6 Sequential Data Assimilation for Nonlinear Dynamics: The Ensemble Kalman Filter

GEIR EVENSEN
Nansen Environmental and Remote Sensing Center, Bergen, Norway

6.1 Introduction

Sequential data assimilation methods have proven useful for many applications in meteorology and oceanography. For example are most operational weather prediction systems applying a sequential data assimilation technique where observations are "assimilated" into the model whenever they are available.

For linear dynamics the optimal sequential technique is the Kalman filter. In the Kalman filter an additional equation for the second-order statistical moment is integrated forward in time to predict error statistics for the model forecast. The error statistics are then used to calculate a variance-minimizing estimate whenever measurements are available.

For nonlinear dynamics the extended Kalman filter may be applied, in which an approximate linearized equation is used for the prediction of error statistics. The implementation of the extended Kalman filter for data assimilation in a multilayer quasi-geostrophic (QG) model has previously been discussed by Evensen (1992). The main result from this work is the finding of an apparent closure problem in the error covariance evolution equation. The extended Kalman filter applies a closure scheme where third- and higher-order moments in the error covariance evolution equation are discarded. This simple closure technique results in a unbounded error variance growth caused by the linearization performed when higher-order moments are neglected. Thus, it has been shown that the error covariance equation is based on a too simplified closure approximation and this may lead to a nonphysical error variance evolution (see e.g., Evensen (1992), Miller et al. (1994), Gauthier et al. (1993), and Bouttier (1994)).

The Ensemble Kalman Filter (EnKF) was introduced by Evensen (1994b) as an alternative to the traditional extended Kalman filter (EKF). It was shown that if the dynamical model is written as a stochastic differential equation, one can derive the Fokker–Planck 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 where the error statistics are predicted by solving the Fokker–Planck equation using Monte–Carlo or ensemble integrations. By integrating an ensemble of model states forward in time it is possible to calculate statistical moments like mean and error covariances whenever such informa-

tion is required. Thus, all the statistical information about the predicted model state which are required at analysis times are contained in the ensemble.

In Evensen (1994b) an analysis scheme was proposed where the traditional update equation used in the Kalman Filter (KF) is applied, 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, e.g., 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 only a few hundred model states may be sufficient for reasonable statistical convergence. For practical ensemble sizes, say $O(100)$, the errors will be dominated by statistical noise, not by closure problems or unbounded error variance growth.

The EnKF has been further discussed and applied with success in a twin experiment in Evensen (1994a), in a realistic application for the Agulhas Current using Geosat altimeter data in Evensen and van Leeuwen (1996) and with the strongly nonlinear Lorenz equations in Evensen (1997).

This paper will briefly outline the ensemble Kalman filter (EnKF) and illustrate its properties with a few simple examples. Further, some preliminary results from an implementation with an OGCM will be presented.

6.2 Extended Kalman filter

It is instructive first to give a brief review of the extended Kalman filter algorithm. The derivation of the extended Kalman filter on matrix form can be found in a number of books on control theory (e.g., Gelb (1974), Jazwinski (1970)).

The evolution of the true state vector is described by a dynamical model,

$$\psi^t{}_k = \boldsymbol{f}(\psi^t{}_{k-1}) + \boldsymbol{q}_{k-1} \tag{1}$$

where $\boldsymbol{f}$ is a nonlinear model operator and $\boldsymbol{q}$ is a stochastic term representing model errors. A forecast is calculated from

$$\psi^f{}_k = \boldsymbol{f}(\psi^a{}_{k-1})\,. \tag{2}$$

The error statistics is described by the error covariance matrix $\boldsymbol{P}_k = \overline{(\psi - \psi^y)(\psi - \psi^y)^T}$, which evolves according to the equation

$$\boldsymbol{P}_k{}^f = \boldsymbol{F}_{k-1}\boldsymbol{P}^a{}_{k-1}\boldsymbol{F}^T{}_{k-1} + \boldsymbol{Q}_{k-1}\,, \tag{3}$$

where $\boldsymbol{Q}_k = \overline{\boldsymbol{q}_k \boldsymbol{q}_k^T}$ is the model error covariance matrix, and $\boldsymbol{F}_k$ is the Jacobi matrix or tangent linear operator

$$\boldsymbol{F}_k = \left. \frac{\partial \boldsymbol{f}(\psi)}{\partial \psi} \right|_{\psi = \psi_k} . \tag{4}$$

The equations (1) and (2) are used to generate a prediction or model forecast and a prediction of the corresponding error covariance estimate.

Whenever observations are available the analyzed estimate ψ^a is calculated as a linear combination of the vector of measurements $\boldsymbol{d}$ and the predicted model state vector ψ^f. The linear combination is chosen to minimize the variance in the analyzed estimate ψ^a and is given by the equation

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

where the Kalman gain matrix $\boldsymbol{K}$ is defined as

$$\boldsymbol{K} = \boldsymbol{P}^f \boldsymbol{H}^T (\boldsymbol{H}\boldsymbol{P}^f \boldsymbol{H}^T + \boldsymbol{R})^{-1} . \tag{6}$$

It is a function of the model state forecast error covariance matrix $\boldsymbol{P}^f$, the data error covariance matrix $\boldsymbol{R}$ and the measurement matrix $\boldsymbol{H}$ that relates the model state to the data.

