$; its gradient with respect to the state at t_0 reads

$$\nabla_{\mathbf{x}_0} \mathcal{J}(\mathbf{x}_{K:0}) = \sum_{k=1}^K \mathbf{M}_{k:0}^T \nabla_{\mathbf{x}_k} \mathcal{J}(\mathbf{x}_k) + \nabla_{\mathbf{x}_0} \mathcal{J}(\mathbf{x}_0). \quad (27)$$

where we have used the fact that $\mathbf{x}_k = \mathcal{M}_{k:0}(\mathbf{x}_0)$, and applied the chain rule for differentiating compositions of functions. The matrix $\mathbf{M}_{k:0}^T$ is the *adjoint operator*, the adjoint of the tangent linear model, and implies a backward integration from t_k to t_0 . The gradient, Equation (27), can be obtained through a set of operations involving both forward nonlinear and backward adjoint integration; see Appendix B for more details on the minimization procedures.

The drastic reduction of the number of control variables, from the entire trajectory $\mathbf{x}_{K:0}$ for the weak constraint into the initial state alone \mathbf{x}_0 for the strong-constraint case, together with the use of the adjoint methods (Lions, 1971), have made possible the successful use and diffusion of strong-constraint variational methods, or some suitable approximation of them, in atmospheric and oceanic DA (see Kalnay, 2002; Talagrand, 2010).

3.2.1 | Statistical interpretation and 4DVar

The form of the cost function reflects how we intend to balance the data and the model relative to each other and, again, the statistical approach offers a natural framework to accomplish this task. A natural choice would be to set the cost function proportional to the conditional posterior pdf, $p(\mathbf{x}_{K:0}|\mathbf{y}_{K:0})$, for which a suitable functional form, desirably smooth and derivable, must be given.

This can be done, for instances, under the hypothesis of Gaussian errors, and it leads to a significant simplification. Recall the Gaussian assumptions made in formulating the KF and KS in Section 3.1: the prior, model, and observational errors are all Gaussian distributed. We can now substitute these Gaussian pdfs into Bayes's Equation (8), to obtain the desired posterior distribution; it is the product of three Gaussian pdfs and is thus itself Gaussian

$$\text{Inp}(\mathbf{x}_{K:0} | \mathbf{y}_{K:0}) \propto - \left[\frac{1}{2} \sum_{k=0}^K \|\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\|_{\mathbf{R}_k^{-1}}^2 + \frac{1}{2} \sum_{k=1}^K \|\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\|_{\mathbf{Q}_k^{-1}}^2 + \frac{1}{2} \|\mathbf{x}_0 - \mathbf{x}^b\|_{\mathbf{B}^{-1}}^2 \right], \quad (28)$$

$$= -\mathcal{J}^{\text{w4DVar}}(\mathbf{x}_{K:0}) \quad (29)$$

which provides the cost function already given in Equation (24). Maximizing $\text{Inp}(\mathbf{x}_{K:0} | \mathbf{y}_{K:0})$, that is, finding the most likely trajectory (the analysis $\mathbf{x}_{K:0}^a$) is equivalent to minimizing the term between the squared brackets. Equation (29) defines the cost function of the weak-constraint 4DVar (e.g., Trémoulet, 2006, 2007; Vidard, Piacentini, & Le Dimet, 2004; Zupanski, 1997). When both the dynamical and observational models are linear (i.e., $\mathcal{M}_{k:k-1} = \mathbf{M}_{k:k-1}$ and $\mathcal{H}_k = \mathbf{H}_k$) the solution, $\mathbf{x}_{K:0}^a$ coincides with the solution of the KS Equations (20)–(22). In the general, nonlinear case, the minimization can be carried out iteratively using the gradient of the cost function as described in Appendix B. The control variable for the minimization is the entire trajectory, $\mathbf{x}_{K:0}$, and the estimate at each time t_k is influenced by all observations in the window, before and after t_k . The gradient (not shown; see Carrassi & Vannitsem, 2010; Trémoulet, 2006) with respect to \mathbf{x}_k includes a term proportional to the innovation, $\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)$, because data are assumed to be available at each time step. In the general case, when the model time discretization is finer than the observational temporal density, the corresponding innovation term disappears. Since the weak-constraint 4DVar solves for the full trajectory $\mathbf{x}_{K:0}$ within the window, it does not need the assumption of uncorrelated model error. Nevertheless, if the model error is assumed uncorrelated in time, the cost function requires only the model error spatial covariances at times t_k and not their temporal correlation, and the model error cost function term (the second one in the right-hand side of Equation (28)) reduces to a single summation over the time steps weighted by the inverse of the model error covariances (Trémoulet, 2006). Even in this case, however, the size of the control variable is huge, $\mathbf{x}_{K:0} \approx \mathcal{O}(m \times K)$ (i.e., the state vector dimension times the temporal steps in the assimilation window), and it is difficult to estimate reliably the model error covariances, \mathbf{Q}_k (c.f., Appendix A). The development of specific solutions have thus been necessary to make the weak-constraint 4DVar feasible in the geosciences (see Griffith & Nichols, 2000; Vidard et al., 2004; Trémoulet, 2007), and its use in operational settings has only been very recent (Fisher, Tremoulet, Auvinen, Tan, & Poli, 2011; Lindskog et al., 2009; Ngodock & Carrier, 2014). The weak-constraint 4DVar in the presence of time correlated model error is described in Carrassi and Vannitsem (2010), while its hybrid ensemble-variational formulation (c.f., Section 4.5) has been studied recently by Amezcuia, Goodliff, and van Leeuwen (2017). An interesting alternative conceptualization of the weak-constraint 4DVar is described in Ye, Kadakia, Rozdeba, Abarbanel, and Quinn (2015) that uses path-integral and annealing methods to solve the problem.

In the more widely used *strong-constraint 4DVar*, it is assumed that the model is perfect, which reduces the cost function to

$$\mathcal{J}^{\text{s4DVar}}(\mathbf{x}_0) = \frac{1}{2} \sum_{k=0}^K \|\mathbf{y}_k - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\mathbf{x}_0)\|_{\mathbf{R}_k^{-1}}^2 + \frac{1}{2} \|\mathbf{x}_0 - \mathbf{x}^b\|_{\mathbf{B}^{-1}}^2, \quad (30)$$

with the \circ symbol representing the composition of operators and the control variable is now “only” the system's state at the beginning of the window, \mathbf{x}_0 . Similarly to the equivalence between the solutions of the weak-constraint 4DVar and of the KS, in the linear case, the solution of the strong-constraint 4DVar at the end of the window, \mathbf{x}_K , will be equivalent to that of a KF, Equations (15)–(19), that had started from the same initial condition and had assimilated the same observations. The gradient of Equation (30) can be obtained by applying Equation (27) and reads

$$\nabla_{\mathbf{x}_0} \mathcal{J}^{\text{s4DVar}}(\mathbf{x}_0) = - \sum_{k=0}^K \mathbf{M}_{k:0}^T \mathbf{H}_k^T \mathbf{R}^{-1} [\mathbf{y}_k - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\mathbf{x}_0)] + \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}^b), \quad (31)$$

which reveals the effects of the adjoint operator, $\mathbf{M}_{k:0}^T$, to project backward the influence of the observation at any t_k . The final analysis is obtained using an iterative process starting by providing a first guess, $\mathbf{x}_0^{i=0}$, and evaluating $\mathcal{J}^{\text{s4DVar}}(\mathbf{x}_0^i)$ and the gradient at each iteration. As explained in Appendix B, the process is then repeated until the convergence criteria are met. Then using the final minimum state at \mathbf{x}_0^a , the analyzed trajectory is obtained over the entire window, $\mathbf{x}_{K:0}^a$, by integrating the deterministic model from t_0 to t_K . This trajectory represents the most likely model solution fitting the observations within the window and given the prescribed initial condition and data uncertainty. An extension of the strong-constraint 4DVar to account for time-correlated observational error in an operational weather forecast context has been introduced by Järvinen, Andersson, and Bouttier (1999).

Appendix C discusses in more details some key properties of the variational methods, while three popular approximations of the KF and variational methods, namely the *extended Kalman filter* (EKF), the *incremental 4DVar*, and the *3DVar* are described in Appendix D. Finally, a schematic illustration of the different problems tackled by the weak-constraint 4DVar, strong-constraint 4DVar, and 3DVar (c.f., Appendix D) is given in Figure 2.

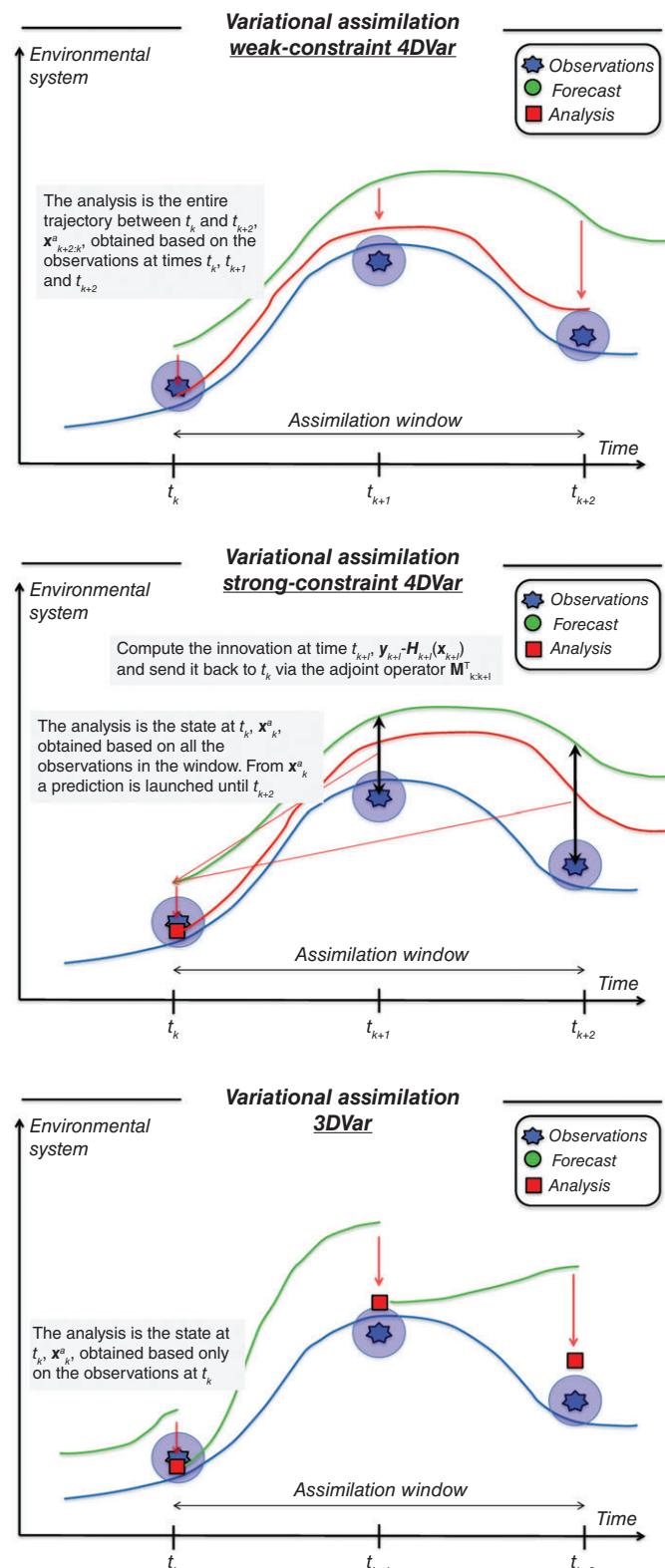


FIGURE 2 Illustration of the three variational problems: **Weak-constraint 4DVar** (w4DVar, top panel), **strong-constraint 4DVar** (mid panel), and **3DVar** (bottom panel). In the w4DVar, the control variable is the entire trajectory within the window (from t_k to t_{k+2} in Figure 2), $\mathbf{x}_{k+2:k}$, so the corrections (continuous red arrow) are computed and utilized at the observation times. The analysis trajectory (red line) is moved toward the observations. In the s4DVar, the control variable is the state at the beginning of the window, \mathbf{x}_k , and the corrections are computed at observation times (black arrows) but then propagated back to t_k (red arrows) using the adjoint model (see equation (31)): Once the analysis at the initial time is computed, it is then used as initial condition to run a forecast until t_{k+2} . In the 3DVar, the corrections are computed and utilized at each observation times t_k sequentially, so that the analysis at t_k , \mathbf{x}_k^a , takes into account only the observations at t_k and is then used as initial condition for a forecast until the next observations at t_{k+1} and so on. Variational methods do not explicitly compute an estimate of the uncertainty (c.f., Appendix B) so the colorful ellipsoids of Figure 1 are not present here

4 | ENSEMBLE METHODS

A Monte Carlo approach is at the basis of a class of algorithms referred to as *ensemble-based methods* (Evensen, 2009b) of which the *ensemble Kalman filter* (EnKF) and the *ensemble Kalman smoother* (EnKS), are the most celebrated examples. The EnKF and EnKS are still based on the Gaussian hypothesis: their analysis update steps are only based on the mean and covariance. The transition density is, however, approximated using an ensemble of realizations of the prior $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, in practice by running an ensemble of trajectories of the model dynamics. The EnKF and its variants (Asch et al., 2016; Evensen, 2009b; and c.f., Section 4.1–4.2) are nowadays among the most popular approaches for DA in high-dimensional systems and have been successfully applied in atmospheric and oceanic contexts. Evidence has emerged that a small number of particles, typically 100, is sufficient in many applications, hence making EnKF feasible in situations where the forward step of the DA is computationally expensive. A version of the EnKF is operational for the atmospheric model at the Canadian Meteorological Centre (CMC) (Houtekamer et al., 2005) Also, the EnKF is used in the ocean system TOPAZ (see Section 5.4) developed at the Nansen Environmental and Remote Sensing Centre (NERSC, Sakov, Oliver, & Bertino, 2012). The meteorological office of Norway is running an operational version of TOPAZ and produces operational forecasts for the Arctic Ocean, sea ice, and ecosystem, for the European Copernicus Marine Services (www.marine.copernicus.eu).

There have been numerous alternative EnKF-like methods, and we will not attempt here to discuss them all; a recent review of the EnKF for atmospheric DA is given by Houtekamer and Zhang (2016). However, the many different formulations fall almost all into two main categories, namely the *stochastic* and the *square-root* (also called *deterministic*) ones, and this is the subject of the next sections. We shall then discuss the EnKS in Section 4.3 and devote Section 4.4 to the description of the important countermeasures that must be put in place to make the EnKF and EnKS work in realistic high-dimensional nonlinear cases. This part of the article closes by presenting an outlook of the current frontier of the hybrid ensemble-based and variational methods in Section 4.5.

4.1 | Stochastic EnKF

The EnKF as introduced by Evensen (1994); Burgers, van Leeuwen, and Evensen (1998); and Evensen (2003) is based on the use of an ensemble representation for error statistics. Thus, instead of storing a full covariance matrix, we can represent the same error statistics using an appropriate ensemble of model states. Given an error covariance matrix, an ensemble of finite size provides an approximation of the error covariance matrix, and, as the size N of the ensemble increases, the errors in the Monte Carlo sampling decrease proportionally to $1/\sqrt{N}$.

The derivations that follow will assume a linear observation operator for the sake of simplicity. However, ensemble-based methods can straightforwardly incorporate a nonlinear observation operator (Evensen, 2009b).

4.1.1 | Representation of error statistics

Given an ensemble of model realizations at a certain time, it is convenient to introduce the ensemble matrix

$$\mathbf{E}^{f,a} = \left[\mathbf{x}_1^{f,a}, \dots, \mathbf{x}_N^{f,a} \right] \in \mathbb{R}^{m \times N}, \quad (32)$$

which contains the N ensemble members in its columns.

The error covariance matrices \mathbf{P}_k^f and \mathbf{P}_k^a for the predicted and analyzed estimate in the KF are defined in terms of the true state in the KF Equation (16–19). However, since the true state is not known, we will instead define the ensemble-anomaly covariance matrices around the ensemble mean. Given the ensemble mean

$$\bar{\mathbf{x}}^{f,a} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n^{f,a}, \quad (33)$$

we define the ensemble-anomaly matrix as

$$\mathbf{X}^{f,a} = \frac{1}{\sqrt{N-1}} \left[\mathbf{x}_1^{f,a} - \bar{\mathbf{x}}^{f,a}, \dots, \mathbf{x}_N^{f,a} - \bar{\mathbf{x}}^{f,a} \right] \in \mathbb{R}^{m \times N}. \quad (34)$$

We can now write the ensemble-based error covariance matrices for the forecast and analysis ensembles as

$$(\mathbf{P}^e)^f = (\mathbf{X}^f)(\mathbf{X}^f)^T, \quad (35)$$

$$(\mathbf{P}^e)^a = (\mathbf{X}^a)(\mathbf{X}^a)^T, \quad (36)$$

where the superscript “*e*” denotes that the quantities (e.g., matrices in Equation (35) and (36)) are estimated based on the ensemble.

Equations (35) and (36) embed an interpretation where the ensemble mean is the best estimate, and the spread of the ensemble around the mean is a natural definition of the error in the ensemble mean. We can interpret an ensemble of model states as a Monte Carlo representation of the pdf $p(\mathbf{x}_0)$ in Equation (7).

4.1.2 | Prediction of error statistics

Given a pdf $p(\mathbf{x}_0)$ at time t_0 , the joint pdf from time t_0 to t_k is given by Equation (7), which involves a multiplication by the transition density defined in Equation (4). It was shown in Evensen (1994) that if the pdf $p(\mathbf{x}_0)$ at time t_0 is represented by an ensemble of model states, then the multiplication with the transition density is equivalent to integrating each realization of \mathbf{x} according to the model equations as defined in Equation (1). Thus, from a finite ensemble of model states, we can compute an ensemble approximation of the joint pdf $p(\mathbf{x}_{k:0})$ for the time interval from t_0 to t_k .

4.1.3 | Analysis scheme

The novel idea of Evensen (1994) was to design an alternative update scheme that could work directly on the ensemble and where it was not necessary to compute the full covariances as defined by Equation (35) and (36). Evensen (1994) showed that, by updating each individual ensemble member according to the standard KF Equation (17) and (18), the updated ensemble will have the correct mean and covariance matrix for the update in agreement with Equation (19) in the standard KF. Burgers et al. (1998) further proved that, in order for the EnKF analysis error covariance, $(\mathbf{P}^e)^a$, to be consistent with that of the KF, it is essential to treat the observations as random variables having a distribution with mean equal to the observed value and covariance equal to \mathbf{R} . Thus, given a vector of observations, $\mathbf{y} \in \mathbb{R}^d$, we define an ensemble of perturbed observations

$$\mathbf{y}_n = \mathbf{y} + \boldsymbol{\epsilon}_n, \quad 1 \leq n \leq N, \quad (37)$$

which we store in the observation matrix

$$\mathbf{Y}_o = [\mathbf{y}_1, \dots, \mathbf{y}_N] \in \mathbb{R}^{d \times N}, \quad (38)$$

which columns are the perturbed measurements $\mathbf{y}_n \in \mathbb{R}^d$. Then, we define the corresponding matrix of the normalized anomaly ensemble of the observations $\mathbf{Y}'_o \in \mathbb{R}^{d \times N}$ as

$$\begin{aligned} \mathbf{Y}'_o &= \frac{1}{\sqrt{N-1}} [\mathbf{y}_1 - \mathbf{y}, \dots, \mathbf{y}_N - \mathbf{y}] \\ &= \frac{1}{\sqrt{N-1}} [\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_N]. \end{aligned} \quad (39)$$

By subtracting any nonzero mean from the N random draws $\boldsymbol{\epsilon}_n$, we ensure that the simulated random measurement errors have zero ensemble mean and thus the random perturbations do not introduce any bias in the update. Next, we define the ensemble covariance matrix of the measurement errors as

$$\mathbf{R}^e = \mathbf{Y}'_o (\mathbf{Y}'_o)^T. \quad (40)$$

In the limit, $N \rightarrow \infty$, of infinite ensemble size, this matrix converges to the prescribed error covariance matrix \mathbf{R} used in the KF.

The analysis step in the EnKF consists of updates performed on each of the ensemble members, as given by

$$\mathbf{x}_n^a = \mathbf{x}_n^f + (\mathbf{P}^e)^f \mathbf{H}^T \left[\mathbf{H}(\mathbf{P}^e)^f \mathbf{H}^T + \mathbf{R}^e \right]^{-1} [\mathbf{y}_n - \mathbf{Hx}_n], \quad 1 \leq n \leq N. \quad (41)$$

With a finite ensemble size, the use of the ensemble covariances introduces an approximation of the true covariances. Furthermore, if the number of measurements is larger than the number of ensemble members, then the matrices $\mathbf{H}(\mathbf{P}^e)^f \mathbf{H}^T$ and \mathbf{R}^e are singular, and we must use a pseudo inversion.

Equation (41) implies the update of the ensemble mean

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + (\mathbf{P}^e)^f \mathbf{H}^T \left[\mathbf{H}(\mathbf{P}^e)^f \mathbf{H}^T + \mathbf{R}^e \right]^{-1} [\bar{\mathbf{y}} - \mathbf{H}\bar{\mathbf{x}}^f], \quad (42)$$

where $\bar{\mathbf{y}} = \mathbf{y}$ since the measurement perturbations have ensemble mean equal to zero. Thus, the relation between the analyzed and predicted ensemble mean is identical to the relation between the analyzed and predicted state in the standard KF, apart from the use of $(\mathbf{P}^e)^f$ and \mathbf{R}^e instead of \mathbf{P}^f and \mathbf{R} .

If the ensemble updates \mathbf{x}_n^a and $\bar{\mathbf{x}}^a$ from Equations (41) and (42) are inserted back into the analyzed covariance Equation (36), it was shown by Evensen (1994) and Burgers et al. (1998) that

$$(\mathbf{P}^e)^a = [\mathbf{I} - \mathbf{K}^e \mathbf{H}] (\mathbf{P}^e)^f, \quad (43)$$

with \mathbf{K}^e being the ensemble-based Kalman gain matrix (c.f., Section 3.1.1 and Equation (17)). The result in Equation (43) proves a consistency between the original KF and the EnKF (c.f., Equation (19)).

Note that the EnKF analysis scheme is approximate since it does not correctly account for non-Gaussian contributions in the predicted ensemble. In other words, the EnKF analysis scheme does not solve the Bayesian update equation for a non-Gaussian pdf. On the other hand, the EnKF analysis scheme is not just a resampling of a Gaussian posterior distribution. Only the updates defined by the right-hand side of Equation (41), which we add to the prior non-Gaussian ensemble, are linear. Thus, the updated ensemble inherits many of the non-Gaussian properties from the forecast ensemble. In summary, we have a computationally efficient analysis scheme where we avoid resampling of the posterior.

4.1.4 | Formulation in terms of the ensemble

Evensen (2003) reformulated the EnKF analysis scheme in terms of the ensemble without reference to the ensemble covariance matrix, which allows for an efficient numerical implementation and alternative interpretation of the method. We follow the slightly updated formulation from Evensen (2009b).

The analysis Equation (41), expressed in terms of the ensemble matrices, is

$$\mathbf{E}^a = \mathbf{E}^f + (\mathbf{X}^f)(\mathbf{X}^f)^T \mathbf{H}^T \left[\mathbf{H}(\mathbf{X}^f)(\mathbf{X}^f)^T \mathbf{H}^T + \mathbf{Y}'_o (\mathbf{Y}'_o)^T \right]^{-1} [\mathbf{Y}_o - \mathbf{H}\mathbf{E}^f], \quad (44)$$

where we replace all the error covariance matrices by their ensemble representations. Evensen (2003) showed that the analysis Equation (44) could be written as

$$\mathbf{E}^a = \mathbf{E}^f \mathbf{T}, \quad (45)$$

with the transformation matrix $\mathbf{T} \in \mathbb{R}^{N \times N}$ defined as

$$\mathbf{T} = \mathbf{I}_N + \mathbf{Y}^T \mathbf{C}^{-1} (\mathbf{Y}_o - \mathbf{H}\mathbf{E}^f). \quad (46)$$

Here, we have defined the observed ensemble-anomaly matrix

$$\mathbf{Y} = \mathbf{H}\mathbf{X}^f \in \mathbb{R}^{d \times N}, \quad (47)$$

and the matrix

$$\mathbf{C} = \mathbf{Y}\mathbf{Y}^T + \mathbf{Y}'_o (\mathbf{Y}'_o)^T \in \mathbb{R}^{d \times d}, \quad (48)$$

while $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ is the identity matrix. In practical implementations, it is common also to use the full-rank exact measurement error covariance matrix \mathbf{R} as an alternative to the product of measurement perturbations $\mathbf{R}^e = \mathbf{Y}'_o (\mathbf{Y}'_o)^T$, although that comes at an additional computational cost unless one assumes \mathbf{R} to be diagonal (see also Hoteit, Pham, Gharamti, & Luo, 2015).

The significant result from Equation (45) is that the EnKF update ensemble becomes a combination of the forecast ensemble members, and is searched within the space spanned by the forecast ensemble. It is clear that the formulation in Equation (45) is a stochastic scheme due to the use of randomly perturbed measurements. The use of perturbed measurements allows for a natural interpretation of the EnKF as a Monte Carlo algorithm, while making it easy to understand and implement. An illustration is provided in Figure 3, which shows the EnKF mean solution and estimated error, as a function of the model grid and for three successive analysis times.

An efficient and stable numerical implementation of the analysis scheme is discussed in Evensen (2009b), including the case in which \mathbf{C} is singular or of low rank, for example, due to the number of measurements being larger than the number of realizations or if measurements are linearly dependent on each other.

In practice, the ensemble size is critical since the computational cost scales linearly with the number of realizations. That is, each realization needs to be integrated forward in time. The cost associated with the ensemble integration motivates the use of an ensemble with the minimum number of realizations that can provide acceptable accuracy. We will see in Section 5.1 how the optimal ensemble size to achieve satisfactory performance is also related to the dynamical properties, notably the degree of instabilities, of the model.

