Smoothing Problems in a Bayesian Framework and Their Linear Gaussian Solutions

EMMANUEL COSME

Université Joseph Fourier/LEGI, Grenoble, France

JACQUES VERRON AND PIERRE BRASSEUR

CNRS/LEGI, Grenoble, France

JACQUES BLUM AND DIDIER AUROUX

Université de Nice Sophia-Antipolis/LJAD, Nice, France

(Manuscript received 29 October 2010, in final form 14 July 2011)

ABSTRACT

Smoothers are increasingly used in geophysics. Several linear Gaussian algorithms exist, and the general picture may appear somewhat confusing. This paper attempts to stand back a little, in order to clarify this picture by providing a concise overview of what the different smoothers really solve, and how. The authors begin addressing this issue from a Bayesian viewpoint. The filtering problem consists in finding the probability of a system state at a given time, conditioned to some past and present observations (if the present observations are not included, it is a forecast problem). This formulation is unique: any different formulation is a smoothing problem. The two main formulations of smoothing are tackled here: the joint estimation problem (fixed lag or fixed interval), where the probability of a series of system states conditioned to observations is to be found, and the marginal estimation problem, which deals with the probability of only one system state, conditioned to past, present, and future observations. The various strategies to solve these problems in the Bayesian framework are introduced, along with their deriving linear Gaussian, Kalman filter-based algorithms. Their ensemble formulations are also presented. This results in a classification and a possible comparison of the most common smoothers used in geophysics. It should provide a good basis to help the reader find the most appropriate algorithm for his/her own smoothing problem.

1. Introduction

Geophysical data assimilation has been historically developed for the purpose of initializing a numerical weather prediction [i.e., to get the best possible estimate of the atmospheric state of the day, based on past and present (the only available) observations]. In terms of estimation theory, initialization is a filtering problem, the Bayesian formulation of which is to find the probability density function (PDF) of a state vector $\mathbf{x}$ at a time index k, $\mathbf{x}_k$, given the observations $\mathbf{y}$ from time 1 to time k, $\mathbf{y}_{1:k}$:

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}). \quad (1)$$

Under certain assumptions, the Kalman filter (Kalman 1960; Cohn 1997) provides the solution. Although geophysical problems rarely verify the appropriate assumptions, the relative simplicity of the Kalman filter algorithm has made it naturally emerge in the community of geophysics (e.g., Parrish and Cohn 1985). It has since undergone many approximations, developments, and tunings to make it applicable and increasingly efficient with high-dimensional, nonlinear systems (Evensen 1994; Dee 1995; Verlaan and Heemink 1997; Houtekamer and Mitchell 1998; Pham et al. 1998; Lermusiaux and Robinson 1999; Mitchell and Houtekamer 2000; Bishop et al. 2001; Houtekamer and Mitchell 2001; Hamill et al. 2001; Lermusiaux 2006; Zheng and Zhu 2008; Houtekamer et al. 2009; Simon and Bertino 2009; Brankart et al. 2009, 2010, 2011; Beal et al. 2010). Recently some fully non-Gaussian and/or nonlinear methods arose in the geophysical data assimilation setup, such as the maximum entropy methods (e.g., Bocquet 2005b), the particle filter (van Leeuwen 2009; Bocquet et al. 2010), or the rank histogram filter (Anderson 2010). These methods are

Corresponding author address: Emmanuel Cosme, Université Joseph Fourier/LEGI, BP53X, 38041 Grenoble CEDEX, France.
E-mail: emmanuel.cosme@hmg.inpg.fr

DOI: 10.1175/MWR-D-10-05025.1

not approximations to the Kalman filter. Although not yet applicable to high-dimensional models, they should become useful for specific, low-dimensional problems in the near future.

In addition to the forecast initialization problem (still topical), many other challenges now exist for geophysical data assimilation. In particular, reanalysis would benefit from data assimilation methods able to process observations subsequent to the states to be estimated, since such observations are available, contrary to the forecast issue. Various geophysical disciplines express their own needs: estimation of parameters such as chemical sources for atmospheric chemistry and air quality (Cosme et al. 2005; Bocquet 2005a,b,c), biological constants for biogeochemistry (Losa et al. 2003), adjustment of ocean forcings (Skandrani et al. 2009), and boundary conditions (Barth et al. 2010) for ocean dynamics. Many of these applications are performed retrospectively, like a reanalysis, so that "future" observations are available. In these cases, a smoother, which takes into account future observations in the estimation process, must do better than a filter. A smoother is also necessary for specific applications, such as multiresolution data assimilation involving graphical models (e.g., Fieguth et al. 1995; Willsky 2002; Zhou et al. 2008). Even for forecast initialization problems, smoothing theory has been introduced to design sequential assimilation schemes able to account for the precise timing of observations (Hunt et al. 2004; Buehner et al. 2010; Sakov et al. 2010). The geophysical community might find other uses of smoothers in the near future.

The smoothing problem consists in estimating a system state or a set of states, given past, present, and future observations. In the first case, the PDF to be found is

$$p(\mathbf{x}_k|\mathbf{y}_{1:K}), \tag{2}$$

where K is fixed and larger than k. In the second case, the PDF is a joint PDF. We restrict the discussion to the case where the states to be jointly estimated form a complete time series, from a time index 0 to K (any other combination of states defines a smoothing problem, as long as some states are estimated with future observations). The PDF to be identified is then

$$p(\mathbf{x}_{0:K}|\mathbf{y}_{1:K}), \tag{3}$$

where $\mathbf{x}_{0:K}$ is the gathering of the state vectors from 0 to K. This is actually called a fixed-interval smoothing problem, as the time interval is fixed here. A related problem is the fixed-lag smoothing one, for which the PDF is

$$p(\mathbf{x}_{k-L:k}|\mathbf{y}_{1:k}). \tag{4}$$

When k takes several consecutive values, each state is estimated using the past, present, and the same number L of future observations. Here L is called the lag. From now on, we shall refer to Eq. (2) as the *marginal* smoothing problem, and to Eqs. (3) and (4) as *joint* smoothing problems.

Both problems are directly connected by the marginalization rule:

$$p(\mathbf{x}_k|\mathbf{y}_{1:K}) = \int p(\mathbf{x}_{0:K}|\mathbf{y}_{1:K})\, d\mathbf{x}_0 \ldots d\mathbf{x}_{k-1} d\mathbf{x}_{k+1} \ldots d\mathbf{x}_K. \tag{5}$$

But to compute an estimate of $\mathbf{x}_k$, the mean for example, using the marginal PDF $[\int \mathbf{x}_k p(\mathbf{x}_k|\mathbf{y}_{1:K})\, d\mathbf{x}_k]$ or the joint PDF $[\int \mathbf{x}_k p(\mathbf{x}_{0:K}|\mathbf{y}_{1:K})\, d\mathbf{x}_{0:K}]$ provides different results in general. In the particular multivariate Gaussian case though, both estimates of the mean are identical: a marginal distribution is simply obtained by removing the vector variables to be marginalized out from the joint distribution. The various optimal linear smoothers are then strictly equivalent, provided that Kalman's linear/Gaussian assumptions are verified. This is rarely the case though. Dealing with nonlinear/non-Gaussian cases is increasingly frequent. While the model's resolution increases, nonlinearities express themselves with increasing importance. Gaussianity is far from being systematic (e.g., consider tracers) and is also altered by nonlinearities. In such a context, the joint and marginal smoothers may yield different estimates, so it is particularly important for anyone to precisely identify the problem to be solved before choosing a smoothing algorithm. Fully nonlinear/non-Gaussian smoothers (e.g., particle smoothers; Fearnhead et al. 2010) have not been developed or used yet in geophysics, but the need might come in the near future.

Several optimal linear smoother algorithms exist, which have been presented and applied to geophysical problems (Cohn et al. 1994; Ménard et al. 1996; van Leeuwen and Evensen 1996; Lermusiaux and Robinson 1999; van Leeuwen 1999; Lermusiaux et al. 2002; Cosme et al. 2010). These algorithms are generally referred to as the fixed-interval sequential smoother, the fixed-lag smoother, the ensemble smoother, the forward–backward [or the Rauch–Tung–Striebel (RTS)] smoother, and the two-filter smoother. They are all based on Kalman's hypotheses and the equations of the Kalman filter, and, beyond their algorithmic differences, differ from the filter only by handling cross covariances in time to correct the state vectors at some times using observations

at future times. In addition, they all fall into one of the two categories mentioned previously: joint or marginal. Sections 3 and 4 describe the smoothers that solve the joint problem and the marginal problem, respectively. For each of them, the resolution strategy is first described in a Bayesian framework. Then the linear Gaussian implementation is recalled. Most of them involve the Kalman filter. Finally, the ensemble formulations are given. (We focus on the formulations of smoothers that naturally provide error estimates. Consequently, variational and representer methods are not discussed here.) Some of them are new. Before coming to the smoothers, section 2 gathers some background information about estimation theory, hidden Markov chains, and Bayesian formulation and Kalman's solution for the filtering problem. To conclude, section 5 provides a summary and a discussion of some important issues for smoothers: algorithm complexity, application of localization, and their adequacy to various geophysical data assimilation problems.

