Classical Methods Tour of Advanced Data Assimilation using Lorenz '96 Model

Final Report for
METO658E, Spring-2004

Takemasa Miyoshi

5/7/2004

Department of Meteorology
University of Maryland, College Park

Abstract

Using Lorenz96 model with 40 variables, classical methods of advanced data assimilation are explained, implemented and examined. The classical methods include full Kalman filter (KF), extended Kalman filter (EKF), full Kalman smoother (KS), its iterative versions, and sawtooth algorithms (Johnston and Kurishnamurthy 2001). A brief explanation of the theoretical background of ensemble Kalman filter (EnKF) is also provided. The methods are tested under the perfect model assumption, and it is shown that KS clearly outperforms KF as expected thanks to the use of the future information. In addition, it is shown that iterative KF works more stably and outperforms KF especially in less dense observations both temporally and spatially. Furthermore, model errors are considered in a very simple way. It is shown that the effect of model errors was significantly reduced by increasing the variance inflation parameter.

Table of Contents

1 Introduction

It is important to find the best estimate of the atmospheric and oceanic states using all available information. In numerical weather prediction (NWP), for example, the better initial state provides the more accurate forecast. For climate studies, more accurate estimation of the atmospheric and oceanic states in past years, i.e. more accurate reanalysis products, provides important fundamental information. Data assimilation systems are designed to provide the most probable estimation of atmospheric and oceanic states, when observation data and a time evolving model are given. Thus, it is important to develop better data assimilation systems in earth sciences for daily NWP as well as climate studies.

The main difficulties in practical data assimilation on atmospheric and oceanic models exist in their high dimensionality, nonlinearity, and imperfectness. In general, the models used in atmospheric and oceanic sciences have millions of dimensionality, which makes it extremely difficult or almost impossible to implement truly optimal data assimilation system. Thus, some simplified methods such as successive correction method and optimal interpolation (OI) have been used for long in the operational NWP. These simplified methods assume constant second order statistics of forecast error, which is actually evolving in time. Recently, the growth of computational capability makes it possible to implement so-called advanced data assimilation methods including four-dimensional variational (4DVAR) data assimilation, which considers time evolution of the error covariance matrix implicitly. Even in this stage, however, the other difficulties, the model nonlinearity and imperfectness, are not considered thoroughly.

Away from the difficulties, Kalman (1960) proposed an optimal data assimilation algorithm known as Kalman filter (KF), which has been used widely in control systems including satellite-tracking calculations. Basically, if a time evolution model and observation data are given, KF solves the optimal estimation in terms of the second order statistics which of course evolves in time. It is known that in linear systems, KF and 4DVAR are equivalent (see for example, Bouttier and Courtier 1999). In large dimensional systems, KF is too expensive to implement, but Evensen (1994) proposed an efficient way to implement KF algorithms using ensemble forecasting, which is known as ensemble Kalman filter (EnKF), and various implementations of EnKF have been proposed so far (Tippet et al. 2003). Thus, it is a current issue to implement KF using computationally effective methods in atmospheric and oceanic sciences, which is why the full KF is cited as a classical method in the present paper. In nonlinear generalization, extended Kalman filter (EKF) is a general extension of the classical KF using tangent linear approximations, whereas EnKF does not linearize the model and have shown some advantages in nonlinear systems (Evensen 1994).

Smoothing problem is also an important issue in data assimilation. Filtering problem is a time forwarding process and only past information is used, whereas smoothing problem includes time backwarding processes and future information is also included. Because of the additional future information, smoothers can provide better estimation than filters. Even in nonlinear systems that forget past information in some finite time scales, smoother significantly outperforms filters because more amount of useful information is included from future. In NWP, smoothing problem seems useless because the latest data are assimilated to produce the latest analysis, that is, future observation is not available. However, if smoothers can create better estimation compared to filtering solutions, smoothers could be a better choice in reanalysis of

past years, where future observations are also available. In this sense, smoothing problem deserves to be considered in atmospheric and oceanic sciences. As a smoother, Kalman smoother (KS) is a well-known algorithm that uses KF in time forwarding and a similar algorithm in time backwarding. Extended Kalman smoother (EKS) is a nonlinear generalization of KS just like EKF is that of KF. At this point, since EKS provides both time forwarding (i.e. EKF) and backwarding algorithms, it is possible that these time forwarding and backwarding processes are repeated iteratively, which is known as iterative EKS (IEKS). It is said that IEKS gives better solution than simple EKF or EKS especially in nonlinear cases. In addition, there is another approach known as sawtooth IEKS (SIEKS) based on alternating expectation conditional maximization (AECM) formalism, whereas IEKS is based on expectation-maximization (EM) formalism. As EKS and EKF is a simplified version of IEKS, SIEKS has its simplified versions: sawtooth EKS (SEKS) and sawtooth EKF (SEKF), which give better estimations in some particular cases (Johnston and Kurishnamurthy 2001).

As a classical methods tour in the present paper, we forget about the problem of high dimensionality. As mentioned above, it is impossible to implement classical methods in high-dimensional systems. Using a low-dimensional system known as Lorenz 1996 model (L96, Lorenz 1996, 1998), classical methods of advanced data assimilation are introduced, implemented, and examined. The classical methods include full KF which is described later, EKF, full KS, an iterative method, and a sawtooth method. The system simulates atmospheric-like chaotic dynamics in one spatial dimensionality. In addition to the perfect model situation, constant and temporally varying biases are considered in a simple way. In realistic cases, a model cannot be perfect, and it is important to consider model errors. In the present paper, model errors are not treated comprehensively, but a simple treatment has shown to stabilize KF, which is described in section 3.

