Analysis Scheme in the Ensemble Kalman Filter

Gerrit Burgers

Royal Netherlands Meteorological Institute, De Bilt, the Netherlands

Peter Jan van Leeuwen

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

Geir Evensen

Nansen Environmental and Remote Sensing Center, Bergen, Norway

(Manuscript received 17 January 1997, in final form 10 June 1997)

ABSTRACT

This paper discusses an important issue related to the implementation and interpretation of the analysis scheme in the ensemble Kalman filter. It is shown that the observations must be treated as random variables at the analysis steps. That is, one should add random perturbations with the correct statistics to the observations and generate an ensemble of observations that then is used in updating the ensemble of model states. Traditionally, this has not been done in previous applications of the ensemble Kalman filter and, as will be shown, this has resulted in an updated ensemble with a variance that is too low.

This simple modification of the analysis scheme results in a completely consistent approach if the covariance of the ensemble of model states is interpreted as the prediction error covariance, and there are no further requirements on the ensemble Kalman filter method, except for the use of an ensemble of sufficient size. Thus, there is a unique correspondence between the error statistics from the ensemble Kalman filter and the standard Kalman filter approach.

1. Introduction

The ensemble Kalman filter (EnKF) was introduced by Evensen (1994b) as an alternative to the traditional extended Kalman filter (EKF), which has been shown to be based on a statistical linearization or closure approximation that is too severe to be useful for some cases with strongly nonlinear dynamics (see Evensen 1992; Miller et al. 1994; Gauthier et al. 1993; Bouttier 1994). If the dynamical model is written as a stochastic differential equation, one can derive the Fokker–Planck or Kolmogorov's equation for the time evolution of the probability density function, which contains all the information about the prediction error statistics. The EnKF is a sequential data assimilation method, using Monte Carlo or ensemble integrations. By integrating an ensemble of model states forward in time, it is pos-sible to calculate the mean and error covariances needed at analysis times.

The analysis scheme that has been proposed in Evensen (1994b) uses the traditional update equation of the Kalman filter (KF), except that the gain is calculated from the error covariances provided by the ensemble of model states. It was also illustrated that a new ensemble representing the analyzed state could be generated by updating each ensemble member individually using the same analysis equation.

The EnKF is attractive since it avoids many of the problems associated with the traditional extended Kalman filter; for example, there is no closure problem as is introduced in the extended Kalman filter by neglecting contributions from higher-order statistical moments in the error covariance evolution equation. It can also be computed at a much lower numerical cost, since usually a rather limited number of model states is sufficient for reasonable statistical convergence. For sufficient ensemble sizes, the errors will be dominated by statistical noise, not by closure problems or unbounded error variance growth.

Corresponding author address: Gerrit Burgers, Oceanographic Research Division, Royal Netherlands Meteorological Institute, P.O. Box 201, 3730 AE De Bilt, the Netherlands.
E-mail: burgers@knmi.nl

The EnKF has been further discussed and applied with success in a twin experiment in Evensen (1994a) and in a realistic application for the Agulhas Current using Geosat altimeter data in Evensen and van Leeuwen (1996).

A serious point that will be discussed here and was not known during the previous applications of the EnKF is that for the analysis scheme to be consistent one must treat the observations as random variables. This assumption was applied implicitly in the derivation of the analysis scheme in Evensen (1994b) but has not been used in the following applications of the EnKF. It will be shown that unless a new ensemble of observations is generated at each analysis time, by adding perturbations drawn from a distribution with zero mean and covariance equal to the measurement error covariance matrix, the updated ensemble will have a variance that is too low, although the ensemble mean is not affected.

A similar problem is present in the ensemble smoother proposed by van Leeuwen and Evensen (1996), although there only the posterior error variance estimate is influenced since the solution is calculated simultaneously in space and time.