The convergence of the solution of the stochastic EnKF for $N \rightarrow \infty$ has been studied by Le Gland, Monbet, and Tran (2009): the rate of such convergence was proven to be $1/\sqrt{N}$, and in the linear dynamics and observation models case, the

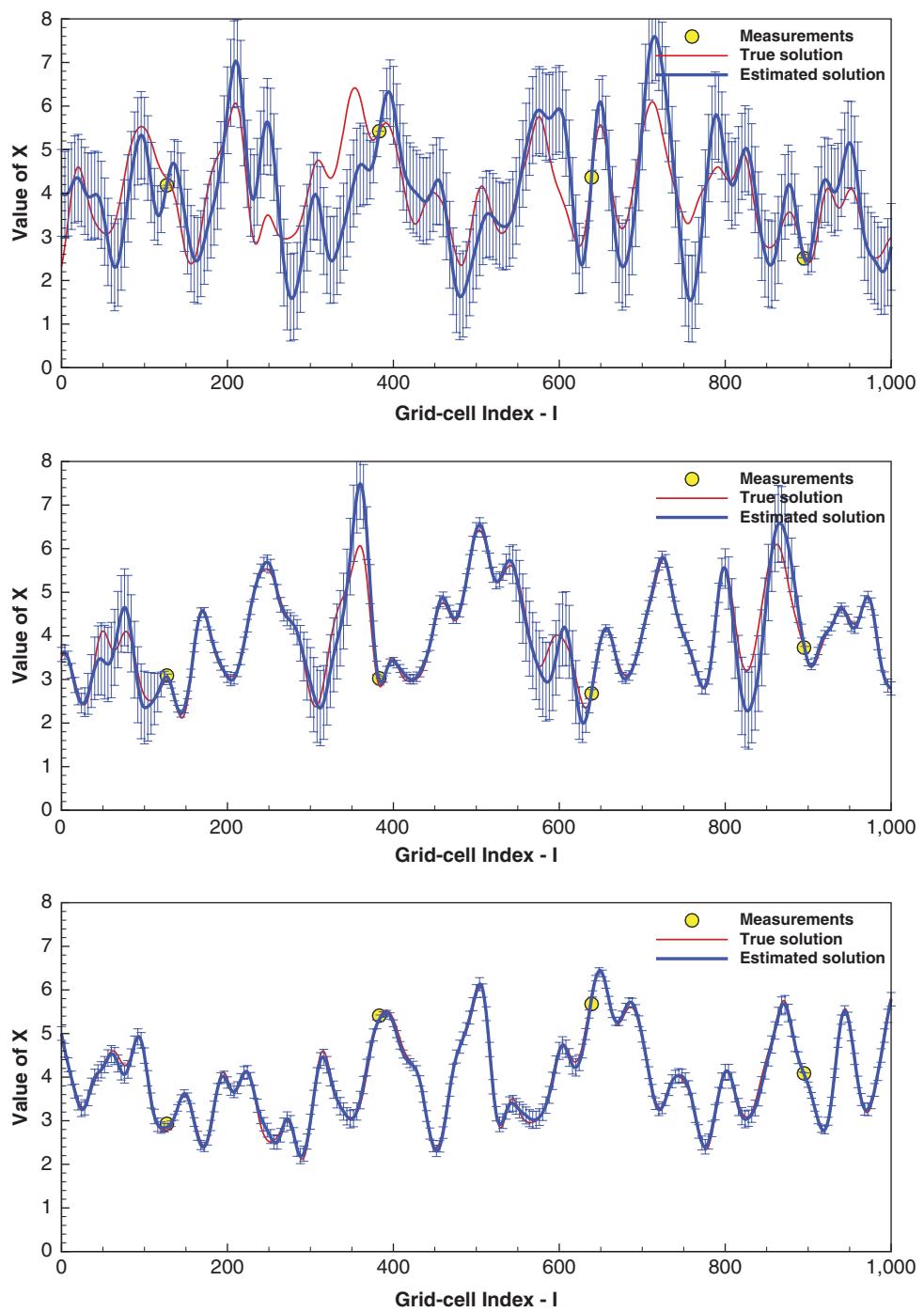


FIGURE 3 A linear advection equation on a periodic domain (traveling wave from left to right) illustrates how the Kalman filter (KF) and/or ensemble Kalman filter (EnKF) estimates the state at three different times, namely, (top) $t = 5$, (middle) $t = 150$, and (bottom) $t = 300$. The plots show the reference true solution (red line), measurements (yellow circles), and error estimate (error bars equivalent to one standard deviation of forecast error variance). The initial condition, observation, and model error variance is 1.0, 0.01, and 0.0001, respectively, and the ensemble size is $N = 500$. In each of the panels we see that the error increases downstream, but also that downstream of an observation its effect is propagated by the wave and the errors are smaller on the right than on the left of the observation point. Similarly, by comparing the three panels for increasing time, we see how the errors decreases as a function of time, indicating a good performance of the filter

solution approaches that of the KF. In the general nonlinear case, and even when at least the observation operator is linear, the asymptotic solution differs from the fully Bayesian filter, and for any finite value of N , the individual ensemble members are not independent.

There are two significant sources of sampling errors in the EnKF, namely (a) the use of a finite ensemble of model realizations and (b) the introduction of stochastic measurement perturbations (Evensen, 2004, 2009b). Besides, stochastic model errors influence the predicted error statistics, which are approximated by the ensemble. The stochastic perturbation of

measurements used in EnKF can be avoided using a square-root implementation of the analysis scheme, which is the topic of the next section.

4.2 | Deterministic square-root schemes

The perturbation of measurements used in the EnKF standard analysis Equation (44) is an additional source of sampling error. However, some papers have introduced methods alternative to the traditional EnKF, that is, *square-root schemes*, which compute the analysis without perturbing the measurements and preserve precisely the posterior variance in the update. Examples include the *ensemble transform kalman filter* (ETKF) by Bishop, Etherton, and Majumdar (2001); Whitaker and Hamill (2002), the *ensemble adjustment kalman filter* (EAKF) by Anderson (2001), and the *maximum likelihood ensemble filter* (MLEF) by Zupanski (2005).

4.2.1 | Updating the mean

In a square-root scheme, the analyzed ensemble mean is computed from the standard KF analysis equation,

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{X}^f \mathbf{Y}^T \mathbf{C}^{-1} (\bar{\mathbf{y}} - \mathbf{H} \bar{\mathbf{x}}^f), \quad (49)$$

with the matrix \mathbf{C} defined in Equation (48) but with $\mathbf{Y}_o' (\mathbf{Y}_o')^T$ replaced by \mathbf{R} , that is,

$$\mathbf{C} = \mathbf{Y} \mathbf{Y}^T + \mathbf{R}. \quad (50)$$

The ensemble mean in the square-root formulation is thus computed without the use of perturbed observations.

4.2.2 | Updating the ensemble perturbations

We derive the deterministic algorithm for updating the ensemble perturbations starting from the traditional analysis Equation (19) for the covariance update in the KF but written in its ensemble-based representation in Equation (43). We then insert the expression for the Kalman gain matrix to obtain

$$(\mathbf{P}^e)^a = (\mathbf{P}^e)^f - (\mathbf{P}^e)^f \mathbf{H}^T \left[\mathbf{H} (\mathbf{P}^e)^f \mathbf{H}^T + \mathbf{R} \right]^{-1} \mathbf{H} (\mathbf{P}^e)^f. \quad (51)$$

Using the ensemble representation of the error covariance matrices, Equations (35) and (36), along with Equations (47) and (50), we can rewrite Equation (51) as

$$(\mathbf{X}^a)(\mathbf{X}^a)^T = (\mathbf{X}^f)[\mathbf{I}_N - \mathbf{Y}^T \mathbf{C}^{-1} \mathbf{Y}] (\mathbf{X}^f)^T = (\mathbf{X}^f)[\mathbf{I}_N + \mathbf{Y}^T \mathbf{R}^{-1} \mathbf{Y}]^{-1} (\mathbf{X}^f)^T. \quad (52)$$

We obtain the second expression by making use of the Sherman–Morrison–Woodbury formula (see Golub & van Loan, 2013, their Section 2.1.4).

We can derive the square-root filters by factorizing the symmetric expression in Equation (52). The simplest, but not a very efficient approach, is to compute the eigenvalue factorization of the matrix $\mathbf{I}_N - \mathbf{Y}^T \mathbf{C}^{-1} \mathbf{Y} = \mathbf{Z} \Lambda \mathbf{Z}^T$, and then the update as

$$\mathbf{X}^a = \mathbf{X}^f \mathbf{Z} \sqrt{\Lambda}, \quad (53)$$

which defines a symmetric factorization of Equation (52). However, this update equation does not preserve the mean (it is biased). It was shown (Livings, Dance, & Nichols, 2008; Sakov & Oke, 2008b; Wang, Bishop, & Julier, 2004) that the *symmetric square root* preserves the zero mean in the updated perturbations. Accordingly, the update in Equation (53) must be replaced by

$$\mathbf{X}^a = \mathbf{X}^f \mathbf{Z} \sqrt{\Lambda} \mathbf{Z}^T, \quad (54)$$

which is another symmetric factorization of Equation (52) (see Hunt, Kostelich, & Szunyogh, 2007). Consequently, if the predicted ensemble members have a non-Gaussian distribution, then the updated distribution retains the shape, although the variance is reduced (see the review in Raanes, Carrassi, & Bertino, 2015, their Section 4). The importance of the symmetry in the square-root methods was first recognized by Ott et al. (2004) and then adopted by Hunt et al. (2007).

The operational use of the square-root schemes requires more focus on numerical stability and efficiency. Here, we assumed that \mathbf{C}^{-1} exists, which is not always the case and particularly not so when the number of measurements is larger than the ensemble size nor when \mathbf{R} is not full rank. For a more elaborate derivation of numerically stable schemes, we refer to the initial publications by Anderson (2001); Whitaker and Hamill (2002); Bishop et al. (2001); and the reviews by Tippett, Anderson, Bishop, Hamill, and Whitaker (2003), and Nerger, Janjić, Schröter, and Hiller (2012). Evensen (2004, 2009a, 2009b)

derived a numerically efficient square-root filter that computes the inversion in the subspace spanned by the measured ensemble perturbations \mathbf{Y} . The scheme works equally well with a nondiagonal measurement error-covariance matrix and in the case when \mathbf{C} is of low rank.

Equation (52) is also known as *right transform* as it applies to the right-hand side of the anomalies, in the N -dimensional ensemble space (Asch et al., 2016). Similarly, *left transform* expressions also exist, that apply to the m -dimensional state-space (see Asch et al., 2016, their Section 6.4, for an extensive treatment of the subject).

A randomization of the analysis update can also be used to generate updated perturbations that better resemble a Gaussian distribution, see Evensen (2004). Thus, we write the symmetric square root solution Equation (54) as

$$\mathbf{X}^a = \mathbf{X}^f \mathbf{Z} \sqrt{\Lambda} \mathbf{Z}^T \boldsymbol{\Phi}^T, \quad (55)$$

where $\boldsymbol{\Phi}$ is a mean-preserving random orthogonal matrix, which can be computed using the algorithms from Pham (2001) or Sakov and Oke (2008a). Note again that the random rotation in the square-root filter, contrary to the measurement perturbation used in EnKF, eliminates all previous non-Gaussian structures from the forecast ensemble. Efficient means to account for stochastic model noise within square-root filters are discussed in Raanes et al. (2015).

4.3 | Ensemble Kalman smoother (EnKS)

The EnKS is a straightforward extension of the EnKF in which we use the output of the latter as a prior in the EnKS. Like the EnKF uses the ensemble covariances in space to spread the information from the measurements, the EnKS uses the ensemble covariances to spread the information in both space and time (also backward). We need to reintroduce here the time index.

Assume we have measurements available at discrete times t_k , $k = 1, \dots, K$. We compute the EnKF solution recursively by assimilating the observations when they are available and then propagate the ensemble until the next measurements (c.f., Section 4.1). The members of the EnKF can be stored at any time instant when we desire a smoother update.

The analysis update at a time t_k (which does not have to be an observation time) from measurements available at a later time t_j , $1 \leq k < j \leq K$, reads

$$\mathbf{E}_k^a = \mathbf{E}_k^f + \mathbf{X}_j^f \mathbf{Y}_j^T \mathbf{C}_j^{-1} (\mathbf{Y}_{0,j} - \mathbf{Y}_j) \quad 1 \leq k < j \leq K, \quad (56)$$

where \mathbf{Y}_j from Equation (47), and \mathbf{C}_j from Equation (48) are evaluated using the ensemble and measurements at t_j . From the right-hand side of Equation (56), we recognize the product $\mathbf{X}_j^f \mathbf{Y}_j^T$ as the covariance between the predicted measurements at t_j and the model state at $t < t_j$.

The update at the time t_k uses precisely the same combination of ensemble members as was defined by \mathbf{T}_k in Equation (46) for the EnKF analysis at the time t_k . Thus, we can compactly write the EnKS analysis as

$$\mathbf{E}_k^{a, \text{EnKS}} = \mathbf{E}_k^{a, \text{EnKF}} \prod_{j=\hat{k}}^K \mathbf{T}_j, \quad (57)$$

where \hat{k} corresponds to the first data time after t_k , and t_K is the last measurement time. It is then a simple exercise to compute the EnKS analysis as soon as the EnKF solution has been found. The computation requires only the storage of the transform matrices $\mathbf{T}_k \in \mathbb{R}^{N \times N}$, for $k = 1, \dots, K$, and the EnKF ensemble matrices for the times when we want to compute the EnKS analysis. Note that the EnKF ensemble matrices are large, but it is possible to store only specific variables at selected locations where the EnKS solution is needed. The equivalence between the formulation of the EnKS given above with the Rauch–Tung–Striebel smoother (c.f., Section 3.1.2) even in the nonlinear, non-Gaussian, case is discussed in Raanes (2016).

An example of implementation of an ensemble smoother is represented by the “no cost smoother” (aka as *Running in place*, RIP) of Kalnay and Yang (2010) that provides explicitly the weights of the ensemble members at the end of the assimilation window, and then applies those same weights throughout the window. The RIP/no-cost smoother has been used to accelerate the spin up of the local ensemble transform kalman filter (LETKF, Hunt et al., 2007), to increase its accuracy and to allow longer time windows. The no cost smoother designed originally designed for the LETKF has been adapted to the ensemble square-root filter (EnSFR) by Wang, Xue, Schenkman, and Min (2013) and is used operationally to accelerate the spinup of the EnSRF forecast of tornadoes and severe storms at the Center for Analysis and Prediction of Storms of the University of Oklahoma (www.caps.ou.edu). A recent study that analyses and compares ensemble smoothers for solving inverse problems can be found in Evensen (2018).

4.4 | Making it work: Localization and inflation

The reduction of dimensionality obtained through the Monte Carlo treatment of the KF that yielded the EnKF comes at a price. Indeed, with a limited number, N , of anomalies, the sample covariance matrix is severely rank deficient. It would be surprising if such sample error covariance matrix could be a good substitute for the, possibly, full-rank true one.

Assume that the true error covariance matrix is \mathbf{B} . Further define \mathbf{P}^e as the sample covariance matrix meant to approximate \mathbf{B} and obtained from N samples, (as in Equation (35)) of the normal distribution with covariance matrix \mathbf{B} . Then, it can be shown that for two distinct entry indices i and j corresponding to distinct locations:

$$\mathbb{E}([P^e - \mathbf{B}]_{ij}^2) = \frac{1}{N-1} ([\mathbf{B}]_{ij}^2 + [\mathbf{B}]_{ii}[\mathbf{B}]_{jj}), \quad (58)$$

with \mathbb{E} indicating the expectation (average) operator. In most spatial geophysical systems, $[\mathbf{B}]_{ij}$ is expected to vanish fast (often exponentially) with the distance between sites corresponding to i and j . By contrast, the $[\mathbf{B}]_{ii}$ are the variances and remain finite, so that

$$\mathbb{E}([P^e - \mathbf{B}]_{ij}^2) \underset{|i-j| \rightarrow \infty}{\sim} \frac{1}{N-1} [\mathbf{B}]_{ii}[\mathbf{B}]_{jj}, \quad (59)$$

meaning that both expressions coincide when the sites separation is sufficiently large. Consequently, the right-hand side arithmetically goes to 0 with the ensemble size N . Unfortunately, since $[\mathbf{B}]_{ij}$ is expected to vanish exponentially with the distance, we would have liked instead $\mathbb{E}([P^e - \mathbf{B}]_{ij}^2)$ to also vanish exponentially with the distance. Hence, with N finite (and usually $N \ll m$) the sample covariance $[P^e]_{ij}$ is potentially a bad approximation of the vanishing true covariance, especially for large distances $|i - j|$. This approximation generates spurious correlations between distant parts of the system, as a manifestation of the rank deficiency of the sample error covariance matrix.

The errors committed with such an approximation are usually referred to as sampling errors. When taking \mathbf{P}^e as the forecast/background error covariance matrix, the accumulation of such errors over the EnKF DA cycles can be detrimental to the stability of the EnKF. Without counter measures, the divergence of the EnKF in high-dimensional geophysical systems is almost systematic. Fortunately, more or less ad hoc fixes meant to address the issue are known: *localization* and *inflation*. They are rather complementary and both of them are often required.

4.4.1 | Localization

Localization fundamentally relies on the exponential decrease of the correlations with the distance in geophysical systems. Under this condition, one can assume that the interdependence of distant parts of a physical system is negligible. Hence, EnKF analyses could be made local (Evensen, 2003; Hamill, Whitaker, & Snyder, 2001; Haugen & Evensen, 2002; Houtekamer & Mitchell, 2001; Ott et al., 2004). Localization comes in two flavors that we describe in the following.

Covariance localization

The first approach, named covariance localization (CL), seeks to regularize the sample error covariance matrix, with the goal to mitigate the rank deficiency of \mathbf{P}^e and the appearance of spurious error correlations. A mathematical means to achieve this objective is to compute the Schur (or Hadamard) product of \mathbf{P}^e with a well chosen smooth correlation function ρ . We assume ρ to have exponentially vanishing correlations for distant parts of the system and to be representative of such dampening of real geophysical systems. The Schur product of ρ and \mathbf{P}^e is defined by

$$[\rho \circ \mathbf{P}^e]_{ij} = [\rho]_{ij} [\mathbf{P}^e]_{ij}, \quad (60)$$

that is, a point-wise matrix multiplication. The Schur product theorem (Horn & Johnson, 2012) ensures that this product is positive semidefinite, and hence still represents a proper covariance matrix. For sufficiently regular ρ , $\rho \circ \mathbf{P}^e$ turns out to be full rank (hence positive definite). Moreover, the spurious correlations should be exponentially damped as is obvious from Equation (60).

To illustrate the CL principle, we consider a multivariate ($m = 100$) Gaussian distribution, with mean $\mathbf{x}^b = 0$. Its covariance matrix \mathbf{B} is given by $[\mathbf{B}]_{ij} = e^{-|i-j|L}$, where $i, j = 1, \dots, m$ and $L = 10$ is the correlation length. An ensemble of $N = 20$ members $\{\mathbf{x}_n\}_{n=1, \dots, N}$ is drawn from this distribution from which we can compute the unbiased sample covariance matrix $\mathbf{P}^e = \frac{1}{N-1} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T$, with the mean $\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n$, which are approximations of \mathbf{B} and \mathbf{x}^b , respectively. The true and the sample covariance matrices are shown in Figure 4. The spurious correlations are obvious in panel Figure 4b, where \mathbf{P}^e is displayed. Let us define ρ through the Gaspari–Cohn (GC) function (Gaspari & Cohn, 1999),

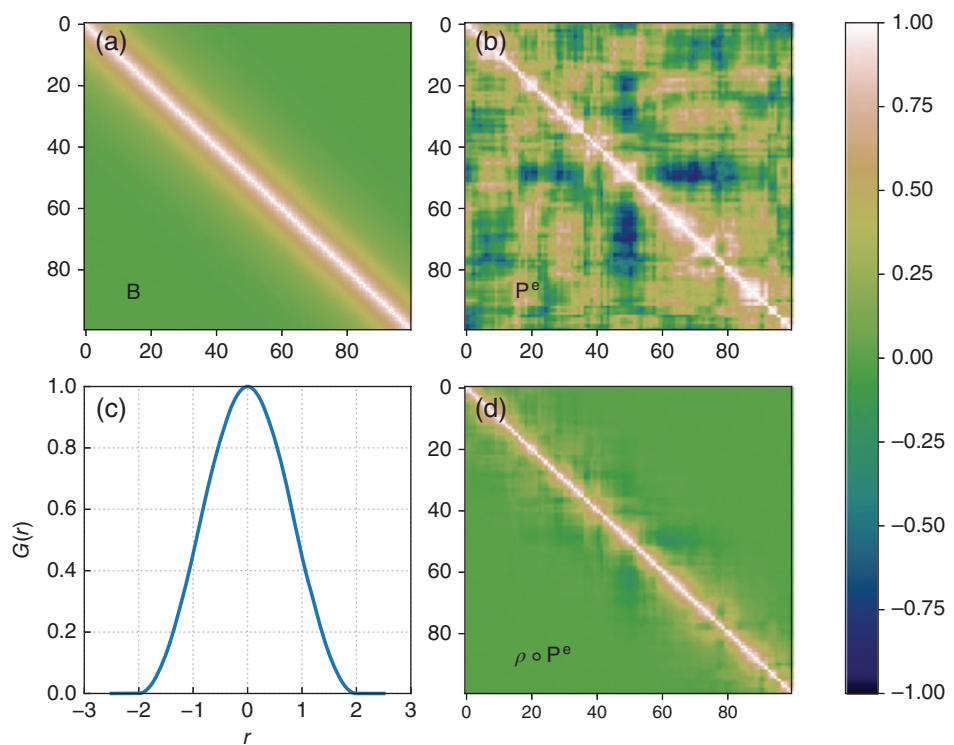


FIGURE 4 Panel a: True covariance matrix. Panel b: Sample covariance matrix. Panel c: Gaspari–Cohn correlation matrix used for covariance localization. Panel d: Regularized covariance matrix obtained from a Schur product

$$G(r) = \begin{cases} \text{if } 0 \leq r < 1: & 1 - \frac{5}{3}r^2 + \frac{5}{8}r^3 + \frac{1}{2}r^4 - \frac{1}{4}r^5, \\ \text{if } 1 \leq r < 2: & 4 - 5r + \frac{5}{3}r^2 + \frac{5}{8}r^3 - \frac{1}{2}r^4 + \frac{1}{12}r^5 - \frac{2}{3}r, \\ \text{if } r \geq 2: & 0, \end{cases}$$

which is a fifth-order piece-wise rational function. The cut off function is defined by $r > 0 \mapsto G(r/c)$, where c is a length scale which is called the localization radius. It mimics a Gaussian distribution but vanishes beyond $r \geq 2c$, which is numerically efficient (in mathematical terms, G is compactly supported). The GC cut off function is shown in Figure 4c. Hence, choosing $c = L$, we define $[\rho]_{ij} = G(|i - j|/L)$. The regularized covariance matrix, $\rho \circ \mathbf{P}^e$ is shown in Figure 4d. As expected, the spurious correlations are tapered.

One of the drawbacks of CL is that it still uses huge regularized covariance matrices (a priori of size $m \times m$). Possible remedies are (a) avoiding the brute-force computation and storage of $\rho \circ \mathbf{P}^e$ exploiting for instance the possible sparsity of ρ , (b) combining CL with *domain localization* (DL) (see below), (c) or transferring the Schur product from state space to observation space for sparse in situ observations. Solution (c) can for instance be used to compute the Kalman gain (c.f., Equation (17)) in a more efficient manner (Greybush, Kalnay, Miyoshi, Ide, & Hunt, 2011; Houtekamer & Mitchell, 2001; Ott et al., 2004).

In a multivariate setting, variable localization was also proposed and tested by Kang et al. (2011), where cross correlations between carbon dioxide concentration and meteorological variables other than wind variables are set to zero.

Domain localization

An alternative to CL is DL. In DL, one performs the global EnKF analysis as a collection of independent local EnKF analyses, each of them assimilating a subset of the observations. Typically, one seeks to update a state at location/grid-cell x . The EnKF analysis is locally carried out by assimilating observations contained in a disk of radius L centered on x . The number of observations in this local domain should be small enough so that the ensemble can accommodate the information content of these observations, that is, that \mathbf{P}^e is locally full rank. This is equivalent to setting the observation error covariance matrix \mathbf{R} to \mathbf{R}_x where all the entries associated to sites outside the disk are large or infinite, which rules out these outer observations. In DL, both the state vector and anomalies are locally updated. A global analysis is then recovered from this collection of local analyses. Because the subdomains overlap, one assumes that the transition between analyses is smooth enough so that the global

updated ensemble remains physically balanced. A way to improve on these transitions is to obtain \mathbf{R}_x by tapering (Schur product) \mathbf{R}^{-1} with a regularizing cut off function, typically the GC, that decreases with the distance from x .

Because the number of local EnKF analyses to be carried out scales linearly with the size of the system m , the method may seem numerically doomed. However, on one hand, all of the analyses can be performed in parallel and, on the other hand, their numerical complexity is significantly reduced as it is driven by the ensemble size and the number of local, not global, observations. The numerical feasibility of the approach has been demonstrated by Ott et al. (2004) and Hunt et al. (2007).

Sakov and Bertino (2011) have shown that CL and DL are not mathematically equivalent in general but that they can become so in regimes where the analysis is dominated by the prior. Another important point to mention is the physical imbalance that may result from the gluing of local analyses and from the true long-distance correlations that have necessarily been discarded in the localization process (Greybush et al., 2011; Kepert, 2009). For chaotic systems, it has been suggested (see Section 5.1) that the ensemble size below which localization is mandatory is equal to the dimension of the unstable and neutral subspace of the dynamics.

4.4.2 | Inflation

Using localization addresses the rank-deficiency issue and gets rid of spurious correlations. Unfortunately, sampling errors are not entirely removed in the process. Residual errors are still often sufficient to make the EnKF diverge for long-term runs (see Grudzien, Carrassi, & Bocquet, 2018a, for a discussion on the need for inflation in reduced order filter). One very simple ad hoc means to account for this residual sampling error is to inflate the error covariance matrix by a multiplicative factor $\lambda^2 \geq 1$, $\mathbf{P}^e \rightarrow \lambda^2 \mathbf{P}^e$, which can also be implemented by inflating the anomalies $\mathbf{x}_n \rightarrow \bar{\mathbf{x}} + \lambda(\mathbf{x}_n - \bar{\mathbf{x}})$ (Anderson & Anderson, 1999; Pham, Verron, & Roubaud, 1998). Note that inflation is not only used to cure sampling errors (Bocquet, 2011; Bocquet, Raanes, & Hannart, 2015; Whitaker & Hamill, 2012), but is also often used to counteract model error impact.

Inflation can also come in an additive, clearly nonequivalent, form: $\mathbf{x}_n \rightarrow \mathbf{x}_n + \boldsymbol{\epsilon}_n$ with $\mathbb{E}[\boldsymbol{\epsilon}_n(\boldsymbol{\epsilon}_n)^T] = \mathbf{Q}$.

As a drawback, inflation often needs to be tuned in order to obtain satisfactory performance of the EnKF, which is numerically costly. Hence, adaptive schemes have been developed to make the task more automatic (Anderson, 2007; Bocquet, 2011; Bocquet, Raanes, & Hannart, 2015; Bocquet & Sakov, 2012; Brankart, Cosme, Testut, Brasseur, & Verron, 2010; Li, Kalnay, & Miyoshi, 2009; Liang et al., 2012; Miyoshi, 2011; Wang & Bishop, 2003; Ying & Zhang, 2015; Zheng, 2009). For instance, a variant of the deterministic EnKF, the DEnKF from Sakov and Oke (2008b), was introduced as an approximation of the EnKF with an implicit inflation.

4.5 | Ensemble-variational methods

4DVar and the EnKF are now both used in operational meteorological centers as well as in academic studies (see Bauer, Thorpe, & Brunet, 2015; Buehner, Morneau, & Charette, 2013). In spite of apparent similar performances in the context of synoptic scale meteorology, they are not equivalent. 4DVar searches for a nonlinear estimation of the maximum a posteriori of the underlying pdf using nonlinear optimization techniques, while the EnKF only performs a linear update (with the notable exception of the MLEF; Zupanski, 2005). The EnKF propagates the statistics of the errors, hence estimating the *errors of the day*, whereas the traditional 4DVar only relies on time-independent statistics of the errors. An important weakness of 4DVar, of technical nature, is the derivation of the adjoint models, which remains a computer sciences challenge, whereas the EnKF estimates the tangent linear of the observation operator and its adjoint within the ensemble subspace. These and other pros and cons of both approaches have been discussed by Lorenc (2003); Kalnay, Li, Miyoshi, Yang, and Ballabrera-Poy (2007); Yang, Corazza, Carrassi, Kalnay, and Miyoshi (2009); and Bocquet and Sakov (2013).