The measurements are related to the true state by

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

with $\mathbf{e}$ the measurement errors. In particular, the true model state is related to the true observations as

$$\boldsymbol{d}^t = \boldsymbol{H}\psi^t . \tag{8}$$

The measurement error covariance matrix is defined as

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

Note that

$$\psi^a - \psi^t = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})(\psi^f - \psi^t) + \boldsymbol{K}(\boldsymbol{d} - \boldsymbol{d}^t) , \tag{10}$$

which is obtained using (5) for replacing ψ^a and then adding $\boldsymbol{K}(\boldsymbol{d}^t - \boldsymbol{H}\psi^t) = 0$ from (8). The error covariance of the analyzed model state vector then becomes

$$\begin{aligned} \boldsymbol{P}^a &= \overline{(\psi^a - \psi^t)(\psi^a - \psi^t)^T} \\ &= \boldsymbol{I} - \boldsymbol{K}\boldsymbol{H}\boldsymbol{P}^f\boldsymbol{I} - \boldsymbol{H}^T\boldsymbol{K}^T + \boldsymbol{K}\boldsymbol{R}\boldsymbol{K}^T \\ &= (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{P}^f \end{aligned} \quad . \tag{11}$$

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

The dynamical model (2) and the error covariance equation (3) together with equations for the analysis scheme, constitutes the so-called extended Kalman filter. Equations (2) and (3) are integrated forward in time and at analysis times the observations are used to update the model solution and its error covariance estimate.

It should be noted that the error covariance equation (3) is only approximate. It results from a linearization of an equation which references infinitely many higher order statistical moments. This statistical linearization lead to serious problems for strongly nonlinear dynamics. In Miller et al. (1994) it was shown that with the Lorenz equations, the error covariance equations under-estimated the predicted error covariance and this resulted in too low gain to keep the model close to the observed state. In Evensen (1992) the extended Kalman filter was used with a nonlinear QG model. Here the problem was that the error covariance equation provided unbounded error variance growth due to the lack of error variance saturation on a climatological level, an effect which requires a higher order closure scheme.

Another major problem of the Kalman filter is related to the storage and computation of the error covariance matrix. If the size of the state vector is n, the size of the error covariance matrix is n^2 and $2n$ model integrations are required to step it forward in time.

These problems with nonlinearities and computational load have lead to the search for alternative methodologies for predicting the error statistics. Currently, there are several approaches where one attempts to evolve the error covariance equation in a reduced state space to save computer time. However, this introduces additional approximations to an already approximate error covariance equation. In the next section the ensemble Kalman filter is introduced as an alternative to the traditional extended Kalman filter.

6.3 Ensemble Kalman Filter

The ensemble Kalman filter as proposed by Evensen (1994b) is now introduced. We will adapt a three stage presentation starting with the representation of error statistics using an ensemble of model states, then an alternative to the traditional error covariance equation is proposed for the prediction of error statistics and finally a consistent analysis scheme is presented.

6.3.1 Representation of error statistics

The error covariance matrices for the predicted and the analyzed estimate, $\boldsymbol{P}^f$ and , $\boldsymbol{P}^a$ are in the Kalman filter defined in terms of the true state as

$$\boldsymbol{P}^f = \overline{(\psi^f - \psi^t)(\psi^f - \psi^t)^T}, \tag{12}$$

$$\boldsymbol{P}^a = \overline{(\psi^a - \psi^t)(\psi^a - \psi^t)^T}, \tag{13}$$

where the overline denotes an expectation value, ψ 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, $\overline{\psi}$,

$$\boldsymbol{P}^f \cong \boldsymbol{P}^f_e = \overline{(\psi^f - \overline{\psi^f})(\psi^f - \overline{\psi^f})^T}, \tag{14}$$

$$\boldsymbol{P}^a \cong \boldsymbol{P}^a_e = \overline{(\psi^a - \overline{\psi^a})(\psi^a - \overline{\psi^a})^T}, \tag{15}$$

where now the overline denote an average over the ensemble. Thus, we can use an interpretation where the ensemble mean is the best estimate and the spreading of the ensemble around the mean is a natural definition of the error in the ensemble mean.

Now, since the error covariances as defined in (14) and (15) are defined as ensemble averages, there will clearly exist infinitively many ensembles with an error covariance equal to $\boldsymbol{P}^f$ and $\boldsymbol{P}^a$. Thus, instead of storing a full covariance matrix, we can represent the same error statistics using an appropriate ensemble of model states. Given an error covariance matrix, an ensemble of limited size will always provide an approximation to the error covariance matrix. However, when the size of the ensemble, *N*, increases the errors in the representation will decrease proportional to $1/\sqrt{N}$. Experience shows that we can represent an error covariance matrix with reasonable accuracy using only about 100–500 members in the ensemble.

Suppose now that we have a number N model states in the ensemble, each of dimension n. Each of these model states can be represented as a single point in an n-dimensional state space. All the ensemble members together will constitute a cloud of points in the state space. Such a cloud of points in the state space can be approximately described using a probability density function

$$\phi(\psi) = \frac{dN}{N}, \tag{16}$$

where dN is the number of points in a small unit volume and N is the total number of points. With knowledge about either ϕ or the ensemble representing ϕ we can calculate the statistical moments (mean, covariances etc.) we need whenever they are needed.

The conclusion so far is that the information contained by a full probability density function can equally well be represented by an ensemble of model states.