There was also another issue pointed out in Evensen (1994b): the *error covariance* matrices for the forecasted and the analyzed estimate, $\mathbf{P}^f$ and $\mathbf{P}^a$, are in the Kalman filter defined in terms of the true state as

$$\mathbf{P}^f = \overline{(\boldsymbol{\psi}^f - \boldsymbol{\psi}^t)(\boldsymbol{\psi}^f - \boldsymbol{\psi}^t)^{\mathrm{T}}}, \tag{1}$$

$$\mathbf{P}^a = \overline{(\boldsymbol{\psi}^a - \boldsymbol{\psi}^t)(\boldsymbol{\psi}^a - \boldsymbol{\psi}^t)^{\mathrm{T}}}, \tag{2}$$

where the overbar denotes an expectation value, $\boldsymbol{\psi}$ is the model state vector at a particular time, and the superscripts *f, a,* and *t* represent forecast, analyzed, and true state, respectively. However, since the true state is not known, it is more convenient to consider *ensemble covariance* matrices around the ensemble mean $\overline{\boldsymbol{\psi}}$,

$$\mathbf{P}^f \simeq \mathbf{P}_e^f = \overline{(\boldsymbol{\psi}^f - \overline{\boldsymbol{\psi}^f})(\boldsymbol{\psi}^f - \overline{\boldsymbol{\psi}^f})^{\mathrm{T}}}, \tag{3}$$

$$\mathbf{P}^a \simeq \mathbf{P}_e^a = \overline{(\boldsymbol{\psi}^a - \overline{\boldsymbol{\psi}^a})(\boldsymbol{\psi}^a - \overline{\boldsymbol{\psi}^a})^{\mathrm{T}}}, \tag{4}$$

where now the overbar denotes an average over the ensemble. It will be shown that if the ensemble mean is used as the best estimate, the ensemble covariance can consistently be interpreted as the error covariance of the best estimate.

This leads to an interpretation of the EnKF as a purely statistical Monte Carlo method where the ensemble of model states evolves in state space with the mean as the best estimate and the spreading of the ensemble as the error variance. At measurement times each observation is represented by another ensemble, where the mean is the actual measurement and the variance of the ensemble represents the measurement errors.

Ensembles of observations were used by Daley and Mayer (1986) in an observations system simulation experiment, and more recently by Houtekamer and De-

rome (1995) in an ensemble prediction system, and by A. F. Bennett (1996) as well (personal communication) as well to derive posterior covariances for the representer method. Recently, Houtekamer and Mitchell (1998) have used ensembles of observations in the application of an ensemble Kalman filter technique.

In the following sections we will present an analysis of the consequences of using the ensemble covariance instead of the error covariance, and then present a modification of the analysis scheme where the observations are treated as random variables. Finally, the differences between the analysis steps of the standard Kalman filter, the original EnKF, and the improved scheme presented here will be illustrated by a simple example. An application of the improved scheme to a more complex example, that of the strongly nonlinear Lorenz equations, is treated in Evensen (1997).

2. The standard Kalman filter

It is instructive first to review the analysis step in the standard Kalman filter where the analyzed estimate is determined by a linear combination of the vector of measurements $\mathbf{d}$ and the forecasted model state vector $\boldsymbol{\psi}^f$. The linear combination is chosen to minimize the variance in the analyzed estimate $\boldsymbol{\psi}^a$, which is then given by the equation

$$\boldsymbol{\psi}^a = \boldsymbol{\psi}^f + \mathbf{K}(\mathbf{d} - \mathbf{H}\boldsymbol{\psi}^f). \tag{5}$$

The Kalman gain matrix $\mathbf{K}$ is given by

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^{\mathrm{T}}(\mathbf{H}\,\mathbf{P}^f \mathbf{H}^{\mathrm{T}} + \mathbf{W})^{-1}. \tag{6}$$

It is a function of the model state error covariance matrix $\mathbf{P}^f$, the data error covariance matrix $\mathbf{W}$, and the measurement matrix $\mathbf{H}$ that relates the model state to the data. In particular, the true model state is related to the true observations through

$$\mathbf{d}^t = \mathbf{H}\boldsymbol{\psi}^t, \tag{7}$$

assuming no representation errors in the measurement operator $\mathbf{H}$, while the actual measurements are defined by the relation

$$\mathbf{d} = \mathbf{H}\boldsymbol{\psi}^t + \boldsymbol{\epsilon}, \tag{8}$$

with $\boldsymbol{\epsilon}$ the measurement errors. The measurement error covariance matrix is defined as

$$\begin{aligned} \mathbf{W} &= \overline{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}} \\ &= \overline{(\mathbf{d} - \mathbf{H}\boldsymbol{\psi}^t)(\mathbf{d} - \mathbf{H}\boldsymbol{\psi}^t)^{\mathrm{T}}} \\ &= \overline{(\mathbf{d} - \mathbf{d}^t)(\mathbf{d} - \mathbf{d}^t)^{\mathrm{T}}}. \end{aligned} \tag{9}$$

As usual, we assume that $\overline{(\mathbf{d} - \mathbf{d}^t)(\boldsymbol{\psi} - \boldsymbol{\psi}^t)^{\mathrm{T}}} = 0$.