2 Theory

2.1 Kalman Filter

KF algorithms consist of two parts: time evolution and innovation. In a linear model $\mathbf{M}$, $\mathbf{x}_i^f = \mathbf{M}\mathbf{x}_{i-1}^a$ and $\mathbf{P}_i^f = \mathbf{M}\mathbf{P}_{i-1}^a\mathbf{M}^T + \mathbf{Q}$ are time evolution equations, where $\mathbf{x}$, $\mathbf{P}$, and $\mathbf{Q}$ denote state vector (model variables), error covariance matrices, and a model error covariance matrix, respectively. The upper subscripts mean analyses (a) and forecasts (f), and the lower subscript denotes time steps. Innovation equations are given as $\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}_i(\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^f)$ and $\mathbf{P}_i^a = [\mathbf{I} - \mathbf{K}_i\mathbf{H}]\mathbf{P}_i^f$, where $\mathbf{K}_i = \mathbf{P}_i^f\mathbf{H}^T[\mathbf{H}\mathbf{P}_i^f\mathbf{H}^T + \mathbf{R}]^{-1}$ is a Kalman gain matrix, which gives an optimal solution of this algorithm (for mathematical proof of the optimality of KF, see for example, Jazwinski 1970). Here $\mathbf{H}$ denotes linear observation operator defined as a mapping from model variables to observation data. For a nonlinear extension, tangent linear approximation is assumed by expanding the nonlinear model M as $M(\mathbf{x}_0 + \delta\mathbf{x}) = M\mathbf{x}_0 + \mathbf{M}_{\mathbf{x}_0}\delta\mathbf{x} + O(\delta\mathbf{x}^2)$, where $\mathbf{M}_{\mathbf{x}_0} = \left.\frac{\partial M}{\partial \mathbf{x}}\right|_{\mathbf{x}_0}$ is the Jacobian of the model M called a tangent linear model. Thus, the following five equations compose EKF algorithm:

$$\mathbf{x}_i^f = M\mathbf{x}_{i-1}^a \tag{2.1.1}$$
$$\mathbf{P}_i^f = \mathbf{M}_{\mathbf{x}_{i-1}^a}\mathbf{P}_{i-1}^a\mathbf{M}_{\mathbf{x}_{i-1}^a}{}^T + \mathbf{Q} \tag{2.1.2}$$
$$\mathbf{K}_i = \mathbf{P}_i^f\mathbf{H}^T[\mathbf{H}\mathbf{P}_i^f\mathbf{H}^T + \mathbf{R}]^{-1} \tag{2.1.3}$$
$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}_i(\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^f) \tag{2.1.4}$$
$$\mathbf{P}_i^a = [\mathbf{I} - \mathbf{K}_i\mathbf{H}]\mathbf{P}_i^f \tag{2.1.5}$$

For an effective implementation of this algorithm, taking square root of analysis error covariance matrix in eq.(2.1.2) as $\mathbf{P}_{i-1}^a = \mathbf{E}_{i-1}^a\mathbf{E}_{i-1}^a{}^T$, we get

$$\mathbf{P}_i^f = \mathbf{M}_{\mathbf{x}_{i-1}^a}\mathbf{E}_{i-1}^a\mathbf{E}_{i-1}^a{}^T\mathbf{M}_{\mathbf{x}_{i-1}^a}{}^T + \mathbf{Q} = \mathbf{M}_{\mathbf{x}_{i-1}^a}\mathbf{E}_{i-1}^a\left(\mathbf{M}_{\mathbf{x}_{i-1}^a}\mathbf{E}_{i-1}^a\right)^T + \mathbf{Q} \tag{2.1.6}$$

In this form, the adjoint model is not needed. Furthermore, this formulation enables to use nonlinear model instead of tangent linear model. This is just a practical treatment, but this has an advantage in the same way as EnKF, and this is called full KF in the present paper. Because the analysis error covariance matrix is positive definite, the real square root exists. Each column of the square root matrix $\mathbf{E}_{i-1}^a$ is considered as a perturbation around $\mathbf{x}_{i-1}^a$, which is an input to the model. Therefore, if nonlinear model is used in the eq.(2.1.6),

$$\begin{aligned}\mathbf{M}_{\mathbf{x}_{i-1}^a}\mathbf{E}_{i-1} &= \mathbf{M}_{\mathbf{x}_{i-1}^a}\left(\delta\mathbf{x}_{i-1}^1 \quad \cdots \quad \delta\mathbf{x}_{i-1}^N\right) \\ &\cong \left(\frac{M(\mathbf{x}_{i-1}^a + \varepsilon\delta\mathbf{x}_{i-1}^1) - M(\mathbf{x}_{i-1}^a)}{\varepsilon} \quad \cdots \quad \frac{M(\mathbf{x}_{i-1}^a + \varepsilon\delta\mathbf{x}_{i-1}^N) - M(\mathbf{x}_{i-1}^a)}{\varepsilon}\right)\end{aligned} \tag{2.1.7}$$

should be computed, where ε is a small number so that tangent linear approximation is valid. Note that eq.(2.1.7) is equivalent to ensemble forecasts.

It is helpful to mention that $\mathbf{E}$ can be a reduced matrix, that is, the rank of the matrix is much less than its full dimension. This is usually the case in atmospheric dynamical systems even in the simple L96 system. The eigenvalue decomposition of real symmetric matrix gives $\mathbf{P}^a = \mathbf{S}\mathbf{D}\mathbf{S}^T$, where $\mathbf{D}$ is a diagonal matrix whose values are real non-negative eivenvalues, and $\mathbf{S}$ is composed of corresponding eigenvectors which are orthogonal to each other. Usually $\mathbf{P}$ shows degeneracy, that is, there are only limited number of positive eivenvalues and remaining are 0. Thus, there exist only limited number of eigenvectors, and the effective dimension of the square root matrix $\mathbf{E}$ is much smaller than the full matrix, which can reduce the number of model integrations. Note that EnKF is based on this sort of discussion, where $\mathbf{E}$ consists of ensemble perturbations. The degeneracy of $\mathbf{P}$ implies that a relatively small number of ensemble members can reproduce $\mathbf{P}$ fairly well. If the effective number of perturbations are significantly smaller than the number of observations, eq.(2.1.3) to calculate Kalman Gain matrix can be significantly simplified as follows:

$$\mathbf{K} = \mathbf{P}\mathbf{H}^T[\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}]^{-1} = \mathbf{E}[\mathbf{I} + (\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1}\mathbf{H}\mathbf{E}]^{-1}(\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1} \tag{2.1.9}$$

where $\mathbf{P} = \mathbf{E}\mathbf{E}^T$, and subscripts are omitted. Eq.(2.1.9) is simpler because $p \times p$ matrix inversion is converted to $m \times m$ matrix inversion, where p and m denote number of observations and effective perturbations (i.e. ensemble members), respectively. The square root of forecast error covariance matrix in eq.(2.1.9) (denoting $\mathbf{E}_i^f$) is straightforward because of eq.(2.1.6), that is, $\mathbf{E}_i^f = \mathbf{M}\mathbf{E}_{i-1}^a$. Thus, except eq.(2.1.5) for innovating $\mathbf{P}$, we do not need to compute covariance matrices explicitly, but we just compute ensemble perturbations and their model integrations. The covariance information is implicitly included by eqs.(2.1.6) and (2.1.9). For the complete EnKF formalism, we need to form an ensemble update algorithm to realize eq.(2.1.5) implicitly, which is beyond the scope of the present paper. The derivation of eq.(2.1.9) is shown in appendix A.