6.3.2 Prediction of error statistics

We start by writing the model dynamics as a stochastic differential equation

$$d\psi_k = \boldsymbol{f}(\psi_{k-1})dt + d\boldsymbol{q}_{k-1}, \tag{17}$$

where $d\boldsymbol{q} \in \Re^n$ is a vector of random white noise with mean zero. This equation is an Itô stochastic differential equation describing a Markov process. The evolution of the probability density for this equation is given by the Fokker–Planck equation

$$\frac{\partial \phi}{\partial t} + \sum_{i=1}^{n} \frac{\partial f_i \phi}{\partial \psi_i} = \sum_{i,j=1}^{n} \frac{Q_{ij}}{2} = \frac{\partial^2 \phi}{\partial \psi_i \partial \psi_j}, \tag{18}$$

where $\boldsymbol{Q} = \overline{\boldsymbol{q}\boldsymbol{q}^T}$ is the covariance matrix for the model errors. A derivation of this equation is given by Jazwinski (1970, p. 129).

The stochastic forcing, $dq_{k\text{-}1}$, introduces a diffusion term that tends to flatten the probability density function (spreading the ensemble) during the integration; that is, the probability decreases and the errors increase.

If this equation could be solved for the probability density function, it would be possible to calculate statistical moments of ϕ like the mean state and the error covariances at different time levels. However, a direct numerical integration of this equation becomes impossible for ocean circulation models.

By taking moments of the Fokker–Planck equation it is however possible to derive equations for the evolution of statistical moments like the mean and the error covariances. This is exactly the procedure used in the Kalman filter. Note also that for linear dynamics and with a Gaussian initial probability density, the probability density will be completely characterized by its mean and covariance matrix for all times. Thus, one can then use exact equations for the evolution of the mean

and the covariance matrix as a simpler alternative than solving the Fokker–Planck equation.

Such moments of the Fokker–Planck equation, including the error covariance equation (3), are easy to derive, and several methods are illustrated by Jazwinski (1970, examples 4.19–4.21).

For a nonlinear model, the mean and covariance matrix will not in general characterize $\phi(\psi,t)$. They do, however, determine the mean path and the dispersion about that path, and it is possible to solve approximate equations for the moments, which is the procedure characterizing the extended Kalman filter.

Another alternative approach for solving the Fokker–Planck equation and predicting the error statistics is to use Monte–Carlo methods. If a probability density function is represented by a large ensemble of model states, it is possible to integrate each member forward in time using the stochastic model (17). Thus, integrating an ensemble of model states becomes equivalent to solving the Fokker–Planck equation using a Monte Carlo method.

The standard approach is to first calculate a best guess initial condition based on information available from data and statistics. An ensemble of initial states is then generated in which the mean equals the best guess initial condition and the variance is specified on the basis of knowledge of the uncertainty in the first-guess initial state. The covariance or smoothness of the ensemble should reflect the true scales of the system.

The effect of external error growth must be included to give reliable estimates for the evolution of errors. In the Kalman filter this can be done rather simply by adding the system error covariance matrix every time step. However, in the Monte–Carlo method, each ensemble member is integrated as a stochastic differential equation and is forced by smooth pseudo-random fields with a specified variance and covariance to simulate the model errors. This will provide a realistic increase in the ensemble variance provided that the estimate of the model error variance is reasonably good.

6.3.3 An analysis scheme

The KF analysis scheme was based on the definitions of $\boldsymbol{P}^f$ and $\boldsymbol{P}^a$ as given by equations (12) and (13). We will now give a new derivation of the analysis scheme where the ensemble covariances are used as defined by (14) and (15). This is convenient since in practical implementations one is doing exactly this, and it will also lead to a consistent formulation of the EnKF.

As will be shown later it is essential that the observations are treated as random variables having a distribution with mean equal to the first-guess observations and covariance equal to $\boldsymbol{R}$. Thus, we start by defining an ensemble of observations

$$\boldsymbol{d}_j = \boldsymbol{d} + \mathbf{e}_j, \tag{19}$$

where j counts from 1 to the number of model state ensemble members. Next we define the ensemble covariance matrix of the measurements as

$$\boldsymbol{R}_e = \overline{\mathbf{e}\mathbf{e}^T}, \tag{20}$$

and of course in the limit of an infinite ensemble, this matrix will converge towards the prescribed error covariance matrix $\boldsymbol{R}$ used in the standard Kalman filter.

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

$$\psi^n{}_j = \psi^f{}_j + \boldsymbol{K}_e(\boldsymbol{d}_j - \boldsymbol{H}\psi^f{}_j) \ . \tag{21}$$

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

$$\boldsymbol{K}_e = \boldsymbol{P}^f{}_e\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}^f{}_e\boldsymbol{H}^T + \boldsymbol{R}_e)^{-1} . \tag{22}$$

Note that equation (21) implies that

$$\overline{\psi^a} = \overline{\psi^f} + \boldsymbol{K}_e(\bar{\boldsymbol{d}} - \boldsymbol{H}\overline{\psi^f}) \ . \tag{23}$$

Thus, the relation between the analyzed and predicted ensemble mean is identical to the relation between the analyzed and predicted state in the standard Kalman filter in equation (5), apart from the use of $\boldsymbol{P}_e$ and $\boldsymbol{R}_e$ instead of $\boldsymbol{P}$ and $\boldsymbol{R}$. 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 equation (23).

In Evensen and van Leeuwen (1996) a more detailed discussion was given on the practical implementation of the analysis scheme. It is possible to avoid the computation and storage of the full error covariance matrix $\boldsymbol{P}_e$, by operating directly on the ensemble members or alternatively the influence functions defined as $\boldsymbol{H}\,\boldsymbol{P}_e$. In addition, for realistic systems with a large number of observations, the inversion in (22) becomes too expensive and poorly conditioned. One can then resort to an approximate algorithm where the analysis is computed grid point by grid point, using only observations within a specified radius from the grid point. This reduces the size of the matrix inversion in addition to allowing for an algorithm where the major storage requirement reduces to keeping the ensemble in memory.

