A nonlinear filter that extends to high dimensional systems

Thomas Bengtsson*, Chris Snyder, Doug Nychka
National Center for Atmospheric Research. Boulder, CO USA.

Abstract

Numerical weather prediction is characterized by high-dimensional, nonlinear systems and poses difficult challenges for real-time data assimilation (updating) and forecasting. The goal of this work is to build on the ensemble Kalman filter (EnsKF) to produce ensemble filtering techniques applicable to non-Gaussian densities in high dimensions.

Two filtering algorithms are presented which extend the ensemble Kalman filter by use of Gaussian mixtures. The first method, referred to as a mixture ensemble Kalman filter (XEnsF), adaptively represents local covariance structures using nearest neighbors. An efficient sampling algorithm is presented for XEnsF, and the filter is shown to be superior to existing methods in simulations on a three-dimensional model. A second algorithm, the local-local ensemble filter (LLEnsF), combines localizations in physical as well as phase space, allowing the update step in high dimensional systems to be decomposed into a sequence of lower-dimensional updates tractable by the XEnsF. Given the same ensemble in a 40-dimensional system, the LLensF update is shown to locally produce more accurate estimates of the state than the EnsKF when the underlying distributions are strongly non-Gaussian. In the 40-dimensional system, a hybrid filter combining the output from LLensF with that of EnsKF is shown to outperform the EnsKF by 5.7%.

Keywords: Non-linear filtering, data assimilation, Bayesian filtering, particle filtering, ensemble Kalman filter, numerical weather prediction, state estimation.

*Supported by National Science Foundation DMS 9815344

1 Introduction

Data assimilation for the ocean and atmosphere are important cases of estimating the state of a system given a sequence of observations and (some) knowledge of the evolution of the system. Because the observations and the forecast model are inexact (and because the evolution of the state depends sensitively on initial conditions), the true state of the system can never be determined precisely. The most complete summary of our knowledge of the system state is therefore given by the probability density function (pdf) of the state conditional on the observations (*Epstein* 1969). In a geophysical context, both forecasting this pdf forward in time and updating the forecast pdf given new observation have formidable obstacles: the dimension of the state-vector in most oceanic and atmospheric models is extremely high, often exceeding 10^6 components, and the systems are significantly nonlinear, leading to the potential for non-Gaussian pdfs.

The present article focuses on ensemble or Monte-Carlo techniques for forecasting and updating of the pdf. One promising approach for high-dimensional geophysical problems is the ensemble Kalman filter (EnsKF; *Evensen* 1994, *Houtekamer and Mitchell* 1998). The EnsKF update, however, depends only on the first and second moments of the ensemble and is thus suboptimal for non-Gaussian pdfs. Our goal here is to build on the EnsKF to produce ensemble techniques applicable to non-Gaussian pdfs in high dimensions.

The algorithms we present approximate the forecast distribution by mixtures of Gaussian distributions. The use of Gaussian mixtures allows (in principle) arbitrary, non-Gaussian pdfs to be handled and reduces updating the pdf given observations to updating each individual Gaussian in the mixture along with its mixing probability (*Anspach and Sorensen* 1972). Gaussian mixtures have been used before as the basis for ensemble assimilation techniques (*Anderson and Anderson* 1999, *Chen and Liu* 2000), but these existing techniques are problematic in high-dimensional systems.

The difficulties with such existing techniques arise in part because the methods used to resample from the posterior pdf are computationally intensive. At a more fundamental level, however, the difficulties are intertwined with the well-known difficulty of estimating pdfs in high dimensions (*Silverman 1986*). Simple estimates suggest that the sample size required to estimate a multivariate pdf with a given accuracy increases exponentially with dimension. For systems with 10^6–10^8 variables, such as global atmospheric forecast models, the huge sample sizes required clearly rule out direct, brute-force attempts to estimate non-Gaussian pdfs. Mixture estimates suffer from the same limitations. In ensemble techniques, these limitations result in extremely large sampling variability and the collapse of the mixture onto a single ensemble member.

To make updating feasible in high dimensions, we suggest three enhancements of these existing techniques.

1. The covariance for each Gaussian in the mixture is based on the sample covariance of a subset of ensemble members that are close in phase space to each center. This makes the mixture adaptive as the estimate of the pdf depends on the structure of the sample in phase space, and helps to capture lower-dimensional "sheets" that are typical of chaotic dynamics.

2. We generalize the implicit sampling scheme of EnsKF, which avoids manipulation of large matrices and is feasible in high dimensions, to mixtures of Gaussian distributions. The extension is straightforward but is not available in the literature.

3. The algorithms allow each observation to influence only state variables that are nearby in physical space. This physically local updating is a common feature of geophysical assimilation schemes, including both optimal interpolation (*Schlatter et al.* 1976) and the EnsKF (*Houtekamer and Mitchell* 1988), but its application to the non-Gaussian filtering problem is novel and nontrivial.

We will show that these three ideas yield a technique that can produce an update with smaller MSE than the EnsKF (given the same forecast ensemble) if the underlying distributions are strongly non-Gaussian.

The paper proceeds as follows. Section 2 presents additional background and notation. This includes an introduction to the atmospheric and oceanic assimilation problem, together with background on the Kalman filter, the EnsKF, and the update for a Gaussian mixture. Readers familiar with these topics may wish to proceed directly to section 3, which outlines two filtering algorithms. These we term the mixture ensemble filter and the local-local ensemble filter. The local-local filter incorporates each of the three enhancements discussed above. Section 4 tests the algorithms on two dynamical systems: the classic Lorenz system (*Lorenz* 1963) and a 40-dimensional system mimicking flow around a latitude circle (*Lorenz* 1996). Although the 40-dimensional system is small compared to numerical weather prediction models, it is easily large enough to challenge existing non-Gaussian techniques. Section 5 discusses strengths and limitations of the new methods.