In the case that observation is given as a nonlinear function of model variables, i.e. observation operator H is nonlinear mapping, it is generally impossible to solve eqs.(2.1.3) and (2.1.5). Usually this problem is treated by linearizing H under the tangent linear assumption. However, by virtue of the square root formulation of covariance matrices, this problem can be solved practically. Eq.(2.1.9) shows, for example, that the ensemble perturbations of $\mathbf{E}_i^f$ are projected to observation space by $\mathbf{H}$, which can be nonlinear mapping just in the same way as the model $\mathbf{M}$ in eq.(2.1.6) could be nonlinear.

In the above discussion, the theory of KF is completed. In practical, however, the error covariance matrix $\mathbf{P}$ is always underestimated, and the filter is eventually getting unstable. Thus, introduction of the model error $\mathbf{Q}$ or variance inflation is necessary for stable filtering. The theoretical explanation for the model error $\mathbf{Q}$ can partially attribute to the model nonlinearity under the perfect model assumption. It is a common technique to add random noise as $\mathbf{Q}$ in eq. (2.1.2). Alternatively, it is also a common technique to multiply a slightly larger number than 1 to the error covariance matrix at the end of the process. Namely, after the process of eq.(2.1.5),

$$\tilde{\mathbf{P}}^a = \mathbf{P}^a \times (1+\delta) \tag{2.1.10}$$

is performed so that the underestimated covariance matrix is simply enlarged by a factor $1+\delta$, where δ is a small number called variance inflation parameter. In the present paper, the model error covariance $\mathbf{Q}$ is ignored, and the variance inflation is considered. The variance inflation factor is an adjustable parameter in the algorithm and plays an important role in the practical filtering performance.

2.2 Kalman Smoother

In smoother, time forwarding process is exactly the same as KF. The time backwarding process is realized by the following algorithm proposed by Rauch-Tung-Striebel (1965) known as RTS algorithm:

$$\mathbf{G}_i = \mathbf{P}_i^a \mathbf{M}_i^T (\mathbf{P}_{i+1}^f)^{-1} = \mathbf{E}_i^a \mathbf{E}_i^{aT} \mathbf{M}_i^T (\mathbf{P}_{i+1}^f)^{-1} = \mathbf{E}_i^a (M_i \mathbf{E}_i^a)^T (\mathbf{P}_{i+1}^f)^{-1} = \mathbf{E}_i^a \mathbf{E}_{i+1}^{fT} (\mathbf{P}_{i+1}^f)^{-1} \quad (2.2.1)$$

$$\mathbf{x}_i^s = \mathbf{x}_i^a + \mathbf{G}_i (\mathbf{x}_{i+1}^s - \mathbf{x}_{i+1}^f) \quad (2.2.2)$$

$$\mathbf{P}_i^s = \mathbf{P}_i^a + \mathbf{G}_i (\mathbf{P}_{i+1}^s - \mathbf{P}_{i+1}^f) \mathbf{G}_i^T \quad (2.2.3)$$

where the subscript s denotes smoothing solutions, and the square roots ($\mathbf{P}_i^a = \mathbf{E}_i^a \mathbf{E}_i^{aT}$ and $\mathbf{P}_{i+1}^f = \mathbf{E}_{i+1}^f \mathbf{E}_{i+1}^{fT}$) simplify eq.(2.2.1) in a similar way as in eq.(2.1.6) of the KF algorithm, thanks to which no further model integration is necessary. However, eq.(2.2.1) requires an $n \times n$ matrix inversion, where n denotes the dimension of model variables, which makes it extremely expensive or almost impossible to compute in high-dimensional systems. Eq.(2.2.1) can be further simplified as

$$\mathbf{E}_i^a \mathbf{E}_{i+1}^{fT} (\mathbf{P}_{i+1}^f)^{-1} = \mathbf{E}_i^a \mathbf{E}_{i+1}^{fT} (\mathbf{E}_{i+1}^f \mathbf{E}_{i+1}^{fT})^{-1} = \mathbf{E}_i^a \mathbf{E}_{i+1}^{fT} (\mathbf{E}_{i+1}^{fT})^{-1} \mathbf{E}_{i+1}^{f\ -1} = \mathbf{E}_i^a \mathbf{E}_{i+1}^{f\ -1} \quad (2.2.4)$$

but even in the case that $\mathbf{E}_{i+1}^f$ is a square matrix, $\mathbf{E}_{i+1}^f$ is not symmetric in usual, thus it is not easy to compute the inverse. In addition, all the covariance matrices or ensemble members need to be stored through the smoothing steps, which requires large amount of memory. Furthermore, eq(2.2.3) has a more complicated form compared to eq.(2.1.5), which also makes it difficult to form an ensemble formalization. Thus, in a practical point of view, it is not easy to implement this algorithm in realistic atmospheric and oceanic systems.

Supposing far away from high-dimensionality, now that we have time forwarding and backwarding algorithms, we can use the future information to estimate the present state. This process is schematically shown in fig. 2.1. $\mathbf{x}_1^s$ is a smoothing solution at the initial time where all observations through the future time N is assimilated in the filtering processes. Thus, to obtain an optimal estimate at the initial time, filters use all the past information, whereas smoothers use additional N-step future information. To initiate the smoother, $\mathbf{x}_N^s = \mathbf{x}_N^a, \mathbf{P}_N^s = \mathbf{P}_N^a$ is used. Fig. 2.2 shows schematic of filter and smoother solutions. Generally, RMSE($\mathbf{x}^s_i$) < RMSE($\mathbf{x}^a_i$) and RMSE($\mathbf{x}^s_i$) < RMSE($\mathbf{x}^s_{i+1}$) are satisfied, where RMSE stands for root mean square error.

