An Ensemble Kalman Smoother for Nonlinear Dynamics

GEIR EVENSEN

Nansen Environmental and Remote Sensing Center, Bergen, Norway

PETER JAN VAN LEEUWEN

Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Utrecht, Netherlands

(Manuscript received 9 February 1999, in final form 1 July 1999)

ABSTRACT

It is formally proved that the general smoother for nonlinear dynamics can be formulated as a *sequential* method, that is, observations can be assimilated sequentially during a forward integration. The general filter can be derived from the smoother and it is shown that the general smoother and filter solutions at the final time become identical, as is expected from linear theory. Then, a new smoother algorithm based on ensemble statistics is presented and examined in an example with the Lorenz equations. The new smoother can be computed as a sequential algorithm using only forward-in-time model integrations. It bears a strong resemblance with the ensemble Kalman filter. The difference is that every time a new dataset is available during the forward integration, an analysis is computed for all previous times up to this time. Thus, the first guess for the smoother is the ensemble Kalman filter solution, and the smoother estimate provides an improvement of this, as one would expect a smoother to do. The method is demonstrated in this paper in an intercomparison with the ensemble Kalman filter and the ensemble smoother introduced by van Leeuwen and Evensen, and it is shown to be superior in an application with the Lorenz equations. Finally, a discussion is given regarding the properties of the analysis schemes when strongly non-Gaussian distributions are used. It is shown that in these cases more sophisticated analysis schemes based on Bayesian statistics must be used.

1. Introduction

The celebrated Lorenz model (Lorenz 1963) has been the subject of extensive studies motivated by its chaotic and strongly nonlinear nature. In the field of data assimilation the model has served as a test bed for examining the properties of various data assimilation methods when used with strongly nonlinear dynamics; see for example, Gauthier (1992), Miller et al. (1994), Evensen and Fario (1997), and in particular Evensen (1997), which gives a more detailed discussion of the ensemble Kalman filter (EnKF) with the Lorenz model in the same setup as will be used here. The results from these studies have been used to suggest properties and possibilities of the methods for applications with oceanic and atmospheric models that may also be strongly nonlinear and chaotic.

A weak constraint ensemble smoother (ES) method was recently proposed by van Leeuwen and Evensen (1996). The method applies ensemble integrations to represent the density for the model evolution in space and time and a variance minimizing estimate can then be calculated. For linear dynamics it would, in the limit of an infinite ensemble size, converge to the generalized inverse solution of the weak constraint problem (which is solved for by the representer method or the traditional Kalman smoother). The method was tested with a quasigeostropic model where it proved to perform poorer than the EnKF. In Evensen (1997), the ES was used with the Lorenz equations and the results were intercompared with those from the ensemble Kalman filter and a weak constraint gradient descent method. It turned out that it performed rather poorly for this example compared to the EnKF.

In this paper the ensemble smoother problem for the Lorenz equations is revisited, using a new formulation of the smoother algorithm that performs better than both the original ES and the EnKF. It may be considered as an extension of the EnKF where information at assimilation times is propagated backward in time.

In the two following sections, the general formulation of the data assimilation problem and the methods used to solve it are discussed. An application of three data assimilation methods, that is, the EnKF, ES, and EnKS, is discussed in section 4. In section 5 the consequences of neglecting non-Gaussian effects are illustrated in a simple example. Finally, a discussion is given in section 6.

Corresponding author address: Geir Evensen, Nansen Environmental and Remote Sensing Center, Edvard Griegsvei 3a, N-5037 Solheimsviken, Norway.
E-mail: Geir.Evensen@nrsc.no

2. Data assimilation in a probabilistic framework

In van Leeuwen and Evensen (1996) the general data assimilation problem was introduced using a Bayesian formalism. Note first that the unknown model state, contained in $\boldsymbol{\psi}$, is viewed as a realization of a random variable that can be described by a probability density $f(\boldsymbol{\psi})$. Note that the vector function $\boldsymbol{\psi}$ is fairly general and can contain a number of continuous functions of space and time, for example, the various velocities and thermodynamic variables in a general circulation model. The observations are contained in a vector, $\mathbf{d}$, and are also treated as random variables with a distribution, $f(\mathbf{d}\,|\,\boldsymbol{\psi})$, which would normally be assumed to be Gaussian. Using the definition of a conditional probability density we can derive the probability density of a model state $\boldsymbol{\psi}$ given a vector of observations as

$$f(\boldsymbol{\psi}\,|\,\mathbf{d}) = \frac{f(\mathbf{d}\,|\,\boldsymbol{\psi})f(\boldsymbol{\psi})}{\int (\cdots)\,d\boldsymbol{\psi}}, \tag{1}$$

where the denominator is just the integrated numerator that normalizes the expression so the integral of $f(\boldsymbol{\psi}\,|\,\mathbf{d})$ becomes one. Thus, the probability density of the data given a model evolution, $f(\mathbf{d}\,|\,\boldsymbol{\psi})$, and the probability density of the model evolution, $f(\boldsymbol{\psi})$, must be known. The value of $\boldsymbol{\psi}$ that maximizes $f(\boldsymbol{\psi}\,|\,\mathbf{d})$ is the maximum likelihood (MLH) estimate for $\boldsymbol{\psi}$.

If one assumes that the model equations describe a first-order autoregressive, or Markov, process, that is, the model is forced randomly as

$$d\boldsymbol{\psi} = \mathbf{g}(\boldsymbol{\psi})\,dt + d\boldsymbol{\beta}, \tag{2}$$

where $\mathbf{g}$ is the nonlinear model operator and $d\boldsymbol{\beta}$ contain random increments with known covariance $\mathbf{Q}$ and zero mean, the probability density, $f(\boldsymbol{\psi})$, for the model solution, can be determined by solving Kolmogorov's equation:

$$\frac{\partial f(\boldsymbol{\psi})}{\partial t} + \sum_{i=1}^{n} \frac{\partial g_i(\boldsymbol{\psi}) f(\boldsymbol{\psi})}{\partial \psi_i} = \sum_{i,j=1}^{n} \frac{Q_{ij}}{2} \frac{\partial^2 f(\boldsymbol{\psi})}{\partial \psi_i\, \partial \psi_j}. \tag{3}$$

A derivation of this equation, which is the fundamental equation for evolution of error statistics, can be found in Jazwinski (1970). The probability density function represents the density of an infinite ensemble of possible model states, each having an associated infinitesimal probability. The width of the probability density function corresponds to the variance of the ensemble and represents the errors in the predicted solution. Given the density at an initial time, $f(\boldsymbol{\psi})$ can be computed by solving Eq. (3).

a. The sequential smoother

It will now be shown that the general smoother in Eq. (1) can be formulated as a sequential method. Assume first that the observations contained in $\mathbf{d}$ are available at a number of different times t_k. For each time t_k, there is now an observation vector, $\mathbf{d}_k$, containing a subset of $\mathbf{d}$. Similarly, the model state $\boldsymbol{\psi}(\mathbf{x}, t)$ is now divided into a number of subsets $\boldsymbol{\psi}_k(\mathbf{x}, t) = \boldsymbol{\psi}(\mathbf{x}, t \in (t_{k-1}, t_k])$. There is an exception for $\boldsymbol{\psi}_0$, which is the initial condition at time t_0.

The general smoother defined by Eq. (1) can now, for the time interval $t \in [t_0, t_k]$, be written as

$$f(\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k\,|\,\mathbf{d}_1, \ldots, \mathbf{d}_k) = \frac{f(\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k) f(\mathbf{d}_1, \ldots, \mathbf{d}_k\,|\,\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k)}{\int (\cdots)\,d\boldsymbol{\psi}}, \tag{4}$$

where the only assumption is that measurements are distributed among a finite number of time instants.

Write now the joint probability function, $f(\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k)$, as

$$\begin{aligned} f(\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k) &= f(\boldsymbol{\psi}_0) f(\boldsymbol{\psi}_1\,|\,\boldsymbol{\psi}_0) f(\boldsymbol{\psi}_2\,|\,\boldsymbol{\psi}_0, \boldsymbol{\psi}_1) \cdots f(\boldsymbol{\psi}_k\,|\,\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{k-1}) \\ &= f(\boldsymbol{\psi}_0) f(\boldsymbol{\psi}_1\,|\,\boldsymbol{\psi}_0) f(\boldsymbol{\psi}_2\,|\,\boldsymbol{\psi}_1) \cdots f(\boldsymbol{\psi}_k\,|\,\boldsymbol{\psi}_{k-1}), \end{aligned} \tag{5}$$

where it has been assumed that the model evolution is a first-order Markov process; that is,

$$f(\boldsymbol{\psi}_k\,|\,\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{k-1}) = f(\boldsymbol{\psi}_k\,|\,\boldsymbol{\psi}_{k-1}). \tag{6}$$

Thus the model state $\boldsymbol{\psi}_k$ can be determined when knowing $\boldsymbol{\psi}_{k-1}$, independently of the prior evolution of the state.

