Toward a nonlinear ensemble filter for high-dimensional systems

Thomas Bengtsson, Chris Snyder, and Doug Nychka

Geophysical Statistics Project, Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, Colorado, USA

Received 30 August 2002; revised 30 September 2003; accepted 10 October 2003; published 24 December 2003.

[1] Many geophysical problems are characterized by high-dimensional, nonlinear systems and pose difficult challenges for real-time data assimilation (updating) and forecasting. The present work builds on the ensemble Kalman filter (EnsKF), with the goal of producing ensemble filtering techniques applicable to non-Gaussian densities and high-dimensional systems. Three filtering algorithms, based on representing the prior density as a Gaussian mixture, are presented. The first, referred to as a mixture ensemble Kalman filter (XEnsF), models local covariance structures adaptively using nearest neighbors. The XEnsF is effective in a three-dimensional system, but the required ensemble grows rapidly with the dimension and, even in a 40-dimensional system, we find the XEnsF to be unstable and inferior to the EnsKF for all computationally feasible ensemble sizes. 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 prior forecasts in a 40-dimensional system, the LLEnsF update produces more accurate state estimates than the EnsKF if the forecast distributions are sufficiently non-Gaussian. Cycling the LLEnsF for long times, however, produces results inferior to the EnsKF because the LLEnsF ignores spatial continuity or smoothness between local state estimates. To address this weakness of the LLEnsF, we consider ways of enforcing spatial smoothness by conditioning the local updates on the prior estimates outside the localization in physical space. These considerations yield a third algorithm, which is a hybrid of the LLEnsF and the EnsKF. The hybrid uses information from the EnsKF to ensure spatial continuity of local updates and outperforms the EnsKF by 5.7% in RMS error in the 40-dimensional system. *INDEX TERMS:* 3220 Mathematical Geophysics: Nonlinear dynamics; 3260 Mathematical Geophysics: Inverse theory; 3299 Mathematical Geophysics: General or miscellaneous; *KEYWORDS:* nonlinear filtering, data assimilation, Bayesian filtering, particle filtering, ensemble Kalman filter, numerical weather prediction

Citation: Bengtsson, T., C. Snyder, and D. Nychka, Toward a nonlinear ensemble filter for high-dimensional systems, *J. Geophys. Res.*, *108*(D24), 8775, doi:10.1029/2002JD002900, 2003.

1. Introduction

[2] 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 observations 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.

[3] The present article focuses on ensemble or Monte Carlo techniques for the 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, and to be generally useful, these techniques should have the potential to extend to high-dimensional systems.

[4] The algorithms we present approximate the forecast distribution by mixtures of Gaussian distributions. The use of Gaussian mixtures allows (in principle) arbitrary, non-

0148-0227/03/2002JD002900$09.00

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 [*Alspach and Sorenson*, 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.

[5] 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 non-Gaussian updating feasible in high dimensions, we suggest three enhancements of these existing techniques.

[6] 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.

[7] 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.

[8] 3. The algorithms allow each observation to influence only state variables that are nearby in physical space in order to reduce the effective dimensionality of the update. 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*, 1998], but as local non-Gaussian updates at different physical locations must be smoothly blended, its application is novel and nontrivial.

[9] We will show that these three ideas, together with information from the EnsKF, yield a hybrid technique that can outperform the EnsKF in a 40-dimensional system. While the improvement relative to the EnsKF is not yet overwhelming, the hybrid technique demonstrates the potential for ensemble-based, non-Gaussian state estimation outside of very low dimensional systems.

[10] 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 three filtering algorithms. The first two of these we term the mixture ensemble filter and the local-local ensemble filter. The local-local filter is then used in conjunction with the EnsKF to give a hybrid filter that 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. Indeed, our tests with this system identify important weaknesses in both the mixture filter and the local-local ensemble filter; for computationally feasible ensemble sizes, neither scheme can outperform the EnsKF. The hybrid filter is then developed to address those weaknesses. Section 5 summarizes both the techniques developed and the experimental results.

2. Background and Notation

2.1. Update/Forecast Cycle

[11] 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 is 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.

[12] In the update step a forecast pdf is updated given a new set of observations via Bayes's 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 2 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 to be inefficient [*Doucet et al.*, 2001].

[13] 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.

[14] 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*, 1971]. 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.

[15] 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 an ensemble but may also carry importance weights attached to each 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 have 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

[16] 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's theorem.

[17] 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 $\mathbf{e}_t \sim \mathrm{N}(0, \mathbf{R})$. Without loss of generality, $\mathbf{R}$ may be assumed diagonal; one can always transform equation (1) to an observation equation with independent and identically distributed errors by multiplying through by $\mathbf{R}^{-1/2}$.