2 Background and notation

2.1 The update/forecast cycle

We will focus on the data assimilation and forecasting problem associated with numerical weather prediction. In this problem, the goal is to modify the forecast pdf for the system once new data is available. The modified pdf is then propagated forward using knowledge of the system dynamics to give a new forecast and subsequently updated again when new observations become available. This process, which we will refer to as a filtering algorithm, consists of two distinct steps: an *update* or data-assimilation step, and a *forecast* step. As mentioned, both the update and forecast steps are challenging to implement in a geophysical context.

In the update step, a forecast pdf is updated given a new set of observations via Bayes theorem. The best known filtering algorithm is in the context of Gaussian distributions and linear system dynamics where the update pdf is described by the Kalman filter recursion (*Kalman* 1960). Unfortunately, analytic solutions to the update step can only be derived for a few special cases, and working explicitly with the state pdf is therefore not practical. As an alternative, various computational techniques have been developed in the last two decades to address more complex problems (see, e.g., *Gilks et al.* 1996). However, as the computational requirements increase rapidly with dimension, calculation of the update pdf can only be envisioned for systems with a small number of degrees of freedom. Furthermore, for problems involving sequential estimation (and propagation), these methods have proven inefficient (*Doucet et al.* 2001).

In the forecast step, a probabilistic forecast is made by evolving the updated pdf forward in time. This is done using known or approximate dynamical laws, typically specified by stochastic differential equations. A statistician may view the forecast step as a transformation-of-variables problem: given a pdf for (the random variable) $\mathbf{X}$, and a transformation $G(\cdot)$, representing the time evolution of a dynamic system, find the pdf of the transformation $G(\mathbf{X})$. Not surprisingly, analytic solutions in the forecast step are rarely available and direct calculation of the forecast pdf in many dimensions is computationally prohibitive.

Some of the difficulties of implementation described above can be surmounted by approximating the pdf with a discrete sample, which we will refer to throughout this paper as an *ensemble*. Given an ensemble sampled from the updated pdf, the forecast ensemble is derived by propagating each ensemble member using $G(\cdot)$ (*Leith* 1974). By elementary probability rules, this yields a sample from the forecast pdf. In this article we will assume that G is known perfectly, although some model errors could be represented by a stochastic process and incorporated into this framework (*Jazwinski* 1970). Updating the forecast ensemble given observations (that is, constructing a sample from the updated pdf) is considerably more complex, especially for non-Gaussian pdfs, and is the focus of this article. The update step for the EnsKF is reviewed in section 2.4, while section 3 presents our algorithms for non-Gaussian pdfs based on Gaussian mixtures.

Outside the geosciences, there is also a rich statistical literature on particle filters (PF) and their variants (*Doucet et al.* 2001). PF are a set of Monte-Carlo techniques for approximating the fully nonlinear, Bayesian update. In their simplest form they represent the forecast pdf with a ensemble but may also carry importance weights attached to each member member, or "particle." The algorithms we consider, in contrast, use ensembles of equally weighted members that can be manipulated as if they were a random sample. PF applications have focused on low dimensional systems and system dynamics that has a random component. In this paper we consider deterministic but chaotic systems, a reasonable framework for problems associated with atmospheric and oceanic data assimilation.

2.2 Notation and the Kalman filter

To set notation, let $\mathbf{x}_t$ denote the state vector of the system at time t and let $\mathbf{y}_t$ be a new vector of observations. Initial knowledge of the system is given by the conditional forecast distribution $p(\mathbf{x}_t|\mathbf{Y}_{t-1})$, where $\mathbf{Y}_{t-1}$ denotes all past data up to and including time $t-1$. The *update* step combines the forecast distribution and the new data, giving the posterior distribution $p(\mathbf{x}_t|\mathbf{Y}_t)$. Calculation of $p(\mathbf{x}_t|\mathbf{Y}_t)$ is an application of Bayes Theorem.

We now outline the standard Kalman filter update, since it forms the basis for all subsequent techniques here. Suppose that a linear observation operator, $\mathbf{H}_t$, relates the unobserved state, $\mathbf{x}_t$, to the data, $\mathbf{y}_t$:

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{e}_t, \tag{1}$$

where $\mathbf{e}_t \sim \mathrm{N}(\mathbf{0}, \mathbf{R})$. Without loss of generality, $\mathbf{R}$ may be assumed diagonal—one can always transform (1) to an observation equation with *i.i.d.* errors by multiplying through by $\mathbf{R}^{-1/2}$.

If we assume that $p(\mathbf{x}_t|\mathbf{Y}_{t-1}) \sim \mathrm{N}(\boldsymbol{\mu}_t^f, \mathbf{P}_t^f)$, then a straightforward application of Bayes theorem yields

$$p(\mathbf{x}_t|\mathbf{Y}_t) = \mathrm{N}(\boldsymbol{\mu}_t^u, \mathbf{P}_t^u) \tag{2}$$

where

$$\boldsymbol{\mu}_t^u = \boldsymbol{\mu}_t^f + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\boldsymbol{\mu}_t^f) \tag{3}$$

and

$$\mathbf{P}_t^u = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_t^f \tag{4}$$

Here, $\mathbf{K}_t$ denotes the Kalman gain matrix and is given by

$$\mathbf{K}_t = \mathbf{P}_t^f\mathbf{H}_t'(\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t' + \mathbf{R})^{-\mathbf{1}}, \tag{5}$$

where a prime superscript denotes matrix transpose.

For completeness, we note here that if the system dynamics are linear then the forecast distribution will again be multivariate normal and the covariance and mean have simple closed forms. However, this aspect will not be used on our discussion as in all subsequent methods we approximate the forecast distribution through the propagation of an ensemble. The creation of the ensemble in the update step is described in the next section.

2.3 Ensemble Kalman filter update

The EnsKF, which has been recently advanced in the geosciences (*Evensen* 1994, *Houtekamer and Mitchell* 1998), is a Monte-Carlo based approach to forecasting and data assimilation. The continuous forecast and update distributions are approximated by a discrete distribution of ensemble members where each member is a point mass assigned equal probability. (The EnsKF may thus be considered a special case of a particle filter.)