Hence, it is tempting to combine both approaches, for theoretical reasons (nonlinear analysis, errors of the day) and technical reasons (no adjoint models). The resulting algorithms are called *ensemble-variational* (EnVar) methods. They have been reviewed in Chapter 7 of Asch et al. (2016) on theoretical grounds and with a focus on operational implementation by Bannister (2017), although the iterative ensemble Kalman smoother (IEnKS) (see Section 4.5.4) is lacking in this latter review. A classification of these methods and the terminology has been proposed by Lorenc (2013).

Here, we distinguish between (a) the hybrid methods, (b) the methods based on an ensemble of data assimilation (EDA), (c) the so called 4DEnVar, and (d) full nonlinear four-dimensional EnVar methods, of which the IEnKS is a deterministic exemplar.

4.5.1 | Hybrid methods

Hybrid can refer to any combination of a variational method and of an EnKF method. Yet, it quite often specifically refers to the hybridizing of a static error covariance matrix with a dynamical one sampled from an ensemble. The idea was first

introduced by Hamill and Snyder (2000). The objective was to use the static covariance of a 3DVar and perform a variational analysis to be used for the analysis step of an EnKF. The effective covariance matrix used for the prior is:

$$\mathbf{B} = \alpha \mathbf{C} + (1 - \alpha) \mathbf{X}^f (\mathbf{X}^f)^T, \quad (61)$$

where \mathbf{C} is the static error covariance matrix (as used in 3DVar or optimal interpolation; Kalnay, 2002), \mathbf{X}^f is the matrix of the forecast ensemble anomalies, and $\alpha \in [0, 1]$ is a scalar that weights the static and dynamical contributions. The updated ensemble can be obtained within the framework of a stochastic EnKF (c.f., Section 4.1) using several stochastically perturbed variational problems. On the other hand, if the mean is estimated in a deterministic EnKF framework (c.f., Section 4.2), the anomalies update has to rely on an approximation due to the necessary reduction of the rank from the hybrid covariance to the new dynamical covariances. There are ways to improve this update, such as using a left-transform update (Bocquet, Raanes, & Hannart, 2015; Sakov & Bertino, 2011), or Lanczos vectors (Auligné, Ménétrier, Lorenc, & Buehner, 2016). Note that, as with the EnKF, it may be necessary to enforce localization of the sample covariances, which would be based on $\mathbf{B} = \alpha \mathbf{C} + (1 - \alpha) \rho \circ [\mathbf{X}^f (\mathbf{X}^f)^T]$, using CL.

4.5.2 | Ensemble of data assimilation

Methods known as *ensemble of data assimilation* (EDA) stem from meteorological prediction centers that operate a 4DVar, and in particular Météo-France and the European Centre for Medium-Range Weather Forecasts (ECMWF). The idea is to introduce dynamical errors that are absent in the traditional 4DVar. In order to build on the existing 4DVar systems, one considers an ensemble of N 4DVar analyses. Each analysis, indexed by i , uses a different first guess \mathbf{x}_0^i , and observations perturbed with $\mathbf{\epsilon}_k^i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ to maintain statistical consistency. Hence, each analysis i carries out the minimization of (c.f., Equation (30))

$$\mathcal{J}_i^{\text{EDA}}(\mathbf{x}_0) = \frac{1}{2} \sum_{k=0}^K \left\| \mathbf{y}_k + \mathbf{\epsilon}_k^i - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\mathbf{x}_0) \right\|_{\mathbf{R}_k^{-1}}^2 + \frac{1}{2} \left\| \mathbf{x}_0 - \mathbf{x}_0^i \right\|_{\mathbf{B}^{-1}}^2. \quad (62)$$

The background covariance \mathbf{B} is typically a hybrid one as in Equation (61) because it still uses the static covariances of the traditional 4DVar and incorporates the sampled covariances from the dynamical perturbations. The procedure yields an updated ensemble, from which it is possible to assess a dynamical part of the error covariance matrix. This is rather close to the idea of the hybrid EnKF-3DVar, but with a 4DVar scheme. It is worth to note that, in the linear case and Gaussian case, EDA is exactly Bayesian, in the sense that it produces an ensemble of independent realizations of the conditional PDF, though not necessarily in the nonlinear or non-Gaussian case (Jardak & Talagrand, 2018; Liu et al., 2017, and references therein). The method has been implemented at Météo-France (Berre, Varella, & Desroziers, 2015; Raynaud, Berre, & Desroziers, 2009, 2011) and at the ECMWF (Bonavita, Isaksen, & Hólm, 2012; Bonavita, Raynaud, & Isaksen, 2011). Like for the hybrid method, it may be necessary to localize the sampled part of the hybrid covariances, using for instance CL. In this context, it may be convenient to enforce CL via techniques such as the α control variable trick (Buehner, 2005; Lorenc, 2003; Wang, Snyder, & Hamill, 2007)—which is mathematically equivalent to CL—or using wavelet truncations (Berre et al., 2015).

4.5.3 | 4DEnVar method

NWP centers which have been using 4DVar were also the ones to offer the most reliable forecasts. But the future of 4DVar as an operational tool is uncertain because of its poor scalability and of the cost of the adjoint model maintenance, although weak-constraint 4DVar has scalable implementations (Fisher & Gürol, 2017). The 4DEnVar method emerged in these centers as a way to circumvent the development of the adjoint of the dynamical model. The key idea is based on the ability of the EnKF to estimate the sensitivities of the observation to the state variables using the full observation model in place of the tangent linear one and of its adjoint in the computation of the Kalman gain.

Liu, Xiao, and Wang (2008) essentially proposed to generalize this idea to the computation of the sensitivities of the observations within a data assimilation window (DAW) to the variables of the initial state. Hence, these sensitivities are associated to the composition $\mathcal{H} \circ \mathcal{M}$ defined over the DAW, rather than to observation operator \mathcal{H} as in the EnKF. As a consequence, the sensitivities of the observations with respect to the degrees of freedom of each anomaly can be approximated with

$$\varepsilon^{-1} \mathcal{H}_k \circ \mathcal{M}_{k:0}(\bar{\mathbf{x}}_0 \mathbf{1}^T + \varepsilon \mathbf{X}^f) \left(\mathbf{I}_m - \frac{\mathbf{1} \mathbf{1}^T}{m} \right). \quad (63)$$

This formula first builds an ensemble of mean $\bar{\mathbf{x}}_0$ with anomalies \mathbf{X}^f scaled by ε ; then propagates this ensemble through $\mathcal{H}_k \circ \mathcal{M}_{k:0}$; rescales the output ensemble by ε^{-1} ; and finally generates an ensemble of centered perturbations. With $0 < \varepsilon \ll 1$, these perturbations can be seen as generated by the tangent linear of $\mathcal{H}_k \circ \mathcal{M}_{k:0}$ (using finite differences), whereas with

$\varepsilon = \sqrt{m-1}$, these perturbations account for the nonlinearity of $\mathcal{H}_k \circ \mathcal{M}_{k:0}$ as in the EnKF. The set of these sensitivity vectors form a basis for the analysis, an idea quite close to that of the reduced-rank 4DVar put forward in oceanography by Robert et al. (2005).

In 4DEnVar, the perturbations are usually generated stochastically, for instance, resorting to a stochastic EnKF (Buehner, Houtekamer, Charette, Mitchell, & He, 2010a; Liu, Xiao, & Wang, 2009). Hence, in addition to avoiding the need for a model adjoint, flow-dependent error estimation is introduced. Given that 4DEnVar has been developed in NWP centers using 4DVar, which relies on building a sophisticated static background error covariance matrix, the background is usually of hybrid nature. If the perturbations are generated by a deterministic scheme, then the method could be seen as similar to the 4D-ETKF (Hunt et al., 2007), where the linear analysis is carried out in a basis made of trajectories (the perturbations) over the DAW obtained from the previous deterministic analysis.

Because of the introduction of the limited set of perturbations, localization is again needed to mitigate the limited rank of the sample error covariance matrix. However, in a 4D ensemble context and in contrast with the EnKF whose update uses observations of given time, localization must be applied to covariances that are not only distant in space but also possibly in time within the DAW. Unfortunately, localization and the dynamics do not commute in general (Bocquet & Sakov, 2014; Fairbairn, Pring, Lorenc, & Roulstone, 2014), yielding inconsistencies in the implementation of a static localization within the flow. A solution is to implement a *covariant* localization that changes in time with the dynamical flow (Bocquet, 2016; Desroziers, Arbogast, & Berre, 2016). For instance, if localization is enforced via DL, one would pull back the influential observations within the DAW using a surrogate, hyperbolic-like, model for the full evolution model and check in which local domains their antecedent would fall in (Bocquet, 2016). This type of solutions has also been put forward in a sensitivity analysis context (Kalnay, Ota, Miyoshi, & Liu, 2012). Note, however, that this issue is circumvented if one possesses the adjoint of the dynamical model.

Many variants of the 4DEnVar are possible depending on the way the perturbations are generated, or if the adjoint model is available or not (Bocquet, 2016; Buehner et al., 2010a; Buehner, Houtekamer, Charette, Mitchell, & He, 2010b; Clayton, Lorenc, & Barker, 2013; Poterjoy & Zhang, 2015; Zhang & Zhang, 2012). Full 4DEnVar operational systems are now implemented or are in the course of being so (Bowler et al., 2017; Buehner et al., 2015a; Buehner et al., 2013; Desroziers, Camino, & Berre, 2014; Gustafsson, Bojarova, & Vignes, 2014; Kleist & Ide, 2015; Lorenc, Bowler, Clayton, Pring, & Fairbairn, 2015).

4.5.4 | The IEnKS

Most of these EnVar methods, with the noticeable exception of the more firmly grounded EDA ones (Section 4.5.2), have been designed heuristically blending theoretical insights and operational constraints. This led to many variants of the schemes, even when this is not mathematically justified. Most of these ideas stemmed from the variational DA community, with the remarkable exception of the RIP, an EnKF with an outer loop (Kalnay & Yang, 2010). By contrast, the iterative ensemble Kalman smoother (IEnKS). (Bocquet & Sakov, 2014), is a four-dimensional EnVar method that is derived from Bayes' rule and where all approximations are understood at a theoretical level. It comes from ensemble-based DA and, specifically, extends the iterative ensemble Kalman filter (EnKS; Sakov, Oliver, et al., 2012) to a full DAW as in 4DVar. The name smoother reminds us that the method smooths trajectories like 4DVar. However, it can equally be used for smoothing and filtering. The name also pays an homage to the iterative KS of Bell (1994) which corresponds to the nonensemble precursor of the method. Basically, the IEnKS can be seen as an EnKF, for which each analysis corresponds to a nonlinear 4DVar analysis but within the reduced subspace defined by the ensemble. Hence, the associated cost function is of the form

$$\mathcal{J}(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{k=L-S+1}^L \frac{1}{2} \|\mathbf{y}_k - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\bar{\mathbf{x}}_0 + \mathbf{X}_0 \mathbf{w})\|_{\mathbf{R}_k^{-1}}^2, \quad (64)$$

where L is the length of the DAW, and S is the length of the forecast in between cycles (in units of $t_{k+1} - t_k$). Because the IEnKS catches the best of 4DVar (nonlinear analysis) and EnKF (flow dependence of the error statistics), both parameters could be critical. In Equation (64), \mathbf{X}_0 is the matrix containing along its columns the normalized (by $\sqrt{N-1}$) anomalies of the ensemble members at the initial time, while \mathbf{w} is the vector of coefficients used to parametrize the state vector in the span of the ensemble, $\mathbf{x}_0 = \bar{\mathbf{x}}_0 + \mathbf{X}_0 \mathbf{w}$.

The minimization of \mathcal{J} can be performed in the ensemble subspace using any nonlinear optimization method, such as Gauss–Newton (c.f., Appendix B), Levenberg–Marquardt or trust-region methods (Bocquet & Sakov, 2012; Mandel, Bergou, Gürol, Gratton, & Kasanický, 2016; Nino Ruiz & Sandu, 2016). The required sensitivities (derivative of the observation with respect to the initial state) are computed using an ensemble forecast within the DAW, as in 4DEnVar. Hence, it does not require the tangent and adjoint models but emulates them, similarly to 4DEnVar. The anomalies update is computed using an

approximate Hessian of this cost function, using again these sensitivities. The ensemble forecast step is similar to the EnKF but over S time units.

The parameters L and S are constrained to satisfy $1 \leq S \leq L + 1$, if all observations are to be assimilated. In the case $S = L + 1$, with a single iteration of the minimization and further less important restrictions, the IEnKS identifies with the 4D-ETKF (Hunt et al., 2004). In the case $S = 1$ and $L = 0$, it coincides with the MLEF (Zupanski, 2005), which can be seen as a precursor to the IEnKF and IEnKS. The IEnKS represents an ideal tool to study deterministic four-dimensional EnVar methods and test new strategies.

In chaotic systems, the IEnKS outperforms any reasonably scalable DA method in terms of accuracy. By construction, it outperforms the EnKF, the EnKS, and 4DVar for smoothing but also filtering. This has been checked numerically on several low-order models (Lorenz models; Lorenz, 1963, 2005; Lorenz & Emanuel, 1998), 2D turbulence and quasi-geostrophic models (Bocquet, 2016; Bocquet & Sakov, 2014). Figure 5 compares the performance of the IEnKS with that of 4DVar, the EnKF, and the EnKS with optimally tuned multiplicative inflation of the background covariance matrix so as to minimize the root mean square error (RMSE), for filtering and smoothing on the 40-variable Lorenz-96 model (Lorenz, 1996). As for the other EnVar methods, localization needs to be used in high-dimensional systems. It can be implemented using covariant DL or CL (Bocquet, 2016).

Note that, as a variational smoother, the IEnKS can theoretically and within its DAW account for errors that are correlated in time. For instance, it can account for a systematic constant in time bias (Bocquet & Sakov, 2013).

To conclude this brief review of the EnVar set of techniques, the simplified Table 1 summarizes some key properties of a selection of the methods we described.

5 | SPECIAL TOPICS

This section discusses four selected topics related to the application of DA to the geosciences. The issues for the DA and the countermeasures connected to the chaotic nature of the atmosphere and ocean are the content of Section 5.1. Section 5.2 discusses the techniques that allow to deal with non-Gaussian variables in the framework of DA methods originally devised to tackle with Gaussian quantities. The DA for chemical species is the subject of Section 5.3: the topic is of great relevance (pollutant dispersion, accidental nuclear release etc.) and has received much attention in recent years. Finally Section 5.4 describes one example of operational DA, the TOPAZ system for ocean prediction.

5.1 | Dealing with chaotic dynamics

Many natural systems, including atmosphere and ocean, are examples of chaotic dissipative dynamics (Dijkstra, 2013). Among the characteristic features of this class of system, the one having the largest impact on prediction is the extreme sensitivity to initial conditions (Lorenz, 1963). Infinitesimal errors are bound to grow exponentially in the short time (and later

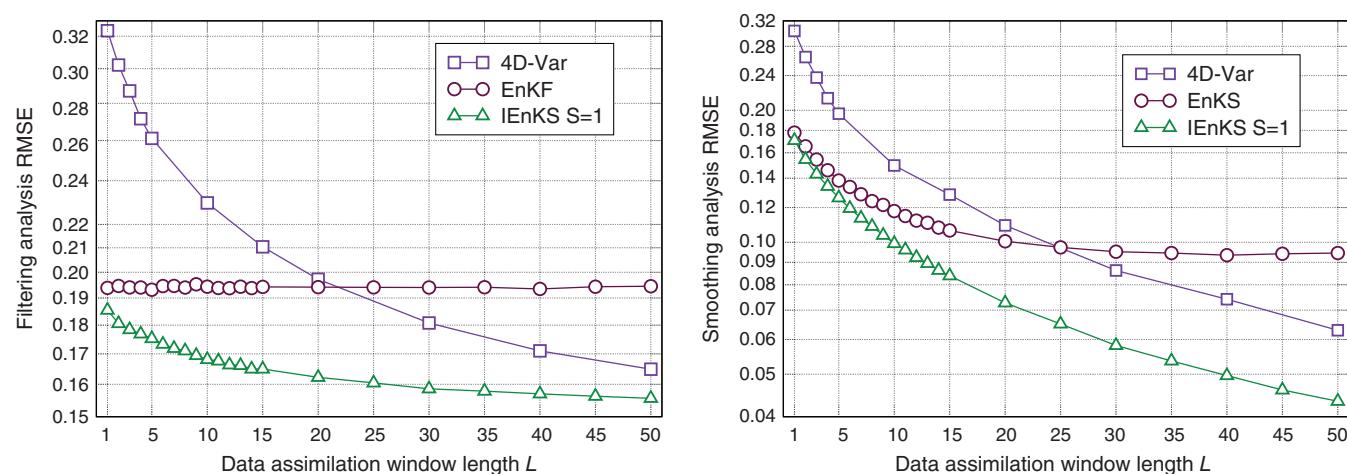


FIGURE 5 Average root mean square error (RMSE) of several data assimilation (DA) methods computed from synthetic experiments with the Lorenz-96 model. The left panel shows the filtering analysis RMSE of optimally tuned ensemble Kalman filter (EnKF), 4DVar, iterative ensemble Kalman smoother (IEnKS) assimilation experiments, as a function of the length of the data assimilation window (DAW). The right panel shows the smoothing analysis RMSE of optimally tuned ensemble Kalman smoother (EnKS), 4DVar, and IEnKS as a function of the length of their data assimilation window. The optimal RMSE is chosen within the window for 4DVar and it is taken at the beginning of the window for the IEnKS. The EnKF, EnKS, and IEnKS use an ensemble of $N = 20$, which avoids the need for localization but requires inflation. The length of the DAW is $L \times \Delta t$, where $\Delta t = 0.05$.

TABLE 1 Comparison of ensemble-variational (EnVar) data assimilation techniques. This table answers the following questions: (a) Is the analysis based on a linear or nonlinear scheme? (b) Is the adjoint of the evolution model required? (c) Is the adjoint of the observation operator required? (d) Is the background flow-dependent? (e) Are the updated perturbations stochastic or deterministic? (f) Are the updated perturbations fully consistent with the analysis, that is, are they a faithful representation of the analysis uncertainty? (g) Is localization of the ensemble analysis required? (h) Is a static background used? To some extent, all algorithms can accommodate a static background; the answer tells whether the published algorithm has a static background. Blank answers correspond to irrelevant questions

Algorithm	Analysis type	Evolution model adjoint required?	Observation operator adjoint required?	Background flow dependence?	Stochastic or Deterministic Perturbations?	Consistent perturbations?	Localization required?	Static background
EnKF	Linear		No	Yes	Both	Yes	Yes	No ^a
3DVar	Nonlinear		Yes	No				Yes
4DVar	Nonlinear	Yes	Yes	No				Yes
EDA with 4DVar	Nonlinear	Yes ^b	Yes ^b	Yes	Sto.	Yes	Part.	Yes ^c
4DEnVar	Linear	No	No	Yes	Sto.	No ^d	Yes	Yes ^c
IEnKS	Nonlinear	No	No	Yes	Det.	Yes	Yes	No ^a
MLEF	Nonlinear		No	Yes	Det.	Yes	Yes	No ^a
4D-ETKF	Linear	No	No	Yes	Det.	Yes	Yes	No ^a

^a But possible with an hybridization of the covariances.

^b The adjoint models could be avoided considering an ensemble of data assimilation (EDA) of 4DEnVar.

^c With an hybridization of the covariances.

^d It depends on the implementation of 4DEnVar; the perturbation are often generated by a concomitant ensemble Kalman filter (EnKF).

saturate), and the rate and directions of such a growth are themselves highly state dependent (Trevisan & Palatella, 2011b). Forecasters have been long aware of this flow-dependent behavior of the forecast uncertainty and “prediction of predictability” (i.e., the prediction of the uncertainty) has become an important issue in NWP (Kalnay & Dalcher, 1987; Lorenz, 1996). The difficulties inherent to the chaotic nature of the climate on one hand, and the social pressure to deliver reliable forecast with associated prediction of the uncertainty, on the other, have motivated a huge bulk of research on predictability (see the review by Vannitsem, 2017, and references therein). In forecasting practice, the problem of predicting uncertainty, thus moving from a “single-realization deterministic forecast” to a “multi-realizations probabilistic forecast,” has been tackled by ensemble prediction schemes (Leutbecher & Palmer, 2008).

Dealing with such a flow-dependent error growth is obviously also a challenge for DA: one must in fact be able to properly track and incorporate this dependency in the DA description of the state estimation error. A situation that is made even worst by the large size of the geophysical models ($m = \mathcal{O}(10^9)$). Nevertheless, the dissipative character (due to friction) of the dynamics, induces an effective dimensional reduction, in the sense that the error dynamics, in deterministic systems, is often confined to a subspace of much smaller dimension, $n_0 \ll m$. This subspace, referred to as the unstable–neutral subspace, is defined as the vector space generated by the backward Lyapunov vectors with non-negative Lyapunov exponents, that is, the space of the small perturbations that are not growing exponentially under the action of the backward dynamics (see Legras & Vautard, 1996, for a topical review). The full phase space can thus be seen as split in a (usually much smaller) unstable–neutral subspace and a stable one (Kuptsov & Parlitz, 2012). For instance, Carrassini, Trevisan, and Ubaldi (2007) have shown how a quasi-geostrophic atmospheric model of $\mathcal{O}(10^5)$ degrees of freedom possesses an unstable–neutral subspace of dimension as small as $n_0 = 24$.

The existence of this underlying splitting of the phase space has enormous impact on the skill of DA with chaotic models. In EnKF-like methods, Carrassini, Vannitsem, Zupanski, and Zupanski (2009); Ng, McLaughlin, Entekhabi, and Ahanan (2011); and Bocquet and Sakov (2014) have studied the relation between the properties of the model dynamics and the performance of DA, and revealed how the knowledge about the former can be used in the design of the square-root EnKF and EnKS: for purely deterministic models, the minimum ensemble size required to achieve good performance must be at least as big as the number, n_0 , of non-negative Lyapunov exponents (Bocquet & Sakov, 2014; Carrassini et al., 2009). Along similar lines, but for variational DA, Pires, Vautard, and Talagrand (1996) showed that the relative projection of the state estimation error on the stable and unstable subspaces depend on the length of the DAW (see Section 4.5.3): for long DAW the stable components are all damped and the error is essentially confined on the unstable subspace.

These results have recently been corroborated by a proper mathematical understanding of the behavior of the KF and KS (c.f., Sections 3.1.1 and 3.1.2) for linear unstable systems in the absence of model noise. It has been proven analytically that the span of the error covariance matrices of the KF and KS tends asymptotically to the unstable–neutral subspace (Gurumoorthy, Grudzien, Apte, Carrassini, & Jones, 2017), independently from the initial condition (Bocquet et al., 2017), and

that the rate of such convergence is faster for the KS (Bocquet & Carrassi, 2017). Remarkably, Bocquet and Carrassi (2017) have shown that, even in the case of a four-dimensional EnVar method (the IEnKS, c.f., Section 4.5.4) applied to a noiseless nonlinear system, the full alignment of the ensemble anomaly subspace with the unstable–neutral subspace is accompanied by the maximum reduction of the state estimation error. Figure 6 illustrates this mechanism for the EnKF using the Lorenz-96 model (Lorenz & Emanuel, 1998).

These results indicate that, in deterministic chaotic systems, the unstable–neutral subspace fully describes (in an average, asymptotic, sense) the uncertainty in the state estimate. This fact is at the basis of a class of DA algorithms known as *assimilation in the unstable subspace* (AUS) pioneered by Anna Trevisan and collaborators (Palatella, Carrassi, & Trevisan, 2013; Trevisan & Uboldi, 2004), in which the unstable–neutral subspace (or a suitable numerical approximation of it) is explicitly used in the DA scheme to parametrize the description (both temporally and spatially) of the uncertainty in the state estimate. Given that usually $n_0 \ll m$, AUS represents an interesting alternative to operate DA efficiently on high-dimensional system (large m) by monitoring only n_0 unstable–neutral directions. Applications to large atmospheric (Carrassi, Trevisan, Descamps, Talagrand, & Uboldi, 2008) and oceanic models (Uboldi & Trevisan, 2006) have confirmed its capabilities. The AUS paradigm has been successfully incorporated in both the EKF (Trevisan & Palatella, 2011a) and the 4DVar (Trevisan, D'Isidoro, & Talagrand, 2010) and its extension to cope with the deviations from the linear error dynamics and perfect model assumptions on which it relies, has been studied recently by Palatella and Trevisan (2015) and Grudzien, Carrassi, and Bocquet (2018b), respectively.

5.2 | Dealing with non-Gaussian variables

There are cases for which the linear analysis is suboptimal. These arise when either the state variables, observations, or model parameters are not Gaussian distributed. We consider first the case in which the distribution of at least one variable is not Gaussian, for example, it exhibits a clear skewness or kurtosis, but the pdf remains continuous.

Ignoring the non-Gaussian nature of a variable will normally make the least squares estimator suboptimal (c.f., Appendix A): either by systematic biases, by underestimating the correlation length scales (Brankart et al., 2012), or more visibly by returning nonphysical values (for example, negative concentrations of tracer variables). In a DA framework, the state variables (called the “control variable” in this context) need not be exactly defined as in the forward model: nonlinear transformations are allowed and will change the outcomes of the minimization problem (Bertino, Evensen, & Wackernagel, 2003).

Transformations of the variables have been practiced extensively in the geostatistical literature under the term Gaussian anamorphosis (also known as “normal-score transform”): a nonlinear function is applied to the cumulative pdf in order to make it Gaussian (Chilès & Delfiner, 2012). In DA, the Gaussian anamorphosis function can either be applied to the state variables $\tilde{\mathbf{x}} = \phi(\mathbf{x})$ or to the observations $\tilde{\mathbf{y}} = \varphi(\mathbf{y})$ or the model parameters or to all of them at the same time. The prerequisite is that the transformed variables $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ should be “more Gaussian” and better suited to a linear analysis update than the original ones. The anamorphosis function must be strictly increasing so that backward transformations from the Gaussian values to the “natural” pdf can be given by ϕ^{-1} and φ^{-1} . Since $E(\phi^{-1}(\tilde{\mathbf{x}})) \neq \phi^{-1}(E(\tilde{\mathbf{x}}))$, Monte Carlo techniques are often employed to

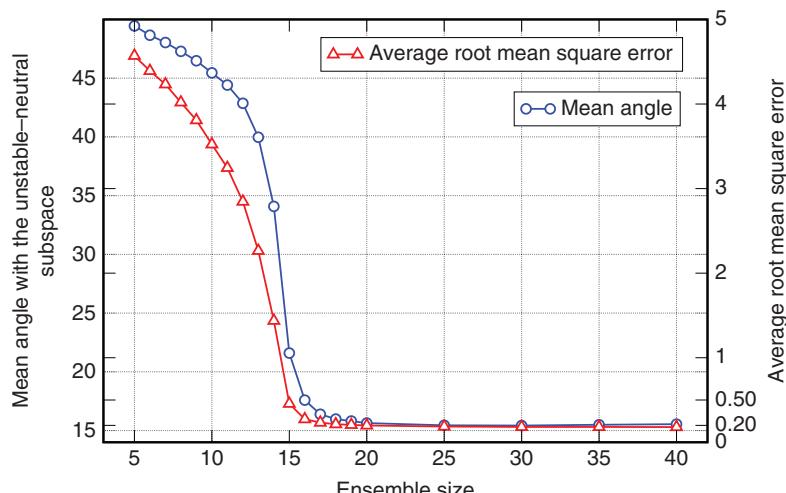


FIGURE 6 Time- and ensemble-averaged angle (in degree) between an anomaly from the ensemble Kalman filter (EnKF) ensemble and the unstable–neutral subspace as functions of the ensemble size N (left y-axis) and corresponding time-averaged root mean square error of the EnKF (right y-axis). The numerical experiments are performed on the Lorenz-96 model with $m = 40$ variables (Lorenz & Emanuel, 1998) and the data assimilation (DA) setup is $\mathbf{H} = \mathbf{I}_d$, $\mathbf{R} = \mathbf{I}_d$, observational frequency $\Delta t = 0.05$ and error variance $\sigma = 1$.

obtain unbiased statistics. This makes the anamorphosis particularly convenient to apply in conjunction with ensemble techniques (Bertino et al., 2003).