The error covariance of the analyzed model state vector is reduced with respect to the error covariance of the forecasted state as

$$\mathbf{P}^a = \overline{(\boldsymbol{\psi}^a - \boldsymbol{\psi}^t)(\boldsymbol{\psi}^a - \boldsymbol{\psi}^t)^{\mathrm{T}}}$$

$$= \overline{[\boldsymbol{\psi}^f - \boldsymbol{\psi}^t + \mathbf{K}(\mathbf{d} - \mathbf{d}^t - \mathbf{H}\boldsymbol{\psi}^f + \mathbf{H}\boldsymbol{\psi}^t)][\boldsymbol{\psi}^f - \boldsymbol{\psi}^t + \mathbf{K}(\mathbf{d} - \mathbf{d}^t - \mathbf{H}\boldsymbol{\psi}^f + \mathbf{H}\boldsymbol{\psi}^t)]^{\mathrm{T}}}$$

$$= (\mathbf{I} - \mathbf{KH})\overline{(\boldsymbol{\psi}^f - \boldsymbol{\psi}^t)(\boldsymbol{\psi}^f - \boldsymbol{\psi}^t)^{\mathrm{T}}}(\mathbf{I} - \mathbf{KH})^{\mathrm{T}} + \mathbf{K}\overline{(\mathbf{d} - \mathbf{d}^t)(\mathbf{d} - \mathbf{d}^t)^{\mathrm{T}}}\mathbf{K}^{\mathrm{T}}$$

$$= (\mathbf{I} - \mathbf{KH})\mathbf{P}^f(\mathbf{I} - \mathbf{H}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}) + \mathbf{KWK}^{\mathrm{T}}$$

$$= \mathbf{P}^f - \mathbf{KHP}^f - \mathbf{P}^f\mathbf{H}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}} + \mathbf{K}(\mathbf{HP}^f\mathbf{H}^{\mathrm{T}} + \mathbf{W})\mathbf{K}^{\mathrm{T}}$$

$$= (\mathbf{I} - \mathbf{KH})\mathbf{P}^f. \tag{10}$$

Note that after inserting Eq. (5) in the definition of $\mathbf{P}^a$, Eq. (7) is used by adding $\mathbf{H}\boldsymbol{\psi}^t - \mathbf{d}^t = 0$. The further derivation then clearly states that the observations $\mathbf{d}$ must be treated as random variables to get the measurement error covariance matrix into the expression.

The analyzed model state is the best linear unbiased estimate. This means that $\boldsymbol{\psi}^a$ is the linear combination of $\boldsymbol{\psi}^f$ and $\mathbf{d}$ that minimizes $\mathrm{Tr}\mathbf{P} = \overline{(\boldsymbol{\psi} - \boldsymbol{\psi}^t)^{\mathrm{T}}(\boldsymbol{\psi} - \boldsymbol{\psi}^t)}$, if model errors and observations errors are unbiased and are not correlated.

3. Ensemble Kalman filter

The analysis scheme in the EnKF was originally based on Eq. (10). If one takes an ensemble of model states such that the error covariance of the forecasted ensemble mean coincides with the ensemble covariance and one performs an analysis on each of the ensemble members, then the *error covariance* of the analyzed *ensemble mean* is given by Eq. (10) as shown in Evensen (1994b). However, the *ensemble covariance* is reduced too much, unless the measurements are treated as random variables. The reason is that in the expression for the analyzed ensemble covariance there will be no analog to the term $\mathbf{K}\overline{(\mathbf{d} - \mathbf{d}^t)(\mathbf{d} - \mathbf{d}^t)^{\mathrm{T}}}\mathbf{K}^{\mathrm{T}} = \mathbf{KWK}^{\mathrm{T}}$ of Eq. (10), and spurious correlations arise because all ensemble members are updated with the the same measurements. The covariance of the analyzed ensemble is then

$$\tilde{\mathbf{P}}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f(\mathbf{I} - \mathbf{KH})^{\mathrm{T}}. \tag{11}$$

This expression contains one factor, $(\mathbf{I} - \mathbf{KH})$, too many. The effect of this term can be illustrated using a simple scalar case with $\mathbf{P}^f = 1$ and $\mathbf{W} = 1$. The analysis variance should then become $\mathbf{P}^a = 0.5$, while Eq. (11) would give $\tilde{\mathbf{P}}^a = 0.25$.