2. From estimation theory to the Kalman filter

a. Background on estimation theory

For a detailed introduction to the Bayesian perspective to data assimilation, the reader should refer to Wikle and Berliner (2007). Here we simply recall the two fundamental rules of estimation theory.

1) Bayes' rule:

   A joint PDF of two random vectors $\mathbf{x}$ and $\mathbf{y}$ can be factored as

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y}). \tag{6}$$

This is, actually, the definition of the conditional PDF $p(\mathbf{x}|\mathbf{y})$. Using the similar equation for $p(\mathbf{y}|\mathbf{x})$, Bayes' rule is obtained:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}. \tag{7}$$

In practice, the denominator is only a normalization factor and it is generally sufficient to consider proportionality: $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x})p(\mathbf{y}|\mathbf{x})$.

2) Marginalization rule:

   This rule is used to compute the PDF of a random vector $\mathbf{x}$ from the joint PDF of two or more vectors (including $\mathbf{x}$):

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{y})\, d\mathbf{y} = \int p(\mathbf{x}|\mathbf{y})p(\mathbf{y})\, d\mathbf{y}. \tag{8}$$

b. Hidden Markov chain

Geophysical data assimilation is often posed as an estimation problem in a hidden Markov chain (e.g., Wikle and Berliner 2007; Ihler et al. 2007), that is, a Markov chain with partially and independently observed nodes. The chain is formed with the sequence $\mathbf{x}_{0:K} = \{\mathbf{x}_0, \ldots, \mathbf{x}_K\}$ of system states to estimate. Each state (but the first, in our presentation) is observed. The transition PDF between the states at times $k-1$ and k, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, is supposed to be known. When a dynamical model is available and the model noise is additive, the model equation is

$$\mathbf{x}_k = \mathcal{M}_{k-1,k}(\mathbf{x}_{k-1}) + \eta_{k-1,k}. \tag{9}$$

Here $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is simply determined from the PDF of the model error $\eta_{k-1,k}$. For the Kalman filter, it is assumed to be Gaussian with a zero mean and a covariance matrix generally noted $\mathbf{Q}_{k-1,k}$, and the model is assumed to be linear.

The measurement process is also Markovian, and the observation PDF at any time k, $p(\mathbf{y}_k|\mathbf{x}_k)$, is known. When an observation operator is available, with additive noise, the measurement equation takes the following form:

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \epsilon_k^o, \tag{10}$$

and $p(\mathbf{y}_k|\mathbf{x}_k)$ is determined from the PDF of the observation error ϵ_k^o. For the Kalman filter, it is assumed to be Gaussian with a zero mean and a covariance matrix generally noted $\mathbf{R}_k$, and the observation operator is assumed to be linear.

The Markov assumption is the rationale for the sequential approaches to filtering and smoothing [see Wikle and Berliner (2007), section 3].

c. Filtering

Filtering follows the very intuitive approach of starting from the (available) prior distribution of the initial state, $p(\mathbf{x}_0)$, and iterating forecast and analysis steps using marginalization and Bayes' rules:

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})\, d\mathbf{x}_{k-1} \text{ (forecast)}, \tag{11a}$$

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \propto p(\mathbf{x}_k|\mathbf{y}_{1:k-1})p(\mathbf{y}_k|\mathbf{x}_k) \text{ (analysis)}. \tag{11b}$$

When the model [Eq. (9)] and the observation operator [Eq. (10)] are linear and their noises are Gaussian with zero mean, this provides the Kalman filter equations. In particular, the analysis is performed with the so-called

TABLE 1. Equations of the Kalman filter. After initialization, the forecast and analysis steps alternate from $k = 1$ to K.

Initialization		
$\mathbf{x}_0^a$ and $\mathbf{P}_0^a$		KF0
Forecast step		
$\mathbf{x}_{k\|1:k-1}^f = \mathbf{M}_{k-1,k}\mathbf{x}_{k-1\|1:k-1}^a$	State propagation	KF1
$\mathbf{P}_{k\|1:k-1}^f = \mathbf{M}_{k-1,k}\mathbf{P}_{k-1\|1:k-1}^a\mathbf{M}_{k-1,k}^{\mathrm{T}} + \mathbf{Q}_{k-1,k}$	Covariance propagation	KF2
Analysis step		
$\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k\mathbf{x}_{k\|1:k-1}^f$	Innovation	KF3
$\mathbf{G}_k = \mathbf{H}_k\mathbf{P}_{k\|1:k-1}^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k$	Innovation error covariances	KF4
$\mathbf{K}_{k\|1:k-1} = \mathbf{P}_{k\|1:k-1}^f\mathbf{H}_k^{\mathrm{T}}\mathbf{G}_k^{-1}$	Kalman gain	KF5
$\mathbf{x}_{k\|1:k}^a = \mathbf{x}_{k\|1:k-1}^f + \mathbf{K}_{k\|1:k-1}\mathbf{d}_k$	Filter analysis	KF6
$\mathbf{P}_{k\|1:k}^a = (\mathbf{I} - \mathbf{K}_{k\|1:k-1}\mathbf{H}_k)\mathbf{P}_{k\|1:k-1}^f$	Filter analysis (covariances)	KF7

best linear unbiased estimate (BLUE) equations. The Kalman filter equations are recalled in Table 1 for the purpose of setting the notations and the background for the smoother algorithms described next. The notation adopted for the time indices in the state vectors and covariance matrices recalls the conditioning of the associated PDF. For example, $\mathbf{x}_{i|1:k}^a$ (which will be met later) represents the (analysis) state estimate at time i updated with the observations $\{\mathbf{y}_1, \ldots, \mathbf{y}_k\}$. The *forecast* and *analysis* superscripts, f and a, are then superfluous, but preserved to make understanding easier. They will be inappropriate only for the ensemble smoother description (section 3c), where other notations are chosen.

For most atmospheric or oceanic applications, the Kalman filter is actually implemented in an ensemble form (Evensen 2003), where the state PDF is represented by a limited (a few tens typically) number M of particles or members, noted $\mathbf{x}^{f,m}$ for the forecast (we drop the time index here for conciseness). One may compute the mean of this ensemble, $\overline{\mathbf{x}}^f = \sum\mathbf{x}^{f,m}/M$, and form the matrix of scaled anomalies $\mathbf{S}^f$, the mth column of which is defined by

$$(\mathbf{S}^f)_m = \frac{\mathbf{x}^{f,m} - \overline{\mathbf{x}}^f}{\sqrt{M-1}}. \tag{12}$$

Theoretically, the covariance matrix can be computed as $\mathbf{P}^f = \mathbf{S}^f\mathbf{S}^{f\mathrm{T}}$, but this is never done in practice. On the contrary, the ensemble Kalman filter (EnKF) takes full advantage of the square root representation of the covariance matrix. At the forecast step, the mean state and matrix propagation (KF1 and KF2 in Table 1) are replaced by the propagation of each ensemble member. The analysis, also performed for each member, can be written as

$$\mathbf{x}^{a,m} = \mathbf{x}^{f,m} + \mathbf{S}^f\boldsymbol{\gamma}^m. \tag{13}$$

The analysis correction then appears as a linear combination of the anomalies. Here $\boldsymbol{\gamma}^m$ is a vector containing the corresponding coefficients. It can be computed following different strategies (e.g., Tippett et al. 2003). For the sake of simplicity here, we mention only the basic form (useful here for the following presentation, but not convenient from the computational viewpoint):

$$\boldsymbol{\gamma}^m = (\mathbf{H}\mathbf{S}^f)^{\mathrm{T}}[(\mathbf{H}\mathbf{S}^f)(\mathbf{H}\mathbf{S}^f)^{\mathrm{T}} + \mathbf{R}]^{-1}(\mathbf{y}^m - \mathbf{H}\mathbf{x}^{f,m}), \tag{14}$$

where $\mathbf{y}^m$ denotes the perturbed observation.

3. Joint smoothing

a. Fixed-interval sequential smoother

This smoother finds the PDF given by Eq. (3) in a sequential way and belongs to the category of fixed-interval smoothers. With k scanning the interval $[0, K]$, the PDF is computed as

$$p(\mathbf{x}_{0:k}|\mathbf{y}_{1:k}) \propto p(\mathbf{x}_{0:k-1}|\mathbf{y}_{1:k-1})p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{y}_k|\mathbf{x}_k). \tag{15}$$

This decomposition is obtained using Bayes' rule, the definition of the conditional PDF and the Markov property of the system. The sequential character of the algorithm appears clearly when this equation is decomposed into a forecast step,

$$p(\mathbf{x}_{0:k}|\mathbf{y}_{1:k-1}) = p(\mathbf{x}_{0:k-1}|\mathbf{y}_{1:k-1})p(\mathbf{x}_k|\mathbf{x}_{k-1}), \tag{16}$$

and an analysis step,

$$p(\mathbf{x}_{0:k}|\mathbf{y}_{1:k}) \propto p(\mathbf{x}_{0:k}|\mathbf{y}_{1:k-1})p(\mathbf{y}_k|\mathbf{x}_k). \tag{17}$$

The forecast equation [Eq. (16)] differs from the filter forecast [Eq. (11a)] by the fact that marginalization on $\mathbf{x}_{k-1}$ is not applied. This is because $\mathbf{x}_{k-1}$ (and actually all the previous state vectors) is estimated jointly with $\mathbf{x}_k$. We note in passing that the dimension of $\mathbf{x}_{0:k}$, the multistep state vector to estimate, increases with k.