[18] 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's theorem yields

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

where

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

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

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

where a prime superscript denotes matrix transpose.

[19] 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 in 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 section 2.3.

2.3. Ensemble Kalman Filter Update

[20] 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.)

[21] To anchor our extensions to the EnsKF, we first describe one of its standard implementations. Let $\{\mathbf{x}_{t,i}^f\}$ for $1 \le i \le m$ 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 of equation (2) to each member. Specifically, the algorithm estimates an approximate gain matrix, $\tilde{\mathbf{K}}_t$, using sample covariances based on the ensemble

$$\mathbf{P}_t^f\mathbf{H}_t' \approx (m-1)^{-1}\sum_{i=1}^{m}\left(\mathbf{x}_{t,i}^f - \bar{\mathbf{x}}_t\right)\left[\mathbf{H}_t\left(\mathbf{x}_{t,i}^f - \bar{\mathbf{x}}_t\right)\right]' \tag{6}$$

$$\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t' \approx (m-1)^{-1}\sum_{i=1}^{m}\left[\mathbf{H}_t\left(\mathbf{x}_{t,i}^f - \bar{\mathbf{x}}_t\right)\right]\left[\mathbf{H}_t\left(\mathbf{x}_{t,i}^f - \bar{\mathbf{x}}_t\right)\right]', \tag{7}$$

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

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

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

[22] Although there are other standard ways to sample from the posterior distribution of equation (2), the scheme in equation (8) is applicable in high dimensions since it does not require the explicit (and computationally expensive) covariance recursion defined in equation (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.

[23] 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 while 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

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

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

With the observation equation (1) as defined in section 2.2, the updated distribution is again a mixture of $\boldsymbol{L}$ multivariate normal distributions:

$$p(\mathbf{x}_t|\mathbf{Y}_t) = \sum_{l=1}^{L} \pi_{t,l}^{u} \mathrm{N}\left(\boldsymbol{\mu}_{t,l}^{u}, \mathbf{P}_{t,l}^{u}\right), \tag{9}$$

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 equations (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

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

3. Ensemble Mixture Filters

[25] This section presents a series of three 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.

[26] The first scheme below 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. We find this scheme to be effective only in very low dimensions for computationally feasible ensemble sizes. The second scheme reduces the computations required in high-dimensional systems by updating only the portion of the state vector that is physically local to the observation location. Beginning from forecast ensembles produced by the EnsKF, this update produces state estimates with smaller RMS error than the EnsKF. When this scheme is cycled, however, the result is inferior to the EnsKF due to the lack of spatial smoothness from one local update to the next and the use of fewer observations in the update at any location. The third scheme combines local updating with information from the EnsKF that ensures spatial smoothness. We show in section 4 that such a hybrid scheme can improve on the EnsKF in a 40-variable system.

3.1. Mixture Covariances Based on Local State-Space Information

[27] 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, i.e., from ensemble members that are close in the state-space to the mixture center. This filter will be termed the mixture ensemble filter, or XEnsF.

[28] The update begins with a forecast ensemble $\{\mathbf{x}_{t,i}^{f}, i = 1, \ldots, m\}$. 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 the state-space, though other norms could be employed.) The covariance associated with each center $\mathbf{x}_{t,i}^{f}$ is then given by $\mathbf{P}_{t,i}^{f}$, the sample covariance for the N nearest neighbors of $\mathbf{x}_{t,i}^{f}$. Finally, the algorithm must produce an updated ensemble that is consistent with the update of the continuous mixture through equation (9). Denoting by $\mathbf{K}_{t,i}$ the Kalman gain matrix, with $\mathbf{P}_{t,i}^{f}$ substituted for $\mathbf{P}_{t}^{f}$, the complete update step is as follows (XensF). Given $\{\mathbf{x}_{t,i}^{f}, i = 1,\ldots, m\}$:

1. Update the mixing probabilities. For l in [1, L],
   - find N nearest neighbors to $\mathbf{x}_{t,l}^{f}$ in state-space;
   - calculate π_{l}^{u} from equation (10) using $\mathbf{P}_{t,l}^{f}$ based on the nearest neighbors.
2. Update the ensemble. For j in [1, m],
   - choose a random index $I \in [1, L]$, where $P(I = i) = \pi_{i}^{u}$;
   - choose one of N nearest neighbors of $\mathbf{x}_{t,I}^{f}$, each with probability $1/N$;

- update according to equation (8) using nearest neighbors:

$$\mathbf{x}^u_{t,j} = \mathbf{x}^f_{t,i} + \mathbf{K}_{t,I}\left(\mathbf{y}_t + \mathbf{e}_j - \mathbf{H}_t\mathbf{x}^f_{t,I}\right),$$

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

[29] 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).