In practice, the stochastic models (13) and (14) are valid for the transformed $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$, so that the EnKF and EnKS equations can be reused with $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ in replacement of \mathbf{x} and \mathbf{y} . Equations (32) and (38) become

$$\mathbf{E}^{f/a} = \left(\tilde{\mathbf{x}}_{[1]}^{f/a}, \tilde{\mathbf{x}}_{[2]}^{f/a}, \dots, \tilde{\mathbf{x}}_{[N]}^{f/a} \right) \in \mathbb{R}^{m \times N}, \quad (65)$$

$$\mathbf{Y}^o = \left(\tilde{\mathbf{y}}_{[1]}, \tilde{\mathbf{y}}_{[2]}, \dots, \tilde{\mathbf{y}}_{[N]} \right) \in \mathbb{R}^{d \times N}. \quad (66)$$

The observations are preferably transformed before perturbing the observations, since the perturbations are more conveniently applied in the Gaussian space (even when using a deterministic version of the EnKF, the \mathbf{R} matrix characterizes more adequately Gaussian observation errors). After the analysis, the inverse transformation ϕ^{-1} is applied to each member of the analyzed ensemble $\mathbf{x}_i^a = \phi^{-1}(\tilde{\mathbf{x}}_i^a)$ before the next propagation step. If one variable is both an observation and a model state variable, then the same transformation should be applied to both so that the observation operator \mathbf{H} keeps a simple form (Amezcu & Van Leeuwen, 2014). Alternatively, using independent transformations can be a practical way to correct for biases (Lien, Kalnay, Miyoshi, & Huffman, 2016; Lien, Miyoshi, & Kalnay, 2016).

There are infinitely many of strictly increasing functions: choosing a good anamorphosis function is therefore an open question. When a set of observations is available or a free model run, it is possible to approximate the anamorphosis function based on the histogram of available data, using for example a piece-wise linear fit (Simon & Bertino, 2009). Adaptive types of anamorphosis functions have also then been explored (Simon & Bertino, 2012), in which the anamorphosis function is fitted to the values present in the ensemble forecast at the assimilation time. These have proven advantageous in twin experiments (Li, Zhou, Hendricks Franssen, & Gómez-Hernández, 2012; Simon & Bertino, 2012), but their stability remains to be tested in real cases, where the true density is unknown and the results can become excessively sensitive to the choice of the tails of the distribution (Simon & Bertino, 2012). In practical cases, simple analytical functions are sometimes preferred like the exponential, gamma, or logit functions, which can be supported theoretically knowing the nature of the problem at hand (Gharamti, Tjiputra, et al., 2017; Simon, Samuels, Bertino, & Mouysset, 2015).

One immediate benefit of the Gaussian anamorphosis is the ability to impose positivity constraints in the EnKF, which is useful for example in ecosystem models (Simon & Bertino, 2009). The use of an anamorphosis function has been demonstrated with the EnKF in hydrological models (Li et al., 2012; Zhou, Li, Hendricks Franssen, & Gómez-Hernández, 2012), NWP models (Kotsuki, Miyoshi, Terasaki, Lien, & Kalnay, 2017; Lien, Kalnay, & Miyoshi, 2013; Lien, Kalnay, et al., 2016; Lien, Miyoshi, & Kalnay, 2016), in joint parameter-state estimation in ocean ecosystem models (Gharamti, Samuels, et al., 2017; Gharamti, Tjiputra, et al., 2017; Simon & Bertino, 2012), for the evaluation of covariance length scales in various ocean modeling applications (Brankart et al., 2012) and with moderate success in a coupled ice–ocean model (Barth et al., 2015). It has been as well introduced in a snow model with an ensemble smoother (Aalstad, Westermann, Chuler, Boike, & Bertino, 2018).

One limitation of the Gaussian anamorphosis as described above is that it does not consider the multi-Gaussian case: only the marginal distributions of the variables are transformed independently from each other. Although this does not guarantee that the joint distributions become multi-Gaussian, the experience still shows that the multi-Gaussian properties are generally improved after transforming the marginals (Amezcu & Van Leeuwen, 2014; Wackernagel, 2003). The anamorphosis framework can, however, be further extended by the use of copulas (Schölzel & Friederichs, 2008).

A more serious limitation in many practical cases is that the Gaussian anamorphosis does not lend itself to discontinuous pdf. These are better addressed in the framework of truncated Gaussians: a threshold is defined on the Gaussian distribution, and all values exceeding the threshold are reset to its precise value, generating an atom of probability distribution. Truncated Gaussians were first applied in optimal interpolation, by introducing a Lagrange parameter in the minimization (Thacker, 2007), only in the cases when assimilated values exceeded the threshold, leading to a two-step interpolation process (detect and constrain). The issue of sampling a multivariate truncated Gaussian in the framework of an EnKF was then explored applying a Gibbs sampler (Lauvernet et al., 2009), under the simplifying assumption of small truncations. Another approach is the use of Quadratic Programming (Janjić, McLaughlin, Cohn, & Verlaan, 2014), more general in the sense that it can accommodate nonlinear constraints. An alternative and inexpensive approach is to “moderate” the strength of DA in order to avoid reaching the discontinuities of the distribution. This has been proposed for assimilation into an ocean model in isopycnic coordinates, for which the values of layer thickness must be positive (Wang, Counillon, & Bertino, 2016). Combinations of Gaussian anamorphosis and truncations have been successfully tested by (Lien et al., 2013; Lien, Miyoshi, & Kalnay, 2016) by setting the Gaussian value of the threshold to the median of the out-of-range Gaussian pdf (zero precipitations in their case).

The above examples are extending the Gaussian framework introduced in Section 3.1 and extend as well the use of the EnKF and EnKS methods. The Gaussian anamorphosis can be included at almost no additional costs, but the costs of the truncated Gaussian methods are potentially a limitation in realistic cases (a 500% increase of the analysis step is reported with the Gibbs sampler (Lauvernet et al., 2009). Still, these methods benefit from the advantages of the Gaussian framework in high dimensions and represent attractive alternatives to more general Bayesian methods that suffer from the curse of dimensionality (see Section 6.1).

5.3 | DA for chemicals

There is a huge literature on the use of DA for chemical constituents of the atmosphere. There are several quasi exhaustive reviews about the advances in this field, such as Carmichael et al. (2008); Zhang, Bocquet, Mallet, Seigneur, and Baklanov (2012); and Bocquet et al. (2015). We here briefly explain what chemical DA is, what the motivations for using DA in this field are, and, finally, which methods are used.

The specific set of chemical reactions that prevails in the atmosphere depends on the species but also on the considered spatial and temporal scales. For instance, one typically distinguishes between global atmospheric chemistry and air quality. The former is concerned with species transported and reacting over long distances and within the whole column of the atmosphere, whereas the latter is more concerned with the complex chemistry within the boundary layer (about 1 km from the ground) at regional scales. As an example of species, greenhouse gases are global pollutants with long life spans considered in the first type of study, even though regional studies of their transport are also of interest. Another example is ozone, which, in the lower troposphere, is studied and forecast in air quality as a harmful pollutant whereas, in the stratospheric layer, or when transported in between continents, is the focus of global studies. Besides the global and regional scales, DA is also nowadays used for chemical studies at urban scale.

Chemistry and transport models are fundamentally multivariate with many species to be accounted for (about a hundred for gaseous species, hundreds when considering aerosols). This also significantly increases the dimensionality of the models compared to atmospheric and ocean models independently. That is why the choice of the control variables (c.f., Section 3.2) is critical in the feasibility of chemical DA.

One also distinguishes between online and offline models. Offline models (also called CTM for *chemical transport model*) only consider emissions and uptakes, chemistry, loss processes and transport driven by meteorological fields that are supposed to be provided. Online models (also called CCMM for *coupled chemistry meteorology model*) couple meteorology with chemistry and transport. CCMMs are therefore much more costly but allow for two-way coupling between meteorology and chemistry. Most of chemical DA studies are performed offline.

One major difference with atmospheric and ocean models is that the dynamics of CTMs are mostly stable. In practice, two model trajectories with distinct initial conditions will coalesce, quite differently from chaotic geofluid dynamics (Haussaire & Bocquet, 2016; Vannitsem, 2017). Hence, the initial conditions are rarely the most important control variables in chemical DA. Quite often, the species emissions are the most influential input parameters, if not control variables. More generally, model error is more influential for chemical DA than for geofluids DA. One consequence of this dynamical stability is that air quality forecasts are rarely useful beyond 48 hr, even with the help of DA. Another consequence is the importance of all other parameters and input of these models and their uncertainty and hence of model error.

For these many reasons, but also to serve the enforcement of regulations, chemical DA is used for (a) chemical forecast, nowcasting and possibly reanalysis, (b) inverse modeling of species emission and uptake, and (c) parameter estimation, such as the boundary conditions, constants of the physical and chemical parametrizations.

The methods used in chemical DA are largely inspired by meteorological DA. Optimal interpolation (Kalnay, 2002) has been largely used from the very beginning. Since the beginning of the 2000s, 4DVar has also been used, which can advantageously be adopted for the simultaneous estimation of inputs and model parameters (Elbern, Strunk, Schmidt, & Talagrand, 2007). Methodological advances in chemical DA also had some influence, with the early development of reduced Kalman and ensemble methods in the 1990s (Constantinescu, Sandu, Chai, & Carmichael, 2007a, 2007b; Hanea, Velders, Segers, Verlaan, & Heemink, 2007; Wu, Mallet, Bocquet, & Sportisse, 2008). Four-dimensional EnVar have also been tested in this field (Emili, Gürol, & Cariolle, 2016; Haussaire & Bocquet, 2016).

5.4 | An example of operational DA and reanalysis: The TOPAZ integrated ice–ocean system

The concept of operational oceanography has gradually arisen in the 1990s under the joint pressure from its users in the navy, the authorities in charge of emergency response (oil spills, search and rescue operations) and the offshore oil and gas industry. Ocean monitoring was becoming possible from remote sensing and increasing in situ observations to fulfill the needs of both the operational and the climate research communities and the development of ocean models was accelerated. The conditions for

ocean DA were then ripe and a community was formed around the international Global Ocean Data Assimilation Experiment (GODAE, now GODAE-OceanView, www.godae-oceanview.org). The EnKF, originally introduced in a QG model, was then extended to a multilayer isopycnal model (Miami Isopycnal Coordinate Model, MICOM) and then its offspring in generalized vertical coordinates HYCOM (HYbrid Coordinate Ocean model Bleck, 2002). The EnKF (Evensen, 2009b), SEEK filter (Pham et al., 1998) as well as the EnKS (Evensen, 2009b) were applied successfully to this model to assimilate ocean remote sensing observations from satellite altimeters and sea surface temperatures in a North Atlantic configuration called DIADEM (Brusdal et al., 2003). In parallel, the assimilation of sea ice concentrations in a coupled ice–ocean model was demonstrated (Lisæter, Rosanova, & Evensen, 2003) as well as the assimilation of remotely sensed ocean color in a coupled physical–ecosystem model (Natvik & Evensen, 2003), both with the EnKF. These experiments were all using similar settings of the EnKF: same number of members, same sources, and statistics of model errors and were therefore encouraging signs that a single integrated system could monitor the ocean, the sea ice and the ocean ecosystem consistently, with the uncertainties of the three components represented by the same ensemble. Such an integrated system for the North Atlantic and the Arctic Oceans was then built and named TOPAZ, including downscaling capabilities to forecast the coastal zones at higher resolution. An illustration showing a typical output from the TOPAZ system is given in Figure 7 that shows the sea surface temperature and sea ice concentration over the Atlantic Ocean. The TOPAZ system is run at a horizontal resolution of 10 km that makes it eddy-permitting and of very large dimension. The scales resolved by the model are visible on the figure. The model resolution is gradually increasing from one configuration to the next one, although without much incidence on the DA source code.

The TOPAZ system has been set in near real-time forecasting mode in January 2003, initially using the perturbed observations EnKF (Evensen, 2003) (c.f., Section 4.1), and then moved on to the DEnKF (Sakov & Oke, 2008b) (c.f., Section 4.2) in January 2010.

The HYCOM ocean model uses isopycnic coordinates in the deep stratified ocean, the following state vector is then the closest possible to the original model prognostic variables:

- Three-dimensional variables, defined for each hybrid vertical layer: layer thickness, u- and v-components of the current velocity, temperature, and salinity.
- Two-dimensional variables: barotropic pressure, u- and v-components of the barotropic velocity, and sea ice characteristics (ice concentration, ice thickness, snow depth).
- Static parameters, two-dimensional bias fields for mean sea level amplitude and mean sea surface temperature.
- If the ecosystem module is coupled:
 - Three-dimensional nutrients concentrations (nitrate, phosphate, and silicates in North Atlantic and Arctic regions)
 - Three-dimensional phytoplankton (diatoms and flagellates) and zooplankton species (meso- and microzooplankton).
 - Three-dimensional detritus variable, as well as oxygen and biogenic silicate.
 - Two-dimensional maps of model parameters (different plankton loss rates).

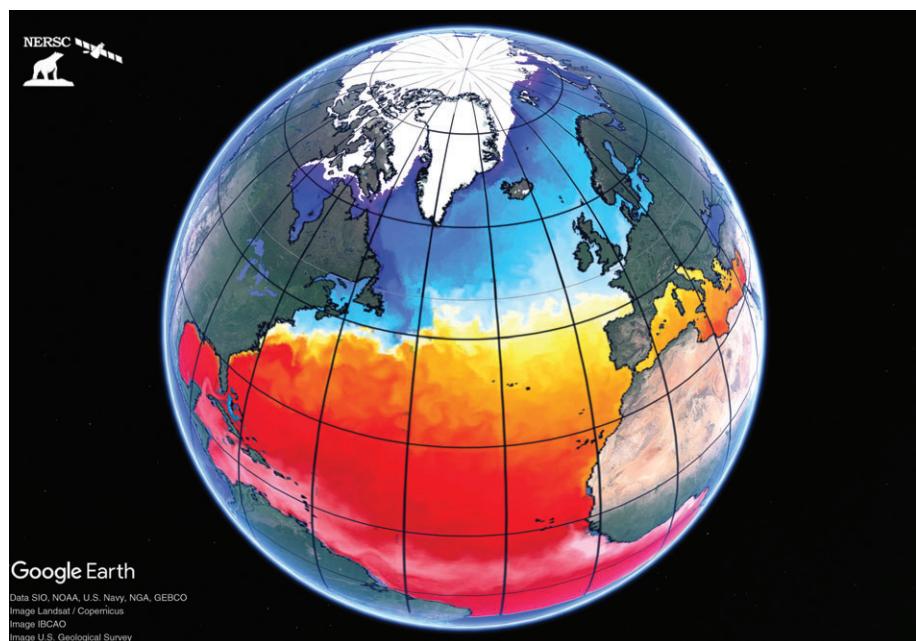


FIGURE 7 Example of sea surface temperature (in color) and sea ice concentration (in white) real-time analysis by the TOPAZ system on November 28, 2009

Recalling that the EnKF analysis is conserving linear equilibrium, the above definition of the state vector in generalized (isopycnic) vertical coordinates is advantageous for preserving the geostrophic balance, which relates linearly the velocity vector averaged over one ocean layer to the gradient of the layer thickness (Evensen, 2003). The use of isopycnic coordinates in the state vector has also the advantage of reducing diapycnal mixing during the linear update in Equation (45) or (55) since the analyzed temperature and salinity tracers are the results of combinations of forecast temperature and salinity from different members but at the same density. The application of the EnKF in the MICOM isopycnic model has also proven to make the assimilation of surface temperature properties more efficient in the context of a climate reanalysis (Counillon et al., 2016).

On the downside, not all the above state variables are Gaussian distributed, so according to the discussion in Section 5.2, biological variables are log transformed (Simon et al., 2015). However variables with discontinuous distributions (sea ice variables, ocean layer thickness) are corrected in case nonphysical values occur after the analysis (Sakov, Counillon, et al., 2012), and TOPAZ is thus exposed to assimilation biases. Yet, the experience has not revealed strong biases.

TOPAZ has used a local analysis (c.f., Section 4.4.1) ever since its inception, a global analysis having proven utterly incapable of controlling millions of state variables with only 100 members. The inflation is implicit in the formulation of the DEnKF (Sakov & Oke, 2008b) and additionally in the form of a factor applied to the **R** matrix, used only in the update of the anomalies (Sakov, Counillon, et al., 2012). The use of a constant multiplicative inflation (c.f., Section 4.4.2) of the forecast anomalies has proven problematic in the case of the Arctic due to the inhomogeneous observation coverage, even with a small inflation of 1%.

The present TOPAZ4 physical system with its horizontal resolution of about 12 km and 28 hybrid vertical layers counts about 80 million unknowns. The data assimilated in a weekly cycle count the along-track sea level anomalies from satellite altimeters, the interpolated sea surface temperatures and sea ice concentrations from satellites as well, the Lagrangian sea ice drift from continuous maximum cross correlation of scatterometer data, and the temperature and salinity profiles from various in situ platforms, including the Argo autonomous buoys and Ice-Tethered Profilers. This amounts to about 400 thousands observations after averaging high-resolution observations onto the model grid. The altimeter and the sea ice drift data are assimilated asynchronously (Sakov, Evensen, & Bertino, 2010), accounting for time-delayed covariances.

The computer costs of the TOPAZ4 system are mainly related to the propagation of the 100 members (1,200 CPU hours per week, but embarrassingly parallel as the 100 members are submitted independently and the HYCOM model is itself parallel, running on 144 CPUs), the analysis step takes 20 hr only, and is also parallelized because all local analysis updates are independent from each other.

The TOPAZ4 reanalysis (Xie et al., 2017) has been conducted over 24 years (covering the altimeter era 1991–2015) and thus 1,250 assimilation steps. During this time, the Arctic observing system has undergone large changes with for example the International Polar Year (IPY, 2007–2009) during which several Ice-Tethered Profilers have been deployed in the Central Arctic, measuring temperature and salinity profiles in areas previously unobserved. The large increase of observations during the IPY is visible in the time series of Figure 8 and does, as expected, reduce the bias and RMSE, but the latter increases again during the 6 months that follow the end of the IPY. This implies that the special observation efforts done on the occasion of the IPY should be the rule rather than the exception for the sake of monitoring the Arctic Ocean. The diagnostics shown for other variables indicate that the reanalysis is stable all through the 1,250 assimilation cycles (Xie et al., 2017) and how much each observation type contributes to the minimization.

The coupled physical-biological reanalysis with assimilation of surface chlorophyll estimates from remotely-sensed ocean color data is a much more difficult endeavor than the physical reanalysis, both on technical and scientific levels (Simon et al., 2015).

6 | WHERE WE ARE AND WHERE WE GO: A LOOK AT THE PERSPECTIVES

DA is nowadays a key ingredient of the atmospheric and oceanic prediction machinery and substantial computational power and economical resources are allocated to its maintenance and updates. What are the challenges DA is facing at this present time? Which solutions are being proposed to cope with the requirements of present days climate research? Figure 9 illustrates this quest by displaying the required DA approach as a function of the model resolution (x-axis) and of the prediction horizon (y-axis).

The constant increase of the numerical model resolution, that is, reducing the model grid size, implies resolving more and more small scale processes (e.g., convection or turbulence) that are often inherently nonlinear and non-Gaussian. The transition toward high-resolution models must thus be accompanied by corresponding DA developments where the Gaussian and linear assumptions are relaxed (see Yano et al., 2018). This identifies one of the more challenging and active lines of research in DA at the current time: fully Bayesian methods. The horizontal dashed line in Figure 9 illustrates the DA method required

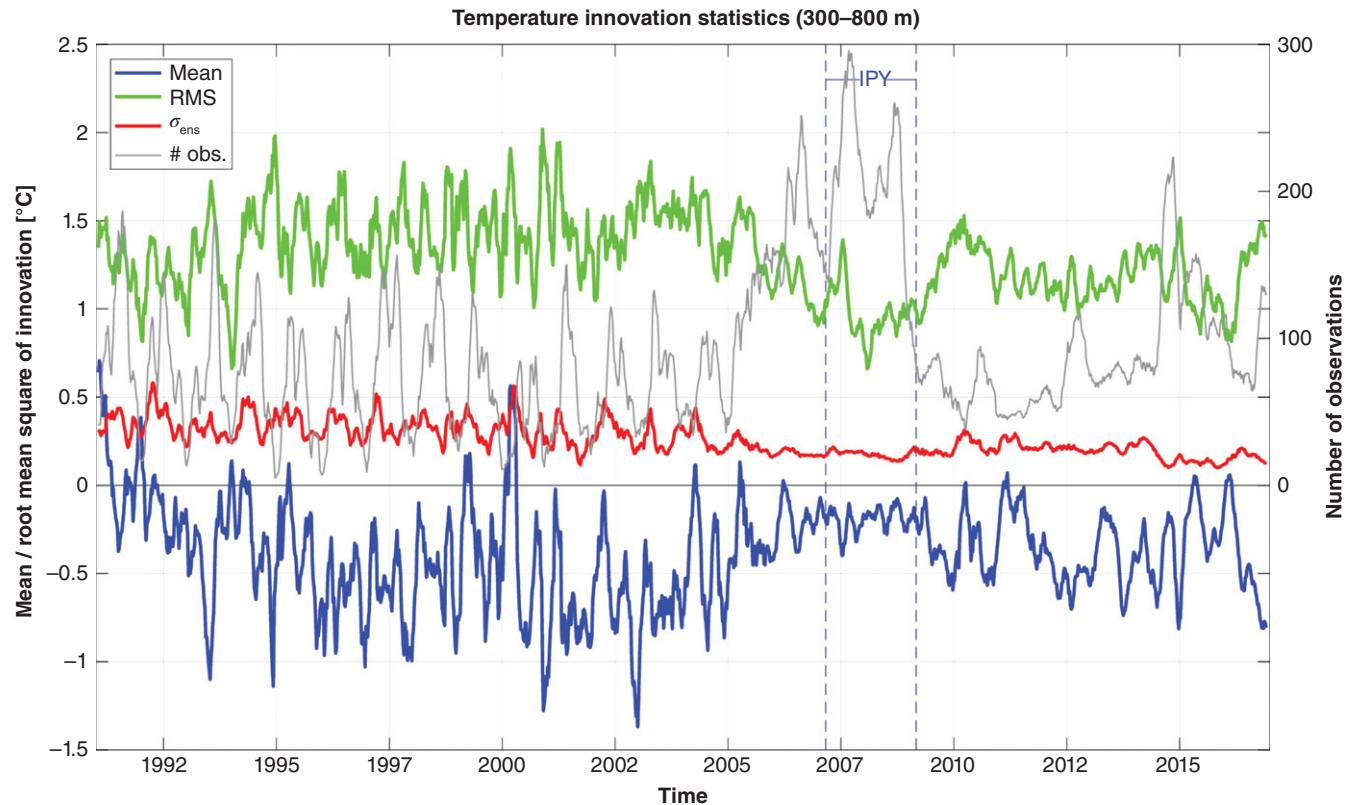


FIGURE 8 Time series of data assimilation diagnostics across the 24-year reanalysis for all temperature profiles in the depths 300–800 m in the whole Arctic. The blue line is the average of all innovations, the green line is the related standard deviation (root mean square error, RMSE), the red line is the ensemble spread and the gray line is the number of temperature observations. The IPY was officially taking place between the two vertical lines, but the observations were increasing progressively

features as a function of the model grid size; PFs, the subject of Section 6.1, are fully Bayesian and nonlinear methods and are placed at the top of the algorithms hierarchy in Figure 9.

Together with the increase in model resolution, the current era is also characterized by a growing interest in long-term forecasts. Indeed, beyond the meteorological time horizon of 2 weeks, predictions over seasonal-to-decadal (s2d) time scales potentially bear higher societal relevance: they can guide adaptation to near-term climate change and related risks (Doblas-Reyes, García-Serrano, Lienert, Biescas, & Rodrigues, 2013). Such long-term predictability arises from the interactions between the atmosphere and the more slowly varying components of the climate system, such as the ocean, land surface, and cryosphere, so that predictions are issued using fully coupled models (Earth system models, ESMs). The fruitful use of DA with coupled models necessitates the development of adequate CDA methods that allow for a consistent and balanced propagation of the informational content of the observations across all model components (see vertical dashed line in Figure 9). As will be discussed in Section 6.2, this is not straightforward with standard DA methods and substantial efforts are currently being undergone to develop CDA strategies.

In addition to the increase in numerical model resolution and of the prediction time horizon, we are also experiencing a dramatic growth and refinement of the observations suppliers. The Earth is now observed over a wide range of spatial and temporal scales, thanks to an increasingly wider variety of sustained observing systems, which includes satellites, but also new, not stationary, ocean measurements such as floats, drifters, and more recently gliders (see Kuznetsov, Ide, & Jones, 2003). The assimilation of data derived from instruments that follow the flow has come to be known as Lagrangian DA (see Ide, Kuznetsov, & Jones, 2002; Nodet, 2006). These methods have gained popularity as they offer a suitable way to consistently incorporate modern observation of the ocean, and have experienced a flourishing stream of improvements in the last decade, driven by addressing two main challenges: the use of indirect measurements of the state variables (see Salman, Kuznetsov, Jones, & Ide, 2006) and the inherent nonlinear character of the underlying dynamics (Apte & Jones, 2013). More recently, the Lagrangian dimension of the DA problem has also involved the model component, and not just the data, with the appearance of numerical models discretised on a spatiotemporal varying mesh (see Rampal, Bouillon, Ólason, & Morlighem, 2016, for an example of Lagrangian sea ice model). This feature represents another methodological challenge for DA, which is no longer only demanded to update the value of the physical variables on the grid points, but also the physical location of grid itself (see Bonan, Nichols, Baines, & Partridge, 2017).