In the linear Gaussian framework, the fixed-interval sequential smoother is quite easy to derive using an augmented state vector approach (e.g., Anderson and Moore 1979). The algorithm is based on the Kalman filter, completed with a set of equations specific to the smoother. These equations are summarized in Table 2.

TABLE 2. Equations of the sequential fixed-interval and fixed-lag smoothers. They come in addition to the filter equations, except KF2 in Table 1, explicitly decomposed into the first two equations here. The smoother analyses are performed just after the filter analysis: the loop on k includes both the forecast and analysis steps (in other words, the two steps alternate as in the Kalman filter). Within each step, the equations (except for the second one) must be applied for every time i concerned by smoothing. For the sequential smoother, $i \in [0, k-1]$; for the fixed-lag smoother, $i \in [k-L, k-1]$.

Forecast step	
$\mathbf{P}^{fa}_{k,i\|1:k-1} = \mathbf{M}_{k-1,k}\mathbf{P}^{aa}_{k-1,i\|1:k-1}$	Cross-covariances propagation
$\mathbf{P}^{f}_{k\|1:k-1} = \mathbf{M}_{k-1,k}(\mathbf{P}^{fa}_{k,k-1\|1:k-1})^{\mathrm{T}} + \mathbf{Q}_{k-1,k}$	
Smoother analysis step	
$\mathbf{K}_{i\|1:k-1} = (\mathbf{H}_k\mathbf{P}^{fa}_{k,i\|1:k-1})^{\mathrm{T}}\mathbf{G}_k^{-1}$	Smoother gain
$\mathbf{x}^{a}_{i\|1:k} = \mathbf{x}^{a}_{i\|1:k-1} + \mathbf{K}_{i\|1:k-1}\mathbf{d}_k$	Smoother analysis (state)
$\mathbf{P}^{aa}_{k,i\|1:k} = (\mathbf{I} - \mathbf{K}_{k\|1:k-1}\mathbf{H}_k)\mathbf{P}^{fa}_{k,i\|1:k-1}$	Smoother analysis (covariance 1)
$\mathbf{P}^{a}_{i\|1:k} = \mathbf{P}^{a}_{i\|1:k-1} - \mathbf{K}_{i\|1:k-1}\mathbf{H}_k\mathbf{P}^{fa}_{k,i\|1:k-1}$	Smoother analysis (covariance 2)

They involve, in particular, cross-covariance matrices, defined as $\mathbf{P}^{fa}_{k,i|1:k-1} = E[\epsilon^{f}_{k|1:k-1}\epsilon^{a}_{i|1:k-1}{}^{\mathrm{T}}]$ where $E[\,]$ is the expectation operator, $\epsilon^{f}_{k|1:k-1}$ the filter forecast error at time k, and $\epsilon^{a}_{i|1:k-1}$ the smoother analysis error at time i when observations from 1 to $k-1$ have been included.

The ensemble approach to fixed-interval sequential smoothing has been introduced and tested on toy models by Evensen and van Leeuwen (2000), on a large-scale oceanic experiment by Brusdal et al. (2003), and recently used by Barth et al. (2010) for an ocean tides application. Extending the ensemble filter to smoothing is technically straightforward. With the matrix definitions of section 2c, it consists of implementing the following retrospective analysis equation, for each ensemble member:

$$\mathbf{x}^{a,m}_{i|1:k} = \mathbf{x}^{a,m}_{i|1:k-1} + \mathbf{S}^{a}_{i|1:k-1}\boldsymbol{\gamma}^{m}, \tag{18}$$

to be performed, at each observation time k, for the past times $i \in [0, k-1]$ [e.g., see Cosme et al. (2010) for the full derivation of this equation]. The retrospective analysis for each time i is initialized with the filter analysis at time i, $\mathbf{x}^{a,m}_{i|1:i}$, when it is computed. At the analysis step $i+1$, $\mathbf{x}^{a,m}_{i|1:i+1}$ is computed using Eq. (18), and so on. It is also possible to compute the smoother estimates offline, which is after the filter pass and without running the dynamical model. In this case, several equivalent implementation strategies exist (Ravela and McLaughlin 2007). Note that the vector $\boldsymbol{\gamma}^{m}$ in Eq. (18) is the same as in Eq. (13), which means that it is available at no extra cost: the smoother has almost the same CPU cost as the filter. But if one wants to save all the smoother estimates, the storage requirement grows quadratically with K, instead of linearly with K for the filter. This defect is generally corrected using a fixed-lag smoother, described next.

b. Fixed-lag smoother

The fixed-lag smoother is close to the sequential fixed-interval smoother described previously. The difference lies in the fact that the size of the state vector to be estimated is kept unchanged through time (except in the first initialization steps). Consequently, the oldest state is ruled out from the estimation process at each forecast step. In the Bayesian framework, this exclusion is obtained through marginalization. This slightly modifies the Bayesian formulation of the problem. The forecast step [Eq. (16)] becomes

$$p(\mathbf{x}_{k-L:k}|\mathbf{y}_{1:k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})\int p(\mathbf{x}_{k-L-1:k-1}|\mathbf{y}_{1:k-1})\,d\mathbf{x}_{k-L-1}, \tag{19}$$

and the analysis step [Eq. (17)] does not change, with $\mathbf{x}_{k-L:k}$ instead of $\mathbf{x}_{0:k}$.

Like the fixed-interval sequential smoother, the "standard" derivation of the fixed-lag smoother in the linear-Gaussian framework is based on an augmented state vector approach, as detailed by Anderson and Moore (1979), Simon (2006), or Zhu et al. (1999). Another original derivation and a table synthetizing the equations can be found in Cohn et al. (1994). It is partly reproduced in Table 2 here.

The ensemble approach to fixed-lag smoothing can be obtained from Evensen and van Leeuwen (2000). Zhu et al. (2003) have performed fixed-lag smoothing experiments with a meteorological data assimilation system based on the physical-space statistical analysis system (PSAS; Cohn et al. 1998). Khare et al. (2008) have examined the benefits of the ensemble Kalman smoother with the Lorenz-96 model and an atmospheric general circulation model. Cosme et al. (2010) have presented a square root form of the fixed-lag smoother and an application with a high-resolution, primitive equations ocean circulation model. The numerical implementation of the fixed-lag smoother strictly follows the fixed-interval sequential smoother, with Eq. (18), but with the retrospective analyses now restricted to the last L time steps: $i \in [k-L, k-1]$. This smoother is straightforward to implement, exhibits a negligible CPU cost (in addition to the filter), and the associated storage requirement grows only linearly with $K(L+1)$ instead of K for the filter. Then, it utterly fulfills the storage constraints of reanalysis. Surprisingly, though, it has

drawn little attention from meteorologists and oceanographers so far.

c. Ensemble smoother

The ensemble smoother was introduced by van Leeuwen and Evensen (1996) to solve the smoothing problem defined by Eq. (3). It has gained interest recently, though not in the exact terms of ''ensemble smoother,'' for four-dimensional or asynchronous ensemble filtering (Hunt et al. 2004; Sakov et al. 2010). The benefit is to use the innovations and state error statistics at the precise observation times within the assimilation window, to compute the filter correction at the end of the window. The ''original'' ensemble smoother is of the fixed-interval type. The strategy is to modify a background, joint PDF of the full sequence of states with the full set of observations, using Bayes' rule:

$$p(\mathbf{x}_{0:K}|\mathbf{y}_{1:K}) \propto p(\mathbf{x}_{0:K})p(\mathbf{y}_{1:K}|\mathbf{x}_{0:K}), \tag{20}$$

which can be reduced, thanks to the Markov hypothesis, to

$$p(\mathbf{x}_{0:K}|\mathbf{y}_{1:K}) \propto p(\mathbf{x}_{0:K}) \prod_{k=1}^{K} p(\mathbf{y}_k|\mathbf{x}_k). \tag{21}$$

In the linear Gaussian framework, the ensuing strategy would be to define an extended (four dimensional) state vector by gathering the state vectors from 0 to K, and form a background [mean, $\hat{\mathbf{x}}^b = (\mathbf{x}_0^b, \ldots, \mathbf{x}_K^{b\,\mathrm{T}})^{\mathrm{T}}$, and covariance matrix] for this vector, using the background at time 0 and the dynamical model; then, the analysis can be performed by implementing either Eq. (20) (global analysis) or Eq. (21) (sequential analysis). In the global option, one has to form the extended observation vector $\hat{\mathbf{y}} = (\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_K^{\mathrm{T}})^{\mathrm{T}}$, the corresponding covariance matrix, and the observation operator, then perform a BLUE analysis. In the sequential option, the observations are used one at a time and update all the components of the extended state vector. The equations are given in Table 3. The sequential analysis equations are close to those of the fixed-interval sequential smoother (Table 2). The main difference lies in the chronology of the operations.