The distribution for the data can be written as

$$\begin{aligned} f(\mathbf{d}_1, \ldots, \mathbf{d}_k\,|\,\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_k) &= f(\mathbf{d}_1\,|\,\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_k) \cdots f(\mathbf{d}_k\,|\,\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_k) \\ &= f[\mathbf{d}_1\,|\,\boldsymbol{\psi}(t_1)] f[\mathbf{d}_2\,|\,\boldsymbol{\psi}(t_2)] \cdots f[\mathbf{d}_k\,|\,\boldsymbol{\psi}(t_k)]. \end{aligned} \tag{7}$$

Here, we assume that (i) the data collected at different times are independent, that is,

$$f(\mathbf{d}_1, \mathbf{d}_2\,|\,\boldsymbol{\psi}) = f(\mathbf{d}_1\,|\,\boldsymbol{\psi}) f(\mathbf{d}_2\,|\,\boldsymbol{\psi}), \tag{8}$$

and (ii) the data at a particular time depend only on the state at this time, that is,

$$f(\mathbf{d}_k\,|\,\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_k) = f[\mathbf{d}_k\,|\,\boldsymbol{\psi}(t_k)]. \tag{9}$$

With some rearrangement, the smoother defined by Eq. (4) can be written as

$$f(\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_k\,|\,\mathbf{d}_1, \ldots, \mathbf{d}_k) = \frac{f(\boldsymbol{\psi}_0) f(\boldsymbol{\psi}_1\,|\,\boldsymbol{\psi}_0) f[\mathbf{d}_1\,|\,\boldsymbol{\psi}(t_1)] f(\boldsymbol{\psi}_2\,|\,\boldsymbol{\psi}_1) f[\mathbf{d}_2\,|\,\boldsymbol{\psi}(t_2)] \cdots f(\boldsymbol{\psi}_k\,|\,\boldsymbol{\psi}_{k-1}) f[\mathbf{d}_k\,|\,\boldsymbol{\psi}(t_k)]}{\int (\cdots)\,d\boldsymbol{\psi}}. \tag{10}$$

The interesting observation is that this expression can be evaluated sequentially and forward in time. Starting with the distribution for the initial conditions, $f(\boldsymbol{\psi}_0)$, we can compute the smoother using the following equations sequentially, one by one:

$$f(\boldsymbol{\psi}_0, \boldsymbol{\psi}_1 | \mathbf{d}_1) = \frac{f(\boldsymbol{\psi}_0) f(\boldsymbol{\psi}_1 | \boldsymbol{\psi}_0) f[\mathbf{d}_1 | \boldsymbol{\psi}(t_1)]}{\int (\cdots) d\boldsymbol{\psi}}, \tag{11}$$

$$f(\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2 | \mathbf{d}_1, \mathbf{d}_2) = \frac{f(\boldsymbol{\psi}_0, \boldsymbol{\psi}_1 | \mathbf{d}_1) f(\boldsymbol{\psi}_2 | \boldsymbol{\psi}_1) f[\mathbf{d}_2 | \boldsymbol{\psi}(t_2)]}{\int (\cdots) d\boldsymbol{\psi}}, \tag{12}$$

$$\vdots$$

$$f(\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_k | \mathbf{d}_1, \ldots, \mathbf{d}_k) = \frac{f(\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{k-1} | \mathbf{d}_1, \ldots, \mathbf{d}_{k-1}) f(\boldsymbol{\psi}_k | \boldsymbol{\psi}_{k-1}) f[\mathbf{d}_k | \boldsymbol{\psi}(t_k)]}{\int (\cdots) d\boldsymbol{\psi}}. \tag{13}$$

Thus, Eq. (11) defines the smoother solution over the interval $t \in [t_0, t_1]$ using data vector $\mathbf{d}_1$. Equation (12) defines the smoother solution over the interval $t \in [t_0, t_2]$ introducing data vector $\mathbf{d}_2$. The information from data vector $\mathbf{d}_1$ is also used through the density, $f(\boldsymbol{\psi}_0, \boldsymbol{\psi}_1 | \mathbf{d}_1)$, from (11). The information from data vector $\mathbf{d}_k$ influences the solution over the entire time interval $t \in [t_0, t_k]$; thus, it is a true smoother. The assumptions used in deriving (11)–(13) are summarized as follows:

1) The model evolution is a first-order Markov process. (This assumption is not needed but has been adopted since the expressions becomes cleaner and the models normally used are describing a Markov process through the forward integration.)
2) The data are distributed over a finite number of discrete times.
3) The data vectors at different times are independent.

These conditions will normally always be satisfied and do not pose any serious restrictions. When satisfied, the sequential smoother should give exactly the same solution as the general smoother defined by Eq. (1).

b. The general filter

It is now interesting to derive the equations for the general filter in the same probabilistic formalism. In the filter there will be no information carried backward in time. Thus, the following notation is used: $\boldsymbol{\psi}(t_k)$ is now the model state at time t_k, and $\boldsymbol{\psi}_k$ is $\boldsymbol{\psi}$ in the interval $t \in (t_{k-1}, t_k]$, as used for the sequential smoother.

The estimate at the ''final'' time t_k is now defined by integrating the sequential smoother (13) over all $\boldsymbol{\psi}$ in the interval $t \in [t_0, t_k)$. Temporarily omitting the normalization, the derivation of the filter equations is as follows:

$$\begin{aligned}
&f[\boldsymbol{\psi}(t_k) | \mathbf{d}_1, \ldots, \mathbf{d}_k] \\
&= \int_{\boldsymbol{\psi}'[t \in (t_0, t_k)]} f(\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_k | \mathbf{d}_1, \ldots, \mathbf{d}_k)\, d\boldsymbol{\psi}' \\
&= \int_{\boldsymbol{\psi}'[t \in (t_0, t_k)]} f(\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{k-1} | \mathbf{d}_1, \ldots, \mathbf{d}_{k-1}) \\
&\qquad \times f(\boldsymbol{\psi}_k | \boldsymbol{\psi}_{k-1})\, d\boldsymbol{\psi}' f[\mathbf{d}_k | \boldsymbol{\psi}(t_k)] \\
&= \int_{\boldsymbol{\psi}'[t \in (t_0, t_k)]} f(\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_k | \mathbf{d}_1, \ldots, \mathbf{d}_{k-1})\, d\boldsymbol{\psi}' \\
&\qquad \times f[\mathbf{d}_k | \boldsymbol{\psi}(t_k)] \\
&= f[\boldsymbol{\psi}(t_k) | \mathbf{d}_1, \ldots, \mathbf{d}_{k-1}] f[\mathbf{d}_k | \boldsymbol{\psi}(t_k)].
\end{aligned} \tag{14}$$

Thus, the general filter can be written as

$$f[\boldsymbol{\psi}(t_1) | \mathbf{d}_1] = \frac{f[\boldsymbol{\psi}(t_1)] f[\mathbf{d}_1 | \boldsymbol{\psi}(t_1)]}{\int (\cdots) d\boldsymbol{\psi}}, \tag{15}$$

$$f[\boldsymbol{\psi}(t_2) | \mathbf{d}_1, \mathbf{d}_2] = \frac{f[\boldsymbol{\psi}(t_2) | \mathbf{d}_1] f[\mathbf{d}_2 | \boldsymbol{\psi}(t_2)]}{\int (\cdots) d\boldsymbol{\psi}}, \tag{16}$$

$$\vdots$$

$$f[\boldsymbol{\psi}(t_k) | \mathbf{d}_1, \ldots, \mathbf{d}_k] = \frac{f[\boldsymbol{\psi}(t_k) | \mathbf{d}_1, \ldots, \mathbf{d}_{k-1}] f[\mathbf{d}_k | \boldsymbol{\psi}(t_k)]}{\int (\cdots) d\boldsymbol{\psi}}. \tag{17}$$

Here the information is carried only forward in time and the state at time t_k is dependent on all the previous data, $\mathbf{d}_1, \ldots, \mathbf{d}_k$. Note also that the estimate at the final time will be identical for the smoother and the filter, as one would expect. At all previous times it will be sub-

optimal since future observations have not been used to define the estimate.

c. A ''lagged'' smoother

It is also possible to define a lagged smoother (see, e.g., Cohn et al. 1994). This is an approach that is normally used for reducing the CPU requirements and it is based on the assumption that the observations at time t_k will only influence the estimate in an interval $t \in [t_k - t_{\text{lag}}, t_k]$. In practical applications t_{lag} would typically be a few times the predictability time. It can be formally derived from the sequential smoother Eq. (13) by integration over $\boldsymbol{\psi}\{t \in [t_0, t_k - t_{\text{lag}})\}$, using the same approach as when deriving the general filter equations.

3. Practical implementation

The formulations in the previous section are of course very complicated to solve when using general density functions and high-dimensional state spaces. On the other hand, if one can assume that they are all Gaussian, which would be the case with linear dynamics, it is possible to solve for the optimal variance minimizing estimate. Note that the variance minimizing estimate will be identical to the MLH estimate as long as the dynamics are linear and the observation errors can be described by a Gaussian. Thus, under this assumption, the methods presented below gives the optimal solution of the general smoother and filter problems. For the computation of all of the analyses below, see Burgers et al. (1998) for the practical implementation and in particular the required perturbation of measurements.

a. Ensemble smoother

The ensemble smoother presented in van Leeuwen and Evensen (1996) attempted to find the $\boldsymbol{\psi}$ that minimized the posterior error variance. Optimally this should have been done by first computing $f(\boldsymbol{\psi}|\mathbf{d})$ from (1) and then finding the mean, or even better the MLH estimate. A more practical approach was chosen based on the following algorithm:

1) The distribution for the initial conditions $f(\boldsymbol{\psi}_0)$ was defined as Gaussian.
2) Kolmogorov's equation, (3), was solved for $f(\boldsymbol{\psi}_0, \boldsymbol{\psi}) = f(\boldsymbol{\psi}|\boldsymbol{\psi}_0) f(\boldsymbol{\psi}_0)$, using a Markov chain Monte Carlo method, or ensemble integration, as described in Evensen (1994b), which provides a representation of the distribution for the model evolution without any observations used.
3) The prior error covariances in space and time, $\mathbf{C}_{\psi\psi}(\mathbf{x}, t, \mathbf{x}', t')$, for the prediction or first guess, were computed from the ensemble representation of $f(\boldsymbol{\psi}_0, \boldsymbol{\psi})$.
4) The first guess solution, $\boldsymbol{\psi}^f_{\text{ES}}(\mathbf{x}, t)$, was defined to be the ensemble mean.
5) The distribution for the observations was assumed to be Gaussian and described by an error covariance matrix, $\mathbf{W}^{-1}$.
6) The analysis, $\boldsymbol{\psi}^a_{\text{ES}}(\mathbf{x}, t)$, could then be computed from

$$\boldsymbol{\psi}^a_{\text{ES}}(\mathbf{x}, t) = \boldsymbol{\psi}^f_{\text{ES}}(\mathbf{x}, t) + \sum_{m=1}^{M} b_m \mathbf{r}_m(\mathbf{x}, t). \tag{18}$$