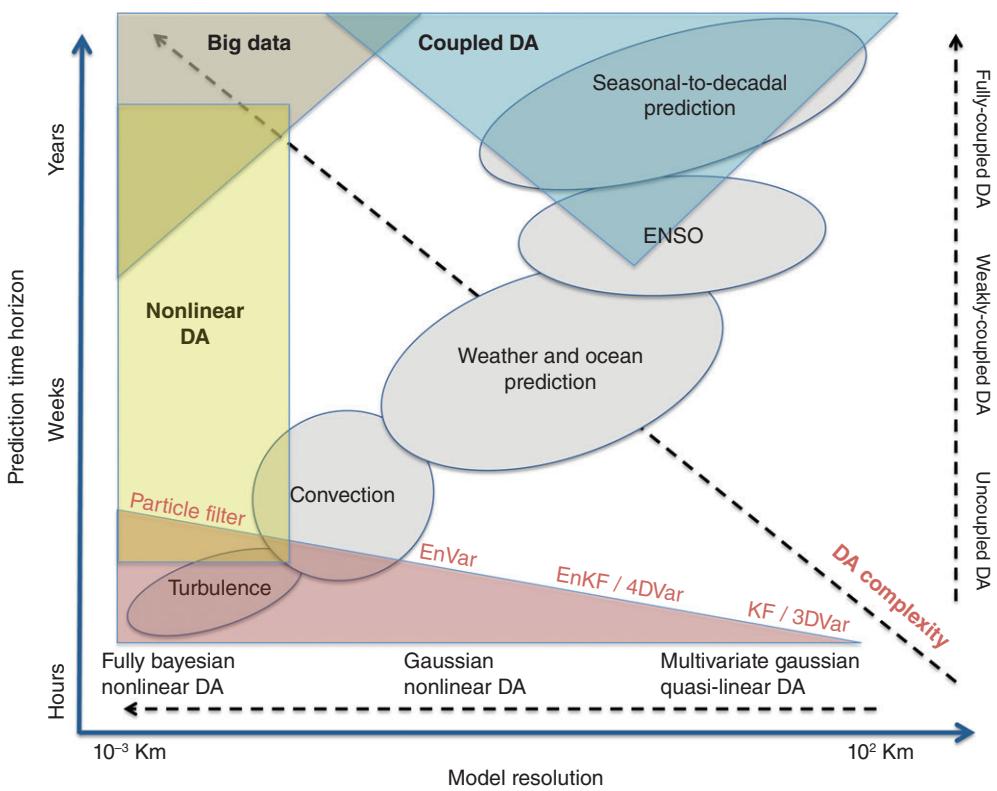


FIGURE 9 Required data assimilation (DA) method versus model resolution and prediction time horizon; examples of corresponding natural phenomena are also shown for illustrative purposes. The degree of sophistication of the DA grows commensurately with the increase in prediction time horizon and the decrease of the model grid size

6.1 | Bayesian DA: PFs

From the Bayesian standpoint, the state estimation problem is best formulated as Equation (3):

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \quad (67)$$

We have seen in Section 2.2 how to exploit and formulate the sequential estimation problem through the conjunction of Bayes' rule and the Chapman–Kolmogorov equations.

Yet, these formulae were claimed to be impractical, at least for high-dimensional models, and we moved on to solutions based on Gaussian approximations. The algebra involved with these approximations is numerically demanding with inversions of covariance matrices in the analysis step, which contrasts with the apparent simplicity of Equation (67).

A direct and brute force approach to the DA problem would be to use Monte Carlo methods and draw N samples from Equation (3). The hope is that in the asymptotic limit, that is, $N \rightarrow \infty$, one would properly estimate the conditional density $p(\mathbf{x}|\mathbf{y})$. In a sequential context, this approach is called the PF, or sequential Monte Carlo. In the following, we show how to justify and implement such simple algorithm but that it unfortunately comes with its own drawbacks.

6.1.1 | Bootstrap PF and resampling

In the following, most pdfs—mainly designated by p —will be identified by their arguments. This is not mathematically rigorous but the notation has the merit of offering sleek expressions.

As a first step, let us see how to solve the filtering problem sequentially with the PF. The forecast pdf $p(\mathbf{x}_k|\mathbf{y}_{k-1})$ from t_{k-1} to t_k is assumed—and this is the main approximation at finite N , that is, with a limited ensemble—to be of the form

$$p(\mathbf{x}_k|\mathbf{y}_{k-1}) = \sum_{n=1}^N \omega_{k-1}^n \delta(\mathbf{x}_k - \mathbf{x}_k^n), \quad (68)$$

which is an empirical distribution with delta-Dirac masses δ , each one centered on an \mathbf{x}_k^n . In Equation (68) the notation \mathbf{y}_k stands for the, ideal, infinite sequence of observations from the far past until time t_k , $\mathbf{y}_k = \{\mathbf{y}_k, \mathbf{y}_{k-1}, \dots, \mathbf{y}_{-\infty}\}$. The $\{\mathbf{x}_k^n\}_{n=1,\dots,N}$ are the particles (i.e., the ensemble members or the samples) and the $\{\omega_{k-1}^n\}_{n=1,\dots,N}$ are positive weights

attached to the particles; \mathbf{x}_k^n is a shortcut for the particle \mathbf{x}_n , as denoted in Section 4, at time t_k . Since Equation (68) represents a pdf, the weights need to be normalized to one, $\sum_{n=1}^N \omega_{k-1}^n = 1$. Hence, those weights tell how probable a particle is.

The analysis at t_k consists in the assimilation of \mathbf{y}_k using Bayes' rule:

$$\begin{aligned} p(\mathbf{x}_k | \mathbf{y}_{k:}) &= \frac{p(\mathbf{y}_k | \mathbf{x}_k)}{p(\mathbf{y}_{k:})} p(\mathbf{x}_k | \mathbf{y}_{k-1:}) \\ &= \sum_{n=1}^N \omega_{k-1}^n \frac{p(\mathbf{y}_k | \mathbf{x}_k)}{p(\mathbf{y}_{k:})} \delta(\mathbf{x}_k - \mathbf{x}_k^n) \\ &\propto \sum_{n=1}^N \omega_{k-1}^n p(\mathbf{y}_k | \mathbf{x}_k^n) \delta(\mathbf{x}_k - \mathbf{x}_k^n). \end{aligned} \quad (69)$$

Equation (69) suggests to define the updated weights as

$$\omega_k^n \propto \omega_{k-1}^n p(\mathbf{y}_k | \mathbf{x}_k^n), \quad (70)$$

where the proportionality factor can be determined afterward by the condition that the normalized updated weights should sum up to 1. Hence, the analysis elegantly sums up to a simple multiplication of the weights by the likelihood of each particle.

The forecast step amounts to applying Chapman–Kolmogorov Equation (9):

$$\begin{aligned} p(\mathbf{x}_{k+1} | \mathbf{y}_{k:}) &= \int d\mathbf{x}_k p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{k:}) \\ &= \sum_{n=1}^N \int d\mathbf{x}_k p(\mathbf{x}_{k+1} | \mathbf{x}_k) \omega_k^n \delta(\mathbf{x}_k - \mathbf{x}_k^n) \\ &= \sum_{n=1}^N \omega_k^n p(\mathbf{x}_{k+1} | \mathbf{x}_k^n). \end{aligned} \quad (71)$$

If the model, defined by its transition density $p(\mathbf{x}_{k+1} | \mathbf{x}_k)$, as in Equation (4), is deterministic, the forecast pdf $p(\mathbf{x}_{k+1} | \mathbf{y}_{k:})$ is, as the update pdf $p(\mathbf{x}_k | \mathbf{y}_{k:})$, in the form of a delta-Dirac density; otherwise it is not so. To obtain $p(\mathbf{x}_{k+1} | \mathbf{x}_k)$ as a delta-Dirac pdf, which would be necessary in order to cycle the algorithm, we need to sample from Equation (71).

One solution to obtain a delta-Dirac pdf from Equation (71) consists, for each particle n , in sampling \mathbf{x}_{k+1}^n from the density $p(\mathbf{x}_{k+1} | \mathbf{x}_k^n)$ by simply forecasting \mathbf{x}_k^n from t_k to t_{k+1} using the (possibly stochastic) model associated with the transition density $p(\mathbf{x}_{k+1} | \mathbf{x}_k)$, which yields:

$$p(\mathbf{x}_{k+1} | \mathbf{y}_{k:}) \approx \sum_{n=1}^N \omega_k^n \delta(\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^n). \quad (72)$$

Alternatively, one can sample particle n from Equation (71) by randomly selecting one of the particles, say n' , with a probability proportional to its importance weight. Once $\mathbf{x}_k^{n'}$ is selected, one can forecast it and define $\mathbf{x}_{k+1}^{n'}$ using, as in the previous case, the model associated with the transition density $p(\mathbf{x}_{k+1} | \mathbf{x}_k^{n'})$, which yields

$$p(\mathbf{x}_{k+1} | \mathbf{y}_{k:}) \approx \frac{1}{N} \sum_{n=1}^N \delta(\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^n). \quad (73)$$

The first option corresponds to the *bootstrap* PF (Gordon, Salmond, & Smith, 1993) or sequential importance sampling (SIS) PF. The second option adds, before forecasting, a *resampling* step that uniformly resets the weights to N^{-1} ; it is called sequential importance resampling (SIR) PF and is depicted in Figure 10. There are several ways to resample the particles given their weights. Popular resampling schemes include multinomial resampling, residual resampling, and stochastic universal (or systematic) resampling, the latter minimizing the sampling noise introduced in the procedure (Douc & Cappé, 2005).

6.1.2 | Importance sampling with a proposal

In statistics, most smart sampling strategies usually come with the possibility to draw samples from an ancillary, more accessible, distribution, naturally called the proposal distribution. However, the weights of these drawn particles must be corrected so that their empirical distribution is an unbiased estimator of the targeted distribution. PFs can also crucially benefit from this approach. Since this subject is more technical, the reader not interested in the details of its implementation is invited to move on to the next subsection Section 6.1.3.

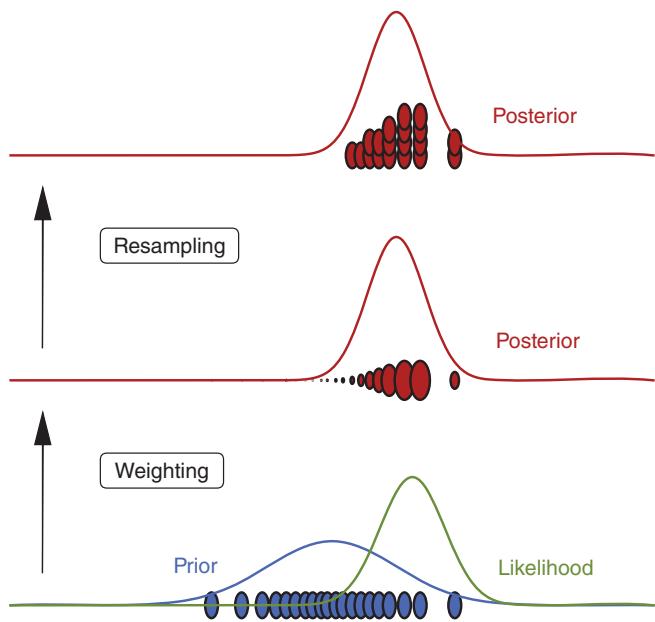


FIGURE 10 Principle of the sequential importance resampling (SIR) particle filter (here $N = 19$). The lower panel curves are the pdfs of the prior and the observation. The initial equal-weight particles are also displayed. The middle panel shows the updated unequal weights of the particles as computed by the likelihood. The upper panel shows the outcome of resampling with multiple copies of several of the initial particles

A richer class of PFs can be formalized if we consider a density of trajectories $p(\mathbf{x}_k|\mathbf{y}_{k:})$ in place of model states (i.e., the density $p(\mathbf{x}_k|\mathbf{y}_k)$), a pdf which is usually associated to a smoothing problem (c.f., Section 2.2). Now, we assume the existence of a smoothing density $q(\mathbf{x}_k|\mathbf{y}_{k:})$ from which it is easier to sample, instead of the “desired” $p(\mathbf{x}_k|\mathbf{y}_{k:})$. It has the following delta-Dirac representation:

$$q(\mathbf{x}_k|\mathbf{y}_{k:}) = \sum_{n=1}^N \omega_k^n \delta(\mathbf{x}_k - \mathbf{x}_k^n). \quad (74)$$

We can exploit this *proposal* density, which is auxiliary to the conditional pdf under study, to compute any statistical moment of the conditional probability $p(\mathbf{x}_k|\mathbf{y}_{k:})$. If ϕ is a generic test function, we have

$$\mathbb{E}[\phi(\mathbf{x}_k)] = \int d\mathbf{x}_k \phi(\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{k:}) \quad (75)$$

$$= \int d\mathbf{x}_k \phi(\mathbf{x}_k) \frac{p(\mathbf{x}_k|\mathbf{y}_{k:})}{q(\mathbf{x}_k|\mathbf{y}_{k:})} q(\mathbf{x}_k|\mathbf{y}_{k:}) \quad (76)$$

$$= \sum_{n=1}^N \omega_k^n \phi(\mathbf{x}_k^n) \frac{p(\mathbf{x}_k^n|\mathbf{y}_{k:})}{q(\mathbf{x}_k^n|\mathbf{y}_{k:})}. \quad (77)$$

Hence, the conditional pdf can be represented using weighted samples:

$$p(\mathbf{x}_k|\mathbf{y}_{k:}) \approx \sum_{n=1}^N \frac{p(\mathbf{x}_k^n|\mathbf{y}_{k:})}{q(\mathbf{x}_k^n|\mathbf{y}_{k:})} \omega_k^n \delta(\mathbf{x}_k - \mathbf{x}_k^n). \quad (78)$$

We further assume that the proposal density factorizes according to

$$q(\mathbf{x}_k|\mathbf{y}_{k:}) = q(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_k) q(\mathbf{x}_{k-1}|\mathbf{y}_{k-1}) \quad (79)$$

such that it is easy to sample \mathbf{x}_k^n from $q(\mathbf{x}_k|\mathbf{x}_{k-1}^n, \mathbf{y}_k)$.

The particle trajectories $\{\mathbf{x}_{k-1}^n, \omega_{k-1}^n\}_{n=1,\dots,N}$ can then be extended to t_k by sampling \mathbf{x}_k^n from $q(\mathbf{x}_k|\mathbf{x}_{k-1}^n, \mathbf{y}_{k:})$ so as to obtain $\{\mathbf{x}_k^n, \omega_k^n\}_{n=1,\dots,N}$. Hence, the smoothing conditional pdf at t_k is

$$p(\mathbf{x}_k|\mathbf{y}_{k:}) = \sum_{n=1}^N \frac{p(\mathbf{x}_k^n|\mathbf{y}_{k:})}{q(\mathbf{x}_k^n|\mathbf{y}_{k:})} \omega_k^n \delta(\mathbf{x}_k - \mathbf{x}_k^n). \quad (80)$$

Assuming Markovian dynamics, the sequential evolution of the smoothing pdf decomposes as

$$p(\mathbf{x}_{k:} | \mathbf{y}_{k:}) \propto p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1:} | \mathbf{y}_{k-1:}), \quad (81)$$

which, together with Equation (79), yields

$$p(\mathbf{x}_{k:} | \mathbf{y}_{k:}) \propto \sum_{n=1}^N \frac{p(\mathbf{x}_k^n | \mathbf{x}_{k-1}^n) p(\mathbf{y}_k | \mathbf{x}_k^n)}{q(\mathbf{x}_k^n | \mathbf{x}_{k-1:}, \mathbf{y}_{k:})} \omega_{k-1}^n \delta(\mathbf{x}_{k:} - \mathbf{x}_{k:}^n). \quad (82)$$

By comparison with the generic Equation (78), the weights should be updated at t_k according to

$$\omega_k^n \propto \frac{p(\mathbf{x}_k^n | \mathbf{x}_{k-1}^n) p(\mathbf{y}_k | \mathbf{x}_k^n)}{q(\mathbf{x}_k^n | \mathbf{x}_{k-1:}, \mathbf{y}_{k:})} \omega_{k-1}^n, \quad (83)$$

up to a normalization to 1 of the updated weights. The filtering solution of the estimation problem is simply obtained without further computation from $\{\mathbf{x}_{k:}^n, \omega_k^n\}_{n=1,\dots,N}$ by marginalization, that is, keeping the states at t_k with the same weights: $\{\mathbf{x}_k^n, \omega_k^n\}_{n=1,\dots,N}$.

Importantly, if we choose $q(\mathbf{x}_k | \mathbf{x}_{k-1:}, \mathbf{y}_{k:}) \equiv p(\mathbf{x}_k | \mathbf{x}_{k-1})$, then we obtain $\omega_k^n \propto p(\mathbf{y}_k | \mathbf{x}_k^n) \omega_{k-1}^n$ and recover the bootstrap PF (c.f., Equation (70)). Furthermore, if we choose $q(\mathbf{x}_k^n | \mathbf{x}_{k-1:}, \mathbf{y}_{k:}) \equiv p(\mathbf{x}_k | \mathbf{y}_k, \mathbf{x}_{k-1})$, then we obtain $\omega_k^n \propto p(\mathbf{y}_k | \mathbf{x}_{k-1}^n) \omega_{k-1}^n$. This corresponds to the *optimal importance proposal* PF (Doucet, Godsill, & Andrieu, 2000). It is optimal only in the sense that it minimizes the theoretical variance of each weight ω_k^n conditional on $\mathbf{x}_{k-1:}^n$ and $\mathbf{y}_{k:}$ over the realizations of \mathbf{x}_k^n . This variance is actually 0.

6.1.3 | Degeneracy of the PF

The algebra required by these PFs is very simple and elegant, and in principle offers a nice and asymptotically exact alternative to the Gaussian approximations to DA. Unfortunately, the PF is plagued by the *curse of dimensionality* as the dimensionality of the model is increased. In a sequential scheme, this curse manifests itself by the degeneracy of the weights: one weight will be close to 1, while the others essentially vanish (Kong, Liu, & Wong, 1994). That is to say, the ensemble of particles collapses onto one single particle, while updating the weights via Equation (70) or Equation (83). This collapse is very inefficient as far as state estimation is concerned. Resampling does help by resetting to uniform weights, but it is, quite often, not enough to counteract this curse. Moreover, this trend essentially grows exponentially with the dimension of the system. More precisely, it has been shown in very simple but instructive Gaussian models that the particle numbers required to avoid degeneracy should scale like the variance of the likelihood (Snyder, Bengtsson, Bickel, & Anderson, 2008):

$$\ln(N) \propto \text{Var}[\ln(p(\mathbf{y} | \mathbf{x})]. \quad (84)$$

Equation (84) has the merit to show that N could scale exponentially with the size of simple systems, but the derivation of such scaling is not straightforward in general. A carefully designed proposal, such as the optimal proposal mentioned in Section 6.1.2, and which can be shown to minimize the variance of the weights, does not change this exponential trend. Yet, it does reduce the constant in the exponent (MacKay, 2003; Snyder, Bengtsson, & Morzfeld, 2015). Numerical investigations confirm this trend (Bocquet, Pires, & Wu, 2010; Slivinski & Snyder, 2016).

6.1.4 | Smarter PFs for the geosciences

Particle filtering is a well-developed field of statistics and engineering (Arulampalam, Maskell, Gordon, & Clapp, 2002; Chen, 2003; Doucet, de Freitas, & Gordon, 2001). It is very successful with low-dimensional models (such as object tracking). The number of contributions to the topic has substantially grown in geophysical DA in recent years (Bocquet et al., 2010; van Leeuwen, 2009). Yet, the curse of dimensionality remains a major obstacle. That is why appealing schemes have been proposed to reduce its impact.

The implicit PF (Chorin & Tu, 2009; Morzfeld, Tu, Atkins, & Chorin, 2012) combines smoothing and particle filtering over a DAW, similarly to 4DVar or the IEnKS. It stands as a possible extension of the optimal proposal PF but over a several-step long DAW.

In order to circumvent the curse of dimensionality, one can restrict the full Bayesian analysis to a limited number of degrees of freedom, while the rest of the control variables are estimated via a Gaussian-based method, typically the EnKF. This strategy has been developed in Lagrangian DA (Slivinski, Spiller, Apte, & Sandstede, 2015).

Another strategy relies on mitigating the degeneracy of the PF by hybridizing with the EnKF (Santitissadeekorn & Jones, 2015). A family of algorithm can be created and parametrized by a mixing coefficient, which can be tuned or adaptively selected (Chustagulprom, Reich, & Reinhardt, 2016; Frei & Künsch, 2013; Stordal, Karlsen, Nævdal, Skaug, & Vallès, 2011).

The equal weight PF and variants thereof (Ades & van Leeuwen, 2015; Zhu, van Leeuwen, & Amezcuia, 2016) build a proposal such that the particles effectively get the same numerical weight out of the analysis, a procedure which is obviously meant to avoid the degeneracy.

6.1.5 | Localization

Similarly to the EnKF, a strategy to mitigate the curse of dimensionality that plagues the PF is to reduce the number of degrees of freedom by making local analyses. However, its implementation is trickier than with the EnKF. Indeed, considering a local DL, a particle will be given distinct local weights, even though they might vary smoothly in space. Hence, there is no natural updated particle that could emerge from the former if the local weights are unequal, as opposed to the local EnKF. One should devise a gluing of local parts of particles through resampling and try to avoid nonphysical discontinuities at the intersections of the local domains. Such strategies have recently been proposed and implemented (Beskos, Crisan, Jasra, Kamatani, & Zhou, 2017; Farchi & Bocquet, 2018; Penny & Miyoshi, 2016; Poterjoy, 2016; Reich, 2013; Robert & Künsch, 2017). Their success is mitigated by the still large number of particles required in each local domain. As a consequence, the size of such domain (localization radius) is generally diagnosed to be significantly smaller than with the local EnKF. Yet, localization is thought to be a necessary ingredient of a successful PF in high-dimensional geophysical systems.

6.2 | Coupled DA

DA algorithms have been conceived mainly for NWP applications and have been usually designed for state estimation in systems with a single dominant dynamical scale and/or for an observational network having a dominant spatiotemporal density. The sustained increase of model resolutions, the deployment of more and more observation platforms, and the use of coupled ESMs for s2d predictions, altogether bind to a deep rethinking of the DA procedures. The design of efficient CDA methods, able to keep simultaneously control of all resolved scales and propagate adequately information across the climate system components, has been recently recognized to have primary importance.

Several research groups and institutions, including weather and climate services, are currently studying and developing CDA (see Penny & Hamill, 2017, for an updated report on these efforts). Early attempts include the case of sparsely observed flow possessing a wide range of scales with a KF-like procedure (Harlim & Majda, 2010), or a study of the performance of the EnKF in a prototypical nonlinear dynamics possessing two scales of motion (Ballabrera-Poy, Kalnay, & Yang, 2009). On the side of variational methods, Lorenc and Payne (2007) have nicely illustrated a modification of the 4DVar, which might be successfully applied to global high-resolution coupled models.

The s2d prediction spans time horizons of up to approximately 10 years, falling between NWP and centennial projections (Doblas-Reyes et al., 2013). Correct initialization of the model is known to improve forecast quality on horizons of several years (Carrassi, Guemas, Doblas-Reyes, Volpi, & Asif, 2016), and for a long time predictions have been initialized with observations of the present climatic state using either the *full field* (FFI) or the *anomaly initialization* (AI) (see Carrassi et al., 2014). FFI makes use of the best possible available estimate of the real state: it reduces the initial error, but the unavoidable presence of model deficiencies causes the model trajectory to drift away from the observations (see Stockdale, 1997). AI assimilates the observed climate anomalies on top of an estimate of the model mean climate. This initial state, at the expense of an initial error of the size of the model bias, is expected to be closer to the model' attractor (see Smith & Murphy, 2007), so that drift is reduced. Comparisons between FFI and AI have revealed respective advantages and drawbacks, the strong regional and model dependency of the results (see Carrassi et al., 2014; Hazeleger et al., 2013; Magnusson, Alonso-Balmaseda, Corti, Molteni, & Stockdale, 2013; Smith, Eade, & Pohlmann, 2013), and the fact that AI is a viable option only when the model and the observed statistics differ largely on their first moments alone (i.e., the bias) (Weber, Carrassi, & Doblas-Reyes, 2015).

However, it was made clear that such a “decoupled” initialization approach induced problems, particularly imbalances at the boundary between the ocean and the atmosphere. To cope with this issue, a solution has shown some success: the *weakly coupled data assimilation* (wCDA). In the wCDA, a coupled model is used to run the predictions but the observations of the different model compartments (atmosphere, ocean, land, and sea ice) are used independently, so that each component is subject to a separate analysis. A first attempt to create a weakly coupled reanalysis has been done in the USA at the National Center for Environmental Prediction (NCEP) (Saha et al., 2010) and at the Japanese Agency for Marine-Earth Science and Technology (JAMSTEC) (Sugiura et al., 2008), based on global ESMs and using 3DVar and 4DVar, respectively. The wCDA reanalysis showed a marked improvement over the standard uncoupled formulation. In the JAMSTEC implementation, the control variable includes the ocean initial conditions plus a set of parameters related to the air-sea fluxes. The approach acted as a proof-of-concept for successfully producing balanced initial conditions for the coupled system and optimal coupling parameters, and enhancing the skills of the s2d prediction. The UK MetOffice has designed a weakly coupled atmosphere-ocean assimilation using the incremental 4DVar (Laloyaux, Balmaseda, Dee, Mogensen, & Janssen, 2016) (c.f., Appendix D)

and the global coupled model, but the corrections for atmosphere and ocean are calculated independently. Similarly, the ECMWF has produced a 20th century reanalysis based on wCDA (Poli et al., 2016). The EnKF in a wCDA setting has been recently used to assimilate ocean observations to initialize s2d predictions with the Norwegian Earth System Model (NorESM) (Counillon et al., 2014).

Atmosphere and ocean are constrained independently using the ensemble-based approach at the Geophysical Fluid Dynamics Laboratory (GFDL) using the EAKF (Zhang, Harrison, Rosati, & Wittenberg, 2007). Using the same framework, Lu, Liu, Zhang, and Liu (2015) have achieved some success in a controlled simulated scenario using *strongly CDA* (sCDA), in which the different model compartments are coupled together also at the analysis times, so that observations on one compartment, say the atmosphere, impact on another, say the ocean. The assimilation reconstructed successfully relevant climate fields over the period of interest and provides automatically the initial conditions to run an ensemble of forecasts. One of the first attempts of sCDA for a coupled ocean and sea ice model has been used operationally in TOPAZ (c.f., Section 5.4), demonstrating that successful assimilation of sea ice concentrations requires a coupled, multivariate, and time-dependent assimilation method (Sakov, Counillon, et al., 2012). A recent interesting result using sCDA is due to Sluka, Penny, Kalnay, and Miyoshi (2016) that shows improvements over wCDA in using only atmospheric observations in a coupled atmosphere–ocean model. CDA with the EnKF to recover the Atlantic meridional overturning circulation (AMOC) with simulated observations in a low-order coupled atmosphere–ocean model has been studied by Tardif, Hakim, and Snyder (2014), and subsequently with data from a millennial-scale simulation of a comprehensive coupled atmosphere–ocean climate model in Tardif, Hakim, and Snyder (2015). These studies suggest that atmospheric observations alone, albeit frequent, do not suffice to properly recover the slowly evolving AMOC. Interestingly, it was shown that, in the lack of enough observations in the ocean, CDA of time-averaged atmospheric measurements can successfully track the AMOC.

A comparison of different CDA approaches in the context of incremental 4DVar (c.f., Appendix D) using an idealized single-column atmosphere–ocean model is given in Smith, Fowler, and Lawless (2015), and revealed the benefit of CDA as being able to produce more balanced analysis fields than its uncoupled counterpart, thus reducing initialization shock and the subsequent predictions. The same idealistic model setup has been used to investigate the impact of the model error and of the window length of the 4DVar showing that while uncoupled DA may reduce the analysis error more than the CDA, the latter better reduces the imbalances and thus reduces the forecasts error.

CDA is one of the main areas of research at present time and more advancements can be expected in the coming years; a review on the current status of the field can be found in Penny and Hamill (2017).

7 | CONCLUSION

The goal of this work is to provide an up-to-date review of DA methods for the geosciences. We hope that it may serve as a first guide for scientists who are confronting with the use of DA methods, by providing a complete outlook of the approaches and of their foundations. This work offers a detailed introduction to the topic aimed at being a starting point from where interested readers may later expand their knowledge.

7.1 | Summary of content

We have introduced the estimation problem, along with the definition of the dynamical and observational models, in Section 2. A statistical, Bayesian, point of view has been adopted to derive the filter and smoother equations. The statistical approach offers notable conceptual and mathematical advantages: it genuinely accommodates the treatment of the uncertainty in terms of probability and the model/observation outputs as realizations of random variables. The assumption of time-uncorrelated model and observational errors has been employed and led to formulate the problem as an hidden Markov model. Nevertheless, such an assumption does not generally hold in geosciences applications and we have thus mentioned methods to overcome it even though their extensive description is beyond our scope here.