If the mean is considered to be the best estimate, then the linearity of the analysis scheme makes it an arbitrary choice whether one update the mean using the first-guess observations, or if one update each of the ensemble members using the perturbed observations. However, it will now be shown that by updating each of the ensemble members using the perturbed observations one also creates a new ensemble having the correct error statistics for the analysis. The updated ensemble can then be integrated forward in time till the next observation time.

Moreover, the covariance of the analyzed ensemble is reduced in the same way as in the standard Kalman Filter. First, note that equations (21) and (23) are used to get

$$\psi^a{}_j - \overline{\psi^a} = (\boldsymbol{I} - \boldsymbol{K}_e\boldsymbol{H})(\psi^f{}_j - \overline{\psi^f}) + \boldsymbol{K}_e(\boldsymbol{d} - \bar{\boldsymbol{d}}) . \tag{24}$$

We then get

$$\begin{aligned} \boldsymbol{P} &= \overline{(\psi^a - \overline{\psi^a})(\psi^a - \overline{\psi^a})^T} \\ &= (\boldsymbol{I} - \boldsymbol{K}_e\boldsymbol{H})\boldsymbol{P}^f{}_e(\boldsymbol{I} - \boldsymbol{H}^T\boldsymbol{K}^T{}_e) + \boldsymbol{K}_e\boldsymbol{R}_e\boldsymbol{K}^T{}_e \cdot \\ &= (\boldsymbol{I} - \boldsymbol{K}_e\boldsymbol{H})\boldsymbol{P}^f{}_e \end{aligned} \tag{25}$$

Clearly the observations **d** must be treated as random variables to get the measurement error covariance matrix into the expression.

6.3.4 Summary

We now have a complete system of equations which constitutes the ensemble Kalman filter (EnKF), and the resemblance with the standard Kalman filter is maintained. This is also true for the forecast step. Each ensemble member evolves in time according to the model dynamics and the ensemble covariance matrix of the errors in the model equations, $\boldsymbol{Q}_e$, converges to $\boldsymbol{Q}$ in the limit of an infinite ensemble size. The ensemble mean then evolves according to an equation of the form

$$\begin{aligned} \overline{\psi_{k+1}} &= \overline{\boldsymbol{f}(\psi_k)} \\ &= \boldsymbol{f}(\overline{\psi_k}) + \text{n.1.} \end{aligned} , \tag{26}$$

where n.l. represents the terms which may arise if $\boldsymbol{f}$ is non-linear. One of the advantages of the EnKF is that the effect of these terms is retained since each ensemble member is integrated independently by the model.

The covariance of the ensemble evolves according to an equation

$$\boldsymbol{P}_e^{k+1} = \boldsymbol{F}\boldsymbol{P}_e^k\,\boldsymbol{F}^T + \boldsymbol{Q}_e + \text{n.1.}, \tag{27}$$

where $\boldsymbol{F}$ 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 of the extra terms n.l. that may appear if $\boldsymbol{f}$ is non-linear. 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 $\boldsymbol{P}_e^{f,a}$ interpreted as the error covariance $\boldsymbol{P}^{f,a}$, and by defining the observation error covariance matrix $\boldsymbol{R}_e = \boldsymbol{R}$ and the model error covariance $\boldsymbol{Q}_e = \boldsymbol{Q}$, 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 certainly interpret the ensemble covariances as error covariances while the ensemble mean is used as the best guess trajectory.

The extended Kalman filter applies the evolution equations (26) and (27) with the n.l. terms neglected. However, the ensemble Kalman filter includes the full effect of these terms and there are no linearizations or closure assumptions applied. In addition, there is no need for a tangent linear operator or its adjoint, and this makes the EnKF very easy to implement for practical applications.

6.4 An example of the analysis scheme

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

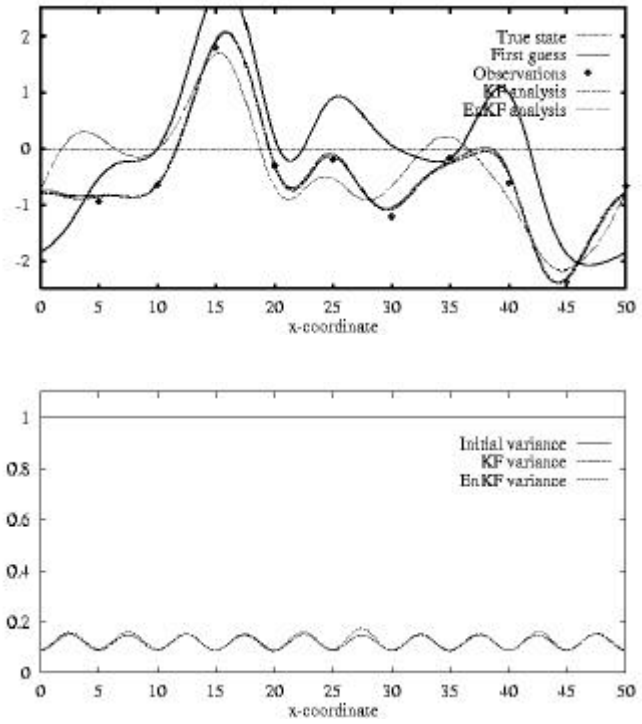

Fig. 6.1. Comparing results from the KF and the EnKF analysis schemes. On top the true reference state, the first-guess, and the analyzed estimate. The lower plot shows the corresponding error variance estimates.