To anchor our extensions to the EnsKF, we first describe one of its standard implementations. Let $\{\mathbf{x}_{t,i}^f\}$ for $1 \leq i \leq m$ denote an m member ensemble representing the distribution $p(\mathbf{x}_t|\mathbf{Y}_{t-1})$. The update step consists of applying an approximate form of the Kalman filter update (2) to each member. Specifically, the algorithm estimates an approximate gain matrix, $\tilde{\mathbf{K}}_t$ using sample covariances based on the ensemble:

$$\begin{aligned}
\mathbf{P}_t^f\mathbf{H}' &\approx (m-1)^{-1}\sum_{i=1}^{m}(\mathbf{x}_{t,i}^f - \bar{\mathbf{x}}_t)[\mathbf{H}(\mathbf{x}_{t,i}^f - \bar{\mathbf{x}}_t)]', && (6)\\
\mathbf{H}\mathbf{P}_t^f\mathbf{H}' &\approx (m-1)^{-1}\sum_{i=1}^{m}[\mathbf{H}(\mathbf{x}_{t,i}^f - \bar{\mathbf{x}}_t)][\mathbf{H}(\mathbf{x}_{t,i}^f - \bar{\mathbf{x}}_t)]', && (7)
\end{aligned}$$

where $\bar{\mathbf{x}}_t$ denotes the forecast ensemble mean. Each member is then updated according to

$$\mathbf{x}^u_{t,i} = \mathbf{x}^f_{t,i} + \tilde{\mathbf{K}}_t \left(\mathbf{y}_t + \epsilon_{t,i} - \mathbf{H}_t \mathbf{x}^f_{t,i} \right), \tag{8}$$

where $\{\epsilon_{t,i}\}$ for $1 \le i \le m$ is a sample from $\mathrm{N}(0, \mathbf{R})$. If $\{\mathbf{x}^f_{t,i}\}$ was sampled from $\mathrm{N}(\boldsymbol{\mu}^f_t, \mathbf{P}^f_t)$, then the EnsKF update converges to that of the KF for large m and linear algebra can be used to verify that $\mathbf{x}^u_{t,i}$ is a sample from the update distribution given in (2) (*Houtekamer and Mitchell* 1998, *Burgers et al.* 1998).

Although there are other, standard ways to sample from the posterior distribution (2), the scheme (8) is applicable in high dimensions since it does not require the explicit (and computationally expensive) covariance recursion defined in (4) or other direct manipulation of the covariance matrices. Instead, the algorithm relies on being able to multiply the Kalman gain matrix by arbitrary vectors and in this way the large matrices are never explicitly constructed or stored.

One further assumption is necessary to make the EnsKF feasible and effective in high-dimensional problems. When the domain of interest encompasses many characteristic spatial scales of the physical system, it is often the case that the covariance of two elements of the state vector will be nearly zero when the physical locations corresponding to those elements are separated by a sufficient distance. Many or most of the elements of the sample covariance matrix are then expected to be small. In most implementations of the EnsKF, covariances at sufficient separation are therefore assumed to decrease smoothly to zero at a certain distance; this increases the computational efficiency of the update and decreases the effects of random error arising from working with a sample covariance (*Houtekamer and Mitchell* 2001, *Hamill et al.* 2001). We refer to this method as *tapering* the sample covariance matrix. Statisticians can understand this modification as a specific way of shrinking the sample covariance matrix elements toward zero for large separation distances but still retaining the positive definite character of the matrix. Delineating the statistical properties that are produced through tapering remains an open question.

2.4 Updating a Gaussian mixture

The Kalman filter update is easily extended to a mixture of Gaussian distributions (*Anspach and Sorensen* 1972). Suppose that $p(\mathbf{x}_t|\mathbf{Y}_{t-1})$ is a mixture of L multivariate normal distributions:

$$\sum_{l=1}^{L} \pi_{t,l}^{f} \mathrm{N}(\boldsymbol{\mu}_{t,l}^{f}, \mathbf{P}_{t,l}^{f}).$$

With the observation equation as defined above, the updated distribution is again a mixture of L multivariate normal distributions:

$$\sum_{l=1}^{L} \pi_{t,l}^{u} \mathrm{N}(\boldsymbol{\mu}_{t,l}^{u}, \mathbf{P}_{t,l}^{u}), \tag{9}$$

where the mean and covariance matrix of each component of the mixture are updated in an analogous manner as in the single Gaussian case. Specifically, one determines $\boldsymbol{\mu}_{t,l}^{u}$ and $\mathbf{P}_{t,l}^{u}$ by substituting $\boldsymbol{\mu}_{t,l}^{f}$ for $\boldsymbol{\mu}_{t}^{f}$ and $\mathbf{P}_{t,l}^{f}$ for $\mathbf{P}_{t}^{f}$ in (3) and (4). The mixing probabilities are updated by calculating

$$\pi_{t,l}^{u} = \frac{\pi_{t,l}^{f} w_l}{\sum_{k=1}^{L} \pi_{t,k}^{f} w_k}, \tag{10}$$

with w_l given by

$$|(\mathbf{H}_t \mathbf{P}_{t,l}^{f} \mathbf{H}_t' + \mathbf{R})|^{-.5} \exp\left[-(1/2)(\mathbf{y}_t - \mathbf{H}_t \boldsymbol{\mu}_{t,l}^{f})'(\mathbf{H}_t \mathbf{P}_{t,l}^{f} \mathbf{H}_t' + \mathbf{R})^{-1}(\mathbf{y}_t - \mathbf{H}_t \boldsymbol{\mu}_{t,l}^{f})\right].$$

3 Ensemble mixture filters

This section presents two non-Gaussian algorithms for the update step. Like the EnsKF, each begins with an ensemble that is a sample from the prior forecast distribution and updates that ensemble to produce (approximately) a sample from the posterior distribution given observations. [The forecast step, as discussed in section 2, would consist of simply propagating each ensemble member to the next observation time using the forecast model.] Unlike the EnsKF, these algorithms are based on Gaussian mixtures.