The original analysis scheme was based on the definitions of $\mathbf{P}^f$ and $\mathbf{P}^a$ as given by Eqs. (1) and (2). We will now give a new derivation of the analysis scheme where the ensemble covariance is used as defined by Eqs. (3) and (4). This is convenient since in practical implementations one is doing exactly this, and it will also lead to a more consistent formulation of the EnKF.

The difference between the original EnKF and the modified version presented here is that the observations are now treated as random variables by generating an ensemble of observations from a distribution with mean

equal to the first-guess observation and covariance equal to $\mathbf{W}$. Thus, we define the new observations

$$\mathbf{d}_j = \mathbf{d} + \boldsymbol{\epsilon}_j, \tag{12}$$

where j counts from 1 to N, the number of model state ensemble members.

The modified analysis step of the EnKF now consists of the following updates performed on each of the model state ensemble members:

$$\boldsymbol{\psi}_j^a = \boldsymbol{\psi}_j^f + \mathbf{K}_e(\mathbf{d}_j - \mathbf{H}\boldsymbol{\psi}_j^f). \tag{13}$$

The gain matrix $\mathbf{K}_e$ is similar to the Kalman gain matrix used in the standard Kalman filter (6) and is defined as

$$\mathbf{K}_e = \mathbf{P}_e^f\mathbf{H}^{\mathrm{T}}(\mathbf{HP}_e^f\mathbf{H}^{\mathrm{T}} + \mathbf{W})^{-1}. \tag{14}$$

Note that Eq. (13) implies that

$$\overline{\boldsymbol{\psi}^a} = \overline{\boldsymbol{\psi}^f} + \mathbf{K}_e(\overline{\mathbf{d}} - \mathbf{H}\overline{\boldsymbol{\psi}^f}). \tag{15}$$

Thus, the relation between the analyzed and forecasted ensemble mean is identical to the relation between the analyzed and forecasted states in the standard Kalman filter in Eq. (5), apart from the use of $\mathbf{P}_e$ instead of $\mathbf{P}$.

Moreover, the covariance of the analyzed ensemble is reduced in the same way as in the standard Kalman filter as given by Eq. (10),

$$\mathbf{P}_e^a = \overline{(\boldsymbol{\psi}^a - \overline{\boldsymbol{\psi}^a})(\boldsymbol{\psi}^a - \overline{\boldsymbol{\psi}^a})^{\mathrm{T}}}$$

$$= (\mathbf{I} - \mathbf{K}_e\mathbf{H})\mathbf{P}_e^f + O(N^{-1/2}), \tag{16}$$

where Eqs. (13) and (15) are used to get

$$\boldsymbol{\psi}^a - \overline{\boldsymbol{\psi}^a} = (\mathbf{I} - \mathbf{K}_e\mathbf{H})(\boldsymbol{\psi}^f - \overline{\boldsymbol{\psi}^f}) + \mathbf{K}_e(\mathbf{d} - \overline{\mathbf{d}}), \tag{17}$$

otherwise the derivation is as for Eq. (10). The finite ensemble size fluctuations, which have on the average zero mean and $O(N^{-1/2})$ rms magnitude, are proportional to $\mathbf{W} - (\mathbf{d} - \overline{\mathbf{d}})(\mathbf{d} - \overline{\mathbf{d}})^{\mathrm{T}}$ and $(\boldsymbol{\psi} - \overline{\boldsymbol{\psi}})(\mathbf{d} - \overline{\mathbf{d}})^{\mathrm{T}}$.

Note that the introduction of an ensemble of observations does not make any difference for the update of the ensemble mean since this does not affect Eq. (15).