For the experiment a 1–dimensional periodic domain in x, with $x \in [0,50]$, is used. We assume a characteristic length scale for the function $\psi(x)$ as $l = 5$. 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 pseudo random functions from a distribution with zero mean, unit variance and a specified covariance given by

$$P(x_1 - x_2) = \exp\left[-\frac{(x_1 - x_2)^2}{l^2}\right]. \tag{28}$$

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

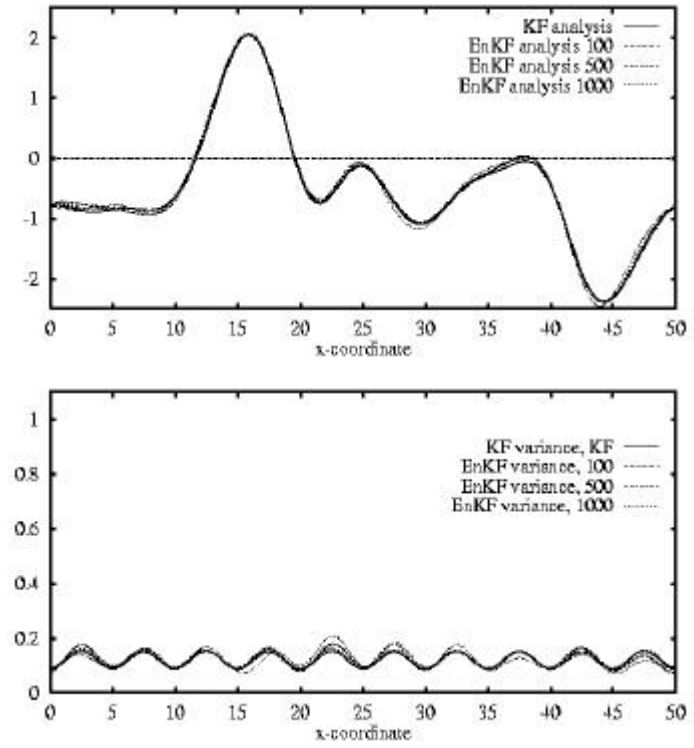

Fig. 6.2. Comparing results from the KF and the EnKF analysis schemes, using different ensemble sizes with 1000, 500, and 100 members. The upper plot shows the analyzed estimates. The lower plot shows the corresponding error variance estimates

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, i.e. we have assumed that the first-guess has an error variance equal to one and covariance functions as specified by (28).

The error covariance matrix for the first-guess is constructed by discretizing the covariance function (28) on the numerical grid to form $\boldsymbol{P}^f$.

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 0.2.

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.

The results from this example are given in Fig. 6.1 . The ensemble estimates are of course the means of the analyzed ensembles. By comparing the KF and the EnKF estimates it is clear that EnKF gives a consistent analysis ψ^a. The lower plot shows the corresponding error variances. The upper line is the initial error variance for the first-guess equal to one. Then there are two error variance estimates corresponding to the EnKF and the standard Kalman filter. Clearly, the EnKF analysis scheme provides an error variance estimate which is very close to the one which follows from the standard Kalman filter.

In Fig. 6.1 we examine the sensitivity of the analysis scheme with respect to the size of the ensemble. Clearly there is not a big difference in the estimates using 100 or 1000 ensemble members.

6.5 A highly nonlinear case: the Lorenz equations

An example is now given using the highly nonlinear and chaotic Lorenz equations. The celebrated Lorenz model has been subject of extensive studies motivated by its chaotic and strongly nonlinear nature. In the field of data assimilation, the model has served as a testbed for examining the properties of various data assimilation methods when used with strongly nonlinear dynamics. The results have been used to suggest properties and possibilities of the methods for applications with oceanic and atmospheric models which may also be strongly nonlinear and chaotic.

6.5.1 Model Equations

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

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x) + q^x, \\ \frac{dy}{dt} &= \rho x - y - xz + q^y, \\ \frac{dz}{dt} &= xy - \beta z + q^z. \end{aligned} \tag{29}$$

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 equation; σ=10, ρ=28 and β=8/3. The terms $q^x(t)$, $q^y(t)$ and $q^z(t)$ are assumed to represent the unknown model errors. Initial conditions for the model are given as

$$\begin{aligned} x(0) &= x_0 + a^x, \\ y(0) &= y_0 + a^y, \\ z(0) &= z_0 + a^z, \end{aligned} \tag{30}$$

where x_0, y_0, and z_0 are the first-guess values of the initial conditions and the terms a^x, a^y and a^z represent the errors in the first-guess initial conditions. If all the error terms were known or equal to zero, these equations would formulate a well-posed problem having a unique solution in a mathematical sense.

Now a set of measurements, $\boldsymbol{d} \in \Re^M$, of the true solution are assumed given and linearly related to the model variables by the measurement equation

$$\boldsymbol{d} = \boldsymbol{L}[x, y, z] + \mathbf{e}, \tag{31}$$

where $\boldsymbol{L} \in \Re^M$ is a linear measurement functional, $\mathbf{e} \in \Re^M$ is a vector of measurement errors and M is the number of measurements.

6.5.2 Discussion of cases

The initial condition for the reference case is given by

$$(x_0, y_0, z_0) = (1.508870, -1.531271, 25.46091)$$

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 initial conditions used are also assumed to have the same variance as the observations. These are the same values that were used in Miller et al. (1994a) and Evensen and Fario (1997).

The following examples are discussed.