The huge dimension of typical DA problems in the geosciences makes the full Bayesian approach computationally unaffordable in many realistic cases, and a parametric description of the pdf is required. The Gaussian hypothesis is thus employed in most DA methods and this has been the content of Section 3, where in particular we have described the popular KF, KS, and the variational approach. The latter class of methods in particular does not rely much on the hypothesis of uncorrelated model error. Section 3 has four complementary Appendices in which more details, properties, and features of these methods are explained. The Monte Carlo formulations of the KF and KS, known as ensemble Kalman filter (EnKF) and ensemble Kalman smoother (EnKS), have made possible the successful extension of the classical KF and KS to high-dimensional nonlinear situations. The EnKF and EnKS, in their stochastic and deterministic (square-root) formulations, are the main subject of Section 4, that also discusses the two popular ad hoc fixes that are functional to the success of the EnKF and EnKS in high

dimension: localization and inflation. Section 4 includes as well an updated survey of the very recent class of hybrid methods known as ensemble variational that are being increasingly applied in operational weather services.

To gain more insight and getting the flavor of the scientific challenges encountered by DA in its application to the geosciences, Section 5 exposes four selected topics: (a) DA for chaotic dynamics such as the atmosphere or the ocean, (b) DA for non Gaussian variables, (c) DA for chemical constituents of the atmosphere, and, (d) an example of operational DA for the ocean prediction.

We have finally presented a prospect of the challenges that DA is facing nowadays in Section 6, with special attention to CDA, needed to perform DA with coupled climate systems, and to PFs, that are experiencing a tremendous trend of development aimed at making computationally viable the use of a fully nonlinear, Bayesian, DA for high-dimensional systems.

7.2 | Forward looking

From its origin in the context of NWP, DA has later expanded to the broad areas of environmental prediction, including seasonal, interannual, and decadal time scales. The current efforts toward the implementation of seamless predictions, where the same high-resolution coupled models are used from short- to long-term forecasts, are also accompanied by advancements in DA. In particular, DA has to be conceived to tackle the nonlinearities emerging from the increase in resolution and the coupling mechanisms giving rise to long-term predictability.

Nevertheless, the ranges of applications of DA have not remained confined to the state (and/or parameters) estimation to initialize prediction, but are progressively involving other problems. A notable example is the detection and attribution of climate change, or climate-related events (Stott et al., 2013), which is the issue of providing evidence for either the existence or the nonexistence of a causal relationship between a hypothetical external forcing (e.g., anthropogenic emission) to a system (e.g., the climate) and an observed response (e.g., increase of temperature), for which novel methods based on DA have proven to be very efficient (Hannart et al., 2016). More generally, DA can be efficiently used to estimate the marginal likelihood of the data, the so called model evidence (Carrassi, Bocquet, Hannart, & Ghil, 2017), which is a key statistical metric to perform model selection (see Carson, Crucifix, Preston, & Wilkinson, 2018, for an application in the context of glacial–interglacial cycle) and calibration or parameter estimation (see Winiarek, Vira, Bocquet, Sofiev, & Saunier, 2011, for the estimation of a radiological plume), or Tandeo, Pulido, and Lott (2015) for the optimization of a subgrid-scale parametrization. DA has been used to reconstruct the climate of the past based on observations proxies (see Dubinkina & Goosse, 2013), and, in the solid Earth science, for seismology applications (see Fichtner, Bunge, & Igel, 2006).

In general, the use of DA has proven that the consistent data-to-model fusion provides a more insightful view on the phenomena of interest, than any of the two components, the model or the data, independently. Future applications of DA within the geosciences and beyond, are expected to be numerous, and to naturally arise by the improvement of our modeling capabilities, as a result of the increased computational power and physical understanding, on one hand, and by the progresses of the observing facilities (such as, but not only, satellite) on the other. DA is nowadays spreading to many emerging disciplines such as neurosciences, genetics, biology, medicine, or even in sociology–demography and traffic managements just to mention a few (see Palatella, Trevisan, & Rambaldi, 2013 and Kadakia et al., 2016 for an example of applications for traffic flow and biology, respectively). This expansion exposes DA to the need of new theoretical principles and novel methodological solutions and provides new contexts for challenging its effectiveness and robustness. DA is thus expected to continue playing a central role to bridge model with data, to maximally exploit their respective informational content.

We hope that this overview may be a first guide for scientists who are confronting with the use of DA methods, and provide them with a complete first outlook of the approaches and of their foundations. The present work is thus to be intended as a detailed introduction to the topic from where interested readers and researchers may later expand their knowledge.

ACKNOWLEDGMENTS

The authors wish to thank Eugenia Kalnay and two anonymous reviewers for their detailed, deep, and critical reviews of the original version of this work. Their comments and suggestions have substantially improved its readability to a wider audience and have also helped in clarifying the discussion in many instances. The authors are thankful to P. N. Raanes (NERSC), A. Farchi (ENPC), and C. Grudzien (NERSC) for their comments, suggestions, and insightful discussions and to J. Xie (NERSC) for providing Figure 8. Finally, the authors also wish to thank R. Davy (NERSC) who provided a critical review of the second version of the manuscript that helped to smooth further the mathematics and to make the work more accessible to the geosciences community at large. A.C. has been funded by the project REDDA (#250711) of the Norwegian Research Council. G.E. has been partly funded by the project Embla of Nordforsk. Cerea is a member of Institut Pierre-Simon Laplace (IPSL).

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

ORCID

Alberto Carrassi  <http://orcid.org/0000-0003-0722-5600>

REFERENCES

- Aalstad, K., Westermann, S., Chuler, T., Boike, J., & Bertino, L. (2018). Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites. *The Cryosphere*, 12, 247–270.
- Ades, M., & van Leeuwen, P. J. (2015). The equivalent-weights particle filter in a high-dimensional system. *Quarterly Journal of the Royal Meteorological Society*, 141, 484–503.
- Amecua, J., Goodliff, M., & van Leeuwen, P. J. (2017). A weak-constraint 4DEnsembleVar. Part I: Formulation and simple model experiments. *Tellus A*, 69(1271), 564.
- Amecua, J., & Van Leeuwen, P. J. (2014). Gaussian anamorphosis in the analysis step of the EnKF: A joint state-variable/observation approach. *Tellus A*, 66, 1–18.
- Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129, 2884–2903.
- Anderson, J. L. (2007). An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, 59, 210–224.
- Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127, 2741–2758.
- Apte, A., & Jones, C. K. (2013). The impact of nonlinearity in Lagrangian data assimilation. *Nonlinear Processes in Geophysics*, 20, 329–341.
- Arulampalam, S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50, 174–188.
- Asch, M., Bocquet, M., & Nodet, M. (2016). *Data assimilation: Methods, algorithms, and applications, fundamentals of algorithms*. Philadelphia, PA: SIAM.
- Auligné, T., Ménétrier, B., Lorenc, A. C., & Buehner, M. (2016). Ensemble-variational integrated localized data assimilation. *Monthly Weather Review*, 144, 677–3696.
- Auroux, D., & Blum, J. (2008). A nudging-based data assimilation method: The back and forth nudging (bfn) algorithm. *Nonlinear Processes in Geophysics*, 15, 305–319.
- Bain, A., & Crisan, D. (2009). *Fundamentals of stochastic filtering* (Vol. 3). New York: Springer.
- Ballabrera-Poy, J., Kalnay, E., & Yang, S.-C. (2009). Data assimilation in a system with two scales—Combining two initialization techniques. *Tellus A*, 61, 539–549.
- Bannister, R. N. (2017). A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143, 607–633.
- Barth, A., Canter, M., van Schaeybroeck, B., Vannitsem, S., Massonnet, F., Zunz, V., ... Beckers, J.-M. (2015). Assimilation of sea surface temperature, ice concentration and ice drift in a model of the Southern Ocean. *Ocean Modelling*, 93, 22–39.
- Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55.
- Bell, B. M. (1994). The iterated Kalman smoother as a Gauss-Newton method. *SIAM Journal on Optimization*, 4, 626–636.
- Bengtsson, L., Ghil, M., & Källén, E. (Eds.). (1981). *Dynamic meteorology: Data assimilation methods*. New York, NY/Heidelberg and Berlin, Germany: Springer-Verlag.
- Bennett, A. F. (1992). *Inverse methods in physical oceanography*. Cambridge, UK: Cambridge University Press.
- Berre, L., Varella, H., & Desroziers, G. (2015). Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4D-Var at Météo-France. *Quarterly Journal of the Royal Meteorological Society*, 141, 2803–2812.
- Bertino, L., Evensen, G., & Wackernagel, H. (2003). Sequential data assimilation techniques in oceanography. *International Statistical Review*, 71, 223–241.
- Beskos, A., Crisan, D., Jasra, A., Kamatani, K., & Zhou, Y. (2017). A stable particle filter for a class of high-dimensional state-space models. *Advances in Applied Probability*, 49, 24–48.
- Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, 129, 420–436.
- Bleck, R. (2002). An oceanic general circulation model in pressure coordinates. *Ocean Modelling*, 37, 55–88.
- Bocquet, M. (2011). Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, 18, 735–750.
- Bocquet, M. (2016). Localization and the iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, 142, 1075–1089.
- Bocquet, M., & Carrassi, A. (2017). Four-dimensional ensemble variational data assimilation and the unstable subspace. *Tellus A*, 69, 1304504.
- Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Zabkar, R., Carmichael, G. R., ... Seigneur, C. (2015). Data assimilation in atmospheric chemistry models: Current status and future prospects for coupled chemistry meteorology models. *Atmospheric Chemistry and Physics*, 15, 5325–5358.
- Bocquet, M., Gurumoorthy, K. S., Apte, A., Carrassi, A., Grudzien, C., & Jones, C. K. (2017). Degenerate Kalman filter error covariances and their convergence onto the unstable subspace. *SIAM/ASA Journal on Uncertainty Quantification*, 5, 304–333.
- Bocquet, M., Pires, C., & Wu, L. (2010). Beyond Gaussian statistical modeling in geophysical data assimilation. *Monthly Weather Review*, 138, 2997–3023.
- Bocquet, M., Raanes, P. N., & Hannart, A. (2015). Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, 22, 645–662.
- Bocquet, M., & Sakov, P. (2012). Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems. *Nonlinear Processes in Geophysics*, 19, 383–399.
- Bocquet, M., & Sakov, P. (2013). Joint state and parameter estimation with an iterative ensemble Kalman smoother. *Nonlinear Processes in Geophysics*, 20, 803–818.
- Bocquet, M., & Sakov, P. (2014). An iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, 140, 1521–1535.
- Bonan, B., Nichols, N. K., Baines, M. J., & Partridge, D. (2017). Data assimilation for moving mesh methods with an application to ice sheet modelling. *Nonlinear Processes in Geophysics*, 24, 515–534.
- Bonavita, M., Isaksen, L., & Hólm, E. (2012). On the use of EDA background error variances in the ECMWF 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 138, 1540–1559.
- Bonavita, M., Raynaud, L., & Isaksen, L. (2011). Estimating background-error variances with the ECMWF ensemble of data assimilation system: Some effects of ensemble size and day-to-day variability. *Quarterly Journal of the Royal Meteorological Society*, 137, 423–434.
- Bowler, N. E., Clayton, A. M., Jardak, M., Lee, E., Lorenc, A. C., Piccolo, C., ... Swinbank, R. (2017). Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations. *Quarterly Journal of the Royal Meteorological Society*, 143, 1280–1302.

- Brankart, J.-M., Cosme, E., Testut, C.-E., Brasseur, P., & Verron, J. (2010). Efficient adaptive error parameterization for square root or ensemble Kalman filters: Application to the control of ocean mesoscale signals. *Monthly Weather Review*, 138, 932–950.
- Brankart, J.-M., Testut, C.-E., Béal, D., Doron, M., Fontana, C., Meinvielle, M., ... Verron, J. (2012). Towards an improved description of ocean uncertainties: Effect of local anamorphic transformations on spatial correlations. *Ocean Science*, 8, 121–142.
- Brusdal, K., Brankart, J., Halberstadt, G., Evensen, G., Brasseur, P., van Leeuwen, P. J., ... Verron, J. (2003). An evaluation of ensemble based assimilation methods with a layered OGCM. *Journal of Marine Systems*, 40–41, 253–289.
- Buehner, M. (2005). Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quarterly Journal of the Royal Meteorological Society*, 131, 1013–1043.
- Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., & He, B. (2010a). Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Monthly Weather Review*, 138, 1550–1566.
- Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., & He, B. (2010b). Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Monthly Weather Review*, 138, 1567–1586.
- Buehner, M., McTaggart-Cowan, R., Beaulne, A., Charette, C., Garand, L., Heilliette, S., ... Zadra, A. (2015a). Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part I: The global system. *Monthly Weather Review*, 143, 2532–2559.
- Buehner, M., Morneau, J., & Charette, C. (2013). Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlinear Processes in Geophysics*, 20, 669–682.
- Burgers, G., van Leeuwen, P. J., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126, 1719–1724.
- Carmichael, G. R., Sandu, A., Chai, T., Daescu, D., Constantinescu, E., & Tang, Y. (2008). Predicting air quality: Improvements through advanced methods to integrate models and measurements. *Journal of Computational Physics*, 227, 3540–3571.
- Carrassi, A., Bocquet, M., Hannart, A., & Ghil, M. (2017). Estimating model evidence using data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143, 866–880.
- Carrassi, A., Ghil, M., Trevisan, A., & Uboldi, F. (2008). Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system. *Chaos*, 18(023), 112.
- Carrassi, A., Guemas, V., Doblas-Reyes, F., Volpi, D., & Asif, M. (2016). Sources of skill in near-term climate prediction: Generating initial conditions. *Climate Dynamics*, 47, 3693–3712.
- Carrassi, A., Hamdi, R., Termonia, P., & Vannitsem, S. (2012). Short time augmented extended Kalman filter for soil analysis: A feasibility study. *Atmospheric Science Letters*, 13, 268–274.
- Carrassi, A., Trevisan, A., Descamps, L., Talagrand, O., & Uboldi, F. (2008). Controlling instabilities along a 3DVar analysis cycle by assimilating in the unstable subspace: A comparison with the EnKF. *Nonlinear Processes in Geophysics*, 15, 503–521.
- Carrassi, A., Trevisan, A., & Uboldi, F. (2007). Adaptive observations and assimilation in the unstable subspace by breeding on the data-assimilation system. *Tellus A*, 59, 101–113.
- Carrassi, A., & Vannitsem, S. (2010). Accounting for model error in variational data assimilation: A deterministic formulation. *Monthly Weather Review*, 138, 3369–3386.
- Carrassi, A., & Vannitsem, S. (2011). State and parameter estimation with the extended Kalman filter: An alternative formulation of the model error dynamics. *Quarterly Journal of the Royal Meteorological Society*, 137, 435–451.
- Carrassi, A., & Vannitsem, S. (2016). Deterministic treatment of model error in geophysical data assimilation. In *Mathematical paradigms of climate science* (pp. 175–213). Springer International Publishing Switzerland: Springer.
- Carrassi, A., Vannitsem, S., Zupanski, D., & Zupanski, M. (2009). The maximum likelihood ensemble filter performances in chaotic systems. *Tellus A*, 61, 587–600.
- Carrassi, A., Weber, R., Guemas, V., Doblas-Reyes, F., Asif, M., & Volpi, D. (2014). Full-field and anomaly initialization using a low-order climate model: A comparison and proposals for advanced formulations. *Nonlinear Processes in Geophysics*, 21, 521–537.
- Carson, J., Crucifix, M., Preston, S., & Wilkinson, R. D. (2018). Bayesian model selection for the glacial–interglacial cycle. *Journal Of The Royal Statistical Society Series C-Applied Statistics*, 67, 25–54.
- Chekroun, M., Simonnet, E., & Ghil, M. (2011). Stochastic climate dynamics: Random attractors and time-dependent invariant measures. *Physica D*, 240, 1685–1700.
- Chen, Z. (2003). Bayesian filtering: From Kalman filters to particle filters, and beyond. *Statistics*, 182, 1–69.
- Chièze, J. P., & Delfiner, P. (2012). *Geostatistics: Modeling spatial uncertainty*. New York, NY: Wiley.
- Chorin, A. J., & Tu, X. (2009). Implicit sampling for particle filters. *PNAS*, 106, 17249–17254.
- Chustagulprom, N., Reich, S., & Reinhardt, M. (2016). A hybrid ensemble transform particle filter for nonlinear and spatially extended dynamical systems. *SIAM/ASA Journal on Uncertainty Quantification*, 4, 592–608.
- Clayton, A. M., Lorenc, A. C., & Barker, D. M. (2013). Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, 139, 1445–1461.
- Cohn, S. E. (1997). An introduction to estimation theory (Special issue—Data assimilation in meteorology and oceanography: Theory and practice). *Journal of the Meteorological Society of Japan*, 75, 257–288.
- Cohn, S. E., & Dee, D. P. (1988). Observability of discretized partial differential equations. *SIAM Journal on Numerical Analysis*, 25, 586–617.
- Constantinescu, E. M., Sandu, A., Chai, T., & Carmichael, G. R. (2007a). Ensemble-based chemical data assimilation. I: General approach. *Quarterly Journal of the Royal Meteorological Society*, 133, 1229–1243.
- Constantinescu, E. M., Sandu, A., Chai, T., & Carmichael, G. R. (2007b). Ensemble-based chemical data assimilation. II: Covariance localization. *Quarterly Journal of the Royal Meteorological Society*, 133, 1245–1256.
- Cosme, E., Verron, J., Brasseur, P., Blum, J., & Auroux, D. (2012). Smoothing problems in a Bayesian framework and their linear Gaussian solutions. *Monthly Weather Review*, 140, 683–695.
- Counillon, F., Bethke, I., Keenlyside, N., Bentsen, M., Bertino, L., & Zheng, F. (2014). Seasonal-to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System Model: A twin experiment. *Tellus A*, 66, 21,074.
- Counillon, F., Keenlyside, N., Bethke, I., Wang, Y., Billeau, S., Shen, M. L., & Bentsen, M. (2016). Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian climate prediction model. *Tellus A*, 68, 1–17.
- Courtier, P., Thépaut, J.-N., & Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120, 1367–1387.
- Daley, R. (1993). *Atmospheric data analysis*. Cambridge, MA: Cambridge University Press.
- de Rosnay, P., Balsamo, G., Albergel, C., Muñoz-Sabater, J., & Isaksen, L. (2014). Initialisation of land surface variables for numerical weather prediction. *Surveys in Geophysics*, 35, 607–621.
- Dee, D., Cohn, S., Dalcher, A., & Ghil, M. (1985). An efficient algorithm for estimating noise covariances in distributed systems. *IEEE Transactions on Automatic Control*, 30, 1057–1065.

- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Bechtold, P. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597.
- Dee, D. P. (1995). On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly Weather Review*, 123, 1128–1145.
- Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131, 3323–3343.
- Desroziers, G., Arbogast, E., & Berre, L. (2016). Improving spatial localization in 4DEnVar. *Quarterly Journal of the Royal Meteorological Society*, 142, 3171–3185 accepted for publication.
- Desroziers, G., Camino, J.-T., & Berre, L. (2014). 4DEnVar: Link with 4D state formulation of variational assimilation and different possible implementations. *Quarterly Journal of the Royal Meteorological Society*, 140, 2097–2110.
- Dijkstra, H. (2013). *Nonlinear climate dynamics*. Cambridge, England: Cambridge University Press.
- Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., & Rodrigues, L. R. (2013). Seasonal climate predictability and forecasting: Status and prospects. *WIREs Climate Change*, 4, 245–268.
- Douc, R., & Cappé, O. (2005). Comparison of resampling schemes for particle filtering. In *Image and Signal Processing and Analysis, ISPA 2005. Proceedings of the 4th International Symposium* (pp. 64–69). IEEE.
- Doucet, A., de Freitas, N., & Gordon, N. (Eds.). (2001). *Sequential Monte Carlo methods in practice*. New York, NY: Springer-Verlag New York Inc.
- Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10, 197–208.
- Dreano, D., Tandeo, P., Pulido, M., Ait-El-Fquih, B., Chonavel, T., & Hoteit, I. (2017). Estimating model error covariances in nonlinear state-space models using Kalman smoothing and the expectation-maximisation algorithm. *Quarterly Journal of the Royal Meteorological Society*, 143, 1877–1885.
- Duane, G. S., Tribbia, J. J., & Weiss, J. B. (2006). Synchronicity in predictive modelling: A new view of data assimilation. *Nonlinear Processes in Geophysics*, 13, 601–612.
- Dubinkina, S., & Goosse, H. (2013). An assessment of particle filtering methods and nudging for climate state reconstructions. *Climate of the Past*, 9, 1141–1152.
- Elbern, H., Strunk, A., Schmidt, H., & Talagrand, O. (2007). Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmospheric Chemistry and Physics*, 7, 3749–3769.
- Emili, E., Gürol, S., & Cariolle, D. (2016). Accounting for model error in air quality forecasts: An application of 4DEnVar to the assimilation of atmospheric composition using QG-Chem 1.0. *Geoscientific Model Development*, 9, 3933–3959.
- Evensen, G. (1992). Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *Journal of Geophysical Research*, 97, 17905–17924.
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99, 10143–10162.
- Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53, 343–367.
- Evensen, G. (2004). Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, 54, 539–560.
- Evensen, G. (2009a). The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*, 29, 83–104.
- Evensen, G. (2009b). *Data assimilation: The ensemble Kalman filter* (2nd ed.). Berlin and Heidelberg, Germany: Springer-Verlag.
- Evensen, G. (2018). Analysis of iterative ensemble smoothers for solving inverse problems. *Computational Geosciences*, 22, 885–908.
- Fairbairn, D., Pring, S. R., Lorenc, A. C., & Roulstone, I. (2014). A comparison of 4DVar with ensemble data assimilation methods. *Quarterly Journal of the Royal Meteorological Society*, 140, 281–294.
- Farchi, A., & Bocquet, M. (2018). Review article: Comparison of local particle filters and new implementations. *Nonlinear Processes in Geophysics*, 2018, 1–63.
- Fichtner, A., Bunge, H.-P., & Igel, H. (2006). The adjoint method in seismology: II. Applications: Traveltimes and sensitivity functionals. *Physics of the Earth and Planetary Interiors*, 157, 105–123.
- Fisher, M., & Andersson, E. (2001). *Developments in 4D-Var and Kalman filtering* (Technical Memorandum 347). Reading, UK: European Centre for Medium-Range Weather Forecasts.
- Fisher, M., & Gürol, S. (2017). Parallelization in the time dimension of four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143, 1136–1147.
- Fisher, M., Leutbecher, M., & Kelly, G. (2005). On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131, 3235–3246.
- Fisher, M., Tremolet, Y., Auvinen, H., Tan, D., & Poli, P. (2011). *Weak-constraint and long-window 4D-Var* (ECMWF Technical Report 655).
- Fitzgerald, R. (1971). Divergence of the Kalman filter. *IEEE Transactions on Automatic Control*, 16, 736–747.
- Fletcher, S. J. (2017). *Data assimilation for the geosciences: From theory to application*. Amsterdam, Holland: Elsevier.
- Frei, M., & Künsch, H. R. (2013). Bridging the ensemble Kalman and particle filters. *Biometrika*, 100, 781–800.
- Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757.
- Gelb, A. (1974). *Applied optimal estimation*. Cambridge, MA and London, UK: MIT Press.
- Gharamti, M., Samuelsen, A., Bertino, L., Simon, E., Korosov, A., & Daewel, U. (2017). Online tuning of ocean biogeochemical model parameters using ensemble estimation techniques: Application to a one-dimensional model in the North Atlantic. *Journal of Marine Systems*, 168, 1–16.
- Gharamti, M. E., Tjiputra, J., Bethke, I., Samuelsen, A., Skjelvan, I., Bentsen, M., & Bertino, L. (2017). Ensemble data assimilation for ocean biogeochemical state and parameter estimation at different sites. *Ocean Modelling*, 112, 65–89.
- Ghil, M., Cohn, S., Tavantzis, J., Bube, K., & Isaacson, E. (1981). Applications of estimation theory to numerical weather prediction. In *Dynamic meteorology: Data assimilation methods*. New York, NY: Springer-Verlag.
- Ghil, M., & Malanotte-Rizzoli, P. (1991). Data assimilation in meteorology and oceanography. *Advances in Geophysics*, 33, 141–266.
- Giffin, A., & Umiezu, R. (2014). The kalman filter revisited using maximum relative entropy. *Entropy*, 16, 1047–1069.
- Golub, G. H., & van Loan, C. F. (2013). *Matrix computations* (Vol. 3). Baltimore, MD: JHU Press.
- Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F*, 140, 107–113.
- Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., & Hunt, B. R. (2011). Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, 139, 511–522.
- Griffith, A. K., & Nichols, N. K. (2000). Flow, turbulence and combustion. *Adjoint methods in data assimilation for estimating model error* (Vol. 65, pp. 469–488). New York, NY: Springer.
- Grudzien, C., Carrassi, A., & Bocquet, M. (2018a). Chaotic dynamics and the role of covariance inflation for reduced rank kalman filters with model error. *Nonlinear Processes in Geophysics*, 2018, 1–25. <https://doi.org/10.5194/npg-2018-4>
- Grudzien, C., Carrassi, A., & Bocquet, M. (2018b). Asymptotic forecast uncertainty and the unstable subspace in the presence of additive model error. *SIAM/ASA Journal on Uncertainty Quantification* Manuscript submitted for publication.
- Gurumoorthy, K. S., Grudzien, C., Apte, A., Carrassi, A., & Jones, C. K. (2017). Rank deficiency of Kalman error covariance matrices in linear time-varying system with deterministic evolution. *SIAM Journal on Control and Optimization*, 55, 741–759.