[30] Note that the sampling from the updated mixture distribution in the XEnsF is a modest elaboration from the EnsKF. To draw a sample from equation (9), the algorithm first samples an integer from 1 to L from the multinomial distribution, with probabilities given by $\pi^u_{t,I}$. Denoting this random index by I, the algorithm then samples from the Ith component of the mixture using equation (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 equation (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.

[31] Simulation results in section 3.2 will demonstrate that the XEnsF outperforms the EnsKF for a three-dimensional nonlinear system when the forecast distributions are significantly non-Gaussian. Although successful in very low-dimensional systems, we have found the XEnsF to break down when applied to high-dimensional systems due to the inherent difficulties of estimating pdfs in 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 onto a single solution after a few forecast-update cycles.

3.2. Local-Local Ensemble Filter

[32] 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.

[33] 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 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 the pdf be Gaussian or a mixture of Gaussians.

[34] 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 $\mathbf{H}_t$ 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 $\mathbf{x}^f_{t,i}$ will only be changed by a new observation if the corresponding row of $\mathbf{P}^f_{t,i}\,\mathbf{H}_t$ 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 $\mathbf{H}_t$) are localized, the observation neighborhood will be of low dimension. We then propose to update using the XEnsF within this observation neighborhood.

[35] 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 follows.

[36] 1. Given $\{\mathbf{x}^f_{t,i},\ i = 1,\ldots, m\}$.

[37] 2. 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.

[38] 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. In particular, it avoids a single high-dimensional update, which typically leads to the collapse onto a single mixture component (or particle, in the case of PFs). 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.

[39] Although the LLEnsF provides a non-Gaussian update in a spatially local neighborhood, the posterior sample states may be disjointed between observation neighborhoods. To address this issue, we next explain how to create a smooth update of the complete state vector.

3.3. Smoothness Considerations and the Hybrid Filter

[40] We shall sample state variables inside and outside the observation neighborhood differently in order to guarantee that the LLEnsF samples adjoining neighborhoods in a manner that respects the prior relationships among state variables. To simplify notation, we drop all superscripts and subscripts as these are clear from the context. It is further convenient to split the state vector into two parts, $\mathbf{x} = [\mathbf{x}'_L\,\mathbf{x}'_G]'$, where $\mathbf{x}_L$ corresponds to $\mathbf{x}^{[k]}$ and $\mathbf{x}_G$ corresponds to state variables outside the observation neighborhood. Now, consider the equality

$$p(\mathbf{x}|\mathbf{Y}) = p(\mathbf{x}_L|\mathbf{x}_G, \mathbf{Y})p(\mathbf{x}_G|\mathbf{Y}), \tag{11}$$

where the joint distribution of the state vector is split into two parts. To obtain an estimate of $p(\mathbf{x}|\mathbf{Y})$, the above