Using an ensemble formulation, the process starts with the time integration of the ensemble over the time interval, without assimilating data, forming a set of background state vectors $\mathbf{x}_{0:K}^{b,1:m}$. Then the two strategies described previously can be adopted.

In the first (global) strategy, adopted by van Leeuwen and Evensen (1996), the background vectors are serially updated using the analysis equation:

$$\mathbf{x}_k^{a,m} = \mathbf{x}_k^{b,m} + \mathbf{S}_k^b \hat{\boldsymbol{\gamma}}^m, \tag{22}$$

TABLE 3. Equations of the ensemble smoother. Initialization is similar to the filter's. In a first step, the background trajectory equations are performed with $k = 1, K$ and $i = 1, k-1$. In a second step, the analysis is performed, following a global or a sequential strategy. In the second option, the process is initialized by $\mathbf{x}_k^a = \mathbf{x}_k^b$, $\mathbf{P}_k^a = \mathbf{P}_k^b$, and $\mathbf{P}_{k,i}^a = \mathbf{P}_{k,i}^b$. The loop is on k first. Within the k loop, i runs from 0 to K.

Background trajectory
$\mathbf{x}_k^b = \mathbf{M}_{k-1,k}\mathbf{x}_{k-1}^b$
$\mathbf{P}_{k,i}^b = \mathbf{M}_{k-1,k}\mathbf{P}_{k-1,i}^b$
$\mathbf{P}_k^b = \mathbf{M}_{k-1,k}(\mathbf{P}_{k,k-1}^b)^{\mathrm{T}} + \mathbf{Q}_{k-1,k}$
Analysis, global strategy
BLUE equations with $\hat{\mathbf{x}}^b = (\mathbf{x}_0^{b\mathrm{T}}, \ldots, \mathbf{x}_K^{b\mathrm{T}})^{\mathrm{T}}$ and $\hat{\mathbf{y}} = (\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_K^{\mathrm{T}})^{\mathrm{T}}$
Analysis, sequential strategy
$\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k\mathbf{x}_{k\|1:k-1}^a$
$\mathbf{G}_k = \mathbf{H}_k\mathbf{P}_{k\|1:k-1}^a\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k$
$\mathbf{K}_{i\|1:k-1} = (\mathbf{H}_k\mathbf{P}_{k,i\|1:k-1}^a)^{\mathrm{T}}\mathbf{G}_k^{-1}$
$\mathbf{x}_{i\|1:k}^a = \mathbf{x}_{i\|1:k-1}^a + \mathbf{K}_{i\|1:k-1}\mathbf{d}_k$
$\mathbf{P}_{k,i\|1:k}^a = (\mathbf{I} - \mathbf{K}_{k\|1:k-1}\mathbf{H}_k)\mathbf{P}_{k,i\|1:k-1}^a$
$\mathbf{P}_{i\|1:k}^a = \mathbf{P}_{i\|1:k-1}^a - \mathbf{K}_{i\|1:k-1}\mathbf{H}_k\mathbf{P}_{k,i\|1:k-1}^a$

where $\mathbf{S}_k^b$ is the matrix of scaled anomalies based on the ensemble $\mathbf{x}_k^{b,1:m}$, following Eq. (12). Contrary to the previously described sequential smoothers [Eq. (18)], where the ''background'' term contains the information from the observations $\mathbf{y}_{1:k-1}$, here the observational information is only and fully contained in the M-vector $\hat{\boldsymbol{\gamma}}^m$ (hence the term ''global''). The latter is computed as

$$\hat{\boldsymbol{\gamma}}^m = \sum_{l=1}^{K} \mathbf{S}_l^{b\mathrm{T}}\mathbf{H}_l^{\mathrm{T}}\mathbf{b}_l^m. \tag{23}$$

Denoting by s_l the number of observations at time l, $\mathbf{b}_l^m$ is an s_l vector, the lth part of a $(\sum_l s_l)$-vector $\hat{\mathbf{b}}^m$, itself the solution of the linear equation:

$$[\hat{\mathbf{H}}\hat{\mathbf{S}}(\hat{\mathbf{H}}\hat{\mathbf{S}})^{\mathrm{T}} + \hat{\mathbf{R}}]\hat{\mathbf{b}}^m = \hat{\mathbf{d}}^m. \tag{24}$$

Here $\hat{\mathbf{d}}^m$ is the vector of innovations extended over the time interval, computed from the observations and the background mth trajectory $\mathbf{x}_{1:K}^{b,m}$. The variable $\hat{\mathbf{S}}$ is the time-extended matrix of anomalies, formed by the arrangement of the K matrices $\mathbf{S}_k^b$ in a column, while $\hat{\mathbf{H}}$ is the corresponding, time-extended observation operator. Note that Eq. (24) must be solved M times. This is the most CPU-consuming calculation in the analysis. However, if one uses an eigenbasis observational update (or square root approach; Tippett et al. 2003; Brankart et al.

TABLE 4. Equations of the RTS smoother. They are applied after a Kalman filter pass over the interval. Here k runs from $K-1$ down to 1.

$$\mathbf{K}_k^s = \mathbf{P}_{k|1:k}^a \mathbf{M}_{k,k+1}^{\mathrm{T}} (\mathbf{P}_{k+1|1:k}^f)^{-1}$$
$$\mathbf{x}_{k|1:K}^s = \mathbf{x}_{k|1:k}^a + \mathbf{K}_k^s (\mathbf{x}_{k+1|1:K}^s - \mathbf{x}_{k+1|1:k}^f)$$
$$\mathbf{P}_{k|1:K}^s = \mathbf{P}_{k|1:k}^a - \mathbf{K}_k^s (\mathbf{P}_{k+1|1:k}^f - \mathbf{P}_{k+1|1:K}^s) \mathbf{K}_k^{s\mathrm{T}}$$

2011), for example a formulation of the ensemble update using an eigenbasis decomposition of the matrix

$$\hat{\mathbf{\Gamma}} = (\hat{\mathbf{H}}\hat{\mathbf{S}})^{\mathrm{T}} \hat{\mathbf{R}}^{-1} (\hat{\mathbf{H}}\hat{\mathbf{S}}) = \hat{\mathbf{U}} \hat{\Lambda} \hat{\mathbf{U}}^{\mathrm{T}}, \tag{25}$$

then this decomposition can be done once for all the ensemble members updates. The cost of this eigenbasis decomposition is of the order of M^3, similar to what it is in the filter (at each analysis step). Thus, the complexity of the ensemble smoother analysis appears approximately K times lower than the EnKF, which includes K analysis steps. The method of van Leeuwen and Evensen (1996) has been used experimentally for oceanic problems by van Leeuwen (1999, 2001).

The sequential strategy (never described elsewhere, to our knowledge) involves a serial processing of the observations, from time 1 to K, after the first ensemble integration with the free model. At any time k, the observations $\mathbf{y}_{1:k-1}$ have been used, providing the updated ensemble $\mathbf{x}_{i|1:k-1}^{a,m}$, with $i = 1, \ldots, K$. The use of observation $\mathbf{y}_k$ leads to an analysis equation of the form of Eq. (18):

$$\boldsymbol{\gamma}^m = (\mathbf{H}_k \mathbf{S}_{k|1:k-1}^a)^{\mathrm{T}} [\mathbf{H}_k \mathbf{S}_{k|1:k-1}^a (\mathbf{H}_k \mathbf{S}_{k|1:k-1}^a)^{\mathrm{T}} + \mathbf{R}_k]^{-1} \times (\mathbf{y}_k^m - \mathbf{H}_k \mathbf{x}_{k|1:k-1}^{a,m}). \tag{26}$$

As with the EnKF, any method can be adopted to compute this term.

4. Marginal smoothing

a. Forward–backward smoother

This smoother was first introduced in the linear Gaussian framework by Rauch et al. (1965) and works over a fixed interval. It is also called the Rauch–Tung–Striebel (RTS) smoother.