An explanation of notation is now provided. The model state, $\boldsymbol{\psi}(\mathbf{x}, t)$, can be recognized as an object, or a vector, containing all the model variables over the spatial and temporal domain that is defined for the smoother. Similarly, each of the M representers or influence functions, $r_m(\mathbf{x}, t)$, are of the same type. There is one representer for each of the M measurements and they define the influence the measurement will have on all the model variables at all locations in space and time. The representers are defined as

$$\mathbf{r}_m(\mathbf{x}, t) = L_{m,[\mathbf{x}',t']}[\mathbf{C}_{\psi\psi}(\mathbf{x}, t, \mathbf{x}', t')]. \tag{19}$$

The vector of coefficients $\mathbf{b}$ can be found by solving the system

$$(\mathbf{R} + \mathbf{W}^{-1})\mathbf{b} = \mathbf{d} - \boldsymbol{L}_{[\mathbf{x}',t']}[\boldsymbol{\psi}_{\text{ES}}(\mathbf{x}', t')]. \tag{20}$$

The mth column in the representer matrix, $\mathbf{R}(:, m) = \boldsymbol{L}_{[\mathbf{x},t]}[\mathbf{r}_m(\mathbf{x}, t)]$, can be constructed by measuring the mth representer.

A component, m, of the linear measurement operator would for a direct measurement of the mth component be written

$$L_{m,[\mathbf{x}',t']}[\boldsymbol{\psi}] = \int_{\mathcal{D}}\int_{\mathcal{T}} \delta(\mathbf{x}_m - \mathbf{x}')\delta(t_m - t')\boldsymbol{\delta}^{\mathrm{T}}_{\psi}\boldsymbol{\psi}(\mathbf{x}', t')\, d\mathbf{x}\, dt$$
$$= \boldsymbol{\delta}^{\mathrm{T}}_{\psi}\boldsymbol{\psi}(\mathbf{x}_m, t_m), \tag{21}$$

where $\mathcal{D}$ and $\mathcal{T}$ define the spatial and temporal domains of the smoother problem. The two first δ functions define the spatial and temporal locations of the measurement and the $\boldsymbol{\delta}^{\mathrm{T}}_{\psi}$ vector defines which variable in $\boldsymbol{\psi}$ is measured. The subscript $[\mathbf{x}', t']$ means it operates on $\mathbf{x}'$ and t'.

This is the traditional variance minimizing analysis, which becomes the MLH estimator when only Gaussian statistics are involved. If the model is nonlinear, the distribution for the model evolution will not be Gaussian, and this estimator is no longer the MLH estimator.

The interesting result from van Leeuwen and Evensen (1996) and Evensen (1997) was that this smoother was outperformed by EnKF, in particular in the example with the highly nonlinear Lorenz equations used by Evensen (1997). Note that the Gaussian assumption is valid for linear dynamics and the ES estimate will then be identical to the traditional Kalman smoother or generalized inverse estimate.

b. Ensemble Kalman filter

The ensemble Kalman filter, which is a sequential method, was proposed by Evensen (1994a,b) and used

in a realistic application by Evensen and van Leeuwen (1996). See also the discussion in Burgers et al. (1998). The EnKF applies a similar approach to the ES for computing the analysis. Actually, it is based on the formulation (15)–(17) where one computes the error covariance $\mathbf{C}_{\psi\psi}(\mathbf{x}, \mathbf{x}', t_k)$ for the prediction at time t_k from an ensemble representation of the density $f[\boldsymbol{\psi}(t_k)|\mathbf{d}_1, \ldots, \mathbf{d}_{k-1}]$. Then, with a Gaussian assumption on the distribution for the observations, the following equations define the analysis in the EnKF:

$$\boldsymbol{\psi}^a_{\mathrm{EnKF}}(\mathbf{x}, t_k) = \boldsymbol{\psi}^f_{\mathrm{EnKF}}(\mathbf{x}, t_k) + \sum_{m=1}^{M_k} b_m \mathbf{r}_m(\mathbf{x}, t_k), \quad (22)$$

where the first guess, $\boldsymbol{\psi}^f_{\mathrm{EnKF}}(\mathbf{x}, t_k)$ at the time t_k, is just the mean of the ensemble prediction at t_k. The vector of coefficients, $\mathbf{b}$, is found as the solution of the system

$$(\mathbf{R}_k + \mathbf{W}_k^{-1})\mathbf{b} = \mathbf{d}_k - \boldsymbol{L}_{k,[\mathbf{x}]}[\boldsymbol{\psi}^f_{\mathrm{EnKF}}(\mathbf{x}, t_k)], \quad (23)$$

where the M_k representers corresponding to the M_k data at time t_k are given as

$$\mathbf{r}_m(\mathbf{x}, t_k) = L_{k,m,[\mathbf{x}']}[\mathbf{C}_{\psi\psi}(\mathbf{x}, \mathbf{x}', t_k)], \quad (24)$$

and the representer matrix becomes $\mathbf{R}_k(:, m) = \boldsymbol{L}_{k,[\mathbf{x}]}[\mathbf{r}_m(\mathbf{x}, t_k)]$.

The subscript k on $\boldsymbol{L}_{k,[\mathbf{x}]}$ denotes that this is the measurement operator for the measurements $\mathbf{d}_k$, at time t_k, while $[\mathbf{x}]$ means that it operates on $\mathbf{x}$. For a direct measurement the mth component at time t_k would be

$$L_{k,m,[\mathbf{x}']}[\boldsymbol{\psi}] = \int_{\mathcal{D}} \delta(\mathbf{x}_m - \mathbf{x}')\boldsymbol{\delta}_\psi^{\mathrm{T}}\boldsymbol{\psi}(\mathbf{x}', t_k)\, d\mathbf{x} = \boldsymbol{\delta}_\psi^{\mathrm{T}}\boldsymbol{\psi}(\mathbf{x}_m, t_k). \quad (25)$$

Thus, as for the ES it is assumed that all the statistics are Gaussian. This would be valid for linear dynamics and then the EnKF estimate would equal the Kalman filter estimate.

c. Ensemble Kalman smoother

It will now be illustrated how a variance minimizing analysis similar to the one used in the EnKF and the ES can be used to solve the sequential smoother formulation from (11)–(13). The approach is very similar to the EnKF since it is sequential in time. The only extension is that we now compute the error covariance $\mathbf{C}_{\psi\psi}(\mathbf{x}, \tau, \mathbf{x}', t_k)$, $\tau \in [t_0, t_k]$ from the prediction

$$f(\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_k|\mathbf{d}_1, \ldots, \mathbf{d}_{k-1}) = \frac{f(\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{k-1}|\mathbf{d}_1, \ldots, \mathbf{d}_{k-1}) f(\boldsymbol{\psi}_k|\boldsymbol{\psi}_{k-1})}{\int (\cdots) d\boldsymbol{\psi}}. \quad (26)$$

The analysis then operates on the previous estimate for $\tau \in [t_0, t_k]$ through the equation

$$\boldsymbol{\psi}^a_{\mathrm{EnKS}}(\mathbf{x}, \tau) = \boldsymbol{\psi}^f_{\mathrm{EnKS}}(\mathbf{x}, \tau) + \sum_{m=1}^{M_k} b_m \mathbf{r}_{k,m}(\mathbf{x}, \tau), \quad (27)$$

where the subscripts k, m denote that this is the mth representer for a dataset at time t_k. The vector of coefficients, $\mathbf{b}$, are solutions of the system

$$(\mathbf{R}_k + \mathbf{W}_k^{-1})\mathbf{b} = \mathbf{d}_k - \boldsymbol{L}_{k,[\mathbf{x},\tau]}[\boldsymbol{\psi}^f_{\mathrm{EnKS}}(\mathbf{x}, \tau)]. \quad (28)$$

Note that the $\mathbf{b}$ found in (28) for the ensemble Kalman smoother (EnKS) at $t = t_k$ will be identical to the one found in the EnKF at time $t = t_k$. The representers can now be written as

$$\mathbf{r}_{k,m}(\mathbf{x}, \tau) = L_{k,m,[\mathbf{x}',t']}[\mathbf{C}_{\psi\psi}(\mathbf{x}, \tau, \mathbf{x}', t')], \quad (29)$$

and the representer matrix is identical to the one used in the EnKF, where the columns are constructed from $\mathbf{R}_k(:, m) = \boldsymbol{L}_{k,[\mathbf{x},\tau]}[\mathbf{r}_{k,m}(\mathbf{x}, \tau)]$. For a direct measurement the mth component of the measurement functional would be written as

$$L_{k,m,[\mathbf{x}',t']}[\boldsymbol{\psi}] = \int_{\mathcal{D}}\int_{\mathcal{T}} \delta(\mathbf{x}_m - \mathbf{x}')\delta(t_k - t')\boldsymbol{\delta}_\psi^{\mathrm{T}}\boldsymbol{\psi}(\mathbf{x}', t')\, d\mathbf{x} = \boldsymbol{\delta}_\psi^{\mathrm{T}}\boldsymbol{\psi}(\mathbf{x}_m, t_k). \quad (30)$$

Thus, even if the analysis has been computed sequentially in time for each observation vector, $\mathbf{d}_k$, the estimate at times $t \le t_k$ contains information from all data up to and including t_k. Thus, it is a true smoother. A few remarks should be made.