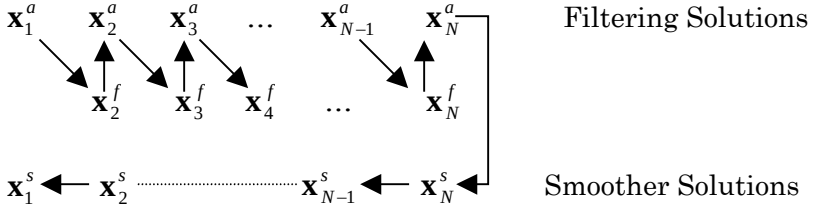

Figure 2.1. Schematic of smoothing processes. The upper subscripts denote filter solutions (a), forecast (f), and smoother solutions (s), respectively, and the lower subscripts denote time steps, where N is smoothing time window.

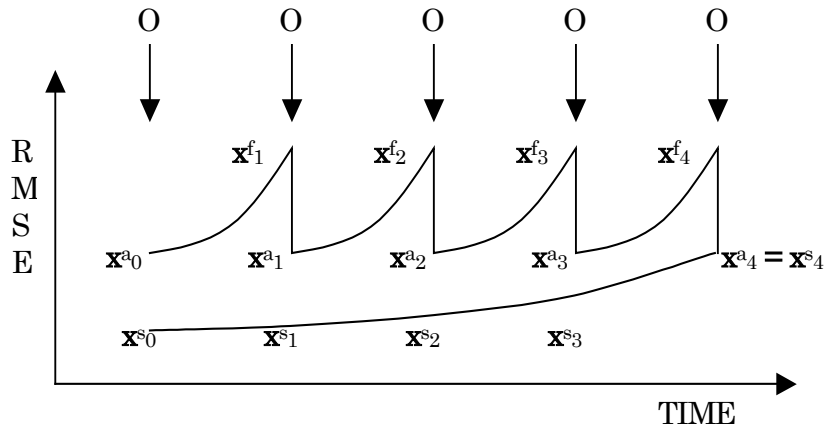

Figure 2.2. Schematic of Kalman filter and smoother solutions. The upper subscripts denote filter solutions (a), forecast (f), and smoother solutions (s), respectively, and O denotes observation. This shows the case of N=4, where N is smoothing time window.

2.3 Iterative Methods

Using the algorithms described above, time forwarding and backwarding can be used iteratively as shown in fig. 2.3, which shows schematic of iterative Kalman smoother (IKS). If the iteration is performed using only past and present observations, the same algorithm becomes iterative Kalman filter (IKF), which is shown in fig 2.4 where the smoothing time window is 1. Note that it is not clear if the iterative solution is better than the non-iterative solution. Theoretically, KF itself is optimal, and it cannot be better using the same amount of information, on condition that the model is linear and perfect, and the error statistics is Gaussian. Model nonlinearity introduces non-optimality of KF, and it is expected that iterative processes reduce the effect of nonlinearity and provide better solution, that is, iterative KF is expected to outperform KF when observation is sparse in time and space.

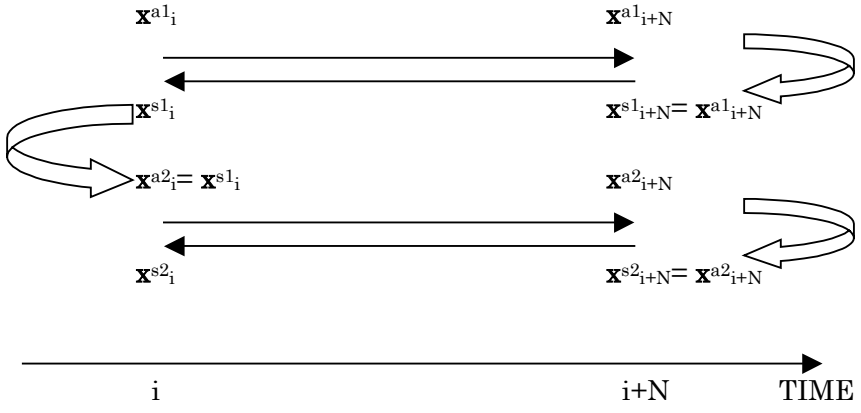

Figure 2.3. Schematic of iterative Kalman smoother (IKS). The upper subscripts denote filter solutions (a) and smoother solutions (s), and their iteration numbers. N is smoothing time window.

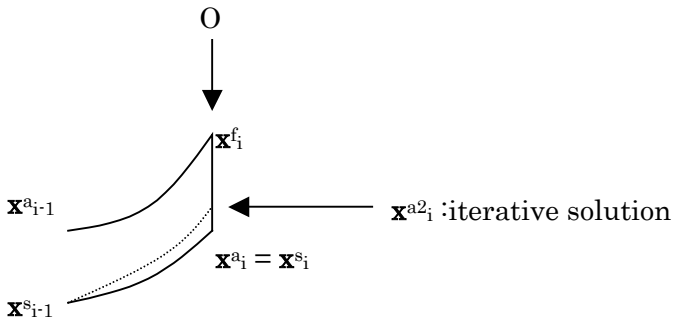

Figure 2.4. Schematic of Iterative Kalman Filter (IKF) with smoothing time window 1. Iterative solution is denoted as upper subscript a2. It is not clear if iterative solution is better than the first filter solution.

2.4 Sawtooth Methods (Johnston and Krishnamurthy 2001)

Extended Kalman filter is based on the expectation-maximization (EM) algorithm, whereas sawtooth extended Kalman filter is based on alternating expectation conditional maximization (AECM) algorithm.

Smoother gain matrix G can be computed without future observation. Expecting the structure of error vector $\mathbf{x}^{a}_{i+1} - \mathbf{x}^{f}_{i+1}$ and error covariance difference $\mathbf{P}^{a}_{i+1} - \mathbf{P}^{f}_{i+1}$ are similar to $\mathbf{x}^{f}_{i+1} - M\mathbf{x}^{f}_{i}$ and $\mathbf{P}^{f}_{i+1} - M\mathbf{P}^{f}_{i}M^{T}$ respectively, we can compute one step smoothing process using one step forecast even if future observation is not available. This process uses only past observations, so this is filtering problem called a sawtooth Kalman filter (SKF). Fig. 2.5 shows schematic of SKF. In the same way, sawtooth Kalman smoothers (SKS) can be formed and their iterative methods are also straightforward.