The first scheme below, which we find to be effective in low dimensions, chooses the mixture centers randomly from the forecast ensemble, and then estimates the covariance for each component of the mixture using ensemble members that are "close" in the state space to the mixture centers. The second scheme extends the first to high-dimensional systems by assimilating observations sequentially (one at a time or in blocks) and updating only the portion of the state vector that is physically local to the observation location.

3.1 Mixture covariances based on local state space information

We first extend the EnsKF to a mixture filter for low-dimensional systems. The basic idea is to update each component of the mixture using "local" sample statistics, that is, from ensemble members that are close in state space to the mixture center. This filter will be termed the mixture ensemble filter, or XEnsF.

The update begins with a forecast ensemble $\{\mathbf{x}^f_{t,i},\ i = 1, \dots, m\}$. To derive a mixture from this ensemble, we choose at random L ensemble members to be the centers of the mixture components; the first L members may be taken as centers for convenience, since there is no preferred order among the ensemble members. Next, we identify from the ensemble the N nearest neighbors to each center. (All our calculations use the Euclidean norm to define distance in state space, though other norms could be employed.) The covariance associated with each center $\mathbf{x}^f_{t,i}$ is then given by $\mathbf{P}_i$, the sample covariance for the N nearest neighbors of $\mathbf{x}^f_{t,i}$. Finally, the algorithm must produce an updated ensemble that is consistent with the update of the continuous mixture through (9). Denoting by $\mathbf{K}_{t,i}$ the Kalman gain matrix with $\mathbf{P}_i$ substituted for $\mathbf{P}^f_t$, the complete update step is as follows:

Mixture Ensemble filter

- Given $\{\mathbf{x}^f_{t,i},\ i = 1, \dots, m\}$
- Update mixing probabilities. For l in $[1, L]$:
  - Find N nearest neighbors to $\mathbf{x}^f_{t,i}$ in state space.
  - Calculate π^u_l from (10) using $\mathbf{P}_l$ based on the nearest neighbors.
- Update ensemble. For j in $[1, m]$:
  - Choose a random index $I \in [1, L]$ where $P(I = i) = \pi^u_i$.
  - Choose one of N nearest neighbors of $\mathbf{x}^f_{t,I}$, each with probability $1/N$; denote this member by $\mathbf{x}^*$.
  - Update according to (8) using nearest neighbors:

    $$\mathbf{x}^u_{t,j} = \mathbf{x}^* + \mathbf{K}_{t,I}(\mathbf{y} + \mathbf{e}^* - \mathbf{H}\mathbf{x}^*),$$

    where e^* is drawn from $\mathrm{N}(0, \mathbf{R})$.

While we have not explored tuning these parameters, the XEnsF requires the choice of the ensemble size m, the number of nearest neighbors N and the number of centers L. For future reference we will refer to this dependence as XEnsF(m,N,L).

Note that the sampling from the updated mixture distribution in the XEnsF is a modest elaboration from the EnsKF. To draw a sample from (9), the algorithm first samples an integer from 1 to L from the multinomial distribution with probabilities given by $\pi^u_{t,l}$. Denoting this random index by I, the algorithm then samples from the Ith component of the mixture using (8) and the nearest neighbors of $\mathbf{x}^f_{t,I}$. It is straightforward to extend the arguments of *Houtekamer and Mitchell* (1998) and *Burgers et al.* (1998) to show that this produces a sample from (9) for $m \to \infty$. The use of this sampling scheme, which as noted in section 2.3 does not require the manipulation of large covariance matrices, is one crucial step toward implementing mixture filters in high dimensions.

Simulation results in the next section will demonstrate that the XEnsF outperforms the EnsKF for a three-dimensional nonlinear system. Although successful in low-dimensional systems, we expect the XEnsF to break down when applied to high-dimensional systems owing to the inherent difficulties of estimating high-dimensional systems. This difficulty is manifest in our experiments by the tendency for the XEnsF update to weight a single center heavily, so that the ensemble collapses on to a single solution after a few forecast-update cycles.

3.2 Local-Local Ensemble filter

In order to address the problems of the XEnsF in high dimensions, we assume that observations only influence the update of state variables that are nearby *in physical space*. This allows the update step to be decomposed into a sequence of lower-dimensional updates that are tractable with the XEnsF. The resulting algorithm then consists of repeated applications of the XEnsF to physically local subsets of the state vector.

To set the stage we first note a well known sequential property associated with the update step. If observations are independent conditional on the state vector then the posterior can be updated sequentially taking each observation in turn. This sequential process will yield the same posterior pdf as what one would obtain using a single and simultaneous update of the full observation vector and of course will not depend on the order that observations are used. This result is a consequence of the factoring of the joint distribution of observations based on conditional independence and does not require the assumption that pdf be Gaussian or a mixture of Gaussians.

We will assume that each component of the state vector is associated with a location and that covariances among the components of $\mathbf{x}$ are localized in the sense that they are close to zero when components are separated by large distances. In addition, we assume that the observations are also localized, by which we mean that each row of $\mathbf{H}$ has a limited number of nonzero elements and those elements correspond to state variables in some region of limited spatial extent. Examining the form of the Kalman gain when the observation is a scalar one notes that a component of $\mathbf{x}_t^f$ will only be changed by a new observation if the corresponding row of $\mathbf{P}_t^f\mathbf{H}$ is nonzero. This leads to the intuition that the update of the state vector based on a single new observation should only affect a subspace of $\mathbf{x}$. We will refer to this portion of the state vector as the *observation neighborhood*. Because of our assumption that covariances (and H) are localized, the observation neighborhood will be of low dimension. We then propose to update using the XEnsF within this observation neighborhood.

The resulting algorithm combines the use of local state-space information in the XEnsF with localization in physical space, and will be denoted the local-local ensemble filter, or LLEnsF. As mentioned above one can choose to update observations sequentially and so the LLEnsF will have an added outer loop over observations. For the kth observation, let $\mathbf{x}^{[k]}$ denote a reduced state vector consisting of only those components of $\mathbf{x}$ contained in the kth observation neighborhood. With this notation, and recalling the dependence of the XEnsF on the tuning parameters m, N, and L, the update step of the LLEnsF may be summarized as:

- Given $\{\mathbf{x}_{t,i}^f,\ i = 1, \ldots, m\}$.
- Loop over observations. For k in $[1, n]$:
  - Apply XEnsF(m, N, L) to update elements of $\mathbf{x}^{[k]}$.