In the forward–backward approach, the smoothed PDF is written as

$$p(\mathbf{x}_k|\mathbf{y}_{1:K}) = \int p(\mathbf{x}_k|\mathbf{x}_{k+1}, \mathbf{y}_{1:K}) p(\mathbf{x}_{k+1}|\mathbf{y}_{1:K})\, d\mathbf{x}_{k+1}, \tag{27a}$$

$$= p(\mathbf{x}_k|\mathbf{y}_{1:k}) \int \frac{p(\mathbf{x}_{k+1}|\mathbf{x}_k)}{p(\mathbf{x}_{k+1}|\mathbf{y}_{1:k})} p(\mathbf{x}_{k+1}|\mathbf{y}_{1:K})\, d\mathbf{x}_{k+1}, \tag{27b}$$

where the Markov hypothesis and Bayes' rule have been used to yield $p(\mathbf{x}_k|\mathbf{x}_{k+1}, \mathbf{y}_{1:K}) = p(\mathbf{x}_k|\mathbf{x}_{k+1}, \mathbf{y}_{1:k})$, and $p(\mathbf{x}_k|\mathbf{x}_{k+1}, \mathbf{y}_{1:k}) = p(\mathbf{x}_k|\mathbf{y}_{1:k}) p(\mathbf{x}_{k+1}|\mathbf{x}_k)/p(\mathbf{x}_{k+1}|\mathbf{y}_{1:k})$, respectively. The right-hand side of Eq. (27b) clearly exhibits the fact that the smoother analysis at k is obtained from the filter analysis at k, corrected with a term that involves the smoother analysis at $k+1$. This induces a natural algorithm formed by a forward filter pass followed by a backward smoother pass.

In the linear Gaussian case, the filter analysis at k comes from a Kalman filter pass over the time interval. The smoother analysis at K is simply the filter analysis, since no future observation is included in the estimation. Then the smoother analyses are calculated recursively from $K-1$ down to 1. This provides the RTS smoother equations, which are summarized in Table 4. The RTS algorithm is appealing because the backward sequence can be solved without using the dynamical model: if the term $\mathbf{M}_{k,k+1} \mathbf{P}_k^a$ has been stored during the filter pass, it is reused in the backward pass. The main issue concerns the inversion of the forecast covariance matrix (first equation in Table 4), that generally requires approximations. The RTS smoother is likely the one that has received the greatest attention from the geophysics community (e.g., Lermusiaux and Robinson 1999; Fukumori 2006).

The ensemble formulation of the RTS smoother logically involves a retrospective analysis (Table 4, second equation) for each member:

$$\mathbf{x}_{k|1:K}^{s,m} = \mathbf{x}_{k|1:k}^{a,m} + \mathbf{K}_k^s (\mathbf{x}_{k+1|1:K}^{s,m} - \mathbf{x}_{k+1|1:k}^{f,m}). \tag{28}$$

It can be easily shown that this ensemble of updates provides the appropriate smoother covariance matrix (Table 4, third equation), and this is still true if $(\mathbf{P}_{k+1|1:k}^f)^{-1}$ represents the pseudoinverse of $\mathbf{P}_{k+1|1:k}^f$ when the latter is singular, which happens in reduced-order approaches. A first method to compute the gain (Table 4, first equation) may rely on a covariance localization strategy (Houtekamer and Mitchell 2001; Hamill et al. 2001). To our knowledge, this has never been explored so far. Localization leads to compute a different gain matrix for each grid point of the numerical model, with a reduced, full-rank forecast covariance matrix. Such a matrix can be numerically inverted, but this must be done for each model grid point, which might be expensive in high-dimensional applications. Another method to compute the smoother gain, described by Fukumori (2002), relies on a *partitioning* approach to filtering and smoothing that consists in approximating the model state by a sum of independent dynamical processes, each of them supposedly

represented by a low-dimensional vector [Fukumori 2002, see his Eq. (3)]. Then the inversion of the covariance matrix involves the inversion of low dimensional matrices [Fukumori 2002, see his Eq. (15)]. The third method we are aware of to compute the smoother gain has been proposed by Lermusiaux and Robinson (1999). It is based on the singular value decomposition (SVD):

$$\mathbf{S}^f_{k+1|k} = \mathbf{U}\Sigma^f\mathbf{V}^{\mathrm{T}}, \tag{29}$$

which allows the simple writing of the smoother gain:

$$\mathbf{K}^s_k = \mathbf{S}^a_{k|k}\mathbf{V}\Sigma^{f^{-1}}\mathbf{U}^{\mathrm{T}}. \tag{30}$$

The SVD is the most CPU-consuming calculation here ($\propto N^2M$, where N is the size of the state vector and assuming $M \ll N$), since $\mathbf{\Sigma}^f$ is an $M \times M$ diagonal matrix. Finally, we suggest that the smoother gain be computed as

$$\mathbf{K}^s_k = \mathbf{S}^a_{k|k}[(\mathbf{S}^f_{k+1|k})^{\mathrm{T}}\mathbf{S}^f_{k+1|k}]^{-1}{\mathbf{S}^f_{k+1|k}}^{\mathrm{T}}. \tag{31}$$

This formula, equivalent to the standard formula [and easily retrieved using Eq. (29)], solves the problem of dimension, since the matrix to be inverted is now of dimension $M \times M$. Assuming this operation is the most CPU consuming in the analysis, this approach makes the cost of the backward pass similar to the filter's, without the model integrations.

b. Two-filter smoother

This smoother was introduced in the linear Gaussian framework by Fraser and Potter (1969) and solves the fixed-interval problem. See also Ménard et al. (1996) for a more synthetic description. Because of its particular complexity, this smoother has never been applied in geophysics to our knowledge.

In the Bayesian framework, the smoothing problem is tackled with the following decomposition, based on Bayes' rule:

$$p(\mathbf{x}_k|\mathbf{y}_{1:K}) \propto p(\mathbf{x}_k|\mathbf{y}_{1:k})p(\mathbf{y}_{k+1:K}|\mathbf{x}_k). \tag{32}$$

The first term on the right-hand side is the filter analysis. The second term has been simplified by invoking the Markov hypothesis, which led to

$$p(\mathbf{y}_{k+1:K}|\mathbf{x}_k, \mathbf{y}_{1:k}) = p(\mathbf{y}_{k+1:K}|\mathbf{x}_k). \tag{33}$$

This joint likelihood function is obtained with a backward-in-time "likelihood" filter that performs a sequence of alternating observational updates:

$$p(\mathbf{y}_{k+1:K}|\mathbf{x}_{k+1}) = p(\mathbf{y}_{k+2:K}|\mathbf{x}_{k+1})p(\mathbf{y}_{k+1}|\mathbf{x}_{k+1}), \tag{34}$$

[easily obtained from Eq. (6) and Markov hypothesis], and propagation steps:

$$p(\mathbf{y}_{k+1:K}|\mathbf{x}_k) = \int p(\mathbf{y}_{k+1:K}|\mathbf{x}_{k+1})p(\mathbf{x}_{k+1}|\mathbf{x}_k)\,d\mathbf{x}_{k+1}. \tag{35}$$

These are initialized with the likelihood function at K, $p(\mathbf{y}_K|\mathbf{x}_K)$.

To see the connection between the backward "likelihood" filter and a backward "standard" filter, let us introduce a (not normalized) probability density g defined for any k and i as

$$g(\mathbf{x}_k|\mathbf{y}_{i:K}) = p(\mathbf{y}_{i:K}|\mathbf{x}_k)p(\mathbf{x}_k). \tag{36}$$

Multiplying Eq. (34) by $p(\mathbf{x}_{k+1})$ leads to

$$g(\mathbf{x}_{k+1}|\mathbf{y}_{k+1:K}) = g(\mathbf{x}_{k+1}|\mathbf{y}_{k+2:K})p(\mathbf{y}_{k+1}|\mathbf{x}_{k+1}), \tag{37}$$

which is very similar, in backward form, to the filter analysis in Eq. (11b). Multiplying Eq. (35) by $p(\mathbf{x}_{k+1})$ leads to

$$g(\mathbf{x}_k|\mathbf{y}_{k+1:K}) = \int g(\mathbf{x}_{k+1}|\mathbf{y}_{k+1:K})p(\mathbf{x}_k|\mathbf{x}_{k+1})d\mathbf{x}_{k+1}. \tag{38}$$

This equation is similar, still in backward form, to the filter forecast Eq. (11a). Consequently, to get the likelihood function $p(\mathbf{y}_{k+1:K}|\mathbf{x}_k)$, a strategy is to (i) define $g(\mathbf{x}_K|\mathbf{y}_K) = p(\mathbf{y}_K|\mathbf{x}_K)p(\mathbf{x}_K)$, (ii) run a backward filter from K to k, and then (iii) compute $g(\mathbf{x}_k|\mathbf{y}_{k+1:K})/p(\mathbf{x}_k)$. The prior PDF $p(\mathbf{x}_k)$ may easily be obtained from $p(\mathbf{x}_K)$ using the marginalization rule involving the backward model $p(\mathbf{x}_k|\mathbf{x}_K)$. In theory, any initial prior $p(\mathbf{x}_K)$ can be assigned, since it is divided out later by $p(\mathbf{x}_k)$, which must not equal zero in the definition interval of $\mathbf{x}_k$. In other words, it must be sufficiently uninformative. A first possibility is that it is uninformative because $p(\mathbf{x}_k)$ has forgotten the final state because of a model error, but this does not apply to the first time steps ($k = K - 1, K - 2, \ldots$); a second possibility is simply that its precursor $p(\mathbf{x}_K)$ is sufficiently uninformative too (e.g., homogeneous).