- Gustafsson, N., Bojarova, J., & Vignes, O. (2014). A hybrid variational ensemble data assimilation for the high resolution limited area model (HIRLAM). *Nonlinear Processes in Geophysics*, 21, 303–323.
- Hamill, T. M., & Snyder, C. (2000). A hybrid ensemble Kalman filter-3D variational analysis scheme. *Monthly Weather Review*, 128, 2905–2919.
- Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129, 2776–2790.
- Hanea, R. G., Velders, G. J. M., Segers, A. J., Verlaan, M., & Heemink, A. W. (2007). A hybrid Kalman filter algorithm for large-scale atmospheric chemistry data assimilation. *Monthly Weather Review*, 135, 140–151.
- Hannart, A., Carrassi, A., Bocquet, M., Ghil, M., Naveau, P., Pulido, M., ... Tandeo, P. (2016). DADA: Data assimilation for the detection and attribution of weather and climate-related events. *Climatic Change*, 136, 155–174.
- Harlim, J., & Majda, A. J. (2010). Filtering turbulent sparsely observed geophysical flows. *Monthly Weather Review*, 138, 1050–1083.
- Harlim, J., Majda, A. J., et al. (2010). Catastrophic filter divergence in filtering nonlinear dissipative systems. *Communications in Mathematical Sciences*, 8, 27–43.
- Haugen, V. E., & Evensen, G. (2002). Assimilation of SLA and SST data into an OGCM for the Indian ocean. *Ocean Dynamics*, 52, 133–151.
- Haussaire, J.-M., & Bocquet, M. (2016). A low-order coupled chemistry meteorology model for testing online and offline data assimilation schemes: L95-GRS (v1.0). *Geoscientific Model Development*, 9, 393–412.
- Hazeleger, W., Guemas, V., Wouters, B., Corti, S., Andreu-Burillo, I., Doblas-Reyes, F., ... Caian, M. (2013). Multiyear climate predictions using two initialization strategies. *Geophysical Research Letters*, 40, 1794–1798.
- Hoke, J. E., & Anthes, R. A. (1976). The initialization of numerical models by a dynamic-initialization technique. *Monthly Weather Review*, 104, 1551–1556.
- Horn, R. A., & Johnson, C. R. (2012). *Matrix analysis*. New York, NY: Cambridge University Press.
- Hoteit, I., Pham, D.-T., Gharamti, M., & Luo, X. (2015). Mitigating observation perturbation sampling errors in the stochastic EnKF. *Monthly Weather Review*, 143, 2918–2936.
- Houtekamer, P., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144, 4489–4532.
- Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129, 123–137.
- Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., & Hansen, B. (2005). Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Monthly Weather Review*, 133, 604–620.
- Hunt, B., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, 230, 112–126.
- Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. J. D. J., Sauer, T., ... Zimin, A. V. (2004). Four-dimensional ensemble Kalman filtering. *Tellus A*, 56, 273–277.
- Ide, K., Kuznetsov, L., & Jones, C. K. (2002). Lagrangian data assimilation for point vortex systems*. *Journal of Turbulence*, 3(053).
- Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., ... Weston, P. (2017). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society* accepted for publication. <https://doi.org/10.1002/qj.3130>
- Janjić, T., McLaughlin, D., Cohn, S. E., & Verlaan, M. (2014). Conservation of mass and preservation of positivity with ensemble-type Kalman filter algorithms. *Monthly Weather Review*, 142, 755–773.
- Jardak, M., & Talagrand, O. (2018). Ensemble variational assimilation as a probabilistic estimator. Part i: The linear and weak non-linear case. *Nonlinear Processes in Geophysics*, In review. <https://doi.org/10.5194/npg-2018-5>
- Järvinen, H., Andersson, E., & Bouttier, F. (1999). Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus A*, 51, 469–488.
- Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. New York, NY: Academic Press.
- Kadakia, N., Armstrong, E., Breen, D., Morone, U., Daou, A., Margoliash, D., & Abarbanel, H. D. (2016). Nonlinear statistical data assimilation for hvc_RA neurons in the avian song system. *Biological Cybernetics*, 110, 417–434.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82, 35–45.
- Kalnay, E. (2002). *Atmospheric modeling, data assimilation and predictability*. Cambridge, England: Cambridge University Press.
- Kalnay, E., & Dalcher, A. (1987). Forecasting forecast skill. *Monthly Weather Review*, 115, 349–356.
- Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., & Ballabriga-Poy, J. (2007). 4-D-Var or ensemble Kalman filter? *Tellus A*, 59, 758–773.
- Kalnay, E., Ota, Y., Miyoshi, T., & Liu, J. (2012). A simpler formulation of forecast sensitivity to observations: Application to ensemble Kalman filters. *Tellus A*, 64(18), 462.
- Kalnay, E., & Yang, S.-C. (2010). Accelerating the spin-up of ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 136, 1644–1651.
- Kang, J.-S., Kalnay, E., Liu, J., Fung, I., Miyoshi, T., & Ide, K. (2011). “Variable localization” in an ensemble Kalman filter: Application to the carbon cycle data assimilation. *Journal of Geophysical Research*, 116, D09110.
- Kepert, J. D. (2009). Covariance localisation and balance in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135, 1157–1176.
- Kleist, D. T., & Ide, K. (2015). An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Monthly Weather Review*, 143, 433–451.
- Kondrashov, D., Sun, C., & Ghil, M. (2008). Data assimilation for a coupled ocean-atmosphere model. Part II: Parameter estimation. *Monthly Weather Review*, 136, 5062–5076.
- Kong, A., Liu, J. S., & Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89, 278–288.
- Kotsuki, S., Miyoshi, T., Terasaki, K., Lien, G. Y., & Kalnay, E. (2017). Assimilating the global satellite mapping of precipitation data with the nonhydrostatic icosahedral atmospheric model (NICAM). *Journal of Geophysical Research*, 122, 631–650.
- Kuptsov, P. V., & Parlitz, U. (2012). Theory and computation of covariant Lyapunov vectors. *Journal of Nonlinear Science*, 22, 727–762.
- Kuznetsov, L., Ide, K., & Jones, C. K. (2003). A method for assimilation of lagrangian data. *Monthly Weather Review*, 131, 2247–2260.
- Lakshmivarahan, S., & Lewis, J. M. (2013). Nudging methods: A critical overview. In S. K. Park & L. Xu (Eds.), *Data assimilation for atmospheric, oceanic and hydrologic applications* (Vol. II, pp. 27–57). Berlin and Heidelberg, Germany: Springer.
- Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K., & Janssen, P. (2016). A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 142, 65–78.
- Lauvret, C., Brankart, J.-M. M., Castruccio, F., Broquet, G., Brasseur, P., & Verron, J. (2009). A truncated Gaussian filter for data assimilation with inequality constraints: Application to the hydrostatic stability condition in ocean models. *Ocean Modelling*, 27, 1–17.
- Law, K., Stuart, A., & Zygalakis, K. (2015). *Data assimilation: A mathematical introduction* (Vol. 62). Cham, Switzerland: Springer. ISBN 978-3-319-20325-6.
- Lawless, A., Nichols, N., Boess, C., & Bunse-Gerstner, A. (2008). Using model reduction methods within incremental four-dimensional variational data assimilation. *Monthly Weather Review*, 136, 1511–1522.
- Le Dimet, F.-X., & Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus A*, 38, 97–110.
- Le Gland, F., Monbet, V., & Tran, V.-D. (2009). *Large sample asymptotics for the ensemble Kalman filter* (Ph.D. thesis). INRIA.
- Legras, B., & Vautard, R. (1996). A guide to Lyapunov vectors. In *Proceedings 1995 ECMWF Seminar on Predictability* (Vol. 1, pp. 143–156). Reading, UK.

- Leutbecher, M., & Palmer, T. N. (2008). Ensemble forecasting. *Journal of Computational Physics*, 227, 3515–3539.
- Lewis, J. M., & Derber, J. C. (1985). The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus A*, 37, 309–322.
- Li, H., Kalnay, E., & Miyoshi, T. (2009). Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135, 523–533.
- Li, L., Zhou, H., Hendricks Franssen, H. J., & Gómez-Hernández, J. J. (2012). Groundwater flow inverse modeling in non-multiGaussian media: Performance assessment of the normal-score ensemble Kalman filter. *Hydrology and Earth System Sciences*, 16, 573–590.
- Liang, X., Zheng, X., Zhang, S., Wu, G., Dai, Y., & Li, Y. (2012). Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble Kalman filter assimilation. *Quarterly Journal of the Royal Meteorological Society*, 138, 263–273.
- Lien, G. Y., Kalnay, E., & Miyoshi, T. (2013). Effective assimilation of global precipitation: Simulation experiments. *Tellus A: Dynamic Meteorology and Oceanography*, 65, 19915.
- Lien, G.-Y., Kalnay, E., Miyoshi, T., & Huffman, G. J. (2016). Statistical properties of global precipitation in the NCEP GFS model and TMPA observations for data assimilation. *Monthly Weather Review*, 144, 663–679.
- Lien, G.-Y., Miyoshi, T., & Kalnay, E. (2016). Assimilation of TRMM multisatellite precipitation analysis with a low-resolution NCEP global forecast system. *Monthly Weather Review*, 144, 643–661.
- Lindskog, M., Dee, D., Tremolet, Y., Andersson, E., Radnoti, G., & Fisher, M. (2009). A weak-constraint four-dimensional variational analysis system in the stratosphere. *Quarterly Journal of the Royal Meteorological Society*, 135, 695–706.
- Lions, J. L. (1971). *Optimal control of systems governed by partial differential equations*. Berlin: Springer.
- Lisæter, K. A., Rosanova, J., & Evensen, G. (2003). Assimilation of ice concentration in a coupled ice-ocean model, using the ensemble Kalman filter. *Ocean Dynamics*, 53, 368–388.
- Liu, C., Xiao, Q., & Wang, B. (2008). An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Monthly Weather Review*, 136, 3363–3373.
- Liu, C., Xiao, Q., & Wang, B. (2009). An ensemble-based four-dimensional variational data assimilation scheme. Part II: Observing system simulation experiments with advanced research WRF (ARW). *Monthly Weather Review*, 137, 1687–1704.
- Liu, Y., Haussaire, J.-M., Bocquet, M., Roustan, Y., Saunier, O., & Mathieu, A. (2017). Uncertainty quantification of pollutant source retrieval: Comparison of Bayesian methods with application to the Chernobyl and Fukushima-Daiichi accidental releases of radionuclides. *Quarterly Journal of the Royal Meteorological Society*, 143, 2886–2901.
- Livings, D. M., Dance, S. L., & Nichols, N. K. (2008). Unbiased ensemble square root filters. *Physica D*, 237, 1021–1028.
- Lorenc, A. (1986). Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112, 1177–1194.
- Lorenc, A. (2013). *Recommended nomenclature for EnVar data assimilation methods*. Retrieved from http://www.wcrp-climate.org/WGNE/BlueBook/2013/individual-articles/01__Lorenc__Andrew__EnVar__nomenclature.pdf
- Lorenc, A. C. (2003). The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 129, 3183–3203.
- Lorenc, A. C., Bowler, N. E., Clayton, A. M., Pring, S. R., & Fairbairn, D. (2015). Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Monthly Weather Review*, 143, 212–229.
- Lorenc, A. C., & Payne, T. (2007). 4D-Var and the butterfly effect: Statistical four-dimensional data assimilation for a wide range of scales. *Quarterly Journal of the Royal Meteorological Society*, 133, 607–614.
- Lorenz, E. (1963). Deterministic non-periodic flow. *Journal of the Atmospheric Sciences*, 20, 130–141.
- Lorenz, E. N. (1996). Predictability—A problem partly solved. In *Proceedings 1995 ECMWF seminar on predictability* (Vol. 1). Reading, UK.
- Lorenz, E. N. (2005). Designing chaotic models. *Journal of the Atmospheric Sciences*, 62, 1574–1587.
- Lorenz, E. N., & Emanuel, K. A. (1998). Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of the Atmospheric Sciences*, 55, 399–414.
- Lu, F., Liu, Z., Zhang, S., & Liu, Y. (2015). Strongly coupled data assimilation using leading averaged covariance (LACC). Part I: Simple model study. *Monthly Weather Review*, 143, 3823–3837.
- MacKay, D. J. C. (2003). *Information theory, inference and learning algorithms*. Cambridge, England: Cambridge University Press.
- Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., & Stockdale, T. (2013). Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dynamics*, 41, 2393–2409.
- Mandel, J., Bergou, E., Gürol, S., Gratton, S., & Kasanický, I. (2016). Hybrid Levenberg-Marquardt and weak-constraint ensemble Kalman smoother method. *Non-linear Processes in Geophysics*, 23, 59–73.
- Miller, R., Ghil, M., & Gauthiez, F. (1994). Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the Atmospheric Sciences*, 51, 1037–1056.
- Mitter, S. K., & Newton, N. J. (2005). Information and entropy flow in the kalman–bucy filter. *Journal of Statistical Physics*, 118, 145–176.
- Miyoshi, T. (2011). The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Monthly Weather Review*, 139, 1519–1535.
- Miyoshi, T., Kalnay, E., & Li, H. (2013). Estimating and including observation-error correlations in data assimilation. *Inverse Problems in Science and Engineering*, 21, 387–398.
- Morozfeld, M., Tu, X., Atkins, E., & Chorin, A. J. (2012). A random map implementation of implicit filters. *Journal of Computational Physics*, 231, 2049–2066.
- Natvik, L. J., & Evensen, G. (2003). Assimilation of ocean colour data into a biochemical model of the North Atlantic. Part 1. Data assimilation experiments. *Journal of Marine Systems*, 40–41, 127–153.
- Nerger, L., Janjić, T., Schröter, J., & Hiller, W. (2012). A unification of ensemble square root Kalman filters. *Monthly Weather Review*, 140, 2335–2345.
- Ng, G.-H. C., McLaughlin, D., Entekhabi, D., & Ahanian, A. (2011). The role of model dynamics in ensemble Kalman filter performance for chaotic systems. *Tellus A*, 63, 958–977.
- Ngodock, H., & Carrier, M. (2014). A 4DVAR system for the Navy Coastal Ocean Model. Part II: Strong and weak constraint assimilation experiments with real observations in Monterey Bay. *Monthly Weather Review*, 142, 2108–2117.
- Ni, B., & Zhang, Q. (2016). Stability of the Kalman filter for continuous time output error systems. *Systems and Control Letters*, 94, 172–180.
- Nicolis, C. (2003). Dynamics of model error: Some generic features. *Journal of the Atmospheric Sciences*, 60, 2208–2218.
- Nino Ruiz, E. D., & Sandu, A. (2016). A derivative-free trust region framework for variational data assimilation. *Journal of Computational and Applied Mathematics*, 293, 164–179.
- Nodet, M. (2006). Variational assimilation of lagrangian data in oceanography. *Inverse Problems*, 22, 245–263.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., ... Yorke, A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A*, 56, 415–428.
- Palatella, L., Carrassi, A., & Trevisan, A. (2013). Lyapunov vectors and assimilation in the unstable subspace: Theory and applications. *Journal of Physics A: Mathematical and Theoretical*, 46, 254020.

- Palatella, L., & Trevisan, A. (2015). Interaction of Lyapunov vectors in the formulation of the nonlinear extension of the Kalman filter. *Physical Review E*, 91(042), 905.
- Palatella, L., Trevisan, A., & Rambaldi, S. (2013). Nonlinear stability of traffic models and the use of lyapunov vectors for estimating the traffic state. *Physical Review E*, 88, 022901.
- Pazo, D., Carrassi, A., & Lopez, J. (2016). Data assimilation by delay-coordinate nudging. *Quarterly Journal of the Royal Meteorological Society*, 142, 1290–1299.
- Penenko, V., & Obraztsov, N. (1976). A variational initialization method for the fields of the meteorological elements. *Soviet Meteorology and Hydrology*, 11, 1–11.
- Penny, S. G., & Hamill, T. M. (2017). Coupled data assimilation for integrated earth system analysis and prediction. *Bulletin of the American Meteorological Society*, 98, ES169–ES172.
- Penny, S. G., & Miyoshi, T. (2016). A local particle filter for high dimensional geophysical systems. *Nonlinear Processes in Geophysics*, 23, 391–405.
- Pham, D. T. (2001). Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, 129, 1194–1207.
- Pham, D. T., Verron, J., & Roubaud, M. C. (1998). A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16, 323–340.
- Pires, C., Vautard, R., & Talagrand, O. (1996). On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus A*, 48, 96–121.
- Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., et al. (2016). Era-20c: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, 29, 4083–4097.
- Poterjoy, J. (2016). A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, 144, 59–76.
- Poterjoy, J., & Zhang, F. (2015). Systematic comparison of four-dimensional data assimilation methods with and without the tangent linear model using hybrid background error covariance: E4DVar versus 4DEnVar. *Monthly Weather Review*, 143, 1601–1621.
- Pulido, M., Tandee, P., Bocquet, M., Carrassi, A., & Lucini, M. (2018). Stochastic parameterization identification using ensemble kalman filtering combined with maximum likelihood methods. *Tellus A*, 70, 1442099.
- Quinn, J. C., & Abarbanel, H. D. (2010). State and parameter estimation using Monte Carlo evaluation of path integrals. *Quarterly Journal of the Royal Meteorological Society*, 136, 1855–1867.
- Raanes, P. N. (2016). On the ensemble rauch-tung-striebel smoother and its equivalence to the ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, 142, 1259–1264.
- Raanes, P. N., Carrassi, A., & Bertino, L. (2015). Extending the square root method to account for additive forecast noise in ensemble methods. *Monthly Weather Review*, 143, 3857–3873.
- Rampal, P., Bouillon, S., Ólason, E., & Morlighem, M. (2016). neXtSIM: A new Lagrangian Sea ice model. *The Cryosphere*, 10, 1055–1073.
- Raynaud, L., Berre, L., & Desroziers, G. (2009). Objective filtering of ensemble-based background-error variances. *Quarterly Journal of the Royal Meteorological Society*, 135, 1177–1199.
- Raynaud, L., Berre, L., & Desroziers, G. (2011). An extended specification of flow-dependent background error variances in the Météo-France global 4D-Var system. *Quarterly Journal of the Royal Meteorological Society*, 137, 607–619.
- Reich, S. (2013). A nonparametric ensemble transform method for Bayesian inference. *SIAM Journal on Scientific Computing*, 35, A2013–A2014.
- Reich, S., & Cotter, C. (2015). *Probabilistic forecasting and Bayesian data assimilation*. Cambridge: Cambridge University Press.
- Robert, C., Durbiano, S., Blayo, E., Verron, J., Blum, J., & Le Dimet, F. X. (2005). A reduced-order strategy for 4D-Var data assimilation. *Journal of Marine Systems*, 57, 70–82.
- Robert, S., & Künsch, H. R. (2017). Localizing the ensemble Kalman particle filter. *Tellus A*, 69, 1282016.
- Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., ... Goldberg, M. (2010). The ncep climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91, 1015–1057.
- Sakov, P., & Bertino, L. (2011). Relation between two common localisation methods for the EnKF. *Computational Geosciences*, 15, 225–237.
- Sakov, P., Counillon, F., Bertino, L., Lisæter, K., Oke, P., & Koralev, A. (2012). TOPAZ4: An ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Science*, 8, 633–656.
- Sakov, P., Evensen, G., & Bertino, L. (2010). Asynchronous data assimilation with the EnKF. *Tellus Series A*, 62A, 24–29.
- Sakov, P., & Oke, P. R. (2008a). Implications of the form of the ensemble transform in the ensemble square root filters. *Monthly Weather Review*, 136, 1042–1053.
- Sakov, P., & Oke, P. R. (2008b). A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters. *Tellus Series A*, 60, 361–371.
- Sakov, P., Oliver, D. S., & Bertino, L. (2012). An iterative EnKF for strongly nonlinear systems. *Monthly Weather Review*, 140, 1988–2004.
- Salman, H., Kuznetsov, L., Jones, C., & Ide, K. (2006). A method for assimilating Lagrangian data into a shallow-water-equation ocean model. *Monthly Weather Review*, 134, 1081–1101.
- Santitissadeekorn, N., & Jones, C. (2015). Two-stage filtering for joint state-parameter estimation. *Monthly Weather Review*, 143, 2028–2042.
- Sasaki, Y. (1970). Some basic formalism in numerical variational analysis. *Monthly Weather Review*, 98, 875–883.
- Schölzel, C., & Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research-introduction to the copula approach. *Nonlinear Processes in Geophysics*, 15, 761–772.
- Simon, E., & Bertino, L. (2009). Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: A twin experiment. *Ocean Science*, 5, 495–510.
- Simon, E., & Bertino, L. (2012). Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: Application to a 1D ocean ecosystem model. *Journal of Marine Systems*, 89, 1–18.
- Simon, E., Samuelsen, A., Bertino, L., & Mouysset, S. (2015). Experiences in multiyear combined state-parameter estimation with an ecosystem model of the North Atlantic and Arctic oceans using the ensemble Kalman filter. *Journal of Marine Systems*, 152, 1–17.
- Slivinski, L., & Snyder, C. (2016). Exploring practical estimates of the ensemble size necessary for particle filters. *Monthly Weather Review*, 144, 861–875.
- Slivinski, L., Spiller, E., Apte, A., & Sandstede, B. (2015). A hybrid particle-ensemble Kalman filter for Lagrangian data assimilation. *Monthly Weather Review*, 143, 195–211.
- Sluka, T. C., Penny, S. G., Kalnay, E., & Miyoshi, T. (2016). Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation. *Geophysical Research Letters*, 43, 752–759.
- Smith, D. M., Eade, R., & Pohlmann, H. (2013). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate Dynamics*, 41, 3325–3338.
- Smith, D. M., & Murphy, J. M. (2007). An objective ocean temperature and salinity analysis using covariances from a global climate model. *Journal of Geophysical Research: Oceans*, 112(C2).
- Smith, P. J., Fowler, A. M., & Lawless, A. S. (2015). Exploring strategies for coupled 4D-Var data assimilation using an idealised atmosphere–ocean model. *Tellus A*, 67, 27025.
- Snyder, C., Bengtsson, T., Bickel, P., & Anderson, J. L. (2008). Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136, 4629–4640.
- Snyder, C., Bengtsson, T., & Morzfeld, T. (2015). Performance bounds for particle filters using the optimal proposal. *Monthly Weather Review*, 143, 4750–4761.

- Stewart, L. M., Dance, S., & Nichols, N. (2008). Correlated observation errors in data assimilation. *International Journal for Numerical Methods in Fluids*, 56, 1521–1527.
- Stockdale, T. N. (1997). Coupled ocean–atmosphere forecasts in the presence of climate drift. *Monthly Weather Review*, 125, 809–818.
- Stordal, A. S., Karlsen, H. A., Nævdal, G., Skaug, H. J., & Vallès, B. (2011). Bridging the ensemble Kalman filter and particle filters: The adaptive Gaussian mixture filter. *Computational Geosciences*, 15, 293–305.
- Stott, P. A., Allen, M., Christidis, N., Dole, R. M., Hoerling, M., Huntingford, C., ... Stone, D. (2013). Attribution of weather and climate-related events. In G. R. Asrar & J. W. Hurrell (Eds.), *Climate science for serving society* (pp. 307–337). Netherlands: Springer.
- Sugiura, N., Awaji, T., Masuda, S., Mochizuki, T., Toyoda, T., Miyama, T., ... Ishikawa, Y. (2008). Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations. *Journal of Geophysical Research: Oceans*, 113(C10).
- Sun, C., Hao, Z., Ghil, M., & Neelin, J. D. (2002). Data assimilation for a coupled ocean–atmosphere model. Part I: Sequential state estimation. *Monthly Weather Review*, 130, 1073–1099.
- Talagrand, O. (1997). Assimilation of observations, an introduction (special issue—Data assimilation in meteorology and oceanography: Theory and practice). *Journal of the Meteorological Society of Japan*, 75, 191–209.
- Talagrand, O. (2010). Variational assimilation. In *Data assimilation: Making Sense of Observations* (pp. 41–67). Berlin Heidelberg: Springer-Verlag.
- Talagrand, O., & Courtier, P. (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quarterly Journal of the Royal Meteorological Society*, 113, 1311–1328.
- Tandeo, P., Pulido, M., & Lott, F. (2015). Offline parameter estimation using EnKF and maximum likelihood error covariance estimates: Application to a subgrid-scale orography parametrization. *Quarterly Journal of the Royal Meteorological Society*, 141, 383–395.
- Tardif, R., Hakim, G. J., & Snyder, C. (2014). Coupled atmosphere–ocean data assimilation experiments with a low-order climate model. *Climate Dynamics*, 43, 1631–1643.
- Tardif, R., Hakim, G. J., & Snyder, C. (2015). Coupled atmosphere–ocean data assimilation experiments with a low-order model and CMIP5 model data. *Climate Dynamics*, 45, 1415–1427.
- Thacker, C. (2007). Data assimilation with inequality constraints. *Ocean Modelling*, 16, 264–276.
- Thompson, P. D. (1969). Reduction of analysis error through constraints of dynamical consistency. *Journal of Applied Meteorology*, 8, 738–742.
- Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., & Whitaker, J. S. (2003). Ensemble square-root filters. *Monthly Weather Review*, 131, 1485–1490.
- Trémoulet, Y. (2006). Accounting for an imperfect model in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 132, 2483–2504.
- Trémoulet, Y. (2007). Model-error estimation in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 133, 1267–1280.
- Trevisan, A., D'Isidoro, M., & Talagrand, O. (2010). Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension. *Quarterly Journal of the Royal Meteorological Society*, 136, 487–496.
- Trevisan, A., & Palatella, L. (2011a). On the Kalman filter error covariance collapse into the unstable subspace. *Nonlinear Processes in Geophysics*, 18, 243–250.
- Trevisan, A., & Palatella, L. (2011b). Chaos and weather forecasting: The role of the unstable subspace in predictability and state estimation problems. *International Journal of Bifurcation and Chaos*, 21, 3389–3415.
- Trevisan, A., & Ubaldi, F. (2004). Assimilation of standard and targeted observations within the unstable subspace of the observation-analysis-forecast cycle. *Journal of the Atmospheric Sciences*, 61, 103–113.
- Ubaldi, F., & Kamachi, M. (2000). Time-space weak-constraint data assimilation for nonlinear models. *Tellus A*, 52, 412–421.
- Ubaldi, F., & Trevisan, A. (2006). Detecting unstable structures and controlling error growth by assimilation of standard and adaptive observations in a primitive equation ocean model. *Nonlinear Processes in Geophysics*, 16, 67–81.
- Ueno, G., & Nakamura, N. (2014). Iterative algorithm for maximum-likelihood estimation of the observation-error covariance matrix for ensemble-based filters. *Quarterly Journal of the Royal Meteorological Society*, 140, 295–315.
- Ueno, G., & Nakamura, N. (2016). Bayesian estimation of the observation-error covariance matrix in ensemble-based filters. *Quarterly Journal of the Royal Meteorological Society*, 142, 2055–2080.
- van Leeuwen, P. J. (2009). Particle filtering in geophysical systems. *Monthly Weather Review*, 137, 4089–4114.
- van Leeuwen, P. J., Cheng, Y., & Reich, S. (2015). *Nonlinear data assimilation* (Vol. 2). Cham, Switzerland; Heidelberg, Germany; New York, NY; Dordrecht, the Netherlands, and London, UK: Springer.
- Vannitsem, S. (2017). Predictability of large-scale atmospheric motions: Lyapunov exponents and error dynamics. *Chaos*, 27(032), 101.
- Vidard, P., Piacentini, A., & Le Dimet, F.-X. (2004). Variational data analysis with control of the forecast bias. *Tellus A*, 56, 177–188.
- Wackernagel, H. (2003). *Multivariate Geostatistics* (3rd ed.). Berlin, Germany: Springer Verlag.
- Wang, S., Xue, M., Schenkman, A. D., & Min, J. (2013). An iterative ensemble square root filter and tests with simulated radar data for storm-scale data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 139, 1888–1903.
- Wang, X., & Bishop, C. H. (2003). A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *Journal of the Atmospheric Sciences*, 60, 1140–1158.
- Wang, X., Bishop, C. H., & Julier, S. J. (2004). Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Monthly Weather Review*, 132, 1590–1605.
- Wang, X., Snyder, C., & Hamill, T. M. (2007). On the theoretical equivalence of differently proposed ensemble-3DVAR hybrid analysis schemes. *Monthly Weather Review*, 135, 222–227.
- Wang, Y., Counillon, F., & Bertino, L. (2016). Alleviating the bias induced by the linear analysis update with an isopycnal ocean model. *Quarterly Journal of the Royal Meteorological Society*, 142, 1064–1074.
- Weber, R. J., Carrassi, A., & Doblas-Reyes, F. J. (2015). Linking the anomaly initialization approach to the mapping paradigm: A proof-of-concept study. *Monthly Weather Review*, 143, 4695–4713.
- Whitaker, J. S., & Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130, 1913–1924.
- Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140, 3078–3089.
- Wiener, N. (1949). Extrapolation, interpolation and smoothing of stationary time series. In *With engineering applications*. Cambridge, MA: MIT Press.
- Wikle, C. K., & Berliner, L. M. (2007). A Bayesian tutorial for data assimilation. *Physica D*, 230, 1–16.
- Winiarek, V., Vira, J., Bocquet, M., Sofiev, M., & Saunier, O. (2011). Towards the operational estimation of a radiological plume using data assimilation after a radiological accidental atmospheric release. *Atmospheric Environment*, 45, 2944–2955.
- Wu, L., Mallet, V., Bocquet, M., & Sportisse, B. (2008). A comparison study of data assimilation algorithms for ozone forecasts. *Journal of Geophysical Research*, 113, D20310.
- Xie, J., Bertino, L., Counillon, F., Lisæter, K. A., & Sakov, P. (2017). Quality assessment of the topaz4 reanalysis in the arctic over the period 1991–2013. *Ocean Science*, 13, 123–144.
- Yang, S.-C., Corazza, M., Carrassi, A., Kalnay, E., & Miyoshi, T. (2009). Comparison of local ensemble transform Kalman filter, 3DVAR, and 4DVAR in a quasigeostrophic model. *Monthly Weather Review*, 137, 693–709.