Note that the size of the observation neighborhood (its radius, for example) must be chosen for the LLEnsF, in addition to m, N, and L.

This algorithm has two important features. First, the mixture filter suitable for non-Gaussian distributions is applied repeatedly to low dimensional components of the state vector. This avoids a single high dimensional update. Secondly, LLEnsF includes the standard ensemble Kalman filter as a special case. This will happen when $L = 1$, $N = m$ and the observation neighborhood includes all components of the state vector.

4 Simulations

We evaluate the filter methods described in the previous section on two nonlinear dynamical systems. Both are sensitive to initial conditions, leading to unstable solutions and error growth. The first system, here denoted L3, is the classic three-dimensional system of *Lorenz* (1963). The second system, denoted L40, consists of 40 state variables that correspond to locations on a latitude circle, so that the spatial localizations discussed previously can be applied, and includes quadratic nonlinearity designed to mimic advection (*Lorenz* 1996). Equations defining the two systems are given in the appendix. The XEnsF algorithm is evaluated on L3 and the LLEnsF is evaluated on L40.

4.1 Simulations for L3

L3 has been studied extensively in the context of data assimilation (see, e.g. *Miller et al.* 1994, *Evensen* 1997, and *Anderson and Anderson* 1999). As can be seen in Figure 1, the system attractor has two lobes or orbits connected near the origin. The trajectories of the system in this saddle region are particularly sensitive to perturbations. Hence, slight perturbations can alter the subsequent path from one lobe to the other. Figure 1 also depicts the error growth exhibited in the system. As sample ensemble points pass through the saddle they rapidly disperse across the two attractor lobes. Thus, even on fairly short time scales the dynamics of this system leads to distinctly non-Gaussian forecast distributions.

To evaluate the effects of the non-linear dynamics on filter performance, forecast lead time δ_t is varied across four levels: $\delta_t = .1, .25, .5, 1$. These lead times provide a range of conditions from approximately linear to fully nonlinear dynamics of the forecast errors. The numerical experiments also vary the number of mixture components ($L = 10, 40$) and ensemble members ($m = 60, 90, 110, 140$), while the number of nearest neighbors was fixed at $N = 25$. The observation operator is taken to be the identity matrix, i.e., $\mathbf{H}_t = \mathbf{I}$, and the observation errors are independent and normally distributed with a variance of 4 ($R_{jj} = 4$). Thus, an informative baseline for the root (posterior) mean squared prediction error is 2 ($\sqrt{R_{jj}}$), the error incurred simply by using the observation vector as a naive update of the state.

Table 1 reports simulation results for assimilating observations over 10000 assimilation cycles, each separated by a time interval of δ_t, using the XEnsF and standard EnsKF. At each observation time the root mean squared error (RMSE) between the sample posterior mean and the true state of the system is calculated for each filter. The prediction error is measured as the median RMSE across all time points. As can be seen from Table 1, the mixture EnsF performs better than the single Gaussian EnsKF for forecast lead times greater than $\delta_t > .1$, with an overall improvement of approximately 20-30% in median RMSE. The improvement is more marked for larger forecast lead times, consistent with the expected increase of nonlinearity and non-Gaussianity as δ_t increases.

δ_t	XensF				EnsKF	
	$L, m = 10, 60$	10,110	40,90	40,140	$m = 40$	120
.1	.59	.60	.48	.47	.38	.37
.25	.72	.71	.49	.52	.72	.69
.5	.93	.90	.69	.69	1.05	1.05
1	1.19	1.14	.93	.90	1.37	1.37

Table 1: Simulation results for the L3 system in terms of median RMSE for the posterior mean. Results are estimated for 10000 assimilation cycles.

The median RMSEs reported in Table 1 are a summary of filter performance across the whole attractor. As an example of the effects of non-Gaussian forecasts on filter performance we took the 250 assimilated states from the EnsKF that were located closest in the saddle region of the attractor. We then performed one forecast cycle with $\delta_t = .5$ and used both the XEnsF as well as the EnsKF to assimilate new data. The median RMSE for the XEnsF with $L = 100$ and $m = 500$ was .73, while the EnsKF with $m = 400$ yielded a median RMSE of 1.64 for a resulting improvement of over 50%. Thus, for forecasts that are distinctly non-gaussian the XEnsF significantly outperforms the EnsKF.

4.2 Simulations for L40

Simulations for L40 use forecasts of length $\delta_t = .4$ and take observations of every other state variable. Thus, at each assimilation cycle we have available the following set of observations: $\{y_1 = x_1 + \epsilon_1, y_2 = x_3 + \epsilon_2, \ldots, y_{20} = x_{39} + \epsilon_{20}\}$. The observational errors are independent and normally distributed with variance .5. These settings are chosen to produce non-Gaussian behavior in the forecast ensembles.

As a baseline of performance, the EnsKF was applied with an ensemble size of $m = 400$. A tapering function that down-weighted the sample covariances between spatially distant state components was used at each assimilation step. The tapering function was defined by (4.10) of *Gaspari* *and Cohn* (2001), with their parameter c chosen such that the covariance of state variables separated by 20 index points or more (e.g., x_1 and x_{21}) is set to zero. Each of the 20 observations were assimilated serially at every time step. Based on posterior mean estimates at every assimilation cycle, the EnsKF produced a time averaged RMSE of .972 across 2000 assimilation steps. The sample variance of the RMSE was $s^2 = .125$, and the median RMSE was .882. The forecast distributions produced by the EnsKF appear to be noticeably non-Gaussian, so there is clearly some potential to improve on the EnsKF.