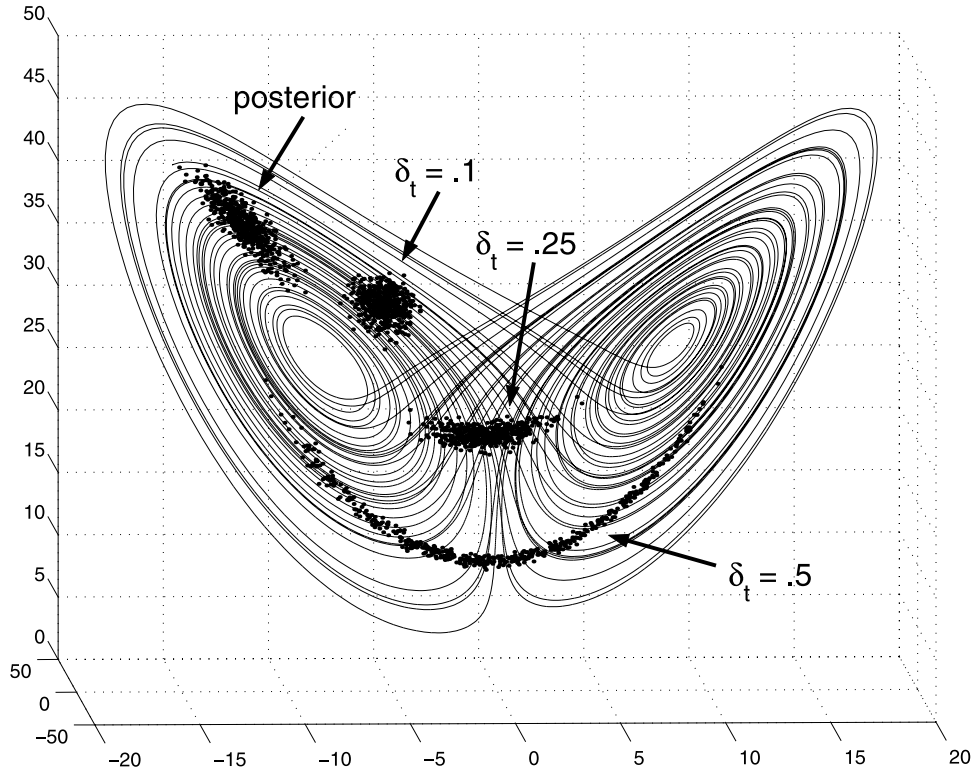

Figure 1. The Lorenz attractor. Non-Gaussian structures appear quickly in this system: 400 points in the upper left-hand corner are sampled from a Gaussian distribution and have been propagated in 0.1, 0.25, and 0.5 time units, respectively.

equality suggests first sampling state variables outside of the observation neighborhood $\mathbf{x}^u_{G,i} \sim p(\mathbf{x}_G|\mathbf{Y})$ and then drawing $\mathbf{z}^u_{L,i} \sim p(\mathbf{x}_L|\mathbf{x}_G, \mathbf{Y})$ by setting $\mathbf{x}_G = \mathbf{x}^u_{G,i}$. The full result is a random draw, where $\mathbf{z}^u_{L,i}$ and $\mathbf{x}^u_{G,i}$ represent a smooth state vector. As (presumably) dimension($\mathbf{x}_G$) > dimension($\mathbf{x}_L$), this sample scheme will be referred to as global-to-local adjustment.

[41] To implement the sequential sample scheme, we suggest combining outputs from EnsKF and LLEnsF using equation (11). Specifically, we use output from EnsKF to represent $p(\mathbf{x}_G|\mathbf{Y})$, while $p(\mathbf{x}_L|\mathbf{x}_G, \mathbf{Y})$ will be based on output from the inner (observation) loop of LLEnsF. With $\mathbf{x}^u_i = [(\mathbf{x}^u_{L,i})'(\mathbf{x}^u_{G,i})']'$ obtained by the sampling algorithm of equation (8), let

$$\begin{bmatrix} \Omega_L & \Omega_{LG} \\ \Omega_{GL} & \Omega_G \end{bmatrix} = \begin{bmatrix} \mathrm{cov}\left(\mathbf{x}^u_{L,i}\right) & \mathrm{cov}\left(\mathbf{x}^u_{L,i}, \mathbf{x}^u_{G,i}\right) \\ \mathrm{cov}\left(\mathbf{x}^u_{G,i}, \mathbf{x}^u_{L,i}\right) & \mathrm{cov}\left(\mathbf{x}^u_{G,i}\right) \end{bmatrix}. \quad (12)$$

Then, using the sample draws $\mathbf{z}^u_i$ from the LLEnsF along with $\mathbf{x}^u_{G,i}$, calculate

$$\begin{bmatrix} \mathbf{z}^u_{L,i} \\ \mathbf{x}^u_{G,i} \end{bmatrix} = \begin{bmatrix} \Omega_{LG}\Omega_G^{-1}\left(\mathbf{x}^u_{G,i} - \mathbf{x}^u_G\right) + \mathbf{A}\left(\mathbf{z}^u_i - \bar{\mathbf{z}}^u\right) \\ \mathbf{x}^u_{G,i} \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{z}}^u \\ 0 \end{bmatrix}, \quad (13)$$

where Ω_G and Ω_L are as defined in equation (12) and $\bar{\mathbf{z}}^u = m^{-1}\Sigma_i \mathbf{z}^u_i$ and $\bar{\mathbf{x}}^u_G = m^{-1}\Sigma_i \mathbf{x}^u_{G,i}$ are sample means. The first term defining $\mathbf{z}^u_{L,i}$ predicts state components inside the observation neighborhood given state components outside the neighborhood, while the second term adds scaled posterior perturbations taken from LLensF. The posterior distribution $p(\mathbf{x}|\mathbf{Y})$ is thus given through equation (13), and $\mathbf{A}$ should be chosen so that the posterior distribution is as "close" as possible to the true non-Gaussian filtering density. In practice, the choice of $\mathbf{A}$ will depend on how informative $\{\mathbf{z}^u_i\}$ and $\{\mathbf{x}^u_i\}$ are relative to the true filtering distribution, and for the systems of interest here a reasonable starting choice is given by $\mathbf{A} = (\Omega_L - \Omega_{LG}\Omega_G{}^{-1}\Omega_{GL})^{1/2}\mathrm{cov}(\mathbf{z}^u_i)^{-1/2}$, setting $\mathrm{cov}(\mathbf{z}^u_{L,i}) = \Omega_L$. (For systems where trace $\{\mathrm{cov}(\mathbf{z}^u_i)\} \ll$ trace$\{\Omega_L\}$ a local-to-global adjustment (described in Appendix B) may be preferable.)