In Kalman's framework, the most uninformative PDF is a Gaussian PDF with infinite covariance matrix (and any mean). This is the choice made by Fraser and Potter (1969) when they set out the two-filter smoother. This results in the equality

$$g(\mathbf{x}_K|\mathbf{y}_K) = p(\mathbf{y}_K|\mathbf{x}_K), \tag{39}$$

demonstrated by applying the BLUE equations with a finite covariance matrix, then letting it tend to infinity.

TABLE 5. Equations of the two-filter smoother (to complete the filter equations). They are applied after a Kalman filter pass over the interval. Here k runs from $K-1$ down to 0.

Initialization of the backward pass
$\mathbf{z}^f_K = 0,\ \mathbf{N}^f_K = 0.$
Backward forecast step
$\tilde{\mathbf{M}}_{k+1,k} = \mathbf{M}^{\mathrm{T}}_{k,k+1}\mathbf{N}^a_{k+1\vert k+1:K}(\mathbf{N}^a_{k+1\vert k+1:K} + \mathbf{N}^a_{k+1\vert k+1:K} \times \mathbf{Q}_{k,k+1}\mathbf{N}^a_{k+1\vert k+1:K})^{-1}$
$\mathbf{z}^f_{k\vert k+1:K} = \tilde{\mathbf{M}}_{k+1,k}\mathbf{z}^a_{k+1\vert k+1:K}$
$\mathbf{N}^f_{k\vert k+1:K} = \tilde{\mathbf{M}}_{k+1,k}\mathbf{N}^a_{k+1\vert k+1:K}\mathbf{M}_{k,k+1}$
Backward analysis step
$\mathbf{N}^a_{k\vert k:K} = \mathbf{N}^f_{k\vert k+1:K} + \mathbf{H}^{\mathrm{T}}_k\mathbf{R}^{-1}_k\mathbf{H}_k$
$\mathbf{z}^a_{k\vert k:K} = \mathbf{z}^f_{k\vert k+1:K} + \mathbf{H}^{\mathrm{T}}_k\mathbf{R}^{-1}_k\mathbf{y}_k$
Smoother analysis
$\mathbf{x}^s_{k\vert 1:K} = (\mathbf{I} + \mathbf{P}^a_{k\vert 1:k}\mathbf{N}^f_{k\vert k+1:K})^{-1}(\mathbf{x}^a_{k\vert 1:k} - \mathbf{P}^a_{k\vert 1:k}\mathbf{z}^f_{k\vert k+1:K})$
$\mathbf{P}^s_{k\vert 1:K} = (\mathbf{I} + \mathbf{P}^a_{k\vert 1:k}\mathbf{N}^f_{k\vert k+1:K})^{-1}\mathbf{P}^a_{k\vert 1:k}$

A concise demonstration is provided in the appendix. Applying the (backward) Kalman filter forecast equations shows that $p(\mathbf{x}_k)$ is also Gaussian with an infinite covariance matrix. Then, the backward filter equations directly provide the desired (Gaussian) PDF $p(\mathbf{y}_{k+1:K}|\mathbf{x}_k) = g(\mathbf{x}_k|\mathbf{y}_{k+1:K})$. The main issue is how to handle an infinite covariance matrix. The information filter (Grewal and Andrews 2001; Simon 2006) is introduced to circumvent this difficulty: the following vectors and matrices are defined:

$$\mathbf{N}^f_{k|k+1:K} = (\mathbf{P}^f_{k|k+1:K})^{-1}, \tag{40a}$$

$$\mathbf{z}^f_{k|k+1:K} = (\mathbf{P}^f_{k|k+1:K})^{-1}\mathbf{x}_{k|k+1:K}, \tag{40b}$$

and initialized to 0. The analysis components $\mathbf{N}^a$ and $\mathbf{z}^a$ are defined similarly. Then the backward Kalman filter equations are written in information form. The transformation also has the property to get rid of the backward model, replaced by the adjoint model. These equations are presented in Table 5. Once the backward likelihood filter forecast at k is available, it is combined with the forward filter analysis [Eq. (32)] to provide the smoother estimate. In the Gaussian case, the latter is simply a BLUE estimate, displayed in Table 5.

5. Summary and discussion

Various smoother algorithms based upon the Kalman filter have been presented and used in geophysics. Five of them are generally put forward: the fixed-interval sequential smoother, the fixed-lag smoother, the ensemble smoother, the forward–backward smoother, and the two-filter smoother. They are derived using different strategies to solve either the joint or the marginal smoothing problem. This paper aimed at making the links between the Bayesian formulations of the smoothing problems, the resolution strategies, and the derived smoother algorithms. It is also an opportunity to gather them together with unified notations, to make comparison easier to the interested reader. Table 6 summarizes the Bayesian formulation of these five smoothers. As many geophysical applications involve an ensemble approach, the ensemble formulation of each algorithm (but the last one, still to be developed to our knowledge) has also been described. Note that we have not mentioned fixed-point smoothing, since it is a particular case of marginal smoothing.

The five algorithms studied here exhibit different characteristics and computational costs. But these costs can be strongly modified by the approximations introduced to deal with the underlying filtering problem, such as an ensemble formulation typically. Here are some important aspects of each algorithm:

- For the fixed-interval sequential smoother, solving the full equations requires $K(K+1)/2$ additional model forecasts and $\mathcal{O}(2KN^3)$ extra operations for the analysis, relative to the filter, with K being the number of analysis updates in the interval and N is the state dimension. For large systems, this is prohibitive, as for

TABLE 6. Summary of the Bayesian formulation of the five smoothers discussed in this paper.

		Joint smoothers	
$p(\mathbf{x}_{0:k}\vert\mathbf{y}_{1:k})$	$\propto$	$p(\mathbf{x}_{0:k-1}\vert\mathbf{y}_{1:k-1})p(\mathbf{x}_k\vert\mathbf{x}_{k-1})p(\mathbf{y}_k\vert\mathbf{x}_k)$	Fixed-interval sequential
$p(\mathbf{x}_{k-L:k}\vert\mathbf{y}_{1:k})$	$\propto$	$\int p(\mathbf{x}_{k-L-1:k-1}\vert\mathbf{y}_{1:k-1})\,d\mathbf{x}_{k-L-1:k-1}p(\mathbf{x}_k\vert\mathbf{x}_{k-1})p(\mathbf{y}_k\vert\mathbf{x}_k)$	Fixed-lag sequential
$p(\mathbf{x}_{0:K}\vert\mathbf{y}_{1:K})$	$\propto$	$p(\mathbf{x}_{0:K})p(\mathbf{y}_{1:K}\vert\mathbf{x}_{0:K})$	Ensemble
		Marginal smoothers	
$p(\mathbf{x}_k\vert\mathbf{y}_{1:K})$	$=$	$p(\mathbf{x}_k\vert\mathbf{y}_{1:k})\int\frac{p(\mathbf{x}_{k+1}\vert\mathbf{x}_k)}{p(\mathbf{x}_{k+1}\vert\mathbf{y}_{1:k})}p(\mathbf{x}_{k+1}\vert\mathbf{y}_{1:K})\,d\mathbf{x}_{k+1}$	Forward–backward
$p(\mathbf{x}_k\vert\mathbf{y}_{1:K})$	$\propto$	$p(\mathbf{x}_k\vert\mathbf{y}_{1:k})p(\mathbf{y}_{k+1:K}\vert\mathbf{x}_k)$	Two filter

the Kalman filter. Using a square root or ensemble approach, the smoother cost drops close to the filter's (Evensen and van Leeuwen 2000).