Experiment A: In this first experiment the distance between the measurements is $\Delta t_{obs} = 0.25$, which is the same as was used in Miller et al. (1994a). Thus, it is possible to compare the results presented here with those presented in

Miller et al. (1994a) using the extended Kalman filter and a strong constraint variational method.

Experiment B: In order to examine the sensitivity with respect to measurement density, an additional experiment is now performed where the distance between the measurements is $\Delta t_{obs} = 0.5$.

The data assimilation estimate from the ensemble Kalman filter is given in Figs. 6.3 and 6.4 for the two cases. The ensemble size is 1000 members. The ensemble Kalman filter seems to do a reasonably good job in tracking the state transitions and also in reproducing the correct amplitudes for the peaks of the solution. There are a few locations where the filter estimate starts diverging from the reference solution, e.g. for $t = 26$ and $t = 34$. Note however that

the ensemble Kalman filter recovers quickly and begins tracing the reference solution again. The error estimate given in the lower plot of Fig. 6.3 shows strong error growth at these particular locations and thus indicates that the ensemble is passing through a region in the state space which is associated with strong instability.

The error estimates show the same behavior as was found by Miller et al. (1994a) 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.

Finally, it should be pointed out that in the ensemble filter a variance minimizing analysis is calculated at measurement times. 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.

6.6 An ensemble Kalman filter for an OGCM: Preliminary results

The ensemble Kalman filter has been implemented with the Miami Isopycnic Coordinate Ocean Model (MICOM) originally developed by R. Bleck at the University of Miami, [see e.g., Bleck and Smith (1990), and Bleck et al. (1992)]. The currently available data assimilation applications for Ocean General Circulation Models (OGCMs) have been based on rather simplistic assimilation schemes. None of these take proper error statistics into account and ad hoc approaches are used for the assimilation. Some examples are: Derber and Rosati (1989) who used

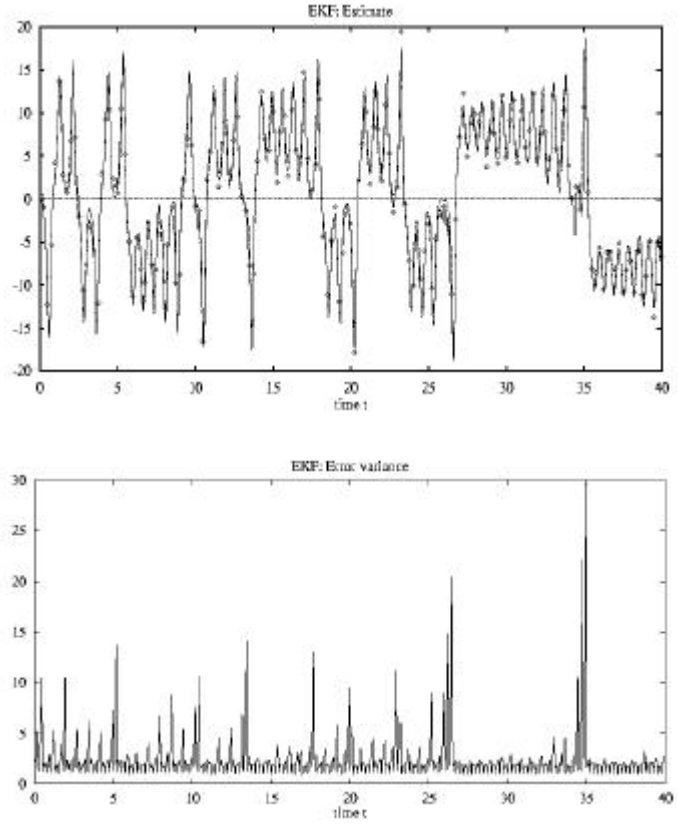

Fig. 6.3. Experiment A (Ensemble Kalman Filter): The upper plot shows the inverse estimate for x as the full line and the "true" reference state as the is the dash-dot line which is mostly hidden behind the estimate. The lower plot shows the corresponding error variance estimate.

an objective analysis technique to update the model temperature in a version of the Cox model (Rosati and Miyakoda, 1988). Mellor and Ezer, (1991) and Ezer and Mellor, (1994), used a univariate optimal interpolation algorithm with vertical projection of surface information in the model by Blumberg and Mellor (1987). Cooper and Haines (1996) used a vertical projection method based on water property conservation in the Cox model (Cox, 1987).

In the works by Mellor and Ezer, (1991), and Ezer and Mellor, (1994), sea surface height (SSH) observations were used to update the vertical density stratification to drive the geostrophic currents associated with gradients in the SSH. The

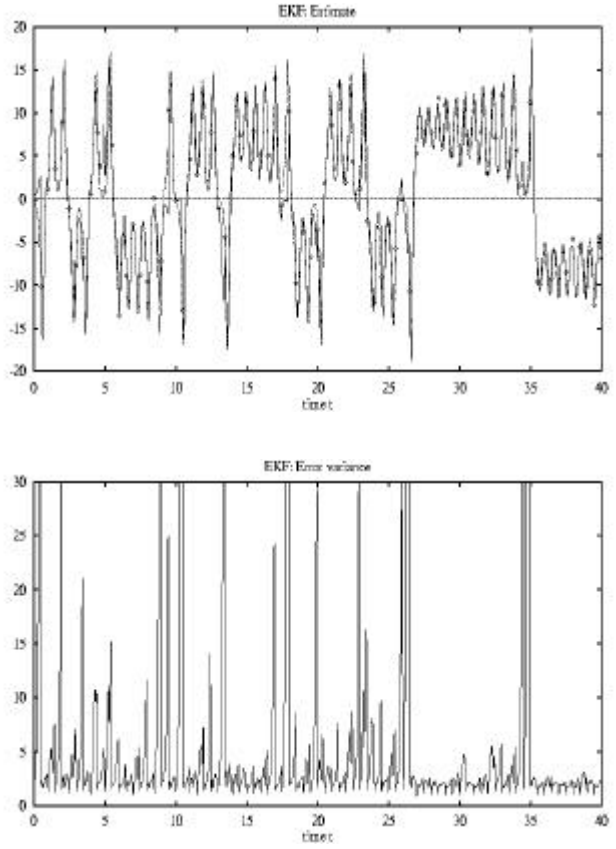

Fig. 6.4. Experiment B (Ensemble Kalman Filter): Same as Fig. 6.3 but for Experiment B.