[42] Note that by the sampling scheme of LLensF, $\mathbf{z}^u_i$ and $\mathbf{x}^u_{G,i}$ are independent, and it follows that the covariance of $\mathbf{z}^u_{L,i}$ and $\mathbf{x}^u_{G,i}$ will equal Ω_{GL}. Thus equation (13) produces samples with exactly the same statistical smoothness properties as EnsKF. In section 4.2 we use the global-to-local adjustment when applying the LLEnsF to a 40-dimensional dynamical system.

4. Simulations

[43] We evaluate the filter methods described in section 3 on two nonlinear dynamical systems. Both are sensitive to initial conditions, leading to aperiodic, chaotic solutions and rapid 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 sys-

Table 1. Simulation Results for the L3 System in Terms of Median RMS Error for the Posterior Mean[a]

	XEnsF				EnsKF	
δ_t	$L = 10$, $m = 60$	$L = 10$, $m = 110$	$L = 40$, $m = 90$	$L = 40$, $m = 140$	$m = 40$	$m = 120$
0.1	0.59	0.60	0.48	0.47	0.38	0.37
0.25	0.72	0.71	0.49	0.52	0.72	0.69
0.5	0.93	0.90	0.69	0.69	1.05	1.05
1	1.19	1.14	0.93	0.90	1.37	1.37

[a]Results are estimated for 10,000 assimilation cycles.

tems are given in Appendix A. The XEnsF algorithm is evaluated on L3, and both the LLEnsF and hybrid filters are evaluated on L40.

4.1. Simulations For L3

[44] L3 has been studied extensively in the context of data assimilation [see, e.g., *Miller et al.*, 1994; *Evensen*, 1997; *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 timescales, the dynamics of this system lead to distinctly non-Gaussian forecast distributions. Although L3 is not a high-dimensional system, the experiments with the XEnsF that follow allow us to confirm the benefits of a fully non-Gaussian scheme and provide a reference for the performance of the spatially local, non-Gaussian updates in L40.

[45] To evaluate the effects of the nonlinear dynamics on filter performance, forecast lead time, δ_t, is varied across four levels: $\delta_t = 0.1, 0.25, 0.5, 1$. 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 $\left(\sqrt{R_{jj}}\right)$, the error incurred simply by using the observation vector as a naive update of the state.

[46] Table 1 reports simulation results for assimilating observations over 10,000 assimilation cycles, each separated by a time interval of δ_t, using the XEnsF and standard EnsKF. At each observation time the root mean square 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 EnsKF 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-Gaussian as δ_t increases.

[47] 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 to 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

[48] Simulations for L40 use forecasts of length $\delta_t = 0.4$ 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 0.5. These settings are chosen to produce non-Gaussian behavior in the forecast ensembles.

[49] 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 equation (4.10) of *Gaspari and Cohn* [1999], 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. On the basis of posterior mean estimates at every assimilation cycle, the EnsKF produced a time-averaged RMSE of 0.972 across 2000 assimilation steps. The sample variance of the RMSE was $s^2 = 0.125$, and the median RMSE was 0.882. The forecast distributions produced by the EnsKF appear to be noticeably non-Gaussian, so there is clearly some potential to improve on the EnsKF.

[50] To provide some quantification and evidence of the non-Gaussian structure of the forecasts produced by the EnsKF, we will focus on a three-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 dynamical 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-square distribution with three degrees of freedom. Applying the Kolmogorov-Smirnov (KS) test [*Hogg and Tanis*, 1993] at each assimilation cycle, i.e., for $t = 1, 2, \ldots, 2000$, the hypothesis of normality was rejected in 1896 cases at the 0.05 critical level. The mean of the KS test statistic was 0.139, well above the 0.001 level of significance of 0.094. Hence there is strong evidence of