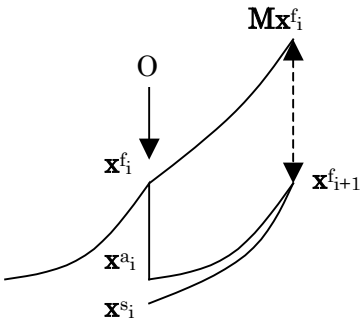

Figure 2.5. Schematic of sawtooth Kalman filter (SKF).

3 Numerical Experiments

3.1 Model description

The L96 model is defined by

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F \,. \tag{3.1.1}$$

Here, $i = 1, \cdots, N_x$, where the boundary is cyclic, i.e. $x_{-1} = x_{N_x-1}$, $x_0 = x_{N_x}$, and $x_{N_x+1} = x_1$. This model behaves chaotically in the case of external forcing $F = 8.0$, in which case the time step of 0.2 non-dimensional units corresponds to about one day in terms of error growth rate of five days (Lorenz 1996). The first term of right hand side simulates "advection", and this model can be regarded as the time evolution of an arbitrary one-dimensional quantity on a constant latitude circle, that is, the subscript i corresponds to longitude. As in Lorenz (1996), I choose $N_x = 40$ and $F = 8.0$. Eq (3.1.1) is solved using Runge-Kutta fourth order scheme with integration time step 0.01, thus, 5 steps correspond to 6 hours. Lag auto-correlation is shown in fig. 3.1.

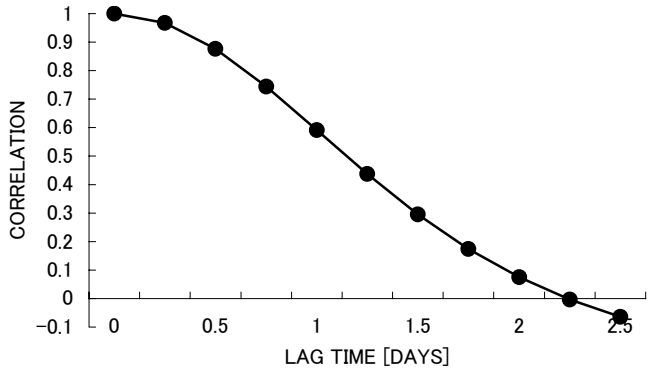

Figure 3.1. Lag correlation of the Lorenz96 model.

3.2 Design of experiments

Basic design is based on a simple observation system simulation experiments (OSSEs) in assuming true state is known. The true state is created by a long-term integration from an arbitrary chosen initial condition. Observations are simulated by simply adding random noise to the true state. Here, observation grid and model grid are the same; observation operator H is simply an identity mapping or reducing observation points. The initial condition for data assimilation can be any arbitrary state, and a long-term mean state is chosen in the present experiment. The initial analysis error covariance matrix can also be any arbitrary real symmetric matrix as long as it is not too small and usually diagonal components are the largest. I chose a matrix whose each row/column is Gaussian shape with fairly large amplitude. The performance of data assimilation is measured by root mean square error (RMSE). Now that true state is known,

the error is easily computed by subtracting true state from analysis state.

It is important to know how data assimilation works in decreasing number of observation both spatially and temporally. Thus, the experiment using 40, 20, and 10 observations in every 6, 12, 24, and 48 hours are performed. Every combination of these spatial and temporal densities is examined. For example, 40 observations in every 6 hours is the densest situation, whereas 10 observations in every 48 hours is the sparsest situation. In this experiment, observation error standard deviation is set to 1.0. In short, the experiment settings are summarized in table 3.1.

For introducing model errors, a constant bias and temporally varying biases are added to true time evolution:

$$\mathbf{x}_{i+1}^t = M\mathbf{x}_i^t + \mathbf{b}_i \quad (3.2.1)$$

$$\mathbf{b}_i = \mathbf{b}_0 + \mathbf{b}_1 \sin(\omega i) \quad (3.2.2)$$

where $\mathbf{b}_0$, $\mathbf{b}_1$, and ω are constant bias, amplitude and frequency of temporally varying bias, respectively. Using this biased true time evolution, the same experiment described above is performed.

Table 3.1. Summary of experiment settings.

Initial state	Longtime mean state
Initial state of the true state	From longtime integration
Model integration time step	$\Delta t = 0.01$
Value of forcing of Lorenz96	F=8.0
Definition of 1 day	0.2 time units (6hr=0.05)
# of observation	40 / 20 / 10
Observation error	1.0
Assimilation cycle	Every 6 / 12 / 24 / 48 hours

3.3 Results

3.3.1 Various methods in perfect model

For the criterion of RMSE, the RMSE of forecast with no assimilation (free run, i.e. doing nothing) is computed. The value is around 5.0. Thus, if the analysis RMSE is less than 5.0, the data assimilation works better than doing nothing. Because observation error is 1.0, 1.0 can be another criterion. For example, if every point has observation and analysis RMSE is larger than 1.0, the direct insertion of observation is better than the data assimilation.

The graphical result how KF works is shown in fig. 3.2, where RMSE (upper panel) and the ratio of RMSE against free run (lower panel) are shown for first 14 days. The left two panels show in the case of 40 observations, and the right two panels show in the case of 20 observations, both at every 6 hours. In the upper two panels, black circle, white circle, green, yellow, red lines show RMSE of free run, that of analysis, that of forecast, square root of the mean of diagonal

components of forecast error covariance matrix, and that of analysis error covariance matrix, respectively. In the lower two panels, the ratio of analysis RMSE against free run is shown in percent. KF performs very well in both cases, though the initial spin-up period is longer in the case of 20 observations.

The result of the full KF is shown in table 3.2. Here, the RMSE is averaged over 365 days where the first 120 steps are eliminated in considering the initial spin-up of the filter. In 40 observations, i.e. observing at every point, the data assimilation works well even in decreasing temporal density of observation to every 48 hours, though the filter is rather unstable in the choice of variance inflation parameter. By decreasing half of the observation, more than 24 hours interval of observation makes KF unstable. In 10 observations, KF is stable only in observing every 6 hours, but the performance is worse than free run, i.e. the data assimilation provides no useful information at all.