- The smoother is entirely sequential. There will be no backward integrations of any model equations.
- The similarity with the EnKF is obvious. At the final time the solution is the EnKF solution.
- The first guess solution for $t \in [t_{k-1}, t_k]$ is the EnKF solution, contrary to the climatology in the ES method.
- The vectors of coefficients $\mathbf{b}$ are identical for the EnKF and the EnKS.
- The method is consistent with prior assumptions and provides posterior error estimates.
- The matrix $\mathbf{R}_k + \mathbf{W}^{-1}$, which needs to be inverted, will have dimension M_k rather than M as in the ES.
- For linear dynamics, the ES and EnKS should give identical solutions.

4. An example

The example from Evensen (1997) will be used to illustrate the new EnKS, and also to intercompare it with the EnKF and ES. Thus, we adapt the Lorenz equations with exactly the same setup as was used in Evensen (1997).

a. Model equations

The Lorenz model consists of a system of three coupled and nonlinear ordinary differential equations (Lorenz 1963):

$$\frac{dx}{dt} = \gamma(y - x), \qquad \frac{dy}{dt} = \rho x - y - xz,$$

$$\frac{dz}{dt} = xy - \beta z. \tag{31}$$

Here $x(t)$, $y(t)$, and $z(t)$ are the dependent variables, and we have chosen the following commonly used values for the parameters in the equations: $\gamma = 10$, $\rho = 28$, and $\beta = 8/3$.

b. Description of experiments

For all the cases to be discussed the initial conditions for the reference case are given by $(x_0, y_0, z_0) = (1.508\,870,\ -1.531\,271,\ 25.460\,91)$ and the time interval is $t \in [0, 40]$. The observations and initial conditions are simulated by adding normal distributed noise with zero mean and variance equal to 2.0 to the reference solution. The variables x, y, and z are measured. The initial conditions used are also assumed to have the same variance as the observations. These are the same values as were used in Miller et al. (1994) and Evensen (1997).

The model error covariance is defined to be diagonal with variances equal to 2.00, 12.13, and 12.31 for the three equations in (31), respectively. These numbers define the error variance growth expected over one unit time in the model, and they correspond to the numbers used in Evensen (1997). The reference case is generated by integrating the model equations including the stochastic forcing corresponding to the specified model error variances. The stochastic forcing is included through a term like $\sqrt{\Delta t}\sqrt{\sigma^2}d\beta$, where σ^2 is the model error variance and $d\beta$ is drawn from the distribution $N(0, 1)$.

In the calculation of the ensemble statistics an ensemble of 1000 members was used. This is a fairly large ensemble but it was chosen so as to prevent the possibility of drawing erroneous conclusions due to the use of too small an ensemble. The same simulation was rerun with various ensemble sizes and the differences between the results were minor; for example, using 500 members gave essentially the same result as the 1000-member case.

c. Assimilation experiment

The three methods discussed above will now be examined and compared in an experiment where the distance between the measurements is $\Delta t_{\text{obs}} = 0.5$, which is similar to experiment B in Evensen (1997).

In the upper plots in Figs. 1–8, the full line denotes the estimate and the dash–dot line is the reference solution. In the lower plots the full line is the estimated standard deviation (from ensemble statistics), while the dash–dot line is the true residuals with respect to the reference solution.

1) ENSEMBLE SMOOTHER SOLUTION

The ensemble smoother solution for the x and z components and their estimated error variance are given in Figs. 1 and 2, respectively. The dotted lines located close to $x = 0$ and $z = 24$, in Figs. 1 and 2, respectively, are the first guesses used by the ES computed as the mean of all the members in the ensemble.

It was found that the ES performed rather poorly with the current data density. Note, however, that even if the fit to the reference trajectory is rather poor, it captures most of the transitions. The main problem is related to the estimate of the amplitudes in the reference solution. The problems for the ES are linked to the appearance of non-Gaussian contributions in the distribution for the model evolution, which can be expected in such a strongly nonlinear case.

Remember that the smoother solution consists of a first guess estimate (mean of the freely evolving ensemble) plus a linear combination of time-dependent influence functions or representers, which are calculated from the ensemble statistics. Thus, the method becomes equivalent to a variance minimizing objective analysis method where the time dimension is included.

In the ensemble smoother the posterior error variances can easily be calculated by performing an analysis for each of the ensemble members and then evaluating the variance of the new ensemble. Clearly, the error estimates are not large enough at the peaks where the smoother performs poorly. This is again a result of neglecting the non-Gaussian contribution from the probability distribution for the model evolution. Thus, the method assumes the distribution is Gaussian and believes it is doing well. Otherwise the error estimate looks reasonable with minima at the measurement locations and maxima in between the measurements. Note again that if a linear model is used the posterior density will be Gaussian and the ensemble smoother will, in the limit of an infinite ensemble size, provide the same solution as would be found using the Kalman smoother or the representer method (Bennett 1992).

2) ENSEMBLE KALMAN FILTER SOLUTION

The ensemble Kalman filter does a reasonably good job at tracking the reference solution with the lower data density, as can be seen in Figs. 3 and 4. One transition is missed in this case at $t = 18$, and there are also a few other locations where the EnKF has problems, for example, $t = 1$, 5, 9, 10, 13, 17, 19, 23, 26, and 34. The error variance estimate is also rather consistent, showing large peaks at the locations where the estimate obviously has problems tracking the reference solution. Note also the similarity between the absolute value of the residual between the reference solution and the estimate, and the estimated standard deviation. To all peaks in the residual there is a corresponding peak for the error variance estimate.

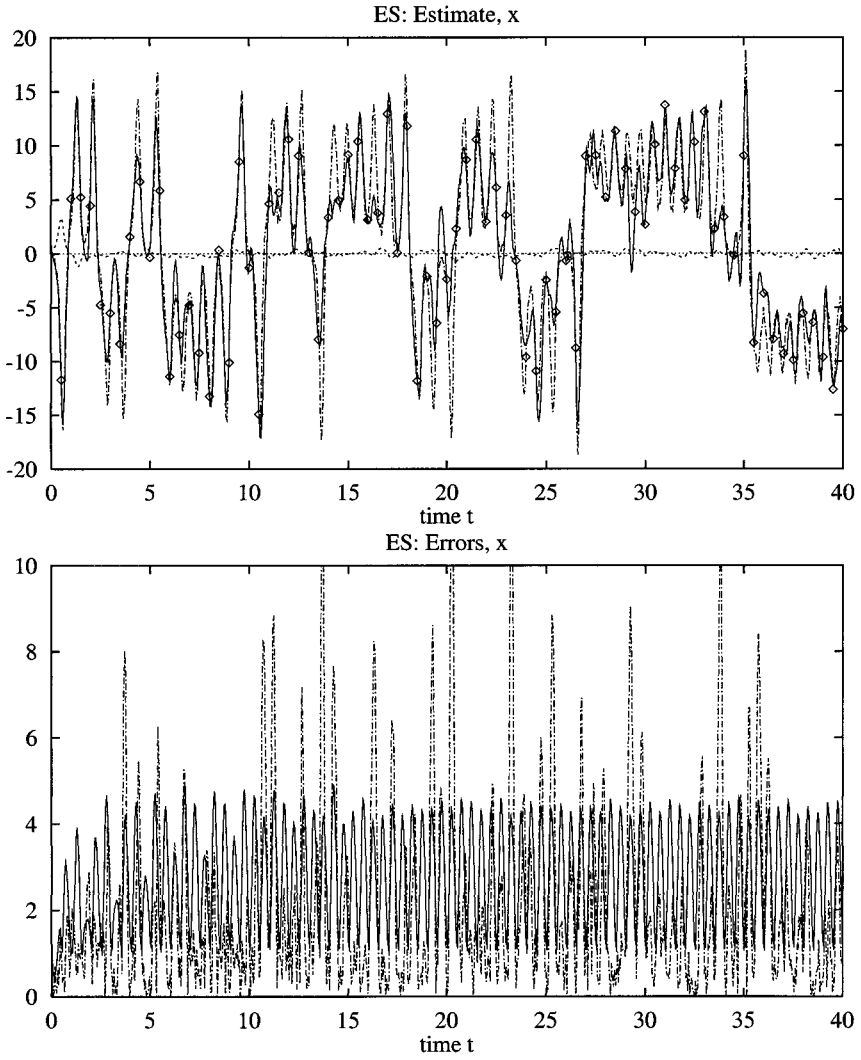

FIG. 1. Ensemble smoother: (top) The inverse estimate (full line) and reference solution (dash–dot line) for x. (bottom) The corresponding estimated standard deviations (full line) and the absolute value of the difference between the reference solution and the estimate, i.e., the real posterior errors (dash–dot line). For this particular plot we also show the first guess used by the ES as the dotted line located close to $x = 0$.

The error estimates show the same behavior as was found by Miller et al. (1994) with very strong error growth when the model solution is passing through the unstable regions of the state space, and otherwise rather weak error variance growth in the more stable regions. Note for example the low error variance when $t \in [28, 34]$ corresponding to the oscillation of the solution around one of the attractors.

A somewhat surprising result is that the ensemble Kalman filter seems to perform better than the ensemble smoother. This is at least surprising based on linear theory, where one has learned that the Kalman smoother solution at the end of the time interval is identical to the Kalman filter solution, and the additional information introduced by propagating the contribution of future measurements backward in time further reduces the error variance compared to the filter solution. Note again that if the model dynamics are linear, the ensemble Kalman filter will give the same solution as the Kalman filter, and the ensemble smoother will give the same result as the Kalman smoother.