Also in the forecast step the correspondence between the standard Kalman filter and the EnKF is maintained. Each ensemble member evolves in time according to a model

$$\boldsymbol{\psi}_j^{k+1} = \mathcal{M}(\boldsymbol{\psi}_j^k) + d\mathbf{q}_j^k, \tag{18}$$

where k denotes the time step, $\mathcal{M}$ is a model operator,

and $d\mathbf{q}$ is the stochastic forcing representing model errors from a distribution with zero mean and covariance $\mathbf{Q}$. The *ensemble covariance* matrix of the errors in the model equations, given by

$$\mathbf{Q}_e = \overline{(d\mathbf{q}^k - \overline{d\mathbf{q}^k})(d\mathbf{q}^k - \overline{d\mathbf{q}^k})^{\mathrm{T}}} = \overline{d\mathbf{q}^k d\mathbf{q}^{k\mathrm{T}}}, \quad (19)$$

converges to $\mathbf{Q}$ in the limit of infinite ensemble size.

The ensemble mean then evolves according to the equation

$$\begin{aligned}
\overline{\psi^{k+1}} &= \overline{\mathcal{M}(\psi^k)} \\
&= \mathcal{M}(\overline{\psi^k}) + \text{n.l.},
\end{aligned} \quad (20)$$

where n.l. represents the terms that may arise if $\mathcal{M}$ is nonlinear. These terms are not present in the traditional Kalman filter. The leading nonlinear term is proportional to the covariance and to the Hessian of $\mathcal{M}$, as shown by Cohn (1993).

The covariance of the ensemble evolves according to

$$\mathbf{P}_e^{k+1} = \mathbf{M}\mathbf{P}_e^k\mathbf{M}^{\mathrm{T}} + \mathbf{Q}_e + \text{n.l.}, \quad (21)$$

where $\mathbf{M}$ is the tangent linear model evaluated at the current time step. This is again an equation of the same form as is used in the standard Kalman filter, except for the extra terms n.l. that may appear if $\mathcal{M}$ is nonlinear. Implicitly, the EnKF retains these terms also for the error covariance evolution.

Thus if the ensemble mean is used as the best estimate, with the ensemble covariance $\mathbf{P}_e^{f,a}$ interpreted as the error covariance $\mathbf{P}^{f,a}$, and by defining the model error covariance $\mathbf{Q}_e = \mathbf{Q}$, the EnKF and the standard Kalman filter become identical. This discussion shows that there is a unique correspondence between the EnKF and the standard Kalman filter (for linear dynamics), and that one can interpret the ensemble covariances as error covariances while the ensemble mean is used as the best-guess trajectory.

For nonlinear dynamics the so-called extended Kalman filter may be used and is given by the evolution Eqs. (20) and (21) with the n.l. terms neglected. This makes the extended Kalman filter unstable is some situations (Evensen 1992), while the EnKF is stable. In addition, there is no need in the EnKF for a tangent linear operator or its adjoint, and this makes the EnKF very easy to implement for practical applications.

An inherent assumption in all Kalman filters is that the errors in the analysis step are Gaussian to a good approximation. After the last data assimilation step, one may continue the model integrations beyond the time that this assumption is valid. The ensemble mean is not the maximum-likelihood estimate, but an estimate of the state that minimizes the rms forecast error. For example, the ensemble mean of a weather forecast will approach climatology for long lead times, which is the "best guess" in the rms sense, although the climatological mean state is a highly unlikely one (Epstein 1969; Leith 1974; Cohn 1993).

The ensemble size should be large enough to prop-

agate the information contained in the observations to the model variables. Going to smaller ensembles, the analysis error becomes larger. Too small ensembles can give very poor approximations to the infinite ensemble case. In those situations, it can be better to go back to optimal interpolation. In the formulation of the EnKF presented here there is a second effect. The finite size flucuations in Eq. (16) tend to make the covariance of the ensemble smaller for smaller ensemble sizes instead of larger. Thus, for too small ensembles, the ensemble covariance underestimates the error covariance substantially. This effect can be monitored by comparing the actual forecasted model data differences to those expected on the basis of the forecasted ensemble covariance. Of course, also a wrong specification of $\mathbf{Q}$ or $\mathbf{W}$ or a systematic error in the model will lead to a difference between the ensemble covariance and the error covariance.

4. An example

An example is now presented that illustrates the analysis step in the original and modified schemes. Further, as a validation of the derivation performed in the previous section the results are also compared with the standard Kalman filter analysis.

For the experiment a one-dimensional periodic domain in x, with $x \in [0, 50]$, is used. We assume a characteristic length scale for the function $\psi(x)$ as $\ell = 5$. The interval is discretized into 1008 grid points, which means there are a total of about 50 grid points for each characteristic length.