The result using EKF is shown in table 3.3 in the same way as table 3.2. Here, the only difference from the previous experiment exists in using tangent linear model instead of nonlinear model in computing eq. (2.1.6). RMSEs are comparable to full KF, although it is expected that the full KF outperforms EKF as temporal density of observation decreases, when model nonlinearity becomes important. In this case, however, the result shows EKF slightly outperforms full KF in 40 observations every 48 hours. Furthermore, EKF is more stable than the previous full KF in the choice of variance inflation parameter. This tendency is clear especially in the case of less observation densities such as 40 observations at every 48 hours.

KS uses future information and is expected to provide better solution than KF. The RMSEs of KF and KS are shown in table 3.3. Here, the RMSE is averaged over 180 days where the first 150 steps are excluded because of the initial spin-up. In this experiment, number of observation is 40 or 20 at every 6 hours. LEKF as well as KF and KS is also shown as an example of EnKF, where LEKF stands for Local Ensemble Kalman Filter (Ott et al. 2004). The detail of LEKF is not explained in considering the scope of the present paper. In the case of 40 observations with variance inflation factor of 0.05, RMSE of 0.11 seems lower limit even if smoothing time window is increased beyond 3. However, if the variance inflation parameter is decreased to 0.01, the assimilation skills are getting better in increasing smoothing time window. Smoothing time window decides how many future data are used in smoother; the longer the time window is, the more observations are used, and the better the performance is expected. Thus, variance inflation $\delta = 0.05$ makes the error covariance overestimated especially in $N > 1$ where substantial decrease of RMSE can be seen in $\delta = 0.01$. In the case of 20 observations, the optimal variance inflation parameter is chosen in each case. The result shows surprisingly better analysis by smoothers compared to filters even in decreasing observations by half.

Analysis RMSE of iterative full KF is shown in table 3.5 in the same way as table 3.2 (365-day average). The algorithm is exactly the same as what figure 2.4 shows, that is, one step smoothing process is performed using only past information. Here, the iteration is only once. In more frequent observation, there is no advantage can be seen. However, there are significant advantages in less dense observations, and the filter is surprisingly much more stable which should be stressed here. For example, in the case of 40 observations at every 48 hours, the RMSE is about 0.5, that was 0.7 in KF. In the case of 20 observations, KF got unstable when observing less than every 24 hours, where iterative KF shows its stability. Furthermore, the case of 20

observations at every 24 hours shows fairly good data assimilation performance. Even in 10 observations, the filter is stable, and data assimilation works better than doing nothing. As mentioned above, the stability of filter depends on the choice of variance inflation parameter, but this method is much more stable in the choice, too.

Sawtooth methods do not work well in this case. Even in observing 40 points every 6 hours, SEKF shows more than 3 of RMSE, where KF showed 0.2. There is no advantage using sawtooth methods in this particular problem.

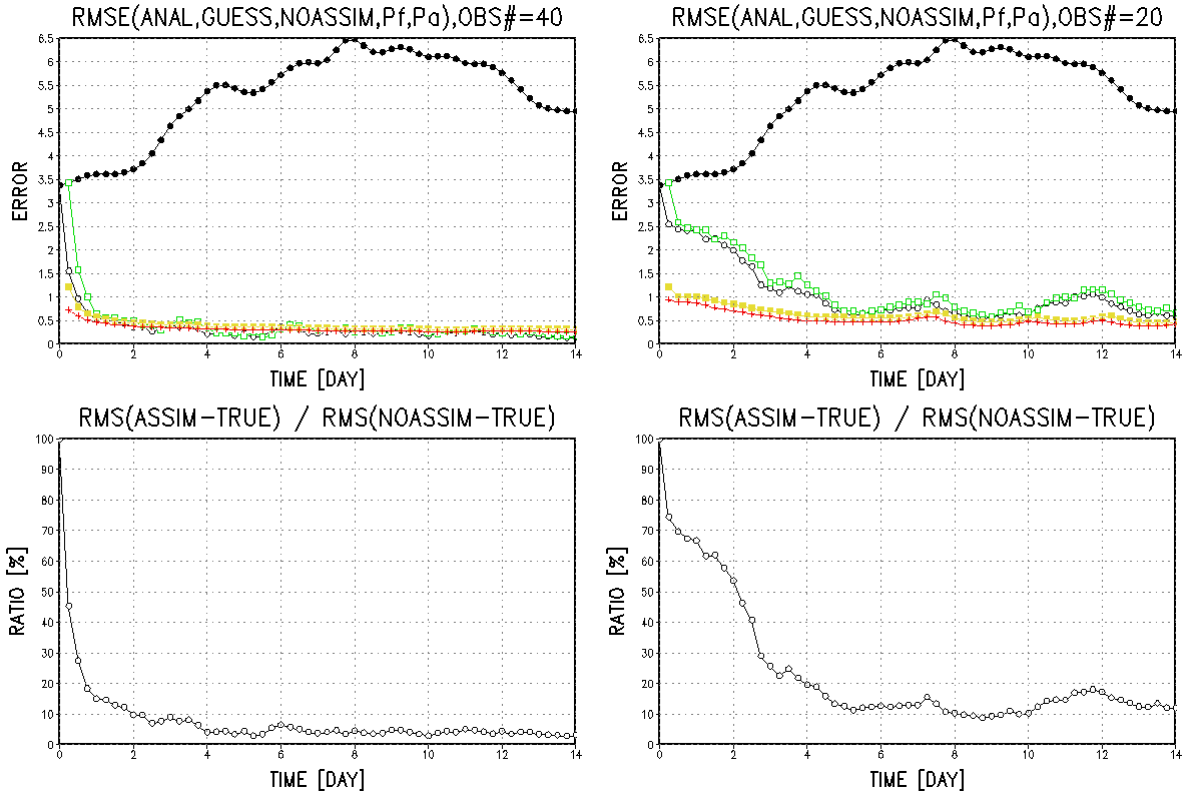

Figure 3.2. RMSE (upper two panels) and ratio of RMSE of analysis against free run (lower two panels) in the case of 40 observations (left two panels) and 20 observations (right two panels) both at every 6 hours. In upper two panels, black circle, white circle, green, yellow, red show RMSE of free run, that of analysis, that of forecast, square root of trace of forecast error covariance matrix, and that of analysis error covariance matrix, respectively. In lower two panels, the ratio of analysis RMSE against free run is shown in percent.