- For the fixed-lag smoother, solving the full equations requires $L - 1$ additional model forecasts and $\mathcal{O}[2(L-1)N^3]$ extra operations for the analysis, L being the lag. Using a square root or ensemble approach, the cost becomes similar to the filter's (Cosme et al. 2010).
- With the ensemble smoother, two strategies are possible to implement the analysis: global or sequential. The latter makes full use of Kalman's assumption of observation errors uncorrelated in time. Its computational complexity is similar to that of the fixed-interval sequential smoother (the same operations are made, but in a different order), and lower than that of the global analysis. In the ensemble approach, it is possible to formulate the analysis step at a lower numerical cost than the filter's.
- The RTS smoother involves the model only in the forward Kalman filter pass. The extra computational cost is due to the analyses of the backward smoothing pass. The main burden concerns the inversion of the forecast error covariance matrix $\mathbf{P}^f_{k+1|1:k}$ (see Table 4). Although impossible with high-dimensional systems and the full Kalman filter, a reduced-rank square root approach affords this inversion, either (i) by using localization to split the high-dimensional (and singular) inversion problem into a large number of low-dimensional (and full rank) problems, or (ii) by decomposing the square root matrix into singular values: the singular value decomposition requires $\mathcal{O}(MN^2)$ operations, similarly to the forecast step of the Kalman filter. Thus, the cost of the RTS smoother is about twice the cost of the filter; or (iii) using a linear algebra formula that reduces the number of operations to $\mathcal{O}(M^3)$ only (i.e., the main cost of a reduced-rank square root filter analysis).
- The two-filter smoother involves a backward integration with the adjoint model. This is a first limitation to the use of this algorithm: ensemble methods are generally appreciated because they do not require the adjoint model. Assuming that the adjoint model has a cost similar to that of the direct model, the total cost should also be near twice the filter cost. Yet, the adjoint model is very often more expensive than the direct model.[1] If it is twice as expensive, then the cost of the smoother is 3 times the filter's. This extra cost does not include the inversion operations. Because this algorithm has never been applied with high-dimensional systems, the computational impact of reduction strategies are unknown.

The application of ensemble methods to high-dimensional systems generally raises the need for spatial localization techniques. The two most common techniques are *covariance localization* (Houtekamer and Mitchell 2001; Hamill et al. 2001) and *local analysis* (e.g., Evensen 2003; Hunt et al. 2007). The former strategy plays on the state covariance matrix $\mathbf{P}^f$ whereas the latter can be formulated with an operation on the observation covariance matrix $\mathbf{R}$; for more details see Sakov and Bertino (2010) and Greybush et al. (2011).

Technically, both can be used for the smoothers of the joint type. Examples are Khare et al. (2008) for covariance localization and Cosme et al. (2010) for local analysis. For covariance localization, the solution is to Schur-multiply the cross-covariance matrices ($\mathbf{P}^{fa}_{k,i|1:k-1}$ in Table 2; $\mathbf{P}^{a}_{k,i|1:k-1}$ in Table 3) as it is done for the covariance matrices involved in the innovation covariance matrices (the filter forecast $\mathbf{P}^f$ in the sequential algorithm; $\mathbf{P}^{a}_{k|1:k-1}$ in the ensemble smoother case). For local analysis, it can be done exactly as with the filter. From a computational viewpoint, local analysis is preferable. As stated by Sakov and Bertino (2010) and Greybush et al. (2011), covariance localization requires the analysis corrections to be calculated with the explicit covariance (and cross covariance for the smoother) matrix, as in the "standard" EnKF scheme (Evensen 2003; Houtekamer and Mitchell 2001). Consequently, a part of the smoother gain must be recomputed for each smoother step. On the contrary, with ensemble square root schemes that lead to expressions such as Eq. (13), and for which only local analysis applies, no significant computation is required in addition to the filter.

For the forward–backward smoother, either localization method may be used for the forward filter pass, but the backward pass does not make use of observations, and local analysis cannot be applied. Thus, covariance localization is probably more appropriate for both passes. Moreover, covariance localization can be a solution to the inversion of the innovation covariance matrix, as suggested in section 4a.

Finally, smoothing relying upon correlations in time, the use of ensemble smoothers also raises the question of localization in time, which is theoretically ensured by the model error term in Kalman's theory. In practice though, with ensemble methods and high-dimensional systems, spurious distant correlations may occur in time as they occur in space. Moreover, localization in space

[1] There can be a factor of 2–4 between the cost of the direct model and its adjoint version. With the ocean model the Nucleus for European Modeling of the Ocean (NEMO), the present version of the adjoint model (including a few approximations) is about twice as expensive as the direct model (A.Vidard 2010, personal communication).

calls localization in time in the presence of strong advection: the ''best'' future observations of the dynamics within a localization domain can be outside the domain. Localization in time has never been addressed to our knowledge. Only Cosme et al. (2010) have shown the essential role of a (simple) model error term in fading the covariances in time. Also, the fixed-lag smoother can be viewed as a time-localized version of the fixed-interval sequential smoother.

The choice of a smoother algorithm must first depend on the physical problem to be solved. Joint smoothers are specifically designed to make reanalyses that are physically and statistically consistent series of states to be used for variability and evolution studies. Marginal smoothers are more appropriate for the identification of pulselike signals, such as volcanic eruptions in ice cores (e.g., Gazeaux et al. 2011) or the accidental release of chemical tracers (Bocquet 2005b,c). Other applications are the estimation of constant-in-time parameters (reaction coefficients for biogeochemistry, atmospheric chemistry, and dynamical diffusion) or identification of a ''good'' dynamical state for initializing a long-term forecast, typically for climate prediction. Multiresolution data assimilation using graphical models also requires a smoother of the marginal type. We are certainly not aware of all the physical problems for which smoothers may be helpful. As mentioned in the introduction, all the smoothing algorithms described here, joint or marginal, are equivalent when Kalman's linear/Gaussian assumptions are verified. In this case, the choice may be based on practical issues. A fixed interval of fixed-lag sequential smoother is particularly easy to implement in the ensemble form, when an EnKF is available. An ensemble smoother is particularly interesting for observation-related studies (e.g., network design and sensitivity to the parameterization of the observation covariance matrix). Once the background ensemble trajectory is available, experiments on the observational update can be performed at will. The ensemble smoother also offers a nice framework for four-dimensional or asynchronous ensemble Kalman filtering (Hunt et al. 2004; Sakov et al. 2010). The two-filter smoother seems unsuited to an ensemble approach, which is probably why it has received limited attention from geophysicists.

Acknowledgments. EC would like to thank Jean-Michel Brankart for fruitful discussions related to this work. The authors are grateful to Chris Snyder and two other anonymous reviews for their comments, which significantly improved the initial manuscript. This work was supported by CNRS/INSU through the LEFE/ASSIM BFN Program (Principal Investigator: Jacques Blum) and by the ANR *Jeunes Chercheurs* Program (PROSSDAG; Principal Investigator: Didier Auroux). The LEGI/MEOM team (EC, JV, PB) is also supported by the French National Space Center, CNES, and the European Commission under Grant Agreement FP7-SPACE-2007-1-CT-218812-MYOCEAN.

APPENDIX

Gaussian Observation Update with Infinite Background Covariance Matrix

By ''infinite covariance matrix,'' we mean a covariance matrix with null inverse matrix or, equivalently, a covariance matrix with infinite eigenvalues. Given a priori $p(\mathbf{x}) \sim \mathcal{N}(0, \mathbf{P}^b)$ and an observation $p(\mathbf{y}|\mathbf{x}) \sim \mathcal{N}(\mathbf{H}\mathbf{x}, \mathbf{R})$, the posterior is $p(\mathbf{x}|\mathbf{y}) \sim \mathcal{N}(\mathbf{x}^a, \mathbf{P}^a)$, with

$$\mathbf{x}^a = \mathbf{P}^b\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}\mathbf{y}, \tag{A1}$$

$$\mathbf{P}^{a-1} = \mathbf{P}^{b-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}. \tag{A2}$$

The logarithm of the Gaussian, posterior distribution may be written as

$$\log[p(\mathbf{x}|\mathbf{y})] \propto (\mathbf{x} - \mathbf{x}^a)^{\mathrm{T}}\mathbf{P}^{b-1}(\mathbf{x} - \mathbf{x}^a) + (\mathbf{x} - \mathbf{x}^a)^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}(\mathbf{x} - \mathbf{x}^a). \tag{A3}$$

After verifying that

$$\lim_{\mathbf{P}^b \to +\infty} \mathbf{P}^{b^{-1}}\mathbf{x}^a = \lim_{\mathbf{P}^b \to +\infty} \mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}\mathbf{y} = 0, \tag{A4}$$

one finds that the first rhs term of Eq. (A3) tends to 0 when $\mathbf{P}^b \to +\infty$. And after showing that

$$\lim_{\mathbf{P}^b \to +\infty} \mathbf{H}\mathbf{x}^a = \lim_{\mathbf{P}^b \to +\infty} \left[\mathbf{I} + \mathbf{R}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}})^{-1}\right]^{-1}\mathbf{y} = \mathbf{y}, \tag{A5}$$

one concludes that Eq. (A3) reads as

$$\log[p(\mathbf{x}|\mathbf{y})] \propto (\mathbf{H}\mathbf{x} - \mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}), \tag{A6}$$

which is exactly the exponent of the Gaussian distribution $p(\mathbf{y}|\mathbf{x})$.

REFERENCES

Anderson, B. D. O., and J. B. Moore, 1979: *Optimal Filtering.* Prentice-Hall, 357 pp.

Anderson, J., 2010: A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.,* **138,** 4186–4198.

Barth, A., A. Alvera-Azcarate, K. Gurgel, J. Staneva, A. Port, J. Beckers, and E. V. Stanev, 2010: Ensemble perturbation smoother for optimizing tidal boundary conditions by assimilation of high-frequency radar surface currents—Application to the German Bight. *Ocean Sci.,* **6,** 161–178.