3) ENSEMBLE KALMAN SMOOTHER SOLUTION

In Figs. 5 and 6 the solutions obtained by the EnKS are shown. Clearly, the estimate is an improvement upon the EnKF estimate. The solution is smoother in time and seems to provide a better fit to the reference trajectory. Looking in particular at the problematic locations in the EnKF solution, these are all recovered in the smoother estimate. Note, for example, the additional transitions in $t = 1$, 5, 13, and 34, in the EnKF solution

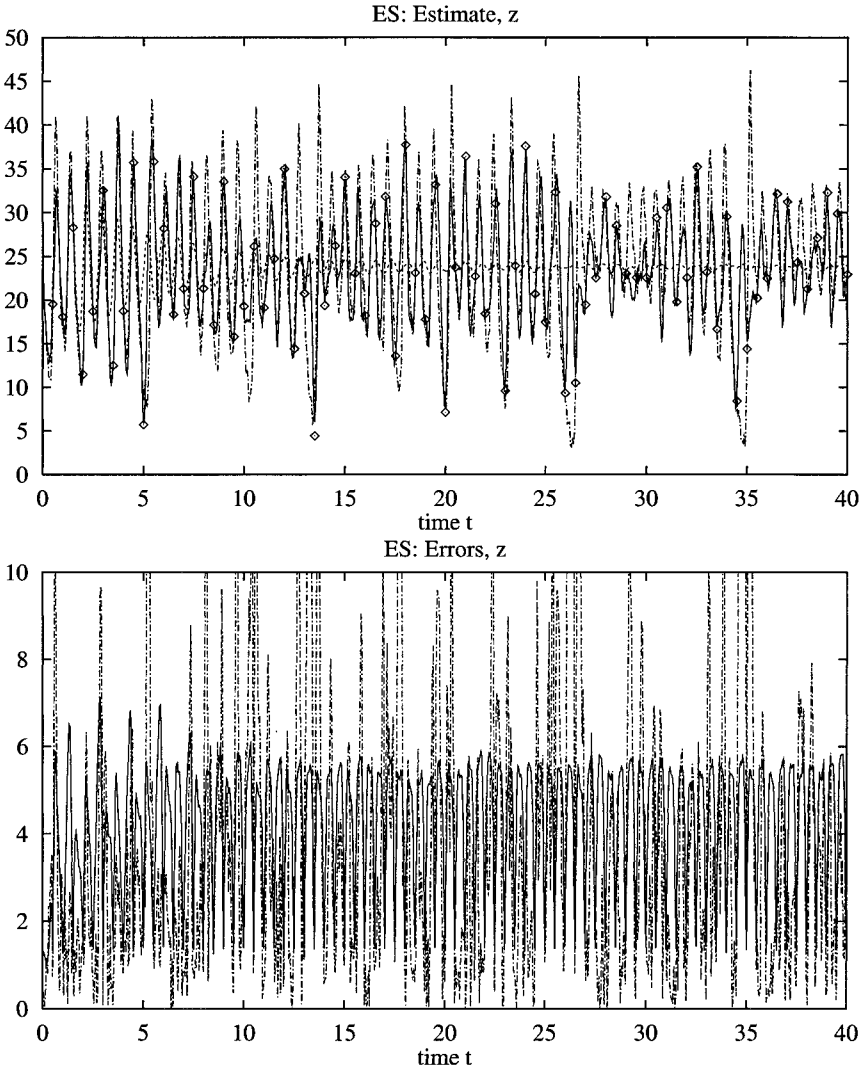

FIG. 2. Ensemble smoother: Same as in Fig. 1 but for the z component. Also here the first guess for the ES is shown as a dotted line close to $z = 24$.

that have now been eliminated in the smoother. The missed transition at $t = 17$ has also been recovered in the smoother solution.

The error estimates are also reduced all through the model domain. In particular the large peaks in the EnKF solution are now significantly reduced. As for the EnKF solution there are corresponding peaks in the error estimates for all the peaks in the residuals, which proves that the EnKS error estimate is consistent with the true errors.

This is a very promising result. In fact the EnKS solution with $\Delta t_{\text{obs}} = 0.5$ seems to do as well or better than the EnKF solution with $\Delta t_{\text{obs}} = 0.25$ (see experiment A in Evensen 1997).

In Figs. 7 and 8 the results for the lagged smoother are shown. A time lag of five time units was used and the results are almost indistinguishable from the full smoother solution. Thus a significant saving of CPU and storage should be possible for more realistic applications when using the lagged smoother.

5. The effect of non-Gaussian statistics

As has been pointed out the analysis schemes used here can be characterized as variance minimizing. Thus, even if the ensemble certainly is non-Gaussian due to the forward integration of nonlinear model equations, only the Gaussian part of the distribution is used. This is in contrast to the work by Miller et al. (1999) where the maximum-likelihood analysis is calculated by actually constructing the density function for the model evolution and then calculating the conditional density in terms of analytical functions. They found that this made a significant improvement on the analysis. However, it is still not clear how this approach can be used in a practical way for high-dimensional state spaces.

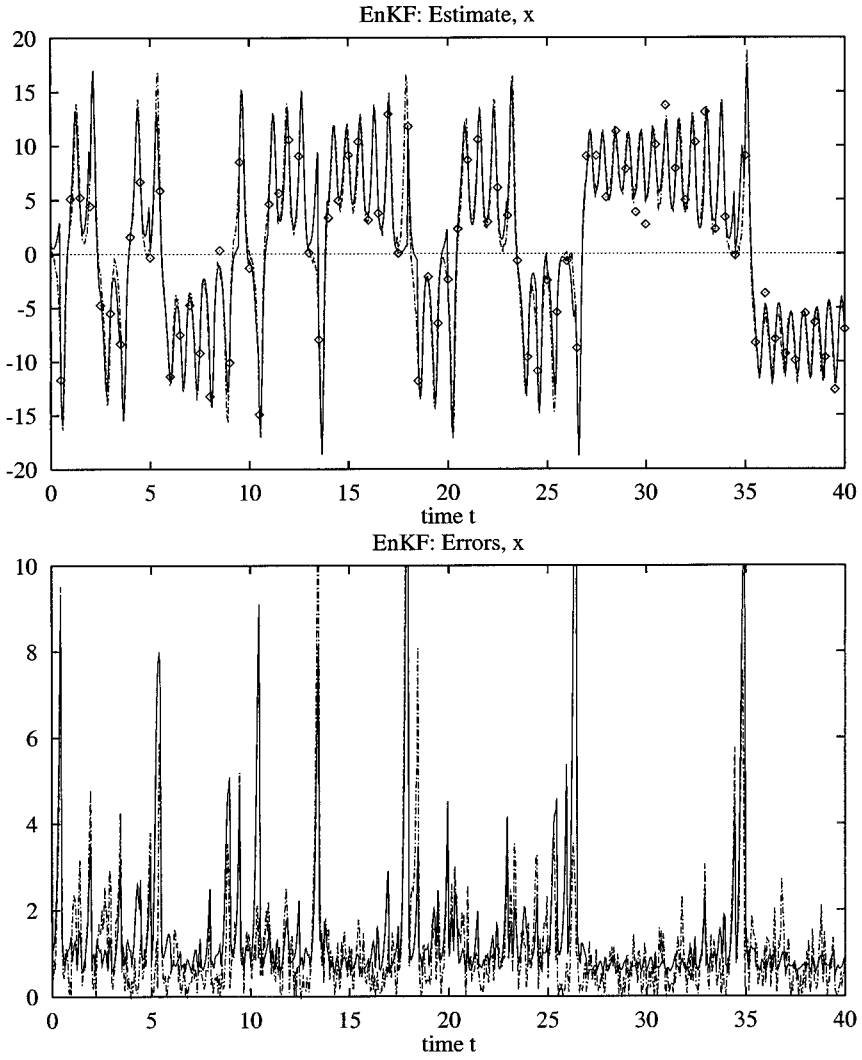

FIG. 3. Ensemble Kalman filter: See explanation in Fig. 1.

It is of interest to examine the properties of the analysis scheme in the ES, EnKF, and EnKS when the distribution of the predicted ensemble is strongly non-Gaussian. Assume now that the forecast distribution is given by the steady-state double-well distribution discussed by Miller et al. (1994). Essentially it describes a state located with equal probability in two potential wells, one located around $x = -1$ and the other around $x = 1$.

The double-well distribution is given as

$$\phi_f(x) = A \exp\left[-\frac{2.0}{\sigma^2}h(x)\right], \tag{32}$$

where σ^2 is the variance of the stochastic forcing in the double-well model (Miller et al. 1994) and A is a normalization coefficient. The function $h(x)$ is defined as

$$h(x) = x^2(x^2 - 2). \tag{33}$$

The distribution $\phi_f(x)$ will now be used as a first guess distribution when examining the analysis schemes. Further, it is assumed that the distribution for a single observation is Gaussian

$$\phi_o(d|x) = A \exp\left[-\frac{(x - d)^2}{\sigma_o^2}\right], \tag{34}$$

where d is the observation and σ_o^2 is the variance of the observation error. The distribution for x given the datum d now becomes

$$\phi(x|d) = A\phi_o(d|x)\phi_f(x) = A \exp[-g(x)], \tag{35}$$

where the function $g(x)$ is

$$g(x) = \frac{2}{\sigma^2}x^4 - \left(\frac{4}{\sigma^2} - \frac{1}{\sigma_o^2}\right)x^2 - \frac{2d}{\sigma_o^2}x + \frac{d^2}{\sigma_o^2}. \tag{36}$$

The MLH estimate is given by the maximum value of $\phi(x|d)$, which can be found from