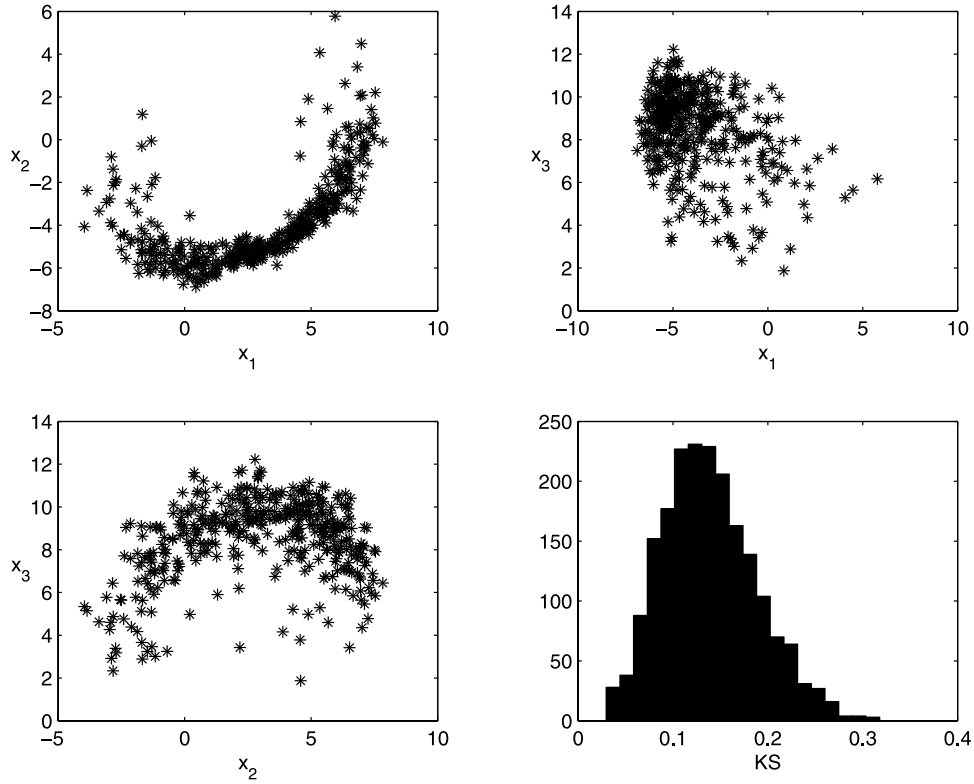

Figure 2. Forecast ensemble members for $\{x_1, x_2, x_3\}$ at a given assimilation time. Bivariate scatterplots depict local non-Gaussian behavior. The ensemble shown produced a Kolmogorov-Smirnov (KS) statistic of 0.134 ($p < 0.001$), while a test of linearity between x_1 and x_2 produced an F statistic of 249.1 ($p < 0.001$). The lower right plot is a histogram of the KS test statistics over 2000 assimilation cycles. See color version of this figure in the HTML.

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 bivariate plots of $\{x_1, x_2, x_3\}$. The lower right plot of Figure 2 shows a histogram of the KS test statistics calculated from the 2000 forecasts produced by the EnsKF.

[51] As can be seen in Figure 2, the relationship between the ensemble members of x_1 and x_2 follows a nonlinear pattern, and the joint distribution of $\{x_1, x_2\}$ is distinctively non-Gaussian. To quantify the degree of nonlinearity 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 0.05 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 nonlinear in a majority of forecast ensembles.

[52] 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 (i.e., the state, observations, and forecast ensembles) from the baseline EnsKF example, the XEnsF was applied to the subvector $\{x_1, x_2, x_3\}$ 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.)

[53] The posterior mean RMSE for the EnsKF across the 2000 assimilation points was 0.827 ($s^2 = 0.383$). On the basis of $L = 400$, $N = 40$, and $m = 400$, the XEnsF improved this by roughly 8%, yielding an RMSE of 0.768 ($s^2 = 0.352$). The improvement is statistically significant ($p < 0.001$). Thus the XEnsF provides, albeit locally and instantaneously, a better estimate of the true state of the system.

[54] Next, we apply the LLEnsF to the same sequence of states and 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 $\mathbf{x}^{[k]} = (x_{k-1 \bmod 40}, x_k, x_{k+1 \bmod 40})$, where $k = 2j - 1$. Using these observation neighborhoods, the LLEnsF was found to be a stable filter; that is, the posterior ensemble mean did not diverge from the true state during any prolonged assimilation sequences. However, LLEnsF did not perform as well as EnsKF, with an approximate 33% increase in RMSE over the 2000 assimilation cycles.

[55] There are two reasons why 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. Hence, although the LLEnsF produces an improved estimate of the state when only spatially local information is used, the EnsKF allows the entire data vector to impact the estimate of the state. The second reason is that samples in adjoining neighborhoods may not be smooth. For example, posterior samples produced in the observation neighborhood $\mathbf{x}^{[1]}$ by assimilating

Table 2. Simulation Results for Hybrid Filters Applied to L40[a]

A	mean(RMSE)	var(RMSE)	median(RMSE)
$\mathbf{B}^{1/2}\{\mathrm{cov}(\mathbf{z}_i^u)\}^{-1/2}$	0.917	$s^2 = 0.100$	0.848
$\sqrt{\mathrm{trace}(\mathbf{B})/\mathrm{trace}\{\mathrm{cov}(\mathbf{z}_i^u)\}}\mathbf{I}_\ell$	0.941	$s^2 = 0.142$	0.854

[a]The time-averaged RMSE, var(RMSE), and median RMSE for 2000 assimilation cycles are shown. The corresponding results for the EnsKF were 0.972, 0.125, and 0.882, respectively.

y_1 may be not be "smooth" with those produced in the observation neighborhood $\mathbf{x}^{[3]}$ by assimilation of y_2.