Using the methodology outlined in the appendix of Evensen (1994b), we can draw smooth pseudorandom functions from a distribution with zero mean, unit variance, and a specified covariance given by

$$\mathbf{P}(x_1 - x_2) = \exp\left[-\frac{(x_1 - x_2)^2}{\ell^2}\right]. \quad (22)$$

This distribution will be called $\Phi(\psi)$ where the functions ψ have been discretized on the numerical grid.

A smooth function representing the true state ψ^t is picked from the distribution Φ, and this ensures that the true state has the correct characteristic length scale ℓ. Then a first-guess solution ψ^f is generated by adding another function drawn from the same distribution to ψ^t; that is, we have assumed that the first guess has an error variance equal to one and covariance functions as specified by Eq. (22).

An ensemble representing the error variance equal to one is now generated by adding functions drawn from Φ to the first guess. Here 1000 members were used in the ensemble. Thus we now have a first-guess estimate of the true state with the error covariance represented by the ensemble.

Since we will compare the results with the standard Kalman filter analysis, we also construct the error co-
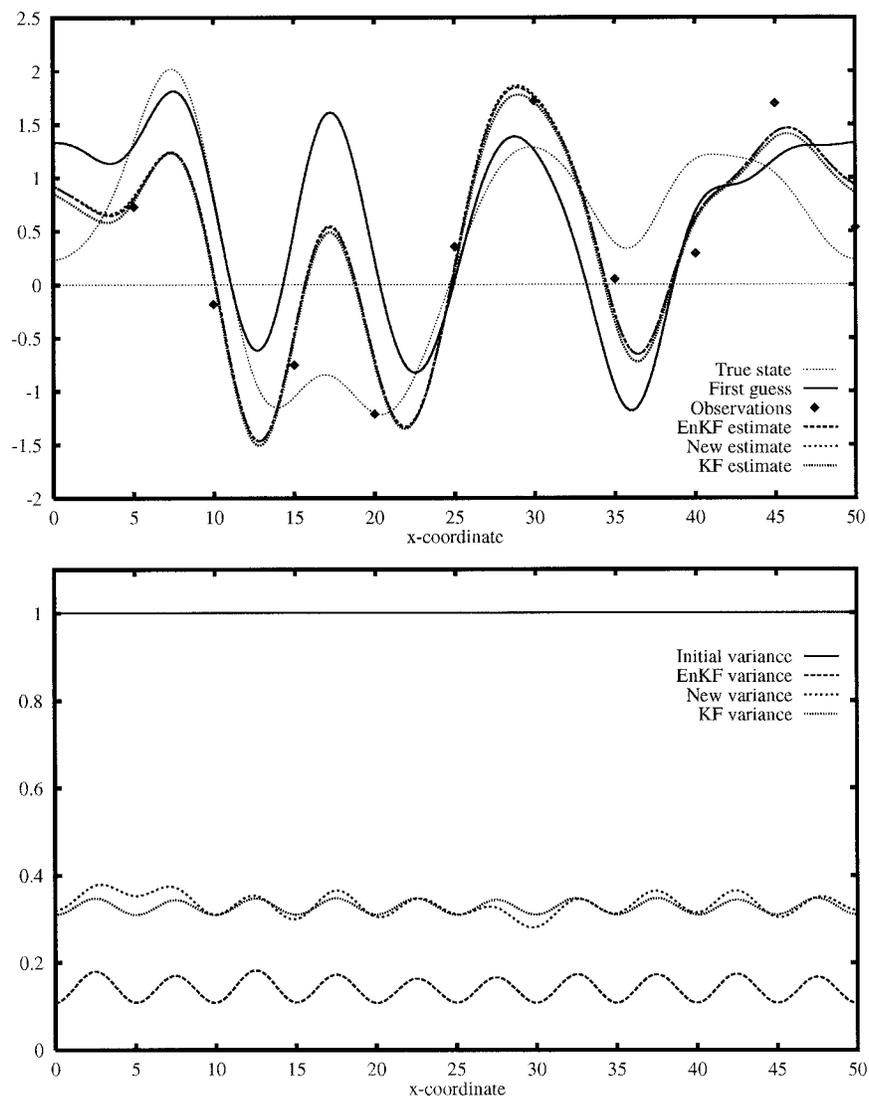
FIG. 1. (Top) The true reference state, the first guess, and the estimates calculated from the different analysis schemes are given. (Bottom) The corresponding error variance estimates.

variance matrix for the first guess by discretizing the covariance function (22) on the numerical grid to form $\mathbf{P}^f$.