density was implicitly updated by actually updating the vertical temperature profiles using estimated correlations between temperature and SSH. Clearly, this is an inconsistent approach, since the vertical density stratification depends on both the temperature and salinity. Updating only temperature will ultimately generate water masses with unrealistic T–S properties.

A method for overcoming this problem has recently been presented by Oschlies and Willebrand (1996), where both temperature and salinity were updated under a constraint of maintaining the T–S properties of the water masses. This approach gave satisfactory results and should now be extended to take proper error statistics into account in the analysis scheme.

In another recent work by Forbes and Brown (1996), a nudging method was used with the Miami Isopycnic Coordinate Ocean Model (MICOM) for assimilation of SSH observations from altimetry. They showed that the nudging approach was capable of pulling the model state towards the observations and that the model, over time, also propagated information from the surface into the lower layers.

Essential when developing an advanced data assimilation system with an OGCM is to estimate proper error statistics for the model prediction. That is, having a model forecast and an observation one needs to know the influence this particular

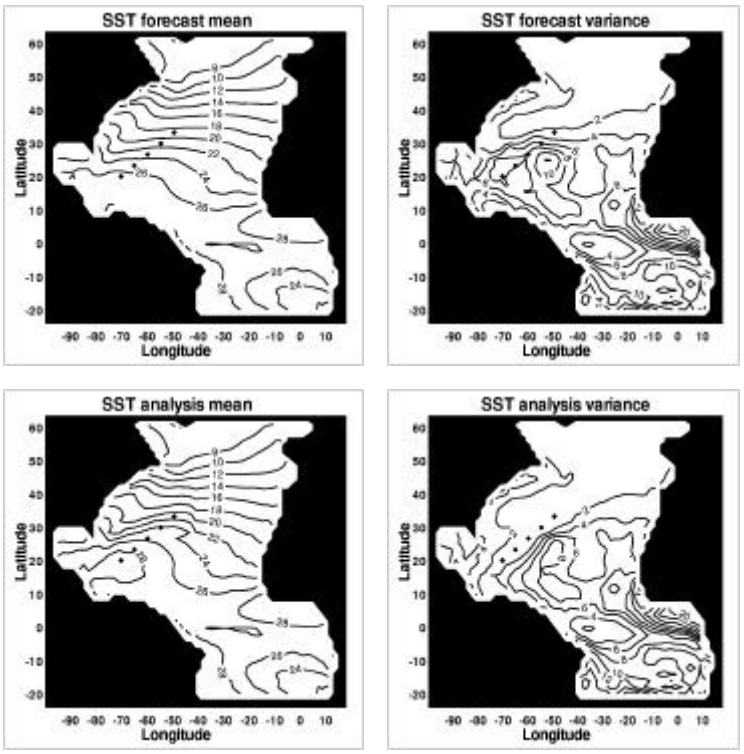

Fig. 6.5. Sea surface temperature prediction and analysis mean and variance.

observation will have on all the prognostic variables in the model. As an example, assume we have one observation of the sea surface or mixed layer temperature at a particular location. If the measured value is greater than the model predicted value this should imply that the model mixed layer temperature is increased by some amount in the analysis scheme. The update should of course be smooth in space. If the model mixed layer temperature is the only variable updated this will lead to an "unbalanced" model state. Clearly, the model density should be decreased accord-

ingly to be consistent with the equation of state. Further if the mixed layer temperature is increasing we would expect a decrease of the mixed layer depth, which requires the full vertical density structure of the model to change.

Important for this study is that MICOM uses density as the vertical coordinate. Thus, in the vertical the model can be considered as a stack of layers each having a constant density. This leads to an interesting interpretation of how the vertical