To provide some quantification and evidence of the non-Gaussian structure of the forecasts produced by the EnsKF we will focus on a 3-dimensional subset of the state-vector involving variables $\{x_1, x_2, x_3\}$. (Since L40 is invariant to translation, any three adjacent state variables will have the same statistical properties.) Letting $\mathbf{z}_{i,t}$ denote the deviation of the ith ensemble member from the mean at time t in the space of $\{x_1, x_2, x_3\}$, we calculate $d_{i,t} = \mathbf{z}'_{i,t}\hat{\Sigma}^{-1}\mathbf{z}_{i,t}$, for $i = 1, 2, \ldots, m$. Here, $\hat{\Sigma}$ denotes the sample covariance of $\mathbf{z}_{i,t}$ (with respect to the subscript i). If the ensembles of $\{x_1, x_2, x_3\}$ follow a multivariate normal distribution, then $d_{i,t}$ will approximately follow a chi-squared distribution with 3 degrees of freedom. Applying the Kolmogorov-Smirnov (KS) test (*Kolmogorov* 1933) at each assimilation cycle, i.e., for $t = 1, 2, \ldots, 2000$, the hypothesis of normality was rejected in 1896 cases at the .05 critical level. The mean of the KS test-statistic was .139, well above the .001 level of significance of .094. Hence, there is strong evidence of frequent departures from multivariate normality. To provide a visual example of the structure of the non-Gaussian ensembles at a given time point, Figure 2 depicts bi-variate plots of $\{x_1, x_2, x_3\}$. The lower right plot in Figure 2 shows a histogram of the KS test-statistics calculated from the 2000 forecasts produced by the EnsKF.

As can be seen in Figure 2, the relationship between the ensemble members of x_1 and x_2 follows a non-linear pattern and the joint distribution of $\{x_1, x_2\}$ is distinctively non-Gaussian. To quantify the degree of non-linearity between x_1 and x_2 we performed an F-test of linearity by regressing the ensembles of x_2 on those of x_1 for the 2000 forecasts. At a .05 critical level, the F-test rejects the hypothesis of linearity between x_1 and x_2 in 83.5% of the 2000 cases. Clearly, the relationship between x_1 and x_2 is decidedly non-linear in a majority of forecast ensembles.

Before applying the LLEnsF as described in section 3 to L40, we performed an intermediate experiment to gauge the potential for improvement relative to the EnsKF, given the non-Gaussian properties of the ensembles. Using the output (that is, the state, observations, and forecast ensembles) from the baseline EnsKF example, the XensF was applied to the sub vector $\{x_1, x_2, x_3\}$ to assimilate y_1 and y_2 at each assimilation time. The quality of the update produced by the XensF was then compared to that of the EnsKF. (Note that the results of XensF were not used to modify the ensemble used in the subsequent forecast and update step.)

The posterior mean RMSE for the EnsKF across the 2000 assimilation points was .827 ($s^2 = .383$). Based on L=400, N=40 and m=400, the XEnsF improved this by roughly 8%, yielding an RMSE of .768 ($s^2 = .352$). The improvement is statistically significant ($p < .001$). Thus, the XEnsF provides, at least locally, a better estimate of the true state of the system.

Next we apply the LLEnsF to the same sequence of observations as in the baseline EnsKF example, and define the observation neighborhoods to consist of three adjoining state variables. Thus, at each assimilation cycle, the scalar observation y_j updates the observation neighborhood $x^{[k]} = (x_{k-1 \mod 40}, x_k, x_{k+1 \mod 40})$, where $k = 2j - 1$. Using these observation neighborhoods, the LLensF was found to be a stable filter but did not perform as well as the EnsKF.

There are at least two reasons the LLensF does not perform as well as the EnsKF in these simulations. The first is that, by assumption, observations affect the update of only three state variables in the LLEnsF, while in the EnsKF each scalar observation can provide information about the entire state vector. The second reason is that the LLEnsF does not guarantee that adjoining observation neighborhoods are sampled in manner that respects the prior relationships among state variables in different neighborhoods. For example, posterior samples produced in the observation neighborhood $\mathbf{x}^{[1]}$ by assimilating y_1 may be not be "smooth" with those produced in the observation neighborhood $\mathbf{x}^{[3]}$ by assimilation of y_2. Thus, it is possible that the LLensF yields posterior sample states that are disjointed between observation neighborhoods.

These limitations of the LLEnsF suggest a hybrid ensemble filter that combines aspects of the LLEnsF and EnsKF. Like both the LLEnsF and the EnsKF, this hybrid processes observations sequentially, but for each observation it calculates two updated ensembles, one from the LLEnsF and another from the EnsKF. The results below use an observation neighborhood of three adjacent state-vector components and XEnsF(400, 400, 40) in the LLEnsF update, and an EnsKF update as in the baseline case. The two updated ensembles are then combined in a simple way: within the LLEnsF observation neighborhood, the EnsKF ensemble is adjusted so that its mean matches the sample mean from the LLEnsF update. In essence, the hybrid ensemble takes its mean from the LLEnsF where that is available (since we know that the LLEnsF update produces smaller RMSE within the observation neighborhood) and uses the EnsKF ensemble otherwise, including outside the LLEnsF observation neighborhood. The use of the EnsKF ensemble outside the observation neighborhood also has the effect of providing some continuity between variables in different neighborhoods.

The hybrid filter arguably makes only modest use of the information from the LLEnsF. Nevertheless, the hybrid filter improves noticeably on the performance of both the LLEnsF and EnsKF. The time-averaged RMSE of the hybrid filter for the 2000 assimilation cycles was .917 ($s^2 = .100$), with median RMSE of .848. The improvement in the posterior mean estimate compared to that produced by the EnsKF is statistically significant ($p < .001$), and corresponds to a 5.7% overall error decrease. Although the LLEnsF and the hybrid filter are not mature assimilation algorithms, these results demonstrate their potential for non-Gaussian systems with many degrees of freedom.

5 Discussion

The results in this work are a proof-of-concept for the use of a mixture filter for data assimilation in systems with strongly non-linear dynamics. The improvement in mean squared error prediction for the low-order model (L3) is significant. Our results also represent a step toward implementation of such techniques in higher dimensions. In particular, an investigation of a more complicated system (L40) suggests that the system often assumes states that are better described and estimated using non-Gaussian distributions.