Beal, D., P. Brasseur, J. Brankart, Y. Ourmières, and J. Verron, 2010: Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: Implications for nonlinear estimation using Gaussian anamorphosis. *Ocean Sci.,* **6,** 247–262.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.,* **129,** 420–436.

Bocquet, M., 2005a: Grid resolution dependence in the reconstruction of an atmospheric tracer source. *Nonlinear Processes Geophys.,* **12,** 219–233.

——, 2005b: Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: Theory. *Quart. J. Roy. Meteor. Soc.,* **131,** 2191–2208.

——, 2005c: Reconstruction of an atmospheric tracer source using the principle of maximum entropy. II: Applications. *Quart. J. Roy. Meteor. Soc.,* **131,** 2209–2223.

——, C. A. Pires, and L. Wu, 2010: Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Wea. Rev.,* **138,** 2997–3023.

Brankart, J., C. Ubelmann, C. Testut, E. Cosme, P. Brasseur, and J. Verron, 2009: Efficient parameterization of the observation error covariance matrix for square root or ensemble Kalman filters: Application to ocean altimetry. *Mon. Wea. Rev.,* **137,** 1908–1927.

——, E. Cosme, C. Testut, P. Brasseur, and J. Verron, 2010: Efficient adaptive error parameterizations for square root or ensemble Kalman filters: Application to the control of ocean mesoscale signals. *Mon. Wea. Rev.,* **138,** 932–950.

——, ——, ——, ——, and ——, 2011: Efficient local error parameterizations for square root or ensemble Kalman filters: application to a basin-scale ocean turbulent flow. *Mon. Wea. Rev.,* **139,** 474–493.

Brusdal, K., J. Brankart, G. Halberstadt, G. Evensen, P. Brasseur, P. van Leeuwen, E. Dombrowsky, and J. Verron, 2003: A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. *J. Mar. Syst.,* **40–41,** 253–289.

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

Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan,* **75** (1B), 257–288.

——, N. S. Sivakumaran, and R. Todling, 1994: A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.,* **122,** 2838–2867.

——, A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Wea. Rev.,* **126,** 2913–2926.

Cosme, E., F. Hourdin, C. Genthon, and P. Martinerie, 2005: Origin of dimethylsulfide, non-sea-salt sulfate, and methanesulfonic acid in eastern Antarctica. *J. Geophys. Res.,* **110,** D03302, doi:10.1029/2004JD004881.

——, J. Brankart, J. Verron, P. Brasseur, and M. Krysta, 2010: Implementation of a reduced-rank, square-root smoother for ocean data assimilation. *Ocean Modell.,* **33,** 87–100.

Dee, D., 1995: On-line estimation error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.,* **123,** 1128–1145.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.,* **53,** 343–367.

——, and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.,* **128,** 1852–1867.

Fearnhead, P., D. Wyncoll, and J. Tawn, 2010: A sequential smoothing algorithm with linear computational cost. *Biometrika,* **97,** 447–464.

Fieguth, P. W., W. C. Karl, A. S. Willsky, and C. Wunsch, 1995: Multiresolution optimal interpolation and statistical analysis of TOPEX/POSEIDON satellite altimetry. *IEEE Trans. Geosci. Remote Sens.,* **33,** 280–292.

Fraser, D. C., and J. E. Potter, 1969: The optimum linear smoother as a combination of two optimum linear filters. *IEEE Trans. Automat. Contrib.,* **14,** 387–390.

Fukumori, I., 2002: A partitioned Kalman filter and smoother. *Mon. Wea. Rev.,* **130,** 1370–1383.

——, 2006: What is data assimilation really solving, and how is the calculation actually done? *Ocean Weather Forecasting: An Integrated View of Oceanography,* E. Chassignet and J. Verron, Eds., Springer, 271–316.

Gazeaux, J., D. Batista, C. Ammann, P. Naveau, C. Jegat, and C. Gao, 2011: Extracting common pulse-like signals from multiple ice core time series. *Comput. Stat. Data Anal.,* in press.

Grewal, M. S., and A. P. Andrews, 2001: *Kalman Filtering: Theory and Practice Using MATLAB*. 2nd ed. Wiley-Interscience, 416 pp.

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.,* **139,** 511–522.

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

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.,* **126,** 796–811.

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **129,** 123–137.

——, ——, and X. Deng, 2009: Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev.,* **137,** 2126–2143.

Hunt, B. R., and Coauthors, 2004: Four-dimensional ensemble Kalman filtering. *Tellus,* **56A,** 273–277.

——, E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D,* **230,** 112–126.

Ihler, A. T., S. Kirshner, M. Ghil, A. W. Robertson, and P. Smyth, 2007: Graphical models for statistical inference and data assimilation. *Physica D,* **230,** 72–87.

Kalman, R. E., 1960: A new approach to linear filter and prediction problems. *J. Basic Eng.,* **82,** 35–45.

Khare, S. P., J. L. Anderson, T. J. Hoar, and D. Nychka, 2008: An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems. *Tellus,* **60A,** 97–112.

Lermusiaux, P. F. J., 2006: Uncertainty estimation and prediction for interdisciplinary ocean dynamics. *J. Comput. Phys.,* **217,** 176–199.

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

——, ——, P. J. H. Haley, and W. G. Leslie, 2002: Advanced interdisciplinary data assimilation: Filtering and smoothing via Error Subspace Statistical Estimation. *Proc. The OCEANS 2002 MTS/IEEE Conf.,* Biloxi, MS, IEEE, 795–802.

Losa, S., G. Kivman, J. Schroter, and M. Wenzel, 2003: Sequential weak constraint parameter estimation in an ecosystem model. *J. Mar. Syst.,* **43,** 31–49.

Ménard, R., S. E. Cohn, L. Chang, and P. M. Lyster, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus,* **48A,** 221–237.

Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.,* **128,** 416–433.

Parrish, D. F., and S. E. Cohn, 1985: A Kalman filter for a two-dimensional shallow-water model: Formulation and preliminary experiments. Tech. Rep., Office Note 304, National Meteorological Center, Washington, DC, 64 pp.

Pham, D. T., J. Verron, and M. C. Roubaud, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.,* **16,** 323–340.

Rauch, H. E., F. Tung, and C. T. Striebel, 1965: Maximum likelihood estimates of linear dynamic systems. *AIAA J.,* **3,** 1445–1450.

Ravela, S., and D. McLaughlin, 2007: Fast ensemble smoothing. *Ocean Dyn.,* **57,** 123–134.

Sakov, P., and L. Bertino, 2010: Relation between two common localisation methods for the EnKF. *Comput. Geosci.,* **15,** 225–237.

——, G. Evenson, and L. Bertino, 2010: Asynchronous data assimilation with the EnKF. *Tellus,* **62A,** 24–29.

Simon, D., 2006: *Optimal State Estimation*. Wiley & Sons, 530 pp.

Simon, E., and L. Bertino, 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 Sci.,* **5,** 495–510.

Skandrani, C., J. Brankart, N. Ferry, J. Verron, P. Brasseur, and B. Barnier, 2009: Controlling atmospheric forcing parameters of global ocean models: Sequential assimilation of sea surface mercator-ocean reanalysis data. *Ocean Sci.,* **5,** 403–419.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.,* **131,** 1485–1490.

van Leeuwen, P. J., 1999: The time mean circulation in the Agulhas region determined with the ensemble smoother. *J. Geophys. Res.,* **104,** 1393–1404.

——, 2001: An ensemble smoother with error estimates. *Mon. Wea. Rev.,* **129,** 709–728.

——, 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.,* **137,** 4089–4114.

——, and G. Evensen, 1996: Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Wea. Rev.,* **124,** 2898–2913.

Verlaan, M., and A. W. Heemink, 1997: Tidal flow forecasting using reduced-rank square root filter. *Stochastic Hydrol. Hydraul.,* **11,** 349–368.

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

Willsky, A. S., 2002: Multiresolution Markov models for signal and image processing. *Proc. IEEE,* **90,** 1396–1458.

Zheng, F., and Y. Zhu, 2008: Balanced multivariate model errors of an intermediate coupled model for ensemble Kalman filter data assimilation. *J. Geophys. Res.,* **113,** C12006, doi:10.1029/2008JC005065.

Zhou, Y., D. McLaughlin, D. Entekhabi, and G. Crystal Ng, 2008: An ensemble multiscale filter for large nonlinear data assimilation problems. *Mon. Wea. Rev.,* **136,** 678–698.

Zhu, Y., R. Todling, and S. E. Cohn, 1999: Technical remarks on smoother algorithms. Tech. Rep., NASA/GSFC/Data Assimilation Office, Office Note 99-02, 45 pp.

——, ——, J. Guo, S. E. Cohn, I. M. Navon, and Y. Yang, 2003: The GEOS-3 retrospective data assimilation system: The 6-hour lag case. *Mon. Wea. Rev.,* **131,** 2129–2150.