There are 10 measurements distributed at regular intervals in x. Each measurement is generated by measuring the true state ψ^t and then adding Gaussian distributed noise with mean zero and variance 0.5. This should give a posterior error variance at measurement locations of about 1/3 for the standard Kalman filter and the modified EnKF, while the original version of the EnKF should give a posterior error variance equal to about 1/9.

The parameters used have been chosen to give illustrative plots. The results from this example are given in Fig. 1. The upper plot shows the true state ψ^t, the first guess ψ^f, and the observations plotted as diamonds. The three curves that almost coincide are the

estimates from the original and the new modified EnKF, and the standard Kalman filter analysis. The ensemble estimates are of course the means of the analyzed ensembles. These three curves clearly show that the EnKF gives a consistent analysis for the estimate ψ^a.

The lower plot shows the corresponding error variances from the three cases. The upper line is the initial error variance for the first guess equal to one. Then there are three error variance estimates corresponding to the original version of the EnKF (lower curve), the new modified EnKF nonsymmetric middle curve, and the standard Kalman filter symmetric middle curve. Clearly, by adding perturbations to the observations, the new analysis scheme provides an error variance estimate that is very close to the one that follows from the standard Kalman filter.

Finally, note also that the posterior variances at the

measurement locations are consistent with what we would expect from a scalar case.

5. Discussion and conclusions

The formulation of the ensemble Kalman filter (EnKF) proposed by Evensen (1994b) has been reexamined with the focus on the analysis scheme. It has been shown that in the original formulation of the EnKF by Evensen (1994b), the derivation of the method was correct but it was not realized that one needs to add random perturbations to the measurements for the assumption of measurements being random variables to be valid. This is essential in the calculation of the analyzed ensemble, which will have a too low variance unless random perturbations are added to the observations.

The use of an ensemble of observations also allows for an alternative interpretation of the EnKF where the ensemble covariance is associated with the error covariance of the ensemble mean. The EnKF then gives the correct evolution of the ensemble mean and the ensemble covariance, provided the ensemble size is large enough, as discussed at the end of section 3.

Note that the only modification needed in existing EnKF applications is that random noise with prescribed statistics must be added to the observations at analysis steps. This can be done very easily by adding a couple of lines in the code, that is, one function call to generate the perturbations with the correct statistics and a line to add the perturbations to the measurements.

There are a couple of reason why the problem with the original analysis scheme has not been discovered earlier. For example, in Evensen (1994b), observations with rather low variance equal to 0.02 were used in the verification example. With prior variance equal to 1 at the measurement locations the theoretical value of the posterior variance is equal to 0.0196, while the original analysis scheme in the EnKF should give 0.00038. Thus the relative difference between them is rather small compared to the prior variance, actually less than 2%, which can be explained by statistical noise caused by using a limited ensemble size.

It should be noted that the results presented here apply equally well to the recently proposed ensemble smoother (van Leeuwen and Evensen 1996). However, for the smoother only the posterior error covariance estimates are affected since a single analysis is calculated only once and simultaneously in space and time.

REFERENCES

Bouttier, F., 1994: A dynamical estimation of forecast error covariances in an assimilation system. *Mon. Wea. Rev.,* **122,** 2376–2390.

Cohn, S. E., 1993: Dynamics of short term univariate forecast error covariances. *Mon. Wea. Rev.,* **121,** 3123–3149.

Daley, R., and T. Mayer, 1986: Estimates of global analysis error from the global weather experiment observational network. *Mon. Wea. Rev.,* **114,** 1642–1653.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus,* **21A,** 739–759.

Evensen, G., 1992: Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.,* **97** (C11), 17 905–17 924.

——, 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 quasi-geostrophic model. *Mon. Wea. Rev.,* **124,** 85–96.

Gauthier, P., P. Courtier, and P. Moll, 1993: Assimilation of simulated wind lidar data with a Kalman filter. *Mon. Wea. Rev.,* **121,** 1803–1820.

Houtekamer, P. L., and J. Derome, 1995: The RPN Ensemble Prediction System. *Seminar Proc. on Predictability,* Vol. II, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 121–146.

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

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.,* **102,** 409–418.

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

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.