An important feature of the XEnsF is the use of local covariances based on nearest neighbors. The local covariances adapt to local linear properties of the attractor and so provide a more accurate representation of the ensemble distribution. Previous work (*Anderson and Anderson* 1999) used scaled versions of the full ensemble covariance around each center in the mixture, and so cannot adapt as easily to local structure in the forecast distribution. One important issue in the mixture approach is the number of nearest neighbors and the localization of the covariance about the mixture center—a large number of nearest neighbors may give a more stable estimate of the covariance but may be too spread out to reflect salient local features. This trade off seems to be balanced in the L3 implementation and is also tuned to some degree for the experiments with L40.

The LLEnsF extends the XensF beyond low-dimensional systems by restricting the update step to low-dimensional, spatially local subspaces of the state vector, so that it is not subject to the statistical problems associated with high-dimensional distributions. The numerical results in this work confirm that there are three-dimensional subspaces of the state space where the mixture takes advantage of non-Gaussian structures. A straight forward implementation of LLEnsF is inferior to the EnsKF because it does not adequately blend the updates in the observation neighborhood with components of the state vector that are unchanged. This is motivation for our hybrid approach which is conservative in how non-Gaussian updates modify the ensemble members. The fact that it can perform better than EnsKF and remain stable is very encouraging.

The local nature of the observation neighborhood can result in updates to the state vector that are discontinuous but locally can be more accurate. In contrast, the ensemble Kalman filter provides a smooth modification to the state vector in the update step but can not handle non-Gaussian features. An important extension of these ideas to develop a statistical framework that seamlessly combines the strength of both of these filters. One benefit of this synthesis would be the ability to quantify uncertainty in the state of the system taking into account multimodality or skewness of the forecast distribution.

A Appendix

The L3 model (*Lorenz* 1963) is defined by three differential equations:

$$\begin{aligned}
\dot{x} &= -\sigma(x_t + y_t), \\
\dot{y} &= rx_t - y_t - x_t y_t, \\
\dot{z} &= x_t y_t - bz_t,
\end{aligned}$$

where the dot represents a derivative with respect to time. The model parameters are set as follows: $\sigma = 10, r = 28,$ and b $= \frac{8}{3}$.

The L40 model (*Lorenz* 1996) is defined by the differential equations

$$\dot{x}_{t,i} = (x_{t,(i+1 \mod k)} - x_{t,(i-2 \mod k)})x_{t,(i-1 \mod k)} - x_{t,i} + F.$$

Here, $k = 40$ and $F = 8$.

Both systems are propagated using a first order Euler method with a time step of .001. This simple numerical scheme facilitates rapid propagation of a large number of ensembles.

References

Alspach, D. L., and H. W. Sorenson, 1972: Nonlinear Bayesian estimation using Gaussian sum approximation. *IEEE Trans. on Automatic Control*, **17** 439–448.

Anderson, J. L.,, and S. L. Anderson, 1999: A Monte-Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.,***127** 2741–2758.

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

Chen, R., and J. S. Liu, 2000: Mixture Kalman filters. *J. R. Statist. Soc.* B, **62** 493–508.

Doucet, A., N. Freitas, and N. Gordon, 2001: *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

Epstein, E. S., 1969: The role of initial uncertainties in prediction. *J. Appl. Meteor.*, **8** 190–198.

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), 10,143–10,162.

Evensen, G., 1997: Advanced data assimilation for strongly nonlinear systems. *Mon. Wea. Rev.,***125** 1342–1352.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.,***125** 723–757.

Gilks, W. R., S. Richardson, and D.J. Spiegelhalter, 1996:. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, Suffolk.

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

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

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,***129** 123–137.

Kolmogorov, A., 1933: Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari.*, **4** 83–91.

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

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

Leith, C. E., 1971: Atmospheric predictability and two-dimensional turbulence. *J. Atmos. Sci.,***28** 145–161.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,***20** 130–148.

Lorenz, E. N., 1996: Predictability: A problem partially solved. in *Proceedings Seminar on Predictability* ECMWF, Reading Berkshire, UK, 1:1–18

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

Schlatter, T. W., G. W. Branstator and L. G. Thiel, 1976: Testing a global multivariate statistical objective analysis scheme with observed data. *Mon. Wea. Rev.,***104** 765–783.

Silverman, B. W., 1986. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 175 pp.