[56] As discussed in section 3.3, these limitations 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. In each observation loop of LLEnsF (with observation neighborhoods as previously defined) we draw $\mathbf{z}_i^u$ from XEnsF(400, 400, 40) and $\mathbf{x}_{G,i}{}^u$ from EnsKF. The two ensembles $\{\mathbf{z}_i^u\}$ and $\{\mathbf{x}_{G,i}^u\}$ are then combined using equation (13) to produce posterior samples. With $\mathbf{B} = \Omega_L - \Omega_{LG}\Omega_G{}^{-1}\Omega_{GL}$, we apply two versions of the hybrid filter by setting $\mathbf{A} = \mathbf{B}^{1/2}\{\mathrm{cov}(\mathbf{z}_i^u)\}^{-1/2}$ and $\mathbf{A} = \sqrt{\mathrm{trace}(\mathbf{B})/\mathrm{trace}\{\mathrm{cov}(\mathbf{z}_i^u)\}}\mathbf{I}_\ell$, where $\mathbf{I}_\ell$ represents the identity matrix of size $\ell = \mathrm{dimension}(\mathbf{x}_L)$.

[57] The first choice of $\mathbf{A}$ yields posterior samples with equivalent second moment statistics to EnsKF. For this choice of $\mathbf{A}$ the two updated ensembles will here be 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 second choice of $\mathbf{A}$ directly uses $\mathbf{z}_i^u$ from LLEnsF but rescales the sample $\{\mathbf{z}_i^u\}$ so that $\mathrm{trace}\{\mathrm{cov}(\mathbf{z}_{L,i}^u)\} = \mathrm{trace}(\Omega_L)$ at every data assimilation. Alternatively, one may consider the hybrid filters to be extensions of the EnsKF. This is a reasonable interpretation, but it ignores the additional generality provided by equation (13).

[58] Table 2 summarizes the results from the hybrid filter using the same states and observations from the baseline EnsKF simulation example. For both choices of $\mathbf{A}$ the improvement in the posterior mean estimate compared to that produced by the EnsKF is statistically significant ($p < .001$, $p < .01$) and corresponds to a 5.7% and 3.2% overall error decrease, respectively.

[59] The results demonstrate the potential of developing non-Gaussian filtering techniques for strongly nonlinear, high-dimensional systems. Clearly, however, work remains to extend the positive results from the L40 experiment to atmospheric systems of realistic dimension. Nevertheless, the key concepts of the hybrid scheme, i.e., the local update of the LLEnsF and the global-to-local adjustment of section 3.3, provide an important foundation for filtering high-dimensional dynamical systems.

5. Summary

[60] This paper has presented a hierarchy of nonlinear ensemble filters, each of which employs mixtures of Gaussian distributions in its update step. These filters range from the XEnsF, which uses a general representation of the prior distribution but is computationally feasible only for systems of very low dimension, to the LLEnsF, in which the XEnsF update is applied only in a spatially local neighborhood of each observation and which suffers from a lack of spatial smoothness between the observation neighborhoods, to a hybrid of the LLEnsF and the ensemble Kalman filter.

[61] A crucial 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 forecast distribution, including error estimates. Accurate representation of error statistics produces a stable filter that does not diverge as t increases, a common problem when devising sequential Bayesian update procedures with fixed sample sizes [*Künsch*, 2001]. 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.

[62] The LLEnsF and hybrid filters extend the XEnsF beyond low-dimensional systems by restricting the update step to spatially local subspaces of the state vector; consequently, it is not subject to the problems of reweighting mixture components (particles) associated with high-dimensional distributions. The numerical results in this work confirm that there are three-dimensional subspaces where the mixture takes advantage of non-Gaussian structures. However, a straightforward 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. By letting more state variables be affected by an observation, the global-to-local adjustment presented in section 3.3 provides smooth updates (from the EnsKF) of larger portions of the state vector yet allows for spatially restricted non-Gaussian updates. This global-to-local adjustment forms the basis for the hybrid of the LLEnsF and the EnsKF. In the 40-variable model of *Lorenz* [1996], whose dimension is sufficiently large that the XEnsF is not feasible, the hybrid method outperforms both the LLEnsF and the EnsKF.