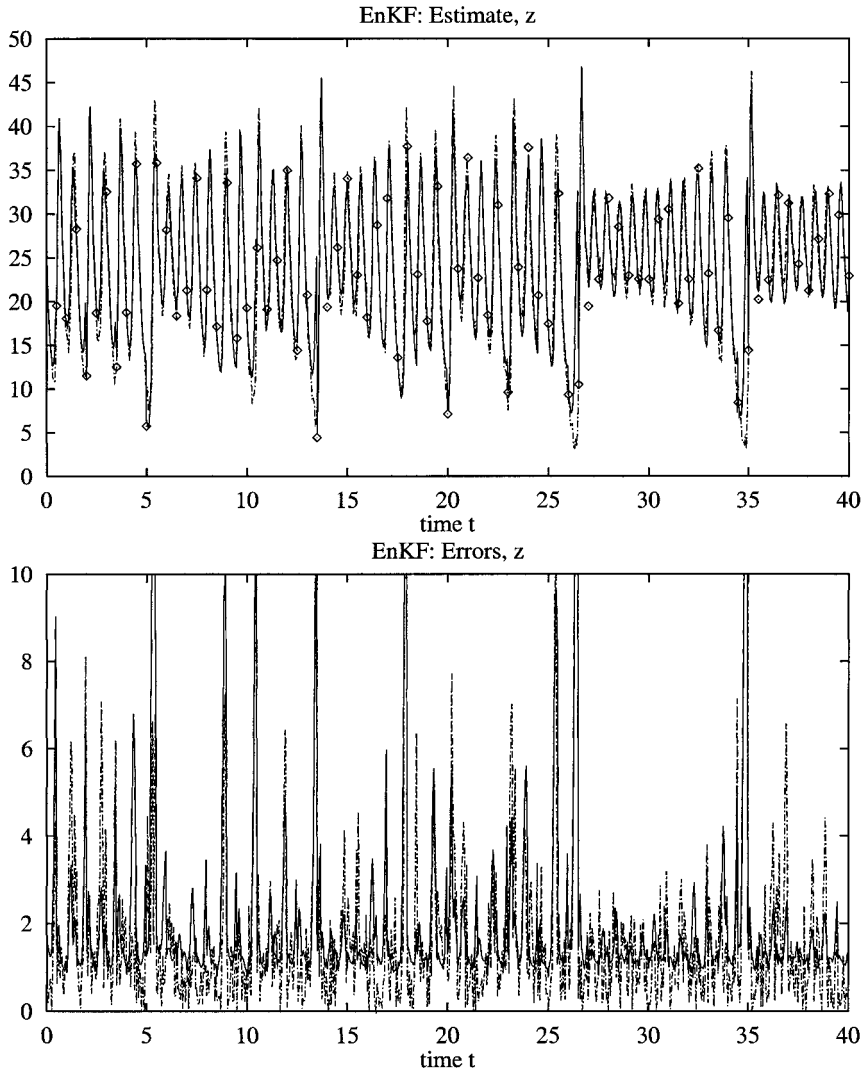

FIG. 4. Ensemble Kalman filter: See explanation in Fig. 2.

$$\frac{\partial \phi(x\,|\,d)}{\partial x} = -\exp[-g(x)]\frac{\partial g(x)}{\partial x} = 0 \quad \text{or} \tag{37}$$

$$\frac{\partial g(x)}{\partial x} = \frac{8}{\sigma^2}x^3 - 2\left(\frac{4}{\sigma^2} - \frac{1}{\sigma_o^2}\right)x - \frac{2d}{\sigma_o^2} = 0. \tag{38}$$

The mean of the posterior distribution is an alternative estimator and is given as

$$\overline{x} = \int_{-\infty}^{\infty} x\phi(x\,|\,d)\,dx, \tag{39}$$

and it is also the variance minimizing estimate from the posterior distribution.

The procedure used for computing the analysis in the previous sections is vastly different. There it is assumed that the first guess distribution and the distribution for the observation are both Gaussian. Thus, they are both represented by their respective (co)variances and a standard variance minimizing analysis is then calculated. In our simple example that would be equivalent to first calculating the mean of $\phi_f(x)$,

$$\overline{x_f} = \int_{-\infty}^{\infty} x\phi_f(x)\,dx, \tag{40}$$

which in this case is equal to zero, and the variance

$$\overline{x_f^2} = \int_{-\infty}^{\infty} x^2\phi_f(x)\,dx. \tag{41}$$

For the observations, the mean, d, and the variance, σ_o^2, from $\phi_o(d\,|\,x)$ are used.

Then the variance minimizing analysis is calculated from

$$\hat{x} = 0 + \frac{\overline{x_f^2}}{\overline{x_f^2} + \sigma_o^2}(d - 0), \tag{42}$$

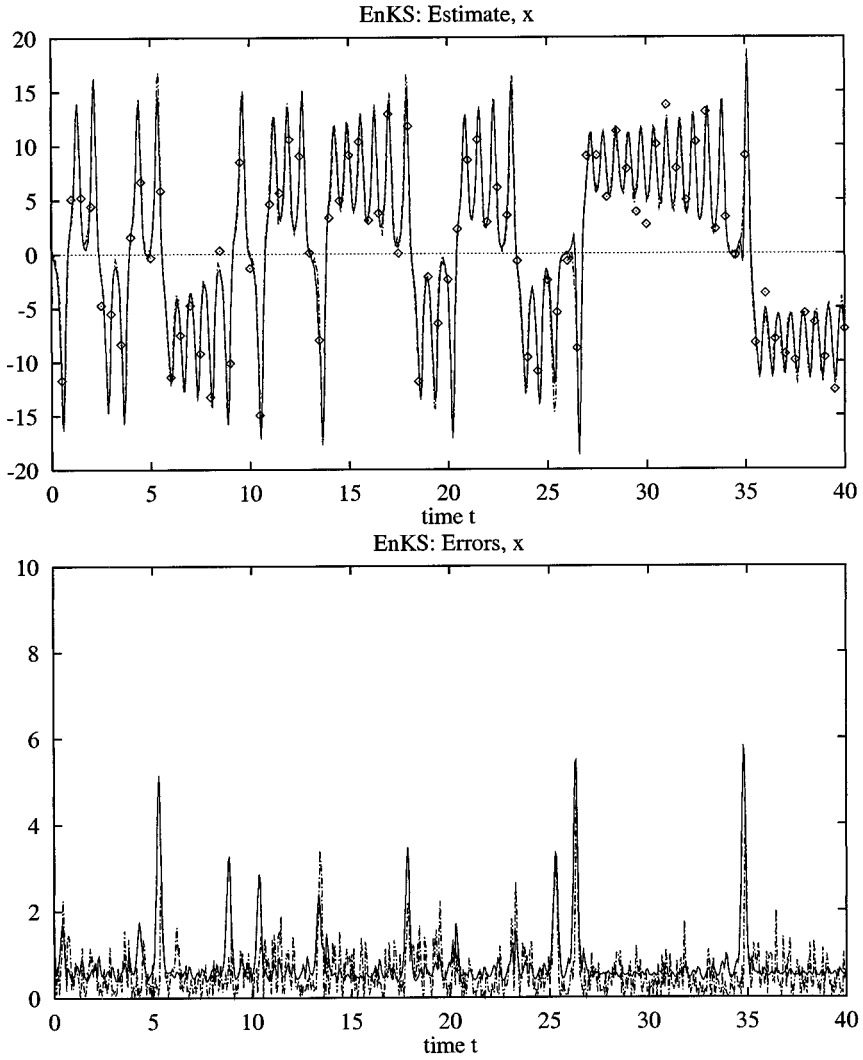

FIG. 5. Ensemble Kalman smoother: See explanation in Fig. 1.

with $\hat{x}$ representing the ES, EnKF, and EnKS estimate. This is different from the estimate found using (39) where the full non-Gaussianities of the prior distributions are taken into account when creating the posterior distribution.

In Fig. 9 an example is given where the double-well distribution indicates that the most likely value for the prediction is either -1 or 1. This is certainly a highly non-Gaussian distribution, which has a mean equal to zero. The observation has the value 0.5, and is relatively inaccurate as can be seen from the wide distribution function. The posterior distribution indicates a large probability for the estimate to be located somewhere in the well centered at $x = 1$. The MLH estimate found from (38) and the mean of the posterior distribution defined by (39) are very similar. The mean is shifted slightly toward zero, which is mainly a contribution from the minor probability that the state would actually be located in the well at $x = -1$. The EnKF-type analysis completely misses the probable solutions and ends up in a very unlikely state that hardly ever occurs. A more accurate observation is shown in Fig. 10 just to illustrate the stronger contribution from the observation on the posterior distribution, and the still very poor estimate from the EnKF analysis scheme. From this example it is clear that the EnKF does not give the optimal solution if a non-Gaussian distribution is involved.