Table 3.2. Analysis RMSE of full KF (365-day average after convergence). δ denotes variance inflation parameter, and "–" means unstable filter.

	Obs# = 40	Obs# = 20	Obs# = 10
6 hours	0.192 ($\delta = 0.05$)	0.368 ($\delta = 0.1$)	5.40 ($\delta = 0.3$)
12 hours	0.298 ($\delta = 0.2$)	0.712 ($\delta = 0.5$)	-
24 hours	0.481 ($\delta = 0.75$)	-	-
48 hours	0.763 ($\delta = 2.0$)	-	-

Table 3.3. Analysis RMSE of EKF. Notations are the same as table 3.2.

	Obs# = 40	Obs# = 20	Obs# = 10
6 hours	0.211 ($\delta = 0.05$)	0.356 ($\delta = 0.1$)	4.85 ($\delta = 0.5$)
12 hours	0.296 ($\delta = 0.2$)	0.765 ($\delta = 0.5$)	-
24 hours	0.471 ($\delta = 0.75$)	-	-
48 hours	0.699 ($\delta = 1.5$)	-	-

Table 3.4. Analysis RMSE of full KF, KS, and LEKF. For smoothers, the smoothing time window (N) is changed from 1 to 6.

Assimilation cycle	RMSE (obs#=40) $\delta = 0.05$	RMSE (obs#=40) $\delta = 0.01$	RMSE (obs#=20) Optimal choice
Filtering	0.193	-	0.358 ($\delta = 0.1$)
Smoothing (N=1)	0.138	0.124	0.198 ($\delta = 0.04$)
Smoothing (N=2)	0.124	0.0895	0.149 ($\delta = 0.02$)
Smoothing (N=3)	0.113	0.0728	0.102 ($\delta = 0.005$)
Smoothing (N=4)	0.110	0.0636	0.0864 ($\delta = 0.005$)
Smoothing (N=5)	0.111	0.0596	0.0790 ($\delta = 0.005$)
Smoothing (N=6)	0.111	0.0508	0.0707 ($\delta = 0.005$)
LEKF	0.246		0.377

Table 3.5. Analysis RMSE of iterative KF. Notations are the same as table 3.2.

	Obs# = 40	Obs# = 20	Obs# = 10
6 hours	0.201 ($\delta = 0.05$)	0.400 ($\delta = 0.1$)	3.67 ($\delta = 0.2$)
12 hours	0.272 ($\delta = 0.05$)	0.469 ($\delta = 0.1$)	3.43 ($\delta = 0.4$)
24 hours	0.371 ($\delta = 0.2$)	0.622 ($\delta = 0.3$)	4.08 ($\delta = 0.45$)
48 hours	0.491 ($\delta = 0.5$)	1.88 ($\delta = 1.2$)	-

3.3.2 Introduction of model errors

The way to introduce model errors is given in eqs. (3.2.1) and (3.2.2), where constant bias and temporally varying biases are considered at the same time. Now, constant bias and temporally varying biases are introduced separately, that is, $\mathbf{b}_0$ is 0 when $\mathbf{b}_1$ has some value, and vice versa. The RMSE is averaged over 90 days where first 150 steps are excluded because of the initial spin-up. For comparison, the result of simple OI/3DVAR is shown. As clearly shown, the simple OI/3DVAR performs very stably in biases. The bias less than 0.2 does not affect at all,

whereas KF is affected very much. Note that 0.2 is the size of the analysis error of KF, which is one fifth of observation error, and it is hard to detect such small biases in realistic situation. KF is very sensitive to both constant and temporally varying biases. Even the bias of 0.1 makes KF significantly worse. Furthermore, even if constant bias is completely known and eliminated, temporally varying bias with amplitude of 0.1 and 3-day cycle makes KF worse significantly. Thus, in implementing KF in a realistic situation, the bias correction is a very important issue. A simple treatment of model biases can be done by just increasing variance inflation parameter. The table 3.6 shows the effect of small model biases can be significantly reduced by increasing variance inflation parameter that is equivalent to overestimating analysis errors. Thus, the increase in analysis errors caused by model errors can be treated as overestimation of analysis errors by large variance inflation parameter. Remember that $\mathbf{Q}$ in eq. (2.1.2) introduces model error statistics in KF algorithms, but $\mathbf{Q}$ is a second order statistics and not first order statistics such as biases, so $\mathbf{Q}$ cannot treat the problem of model biases.

Table 3.6. Analysis RMSE of KF in introducing model biases. The bias is defined in eqs. (3.2.1). The RMSEs are 180-day (720-step) average values, where the first 150 steps are excluded, where 1 step corresponds to 6 hours.

BIAS			RMSE			
$\mathbf{b}_0$	$\mathbf{b}_1$	ω^{-1} [Days]	KF $\delta = 0.05$	KF $\delta = 0.1$	KF $\delta = 0.5$	Simple OI / 3DVAR
0	0	-	0.18	0.20	0.38	0.65
	0.1	1	0.22	0.22	0.39	0.66
		3	2.99	0.57	0.41	0.65
		5	3.37	1.41	0.44	0.66
		10	4.55	1.30	0.46	0.66
		15	4.11	1.78	0.45	0.66
		30	4.17	3.37	0.47	0.65
0.01	0	-	0.30	0.23	0.38	0.65
0.1	0	-	3.69	3.08	0.50	0.66
0.2	0	-	5.35	4.41	0.77	0.67
0.5	0	-	6.44	5.90	1.61	0.74

4 Summary and Future plan

4.1 Main findings

Full KF, EKF, full KS, iterative KF, sawtooth KF are implemented and compared. The main findings in this comparison can be summarized as follows:

1. RMSE of full KF using densest observations was 0.2
2. EKF is more stable than full KF
3. Full KS outperforms KF (the best performance shows RMSE ~0.05)
4. Iterative KF outperforms KF in less dense observations both spatially and temporally
5. Iterative KF is much more stable than full KF
6. Sawtooth KF does not work at all in this particular case

Introduction of model errors clarify the following findings:

1. KF is very sensitive to model errors, whereas OI/3DVAR is very stable
2. Even time mean bias is eliminated, temporally varying bias with frequencies more than 3 days affects very much
3. Increasing variance inflation parameter can reduce the effect of model errors

4.2 Future plans

Tippet et al. (2003) summarized various implementations of ensemble square root Kalman filters (EnSRF) that can be applied in realistic atmospheric models, which includes sequential method, ensemble transform Kalman filter (ETKF), and ensemble adjustment Kalman filter (EAKF). LEKF, which is one of the various implementations of EnSRF, proposed by Ott et al. (2004) has been already implemented and the result is shown in table 3.4, though the precise description was omitted in considering the scope of the present paper. In future, beyond the classical methods, these more realistic methods which can be used in high dimensional systems can be implemented and compared. As Tippet et al. (2003) mentioned, it is not clear if any implementation of efficient Kalman filter outperforms other methods. Thus, it is worth implementing and comparing different implementations of EnKF even in the simple L96 model. Of course, after the precise investigation using the simple model, it is desired to implement in more realistic models.

Another direction is on the model errors. In realistic situations, the model cannot be perfect; the source of model error is abundant. In addition, the estimation of model errors is not easy, and considering model errors in data assimilation algorithm is very important. Thus, it is important to discuss more precisely how to treat model errors. Dee and Da Silva (1998) discussed data assimilation in the presence of model errors. KF provides important information in real-time estimation of model errors. In developing EnKF, including such algorithm has much practical importance.

Acknowledgement

The author is grateful to Dr. Joaquim Ballabrera who directed the course and gave lots of useful suggestions and constructive discussions throughout the course. The author also thanks the classmates Shu-Chih Yang and John Harlim for many useful discussions in the classes. The Japanese Government Long-term Fellowship Program supported the author to conduct the graduate study at the Department of Meteorology, University of Maryland, College Park.

Appendix

A. Derivation of eq. (2.1.9)

Using a formulae of the linear algebra:

$$[X_1 + X_{12}X_2^{-1}X_{21}]^{-1} = X_1^{-1} - X_1^{-1}X_{12}[X_2 + X_{21}X_1^{-1}X_{12}]^{-1}X_{21}X_1^{-1} \quad \text{(A.1)}$$

the RHS of eq.(2.1.3) is transformed as follows:

$$\begin{aligned}
&\mathbf{PH}^T[\mathbf{HPH}^T + \mathbf{R}]^{-1} = \mathbf{E}(\mathbf{HE})^T[\mathbf{HE}(\mathbf{HE})^T + \mathbf{R}]^{-1} \\
&= \mathbf{E}(\mathbf{HE})^T[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{HE}[\mathbf{I} + (\mathbf{HE})^T\mathbf{R}^{-1}\mathbf{HE}]^{-1}(\mathbf{HE})^T\mathbf{R}^{-1}] \\
&= \mathbf{E}(\mathbf{HE})^T\mathbf{R}^{-1} - \mathbf{E}(\mathbf{HE})^T\mathbf{R}^{-1}\mathbf{HE}[\mathbf{I} + (\mathbf{HE})^T\mathbf{R}^{-1}\mathbf{HE}]^{-1}(\mathbf{HE})^T\mathbf{R}^{-1} \\
&= \mathbf{E}[\mathbf{I} - (\mathbf{HE})^T\mathbf{R}^{-1}\mathbf{HE}[\mathbf{I} + (\mathbf{HE})^T\mathbf{R}^{-1}\mathbf{HE}]^{-1}](\mathbf{HE})^T\mathbf{R}^{-1} \\
&= \mathbf{E}[[\mathbf{I} + (\mathbf{HE})^T\mathbf{R}^{-1}\mathbf{HE}] - (\mathbf{HE})^T\mathbf{R}^{-1}\mathbf{HE}][\mathbf{I} + (\mathbf{HE})^T\mathbf{R}^{-1}\mathbf{HE}]^{-1}(\mathbf{HE})^T\mathbf{R}^{-1} \\
&= \mathbf{E}[\mathbf{I} + (\mathbf{HE})^T\mathbf{R}^{-1}\mathbf{HE}]^{-1}(\mathbf{HE})^T\mathbf{R}^{-1}
\end{aligned} \quad \text{(A.2)}$$

References

Bouttier F., 1996: The Kalman Filter. Proc. Seminar on Predictability, Vol. 1, ECMWF, Reading, Berkshire, UK, 221-245.

Bouttier F. and P. Courtier, 1999: Data assimilation concepts and methods. Meteorological Training Course Lecture Series, ECMWF, 75pp.

Dee, D. P. and A. M. Da Silva, 1998: Data Assimilation in the presence of forecast bias. Q. J. R. Meteor. Soc., 124, 269-295.

Evensen, G., 1994: Sequential Data Assimilation with a Nonlinear Quasi-geostrophic Model using Monte Carlo Methods to Forecast Error Statistics. J. Geophys. Res., 99(C5), 10143-10162.

Jazwinski, A. H., 1970: Stochastic processes and filtering theory. Academic Press, 376pp.

Johnston, L. A. and V. Krishnamurthy, 2001: Derivation of a Sawtooth Iterated Extended Kalman Smoother via the AECM Algorithm. IEEE Transactions on Signal Processing, 49, 1899-1909.

Kalman, R. E., 1960: A New Approach to Linear Filtering and Prediction Problems. Transaction of ASME, D, Journal of Basic Engineering, 82, 35-45.

Kalnay, E., 2002: Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press.

Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal, 1997: Reduced Space Optimal Analysis for Historical Data Sets: 136 years of Atlantic Sea Surface Temperatures. J. Geophys. Res., 102(C13), 27835-27860.

Lorenz, E. N., 1996: Predictability: A problem partly solved. Proc. Seminar on Predictability, Vol. 1, ECMWF, Reading, Berkshire, UK, 1-18.

Lorenz, E. N. and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. J. Atmos. Sci., 55, 399-414.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, J. A. Yorke, 2004: A Local Ensemble Kalman Filter for Atmospheric Data Assimilation. Tellus, in press.

Rauch, H. E., F. Tung, and C. T. Striebel, 1965: Maximum Likelihood Estimates of Linear Dynamic Systems. AIAA Journal, 3, 1445-1450.

Tippet M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble Square Root Filters. Mon. Wea. Rev., 131, 1485-1490.