[63] We emphasize that the hierarchy of filtering schemes given here was fundamentally necessitated by computational limitations. It is likely that, given a sufficiently large ensemble, the XEnsF would be effective for L40 or higher dimensional systems as it treats the full Bayesian update in the most general and unapproximated way. Nevertheless, our experience shows, and simple arguments [*Silverman*, 1986] suggest, that the ensemble size required by the XEnsF will grow rapidly, perhaps exponentially, with the system dimension. The additional approximations in the hierarchy from the XEnsF to the LLEnsF and then to the hybrid filter are thus necessary to allow the update to proceed with feasible ensemble sizes without being overly contaminated by sampling errors.

[64] It is further worth noting that our evaluation of the various schemes has considered only the quality of the first moment in terms of ensemble mean error. While this is a natural starting point, future work should evaluate the additional information in the ensemble concerning higher moments and other features of the forecast distribution. Although the EnsKF is not explicitly designed to capture higher moments, some information on these is implicitly carried through the update step as the forecast ensemble is only subjected to a linear transformation. Our results, in which the hybrid filter performs best yet inherits its higher moments from the EnsKF, suggest that the EnsKF does provide robust information on higher moments when compared to the non-Gaussian scheme outside of very low dimensional systems.

[65] Further refinements of the schemes presented here are clearly necessary in order to realize more substantial improvements over the EnsKF. In our view, the ideas of sequential, spatially local updating and the global-to-local adjustment provide a basis for such refinements. Together, these ideas provide a framework for combining potentially more accurate non-Gaussian update procedures with more robust schemes such as the EnsKF that are based on the Gaussian update. A more sophisticated filter will likely rely on efficient, sequential identification of low-dimensional subspaces where non-Gaussian densities can be accurately represented and filtered using finite ensemble sizes.

Appendix A

[66] The L3 model [*Lorenz*, 1963] is defined by three differential equations:

$$\dot{x} = -\sigma(x_t + y_t),$$

$$\dot{y} = rx_t - y_t - x_t y_t,$$

$$\dot{z} = x_t y_t - bz_t,$$

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

[67] The L40 model [*Lorenz*, 1996] is defined by the differential equations

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

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

[68] Both systems are propagated using a first-order Euler method with a time step of 0.001. This simple numerical scheme facilitates rapid propagation of a large number of ensembles.

Appendix B

[69] By reversing the order of the conditioning, i.e., switching the roles of $\mathbf{x}_L$ and $\mathbf{x}_G$ in equation (11), a local-to-global adjustment can be devised, resulting in the following specific sampling scheme:

$$\begin{bmatrix} \mathbf{z}^u_{L,i} \\ \mathbf{x}^u_{G,i} \end{bmatrix} = \begin{bmatrix} \mathbf{z}^u_i \\ \mathbf{x}_G + \Omega_{LG}\Omega_L^{-1}\mathbf{A}\left\{\left(\mathbf{z}^u_i - \mathbf{z}^u\right) - \left(\mathbf{x}^u_{L,i} - \mathbf{x}^u_L\right)\right\} \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{\delta} \end{bmatrix}, \tag{B1}$$

where $\mathbf{A}$ and $\bar{\delta}$ must be chosen appropriately. Selecting $\mathbf{A} = \Omega_L^{1/2}\{\mathrm{cov}(\mathbf{z}_i^u)\}^{-1/2}$ yields $\mathrm{cov}(\mathbf{z}^u_{L,i}, \mathbf{x}^u_{G,i}) = \Omega_{LG}$.

[70] **Acknowledgment.** This research was supported in part by National Science Foundation grant DMS 9815344.

References

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

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

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

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

Doucet, A., N. Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, 2001.

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

Evensen, G., 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, 1994.

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

Gaspari, G., and S. E. Cohn, Construction of correlation functions in two and three dimensions, *Q. J. R. Meteorol. Soc.*, *125*, 723–757, 1999.

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

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

Hogg, R. V., and E. A. Tanis, *Probability and Statistical Inference*, 4th ed., 731 pp., Macmillian, Old Tappan, N. J., 1993.

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

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

Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, 376 pp., Academic, San Diego, Calif., 1970.

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

Künsch, H. R., State space and hidden Markov models, in *Complex Stochastic Systems*, *Monogr. Stat. Appl. Probab.*, vol. 87, pp. 109–173, Chapman and Hall, New York, 2001.

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

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

Lorenz, E. N., Predictability: A problem partially solved, in *Proceedings of the Seminar on Predictability*, vol. 1, pp. 1–18, Eur. Cent. for Medium-Range Weather Forecasts, Reading Berkshire, U.K., 1996.

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

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

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

T. Bengtsson, D. Nychka, and C. Snyder, Geophysical Statistics Project, Climate and Global Dynamics Division, National Center for Atmospheric Research, P.O. Box 2000, Boulder, CO 80307-3000, USA. (tocke@stat.berkeley.edu; nychka@ncar.edu; chris@ucar.edu)