- Yano, J., Ziemiański, M. Z., Cullen, M., Termonia, P., Onvlee, J., Bengtsson, L., ... Wyszogrodzki, A. A. (2018). Scientific challenges of convective-scale numerical weather prediction. *Bulletin of the American Meteorological Society*, 99, 699–710. <https://doi.org/10.1175/BAMS-D-17-0125.1>
- Ye, J., Kadakia, N., Rozdeba, P., Abarbanel, H., & Quinn, J. (2015). Improved variational methods in statistical data assimilation. *Nonlinear Processes in Geophysics*, 22, 205–213.
- Ying, M., & Zhang, F. (2015). An adaptive covariance relaxation method for ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 141, 2898–2906.
- Zhang, M., & Zhang, F. (2012). E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. *Monthly Weather Review*, 140, 587–600.
- Zhang, S., Harrison, M., Rosati, A., & Wittenberg, A. (2007). System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather Review*, 135, 3541–3564.
- Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., & Baklanov, A. (2012). Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects. *Atmospheric Environment*, 60, 656–676.
- Zheng, X. G. (2009). An adaptive estimation of forecast error covariance parameters for Kalman filtering data assimilation. *Advances in Atmospheric Sciences*, 26, 154–160.
- Zhou, H., Li, L., Hendricks Franssen, H.-J., & Gómez-Hernández, J. J. (2012). Pattern recognition in a bimodal aquifer using the normal-score ensemble Kalman filter. *Mathematical Geoscience*, 44, 169–185.
- Zhu, M., van Leeuwen, P.-J., & Amezcua, J. (2016). Implicit equal-weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, 142, 1904–1919.
- Zupanski, D. (1997). A general weak constraint applicable to operational 4DVAR data assimilation systems. *Monthly Weather Review*, 125, 2274–2292.
- Zupanski, M. (1996). A preconditioning algorithm for four-dimensional variational data assimilation. *Monthly Weather Review*, 124, 2562–2573.
- Zupanski, M. (2005). Maximum likelihood ensemble filter: Theoretical aspects. *Monthly Weather Review*, 133, 1710–1726.

How to cite this article: Carrassi A, Bocquet M, Bertino L, Evensen G. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Clim Change*. 2018;9:e535. <https://doi.org/10.1002/wcc.535>

APPENDIX A: SOME PROPERTIES OF THE KALMAN FILTER AND KALMAN SMOOTHING

Although the straightforward use of the KF and KS in geosciences is obviously hampered by the computational limitations and by the inconsistency of their statistical/dynamical hypotheses (Gaussianity and linearity), yet they represent the backbone of many practical DA algorithms. The history of the use of KF-like methods in geosciences is the one of a never-ending search for suitable approximations that, even if suboptimal, can still work satisfactorily in a nonlinear, non-Gaussian, and high-dimensional setting. We have seen in Section 4 how the KF has served as a key conceptual and factual framework upon which several successful operational DA methods have been built. This appendix reviews some of the key properties and issues of the KF and KS. Our discussion here mainly pertains to the KF, but most of the conclusions apply to the KS too.

Time-dependent prior

The KF analysis will be statistically closer to either the observations or the prior depending on their respective accuracy, that is, on our belief about them as estimated via the covariances \mathbf{R}_k and \mathbf{P}_k^f , respectively. In the geosciences, the number of observations, albeit large, is usually insufficient to fully cover the state space ($d \ll m$) so that much of the information is spread from observed to unobserved areas is controlled by the prior. Having an informative, accurate, and reliable prior is thus of great importance. As mentioned at the end of Section 2.1, the situation $d \ll m$ is endemic in NWP, and the use of a short-range numerical forecast in a cyclic DA procedure has been key to the success of DA in that context (see Daley, 1993; Kalnay, 2002). The KF recursion provides a time-dependent estimate of the prior (its mean, \mathbf{x}_k^f , and covariance, \mathbf{P}_k^f) that is highly desirable in environmental systems that are usually chaotic, so that the actual error associated with \mathbf{x}_k^f is itself strongly time dependent. We have seen in Section 5.1 that this property of the chaotic dynamics, while representing a challenge to the state estimation process, can also be exploited explicitly in the design of DA algorithms for this class of systems.

Filter divergence

Filter divergence is the name used to refer to the situation in which the solution of the KF deviates dramatically from the true signal that it was supposed to track, and the KF is not longer able to pull back its solution close to the truth (Fitzgerald, 1971; Harlim, Majda, et al., 2010). Filter divergence is often the result of progressive and repeated underestimation of the actual error, $\mathbf{P}_k^a < \mathbf{P}_k^{\text{truth}}$ (the matrices order relationship is that of the cone of the positive semidefinite matrices, Bocquet et al., 2017). Under the action of the dynamics, Equation (16), the analysis error covariance, \mathbf{P}_{k-1}^a , is transformed into the forecast one at the next time, \mathbf{P}_k^f . If the dynamical model is chaotic (or just unstable) then at least one of the eigenvalues of $\mathbf{M}_{k:k-1}$ is larger than one and $\mathbf{P}_k^f \geq \mathbf{P}_{k-1}^a$: the estimated error grows during the forecast phase. However, for generic stable dynamics such an

error growth is not guaranteed. At the analysis times, the term $(\mathbf{I}_k - \mathbf{K}_k \mathbf{H}_k)$ in Equations (18) and (19) represents the forcing due to the observations, and it has a stabilizing effect since its eigenvalues are bounded to be lower or equal to one (Carrassi, Ghil, Trevisan, & Uboldi, 2008). This implies that the estimated analysis error covariance is always smaller or equal to the forecast (prior) error covariance, $\mathbf{P}_k^a \leq \mathbf{P}_k^f$. The overall fate of the KF error covariance comes by the balances between the (possible) growth during the forecast phases and the (certain) decrease at analysis times. If the dynamics is not able to counteract the covariance decrease occurring at analysis times, the KF error covariance will progressively decrease, and once $\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T \ll \mathbf{R}_k$ the filter solution may start to ignore the observations. This is not an issue in itself, as long as the actual error is also decreasing and the KF solution is properly tracking the desired signal. Nevertheless, when this is not the case, that is, when the KF error estimates decrease but the actual error does not, the KF solution starts to deviate from the observations, eventually diverging completely from the true signal.

Several factors may be at the origin of filter divergence, notably in the misspecification in the DA setup, such as a too strong influence from the measurements (from wrongly specified error statistics, neglected measurement error covariances, etc.), or wrongly specified or neglected model errors. Filter divergence also occurs in ensemble-based DA (see Section 4) and we have seen in Section 4.4 which countermeasures, inflation and localization, have been placed in order to deal with this issue in real applications.

A diagnostic tool

A remarkable property of the KF, originating from the linear and Gaussian assumptions, is that the error covariances, \mathbf{P}_k^f and \mathbf{P}_k^a do not depend on the observation values: they are thus unconditional covariances. This is a direct consequence of the first and second moments of the system's state pdf being independent from each other (and uncoupled with higher order moments); a behavior that no longer holds in nonlinear, non-Gaussian, scenarios. Another peculiar feature, which serves to monitor the goodness of the hypotheses, is that the innovation vector sequence, $\mathbf{v}_k = \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^f$, is Gaussian and uncorrelated in time (Jazwinski, 1970): one can thus keep checking the innovations and, possibly, to implement corrections (Daley, 1993).

Bias and covariance estimation

The optimality of the KF relies upon the veracity of its assumptions: the linearity of the model and observational operator, and the Gaussianity of the true error pdfs. Any mismatch between the real conditions on which the KF operates and its working hypotheses will negate its optimality. Nevertheless, even when the hypotheses are correct, the KF will still depend on the correct specification of its statistical inputs: the model and observational error covariances. The initial conditions, \mathbf{x}_0^a and \mathbf{P}_0^a , are also input but their impact on the filter performance is discussed separately in the following paragraph.

Recall from Equations (13–14) that the model and observational error are assumed unbiased and Gaussian, $\eta_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ and $\epsilon_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$. If either the actual model or observational errors are biased (or both), the KF analysis will be biased too, unless those biases are removed from the forecast before the analysis update, Equations (18–19). These biases can be estimated recursively in time, along with the system's state, using an approach known as state augmentation in which the state is formally augmented with the bias term (e.g., Dee, 2005). The state augmentation strategy is also the classical choice to deal with the simultaneous model state and parameter estimation (Jazwinski, 1970).

Likewise, discrepancies can also be present between the actual model and observation error covariances and those stipulated in the filter setup. In contrast to the bias, the covariances cannot be corrected using the state augmentation approach, and an additional procedure is required. A possibility is again on the use of the innovations: when all error covariances entering the KF are correct, the innovations are distributed according to $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \Sigma_k)$, with $\Sigma_k = \mathbf{H}_k (\mathbf{M}_{k:k-1} \mathbf{P}_{k-1}^a \mathbf{M}_{k:k-1}^T + \mathbf{Q}_k) \mathbf{H}_k^T + \mathbf{R}_k$ (Cohn, 1997). It is then possible, in principle, to estimate the “best” \mathbf{Q}_k and/or \mathbf{R}_k as those maximizing the conditional probability, $p(\mathbf{v}_k | \mathbf{Q}_k, \mathbf{R}_k)$, where the innovation is treated as a random variable (Dee, 1995). Given the large dimension of \mathbf{Q}_k and \mathbf{R}_k , such a maximum likelihood approach can only be feasible if \mathbf{Q}_k and \mathbf{R}_k are parametrized based on a very small number of parameters.

In any case, suitable parametrizations of the covariance matrices are necessary, particularly for model error, given the huge size of the geophysical models and the wide range of possible error sources. The former problem implies the need to estimate large matrices based on a limited number of available observations. The second is related to the multiple sources of model error, such as incorrect parametrization, numerical discretization, and the lack of description of some relevant scale of motion, which makes it difficult to set a unified parametrization. Recent works have proposed efficient combinations of Bayesian estimation procedures with Monte Carlo approximation to estimate both the observational and model error covariances (Dreano et al., 2017; Liu et al., 2017; Pulido, Tandeo, Bocquet, Carrassi, & Lucini, 2018; Ueno & Nakamura, 2014, 2016).

The computation of the Kalman gain, Equation (17), requires the inversion of the matrix $(\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \in \mathbb{R}^{d \times d}$. To make it computationally tractable, \mathbf{R}_k is often assumed to be diagonal and full rank, that is, observations are assumed to be spatially uncorrelated. The estimation of \mathbf{R}_k is reduced to the task of specifying only its diagonal. It can also negatively affect the filter's performance when observations are spatially correlated, which is typically for remotely sensed data. The impact of neglecting observational error correlations, as well as approaches to include them efficiently in the DA setup, has been studied in several works (see Miyoshi, Kalnay, & Li, 2013; Stewart, Dance, & Nichols, 2008).

Dependence on the initial condition

The criticality of the choice of the initial error covariance, \mathbf{P}_0 , is related to the filter's stability, intended as the convergence of its solutions to an asymptotic sequence, independently of the initial conditions (Gelb, 1974). Stability is a very desirable practical property: a stable filter will always tend to a steady solution and all unwanted errors in the specification of the initial conditions, \mathbf{x}_0 and \mathbf{P}_0 , do not alter its final output. Nevertheless, optimality of the filter alone does not guarantee stability but, for a stochastically-driven dynamical system as in Equation (13), it also requires the filter to be (uniformly) *observable* and *controllable* (see Cohn & Dee, 1988; Jazwinski, 1970; Kalman, 1960). Roughly, observability is the condition that, given sufficiently many observations, the initial state of the system can be reconstructed by using a finite number of observations (Quinn & Abarbanel, 2010). To see this, let consider the case of a discrete, autonomous (i.e., constant, $\mathbf{M}_k = \mathbf{M}$), and deterministic dynamical model $\mathbf{x}_k = \mathbf{M}\mathbf{x}_{k-1}$ of dimension n , that is observed n -times, without error, with scalar measurements and a linear operator, so that $y_k = \mathbf{Hx}_k$ (the operator \mathbf{H} is in this case a n -dimensional row vector). Starting from the initial condition at t_0 , we have $y_0 = \mathbf{Hx}_0$, $y_1 = \mathbf{Hx}_1 = \mathbf{HMx}_0$, and so on until, $y_{n-1} = \mathbf{Hx}_{n-1} = \mathbf{HM}^{n-1}\mathbf{x}_0$, that can be written compactly as

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{bmatrix} = \begin{bmatrix} \mathbf{H} \\ \mathbf{HM} \\ \vdots \\ \mathbf{HM}^{n-1} \end{bmatrix} \mathbf{x}_0 = \boldsymbol{\Psi}^T \mathbf{x}_0.$$

We see therefore that, if one wants to determine uniquely the initial state, \mathbf{x}_0 , based on the observations, the $n \times n$ matrix $\boldsymbol{\Psi}$ must be invertible, that is to say its rank must be equal to n , or equivalently its determinant must be nonzero. In this case, the system is said to be observable by the sequence of observations $y_0 \dots y_{n-1}$. As an example, consider the simple 2×2 system

$$\mathbf{M} = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$$

such that the dynamics of the first component depends on both the first and second components, while the second component depends only on itself. It is easy to show that observing the first component alone (i.e., $\mathbf{H} = [1 \ 0]$, a two-dimensional row vector), the corresponding 2×2 matrix $\boldsymbol{\Psi}$ has determinant equal to 1, is therefore invertible, and the system is observable. On the other hand, if it is second component to be observed (i.e., $\mathbf{H} = [0 \ 1]$), the determinant of $\boldsymbol{\Psi}$ is zero and the system is not observable. This result is physically interpretable such that, given that the first component carries also information about the second, but not vice versa, its observation is more effective in informing about the full two-dimensional system. Similarly to observability, controllability can be described as the ability to move the system from any initial state to a desired one over a finite time interval, and is related to the properties of the system noise, \mathbf{Q}_k (see Gelb, 1974, for a complete discussion on observability and controllability with several examples).

The KF stability and convergence for purely deterministic systems (i.e., like in Equation (13) but with $\mathbf{Q}_k = \mathbf{0}$), under the sole condition of uniform observability has been recently proved by Ni and Zhang (2016) and further characterized in terms of the stability properties of the dynamics by Carrassi, Ghil, et al. (2008) and Gurumoorthy et al. (2017). The generalization to the case of degenerate (rank deficient) initial condition error covariance is given in Bocquet et al. (2017), thus corroborating reduced-rank formulations of the KF based on the system's unstable modes (Trevisan & Palatella, 2011a).

APPENDIX B: MINIMIZATION PROCESS IN VARIATIONAL METHODS

With the gradient, Equation (27), in hand, the minimization is iteratively solved searching for the state vector, \mathbf{x}_0^i , at the i th iteration that satisfies $\mathcal{J}(\mathbf{x}_0^i) < \mathcal{J}(\mathbf{x}_0^{i-1})$ (i.e., the amplitude of the cost function decreases from iteration $i - 1$ to i), and the process is repeated until a prespecified convergence criterion (a threshold on the amplitude of the gradient, or on the difference

of the cost function at two successive iterations, $\mathcal{J}(\mathbf{x}_0^{i+1}) - \mathcal{J}(\mathbf{x}_0^i)$) is verified. The new state at each iteration is updated as $\mathbf{x}_0^i = \mathbf{x}_0^{i-1} + \gamma^{i-1} \mathbf{v}^{i-1}$, with \mathbf{v}^i being the searching direction and γ^{i-1} the step size. The various minimization algorithms differ on how \mathbf{v}^i and the step size along it are chosen. When the searching direction is chosen to have an angle greater than 90° with respect to the gradient (i.e., $(\mathbf{v}^i)^T \nabla_{\mathbf{x}} \mathcal{J}(\mathbf{x}^i) < 0$), the minimization procedures are referred to as *descent methods*. The most common and straightforward descent methods are the *steepest* and the *Newton method*.

In the former, \mathbf{v}^i is taken as opposite to the gradient. This strategy works very well when the cost function is uniformly strictly convex (i.e., the Hessian of the cost function is positive definite and it has at most one global minimum), in which case the gradient at any arbitrary point always heads to the absolute (and unique) minimum of the cost function, and the rate of convergence is linear. The computational cost of each iteration is relatively low, but the linear convergence can be so slow that the difference, $\mathbf{x}_0^i - \mathbf{x}_0^{i-1}$, becomes smaller than computer precision. Furthermore, the assumption of a globally convex cost function is critical in geosciences applications (Miller, Ghil, & Gauthiez, 1994; Pires et al., 1996).

To cope with this, Newton's method assumes that the cost function can be locally approximated by a quadratic expansion around the state point, \mathbf{x}^i , $\mathcal{J}(\mathbf{x}) \approx \mathcal{J}^{\text{Newt}}(\mathbf{x}) = \mathcal{J}(\mathbf{x}^i) + \nabla_{\mathbf{x}} \mathcal{J}(\mathbf{x}^i)(\mathbf{x} - \mathbf{x}^i) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^i)^T \nabla_{\mathbf{x}}^2 \mathcal{J}(\mathbf{x}^i) (\mathbf{x} - \mathbf{x}^i)$. The state at i th iteration is found by setting the gradient of this approximation to zero, which gives $\mathbf{v}^i = -\nabla_{\mathbf{x}}^{-2} \mathcal{J}^{\text{Newt}}(\mathbf{x}^i) \nabla_{\mathbf{x}} \mathcal{J}^{\text{Newt}}(\mathbf{x}^i)$; the search direction is equal to the opposite of the Hessian matrix of the cost function multiplied by its gradient. At the minimum, the Hessian of the cost function is positive definite so that the search direction verifies the condition of being oriented with an angle greater than 90° from the gradient. In practice, and in contrast to steepest descent, the Newton method uses also the local information about the curvature of the cost function in order to better point toward its minimum. Although the convergence of the Newton's method is quite rapid, its operational use in geophysical DA is rendered difficult by the need to invert the Hessian matrix, which is usually huge size and ill conditioned.

Minimization algorithms used operationally are a trade off between efficiency and computational limitation and have features that mimic those of the two main algorithms just described. A throughout description of the state-of-art minimization methods goes beyond the scope of this article but interested readers can find more details in, for example, Fisher and Andersson (2001) or Asch et al. (2016).

APPENDIX C: COMMENTS ON THE VARIATIONAL METHODS

The Gaussian hypothesis has not just allowed to get an analytic expression for the cost function, Equation (29) or (30), but it also offered a statistical, and physically plausible, interpretation of the analyzed trajectory. Given the unimodality of the Gaussian pdf, the most likely state is also the mean of the pdf, that is to say the minimum variance estimate. Without unimodality the mean state, while still having minimum variance, may well be of scarce relevance (it may fall in very low probability region) or not have physical plausibility at all.

In deriving the 4DVar, either in the weak or strong constraint formulations, no assumptions have been made about the characteristics of the dynamical and observational models: they can be assumed nonlinear and so they are in many real applications. Nevertheless, whether or not the latter is actually the case, it has enormous consequences on the accuracy of the 4DVar analysis, as well as on the complexity of the algorithms used to solve it. When both models are linear, all errors are Gaussian and independent, the 4DVar cost function is quadratic. If furthermore the Hessian of the cost function is spherical, the gradient will depend linearly on the control variable and will correctly point to the cost function (unique) global minimum. In this linear case, the 4DVar solution will match exactly the mean solution of a KS, Equation (21), to which the same input statistics are provided, and it will thus represent an alternative way to get the best mean estimate without the explicit need to compute inverse matrices as in the KS (Fisher, Leutbecher, & Kelly, 2005).

In the general nonlinear case, however, the exact minimum-variance solution may not be obtained. The approximate analysis will be the outcome of the minimization process and the degree of its accuracy will strongly depend on the degree of nonlinearity in the dynamical and observational models, even if the initial condition and observational error are Gaussian. The cost function will no longer be quadratic and it may possess multiple minima to which the minimization procedure can wrongly be trapped. A number of fixes have been proposed and put in place to overcome this issue, so as to render the cost function "more quadratic," notably by the use of a precondition under the form of a suitable invertible control variable transformation (Zupanski, 1996). It is beyond our scopes to expand further on this subject, but the readers can find more details in the literature (Asch et al., 2016; Talagrand, 2010).

The variational approach does not automatically solve the complete Gaussian estimation problem: it does not provide the two moments, the mean and the covariance, of the posterior distribution, but only the first one. It is possible to show that the analysis error covariance is indeed given, exactly/approximately for the linear/nonlinear case, respectively, by the inverse Hessian matrix of the cost function; at its minimum the (inverse) Hessian must be positive definite (see Appendix

B), consistently with a feature of a covariance matrix. Nevertheless, estimating the Hessian matrix for a realistic geophysical application is extremely difficult, and usually the same (fixed in time) error covariance matrix is used to characterize the background errors at the beginning of each DA cycle. When solving the s4DVar, the background error covariance is implicitly evolved within the window so that, effectively, a dynamically evolved estimate of the prior error is used at the observation times (Pires et al., 1996), but such an updated covariance is not explicitly accessible to initialize the next cycle.

This inherent limitation of the variational approach marks a key distinction with respect to sequential methods such as the KF or KS, that provide a time-dependent description of the uncertainty associated to the state estimate. This aspect has largely, but not solely, contributed to the popularity of KF-like approaches for DA with chaotic models (Vannitsem, 2017) where a time-dependent description of the estimation error is highly desirable (c.f., Section 5.1). We have seen in Section 4.5 that the recent promising efforts toward hybrid variational-ensemble methods are also aiming to cope with this issue, thus endowing the 4DVar with a flow-dependent estimate of the error covariance (Buehner et al., 2015a; Kleist & Ide, 2015; Lorenz et al., 2015).

APPENDIX D: SOME POPULAR APPROXIMATIONS

We describe briefly some of the early successful approximations of the KF and of the variational approach that have made their implementation possible in the geosciences.

Extended Kalman filter

The EKF represents a first-order expansion of the KF and extends its use to nonlinear dynamics (Jazwinski, 1970). Like KF, it is sequential: the system's state and associated error covariance are updated at discrete observation times and evolved in between them. In the EKF, the mean state estimate is propagated by the full nonlinear model, but the error covariance evolution is approximated using the tangent linear one. The linearization is taken around the nonlinear model solution, so that the Jacobian of the model is evaluated upon it and it is thus state dependent.

As with the standard KF for linear dynamics, the EKF also assumes that errors are all Gaussian distributed. Nevertheless, under the action of the nonlinear dynamics, even a possible initial Gaussian error covariance will not stay Gaussian, and the EKF will only provide an approximate description of the actual estimation error distribution. In general, the accuracy of the EKF scales with the degree of nonlinearity in the model (Miller et al., 1994). For instance, Evensen (1992) implemented the EKF with a multilayer ocean model finding that the tangent linear operator led to unbounded error growth since the nonlinear saturation that should occur at climatological level is contained in higher order moments of the error covariance equations, and those are all neglected in the closure used in the EKF.

The EKF has been successful in a number of pioneering applications of DA for meteorology (Dee, Cohn, Dalcher, & Ghil, 1985; Ghil, Cohn, Tavantzis, Bube, & Isaacson, 1981) and oceanography (Ghil & Malanotte-Rizzoli, 1991). It has also been used in one of the early study of CDA (c.f., Section 6.2) with an atmosphere–ocean model of intermediate complexity (Sun, Hao, Ghil, & Neelin, 2002). The joint state and parameter estimation is possible with the EKF using the state-augmentation approach and its efficiency for this purpose has been demonstrated in the context of DA for seasonal forecasts (Kondrashov, Sun, & Ghil, 2008) or land surface DA (see de Rosnay, Balsamo, Albergel, Muñoz-Sabater, & Isaksen, 2014). A formulation of the EKF for parameter estimation in the presence of time-correlated model error has been proposed by Carrassi and Vannitsem (2011) and later applied to a soil model (Carrassi, Hamdi, Termonia, & Vannitsem, 2012).

Along with the linear assumption on which it is built, another limitation of the EKF is due to the enormous computational requirements of the error covariance propagation. This involves the storage of full covariance matrices, the derivation of the tangent linear model, and its application a number of times twice the state vector dimension (Asch et al., 2016).

Incremental 4DVar and 3DVar

The incremental formulation (Courtier, Thépaut, & Hollingsworth, 1994) employs a linearization of the problem around the background trajectory: both the dynamical and observational models are linearized and the cost function of the incremental (strong constraint) 4DVar reads

$$\mathcal{J}^{\text{s4DVar-Incr}}(\delta\mathbf{x}_0) = \frac{1}{2} \sum_{k=0}^K \|\mathbf{v}_k - \mathbf{H}_k \mathbf{M}_{k:0} \delta\mathbf{x}_0\|_{\mathbf{R}_k^{-1}}^2 + \frac{1}{2} \|\delta\mathbf{x}_0\|_{\mathbf{B}^{-1}}^2, \quad (\text{A1})$$

where the increment, $\delta\mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}^b$, is now the control variable for the minimization, and $\mathbf{v}_k = \mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k) = \mathbf{y}_k - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\mathbf{x}_0)$ is the innovation vector (c.f., Appendix A).

The cost function is now quadratic, it possesses a unique absolute minimum, and it can be minimized much more easily. The minimization can be carried out by first computing the innovations (outer loop) using the nonlinear models, \mathcal{H} and \mathcal{M} . In the inner loop, Equation (85) is evaluated using the linearized models, \mathbf{H} and \mathbf{M} , and then the gradient using the adjoint \mathbf{M}^T . This procedure returns the analysis increment, $\delta\mathbf{x}_0$, to be used for the next outer loop and so on until convergence. The incremental 4DVar allows thus to deal with small nonlinearities in an incremental way, given that the linearized models are cyclically updated when a new outer loop trajectory is computed. Usually a simplified version of the model (coarser resolutions, simplified physics, etc.) is used in the inner loop (Lawless, Nichols, Boess, & Bunse-Gerstner, 2008), and this feature along with the quadratic form of the cost function, have been pivotal for the operational implementation of the incremental 4DVar (see Courtier et al., 1994).

The 3DVar is a special case of the 4DVar where the time dimension is removed and only the observations at the analysis time are assimilated (see Figure 2 and Kalnay (2002)). In this case, the control variable is the state at t_0 , \mathbf{x}_0 , like for the strong-constraint 4DVar, but in contrast to it, only the observations at t_0 are used in the update. In operational implementations of the 3DVar, all observations within a specific interval, $[t_0 - \Delta t, t_0 + \Delta t]$ (typically $\Delta t = 3$ hr), are used to update \mathbf{x}_0 . It is also worth to mention the first guess at appropriate time 3D-Var (FGAT 3D-Var), in which the unity operator for the resolvent of the tangent linear system and its adjoint is used. In practice, FGAT 3D-Var has the form of a 4DVar but it reduces the necessary computations to those of a 3DVar Fisher and Andersson (2001).