In van Leeuwen and Evensen (1996) it was pointed out that the mean of the posterior distribution and the posterior variance could easily be calculated in the case where the first guess distribution is represented by an ensemble, using a frequency interpretation. The estimate could then be given as

$$E[\boldsymbol{\psi}] = \left[\sum_{i=1}^{N} \boldsymbol{\psi}_i f(\mathbf{d}\,|\,\boldsymbol{\psi}_i)\right] \Big/ \left[\sum_{i=1}^{N} f(\mathbf{d}\,|\,\boldsymbol{\psi}_i)\right], \quad (43)$$

and the variance becomes

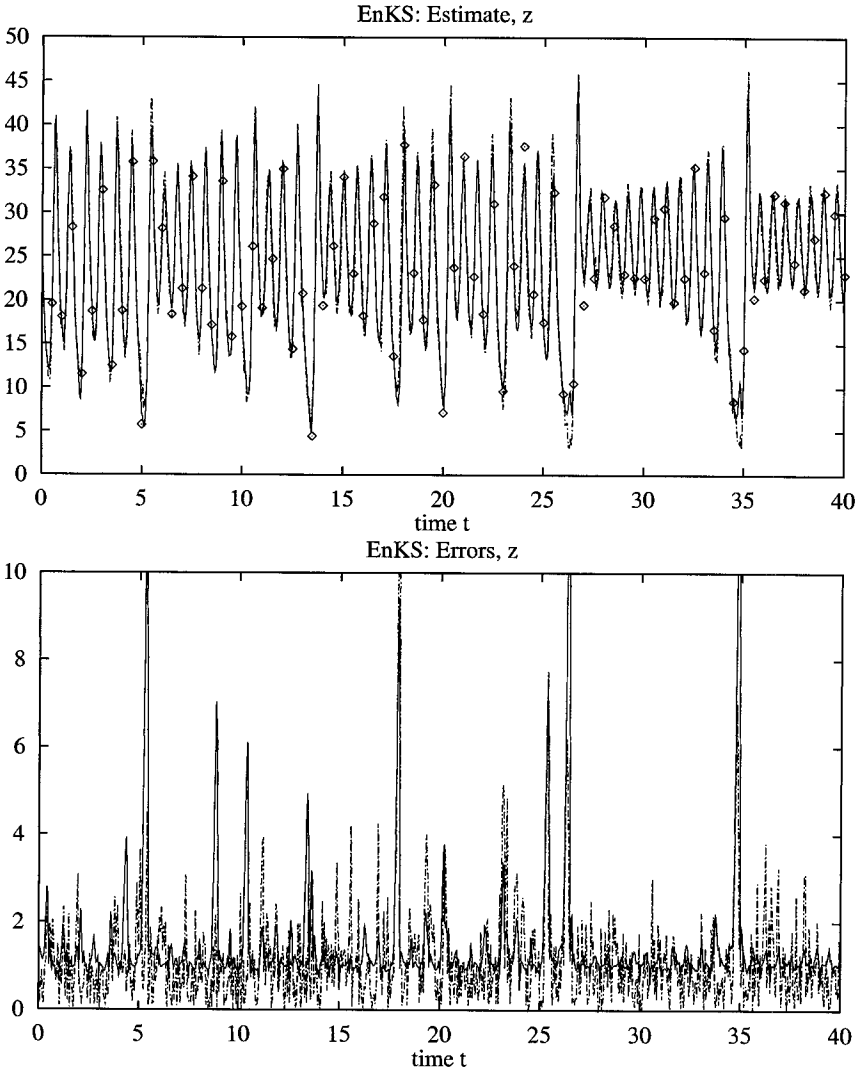

FIG. 6. Ensemble Kalman smoother: See explanation in Fig. 2.

$$E[(\psi - E[\psi])^2] = \frac{\sum_{i=1}^{N} (\psi_i - E[\psi])^2 f(\mathbf{d} \mid \psi_i)}{\sum_{i=1}^{N} f(\mathbf{d} \mid \psi_i)}, \quad (44)$$

Thus, the estimate is a weighted average of all the ensemble members, where the weights are determined by the distance between the members and the data. That is, ensemble members that are close to the data are weighted higher than those in less agreement with the data. However, one also needs an efficient method for creating the new ensemble representing the posterior distribution. One of the strengths of the schemes used in this paper is that the posterior ensemble automatically comes out of the analysis (Burgers et al. 1998).

6. Discussion

A new formulation for an ensemble-based smoother, which can be used in a sequential approach, has been presented. The method, which is named the ensemble Kalman smoother, has been tested with the Lorenz equations and intercompared with results from the ensemble Kalman filter by Evensen (1994b) and the ensemble smoother presented by van Leeuwen and Evensen (1996). The new formulation turned out to provide a significant improvement over the EnKF and the ES solutions. It can be considered as an extension of the EnKF since the EnKF solution is used as the first guess for the analysis, which is propagated backward in time by using the ensemble covariances. This is done in a sequential approach every time new measurements are available. Thus, the method can be characterized as a

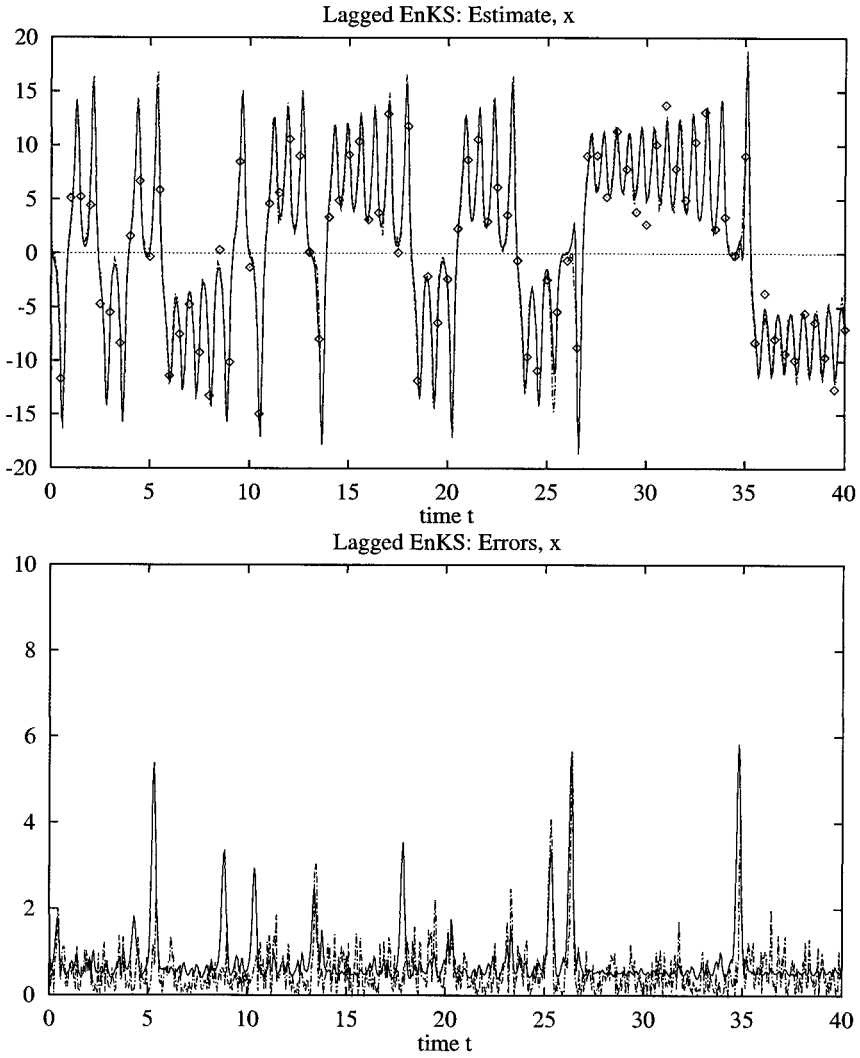

FIG. 7. Lagged ensemble Kalman smoother: See explanation in Fig. 1.

sequential smoother. No backward integrations in time and no adjoint operators are needed.

It was also shown that the EnKS is identical to the ES for linear dynamics. The main advantage for the EnKS when used with nonlinear dynamics is that it relies on the EnKF prediction as the first guess estimate for each analysis. This is clearly an advantage considering that the EnKF also outperformed the ES in Evensen (1997).

The ES provides a variance-minimizing estimate by rejecting the non-Gaussian part of the density of the pure model evolution. That is, it uses the model climatology as the first guess and computes the time- and space-dependent influence functions for each measurement. The problem is that the climatology becomes very bad as a first guess for nonlinear and unstable dynamics. The error variance of the climatology (ensemble mean) will be very large and much accurate data will be needed to pull the estimate toward an accurate representation of the reference solution or true state. Further, in a system like the Lorenz equations we can also expect that the distribution for the model evolution becomes strongly non-Gaussian (see e.g. Miller et al. 1999), and thus the variance minimizing estimate becomes very inaccurate.

The EnKF has proven to do a good job in tracking the reference solution for the Lorenz equations. When using the EnKF solution as a first guess, the EnKS is capable of providing a significant improvement over the EnKF estimate. The EnKS estimate eliminates the discontinuities or steps normally obtained with all sequential data assimilation methods. Thus, if the objective is to compute a climatology over a certain time period, the smoother solution is certainly preferable. On the other hand, in a forecasting mode there is no point in using the smoother since the EnKF and the EnKS will give the same estimate at the latest time, which would be used as an initial condition for the forecast.

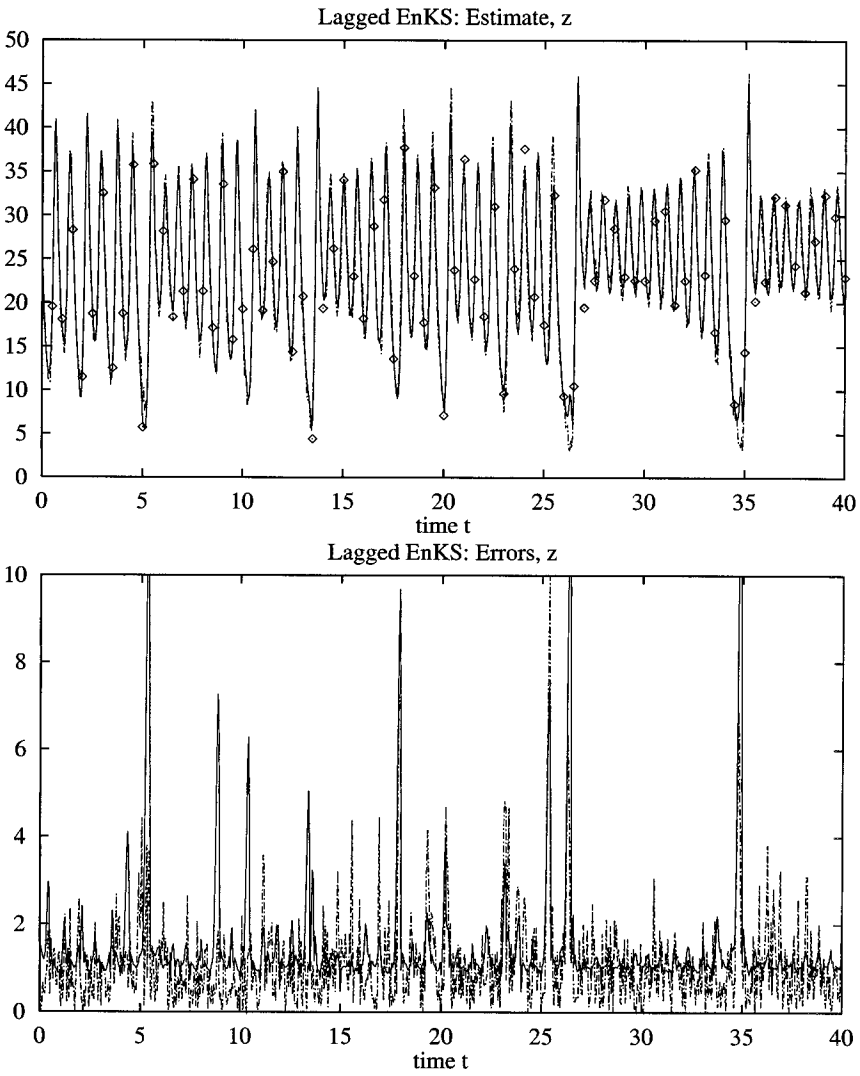

FIG. 8. Lagged ensemble Kalman smoother: See explanation in Fig. 2.