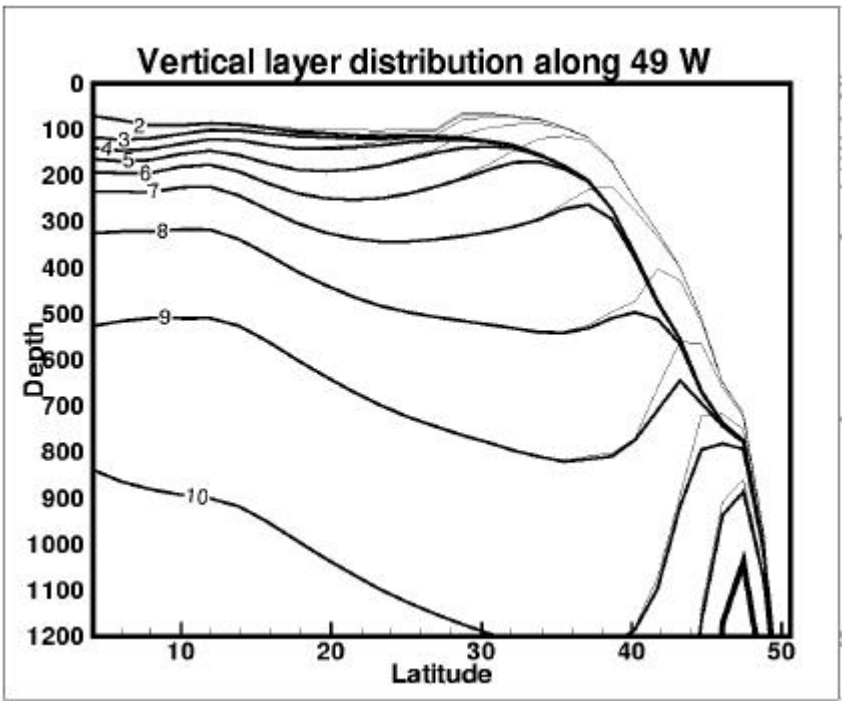

Fig. 6.6. Vertical section of layer thicknesses from South to North along longitude 49 W. The thick lines are the prediction before analysis and the thin lines denote the analysis.

stratification should be updated by the analysis scheme. Instead of updating the thermodynamic variables themselves one rather change the locations of the model layer interfaces, which gives the same effect. Clearly, it becomes complicated to construct Otimal Interpolation schemes for performing such an analysis since at various locations and times different layer interfaces should be altered. As an example, the layers with low density water masses will outcrop below the mixed layer when going northward from the equator. However the ensemble Kalman filter holds all the required information about the space and time dependent error covariances between different model variables which are used to calculate the actual influence functions for the measurements.

Here, a preliminary example is presented from an ensemble Kalman filter implementation with a coarse resolution version of MICOM for the North Atlantic. A 256 member initial ensemble is created by perturbing layer interfaces in a model state resulting from a 10 years spin up run. Each of the ensemble members are generated by adding smooth pseudo random fields with a specified covariance and mean equal to zero to the layer interfaces. Thus each of the ensemble members has a different vertical stratification at the initial time.

The ensemble is then integrated 10 years in a pure prediction mode with no assimilation of observations. Monthly means form the basis for the atmospheric forcing, however, to simulate realistic variability in the forcing fields and to account for errors in the monthly averages, the forcing was perturbed using pseudo-random fields. The perturbations were correlated in time to simulate the variability on a few days time scale of the real atmosphere, and of course different perturbations were used for each member of the ensemble. By perturbing the forcing fields we also include a realistic representation of the forcing errors which form an important part of the model errors. Note that time correlated model errors can be used in the ensemble Kalman filter contrary to the standard Kalman filter.

This 10 year ensemble integration can be considered as a spin up of the ensemble before starting a real data assimilation experiment. Also the predicted ensemble allows us to examine in detail the ensemble statistics including the covariances developing between the different model variables.

In Fig. 6.5 the effect of an analysis step using the ensemble Kalman filter is shown. It is assumed that five observations of the SST are available (shown as diamonds in the plots). Using these observations which all have values equal to the ensemble mean plus 1.5 degrees an analysis is calculated. The upper left plot shows the smooth 10 years ensemble mean SST prediction in January and the lower left plot shows the resulting SST analysis. Clearly, the temperature is increased in the area close to the observations. The upper right plot shows the predicted error variance for the SST, and the lower right plot shows the variance for the analysis having a distinct reduction near the observations. These are results as would be expected.

Of greater interest is the influence the SST data will have in the vertical. In Fig. 6.6 the vertical layer distribution is shown before and after the analysis. The vertical section is taken from South to North along 49°W. The layer interfaces are shown with the thick lines before the analysis and the thin lines after the analysis. The contour labels is the number of the layer below that particular interface. The layer distribution shows the deepening of the mixed layer in the North which is MICOMs way of accounting for convection. Clearly, the increase in the SST, or model mixed layer temperature, has lead to a shallowing of the mixed layer. Further, the vertical influence is obvious for all of the 11 layers in the model, which has been moved upward in the water column. That is, the pycnocline has been lifted and the vertical density structure has been changed in a dynamically balanced manner.

A more extensive discussion of the EnKF implementation with MICOM and an application will be reported in a manuscript in preparation.

6.7 Summary

A general overview discussion has been given on sequential data assimilation for nonlinear ocean models. The emphasis is laid on the ensemble Kalman filter (EnKF) which has shown to be a promising approach suitable for the nonlinear models used in oceanography. The theoretical basis for the EnKF was discussed in detail and the capability of the filter to handle strongly nonlinear dynamics was demonstrated in an example with the Lorenz equations. Preliminary results were also presented from an implementation of the EnKF with the MICOM ocean general circulation model. This is the first reported implementation of an advanced sequential data assimilation technique, where proper evolution of error statistics are taken into account, with a realistic OGCM. It was illustrated how the ensemble statistics contain the required information for generating a dynamically balanced analysis and thereby also project surface information in the vertical in a consistent manner.

A general conclusion so far is that the EnKF should be further implemented and explored with other dynamical models. With a relatively acceptable numerical load which corresponds to about 100–250 model integrations, the method can be used for operational oceanography on extant computer resources. The parallel computers are also perfectly suited for ensemble integrations where each member can be integrated independently on a separate processor.

Acknowledgments

The work was supported by European Commission through the Environment and Climate program under contract ENV4-CT95-0113 (AGORA) and by the Nordic Council of Ministers contract FS/HFj/X-96001. Further, it received support from the Norwegian Super Computing Committee (TRU) through a grant of computing time.