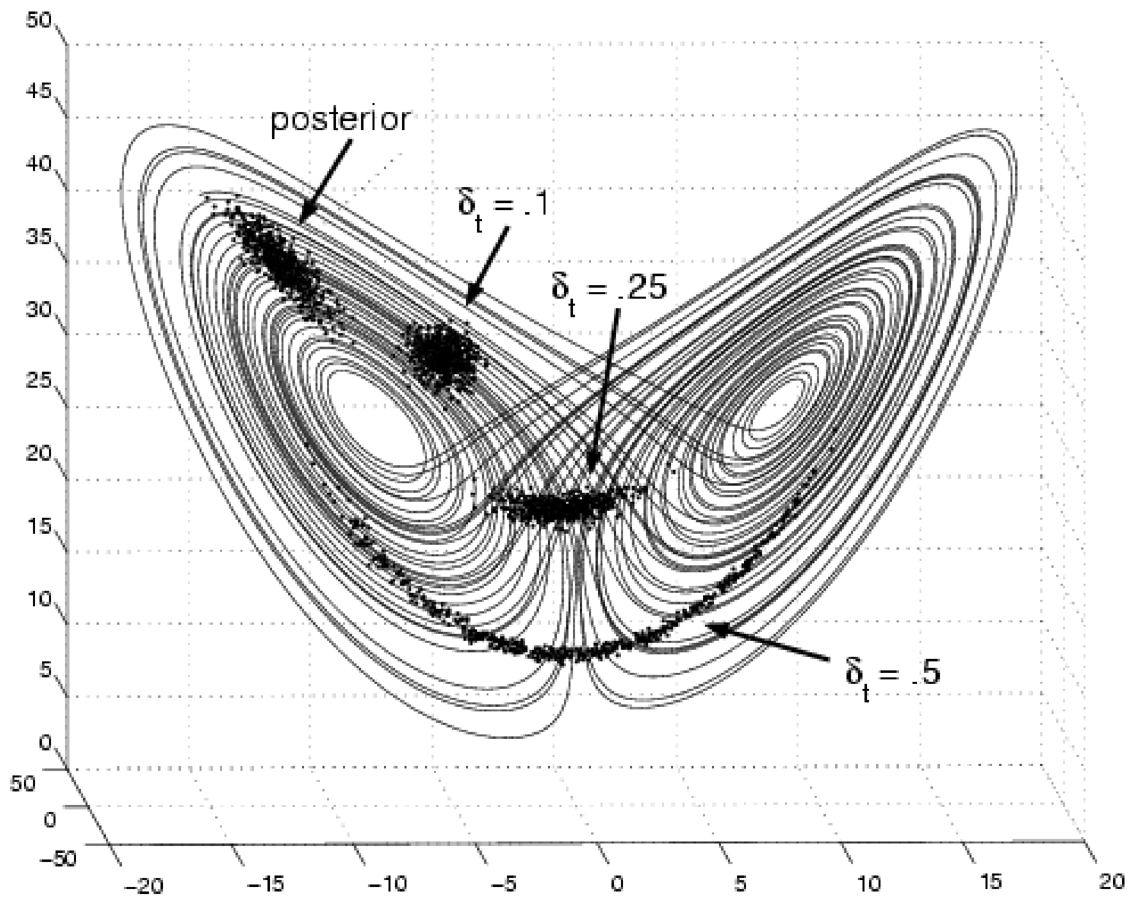

Figure 1: The Lorenz attractor. Non-Gaussian structures appear quickly in this system: 400 points in the upper left-hand corner are sampled from a Gaussian distribution, and have been propagated .1, .25, and .5 time units, respectively.

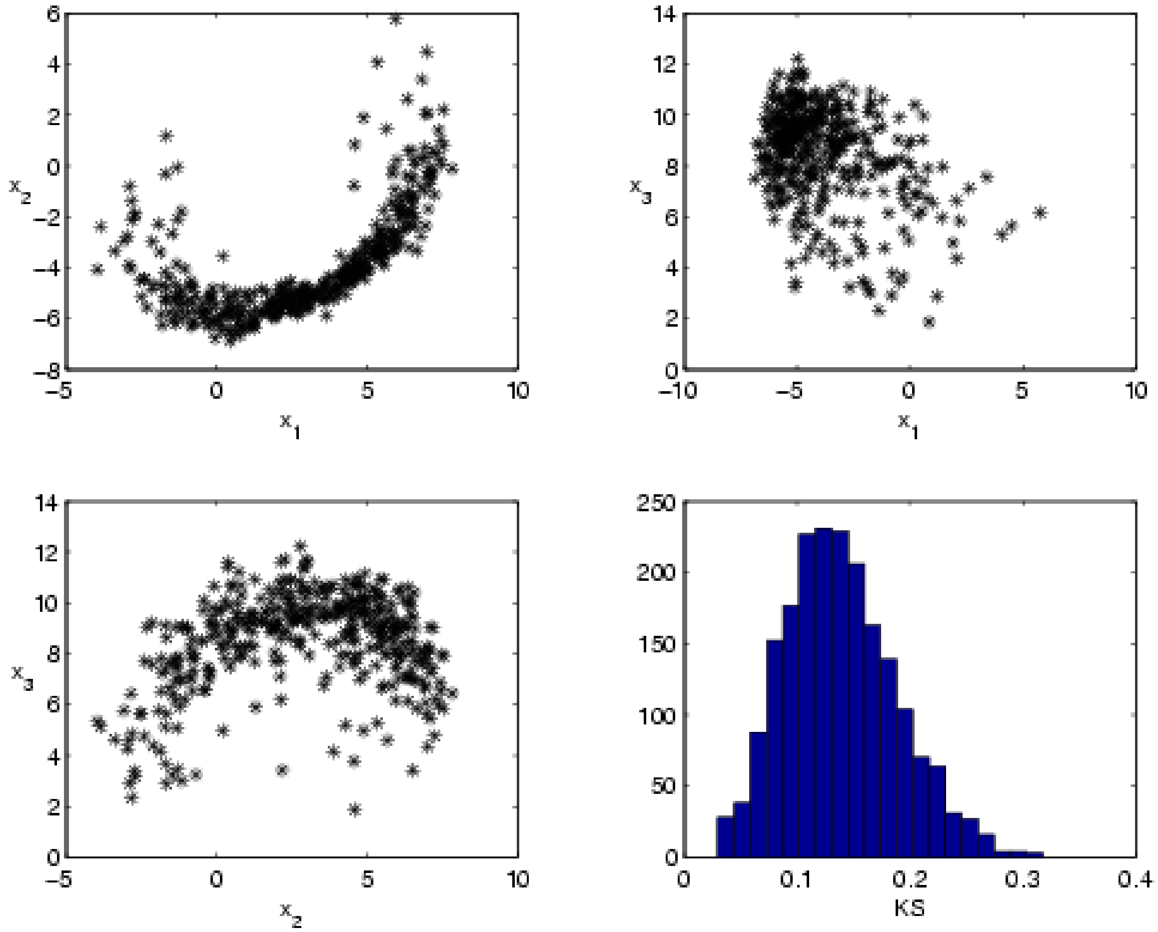

Figure 2: Forecast ensemble members for $\{x_1, x_2, x_3\}$ at a given assimilation time. Bivariate scatterplots depict local non-Gaussian behavior. The ensemble shown produced a KS statistic of .134 ($p. < 001$), while a test of linearity between x_1 and x_2 produced an F-statistic of 249.1 ($p < .001$). The lower right plot is a histogram of the KS test statistics over 2000 assimilation cycles.