The CPU requirements for the new EnKS are similar to those needed for the EnKF. In practical applications one could imagine that the analysis is to be computed at discrete times during a simulation, say every 6 h, which would be typical for an atmospheric model application. To compute the analysis one would need to store the full ensemble at these time instants. Then the representers or influence functions for the observation at the data time and each of the previous analysis times are computed and the estimate is updated.

The storage requirements may become huge for long time intervals with many analysis times. Fortunately, it is possible to use a lagged version of the EnKS, which is based on an assumption about the covariances in time. If one can assume that the time correlation in the ensemble statistics approaches zero over a certain time interval, for example, a few times the predictability time, the lagged smoother can be used to obtain a significant saving of CPU time.

It should be noted that the EnKS is not the only smoother method that benefits from the sequential formulation presented here. There should also be room for a simplification when using the representer method (Bennett 1992) in some applications. The most obvious approach would be to perform the analysis in the same way as is done in the EnKS, except that the influence functions for all prior times are computed by integrations of the adjoint model rather than using the ensemble statistics directly. This would lead to an implementation of the representer method where measurements are processed sequentially in time. For practical implementations a lagged version may also be used, thus making the representer methods more suitable for operational forecasting systems. The similarity between a sequential

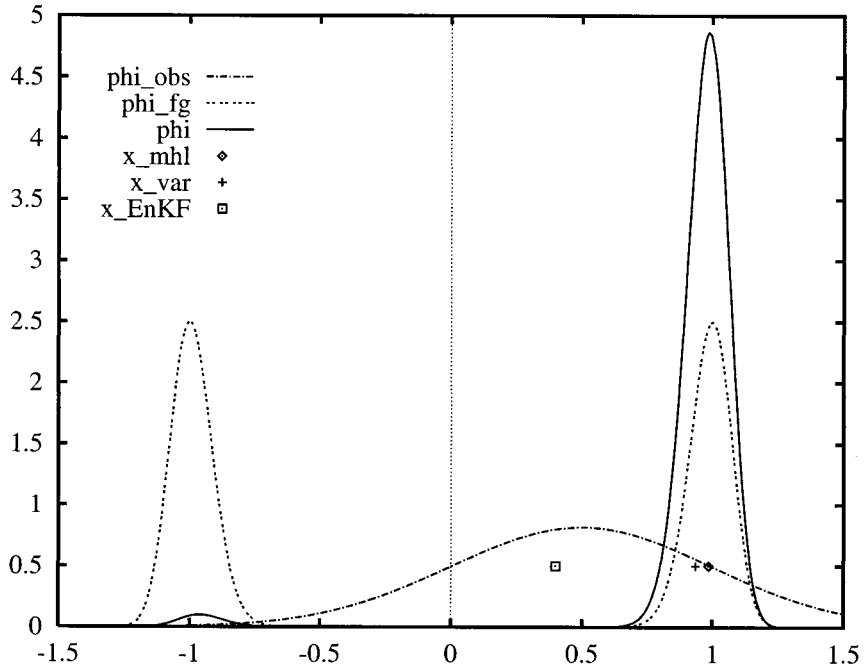

FIG. 9. The dotted line is the double-well distributon defining the first guess, the dash–dot line is the distribution for a measurement, and the solid line is the posterior distribution. The diamond is the MLH estimate, the plus is the mean of the posterior distribution, and the square is the estimate as would be computed from the schemes used in the previous examples.

representer method and the EnKS also allows for a joint implementation where, for example, the representer matrix at time t_k could be obtained from ensemble predictions. One can then solve for the coefficients **b** corresponding to the data at time t_k, and the analysis at previous times could be computed by one backward integration of the adjoint model followed by a forward integration of the model itself.

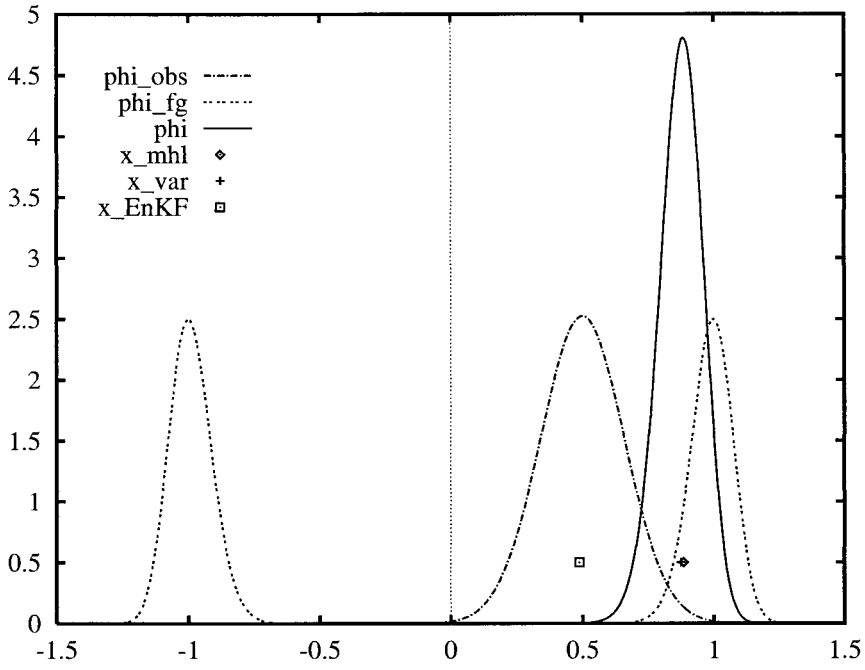

FIG. 10. Same as Fig. 9 but with a more accurate observation.

There is also an apparent similarity between the EnKS and an augmented Kalman filter. As was pointed out by an anonymous reviewer, the lagged EnKS can be considered as an EnKF with an augmented model state consisting of the present state vector along with the past state vectors back to the predefined fixed lag. Thus, the notion of a sequential smoother is not so strange after all.

The use of a variance minimizing estimator for non-Gaussian statistics was also discussed. There is clearly room for improvement of the analysis scheme currently used to take into account properly the non-Gaussian contributions. The maximum-likelihood estimate is normally recognized as the optimal estimate, but it becomes extremely complicated to find it for high-dimensional systems (see Miller et al. 1999). An alternative might be to elaborate further on the direct ensemble method proposed by van Leeuwen and Evensen (1996).

A final remark should be made. The posterior distributions from the examples shown in Figs. 9 and 10 are nearly Gaussian, even if the priors are strongly non-Gaussian. This could indicate that the assimilation of measurements have a regularization effect and may result in more Gaussian-shaped distributions. One could speculate that this is also one reason why the EnKF performs better than the ES. If the time interval between measurements is small, all the members will stay at the same attractor and the estimate is then represented by a nearly Gaussian distribution. Further, this argument could also support the use of the formulas (43) and (44) for computing the analysis and then creating a new Gaussian ensemble by sampling a multivariate Gaussian distribution with the mean and (co)variance defined by these formulas. However, it would be preferrable if an ensemble that correctly represents the posterior distribution could be computed.

This discussion illustrates how complicated the data assimilation problem becomes when strongly nonlinear dynamics are involved. There are several solution methods for the inverse problem that for nonlinear models result in different solutions. It can be formally shown that all these methods give identically the same solution when used with linear dynamics.

Acknowledgments. G. Evensen was supported by the European Commission through the Environment and Climate project AGORA under Contract ENV4-CT95-0113 and the MAST-III project DIADEM under Contract MAS3-CT98-0167. P. J. van Leeuwen was supported by the Dutch National Research Program, Grant 013001237.10, and the MAST-III project DIADEM under Contract MAS3-CT98-0167.

REFERENCES

Bennett, A. F., 1992: *Inverse Methods in Physical Oceanography*. Cambridge University Press, 346 pp.

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

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

Evensen, G., 1994a: Inverse methods and data assimilation in nonlinear ocean models. *Physica D,* **77,** 108–129.

——, 1994b: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

——, 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.,* **125,** 1342–1354.

——, and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.,* **124,** 85–96.

——, and N. Fario, 1997: A weak constraint variational inverse for the Lorenz equations using substitution methods. *J. Meteor. Soc. Japan,* **75,** 229–243.

Gauthier, P., 1992: Chaos and quadri-dimensional data assimilation: A study based on the Lorenz model. *Tellus,* **44A,** 2–17.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130–141.

Miller, R. N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.,* **51,** 1037–1056.

——, E. F. Carter, and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus,* **51A,** 167–194.

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