A Non-Gaussian Ensemble Filter Update for Data Assimilation

JEFFREY L. ANDERSON

NCAR Data Assimilation Research Section, Boulder, Colorado*

(Manuscript received 19 October 2009, in final form 8 June 2010)

ABSTRACT

A deterministic square root ensemble Kalman filter and a stochastic perturbed observation ensemble Kalman filter are used for data assimilation in both linear and nonlinear single variable dynamical systems. For the linear system, the deterministic filter is simply a method for computing the Kalman filter and is optimal while the stochastic filter has suboptimal performance due to sampling error. For the nonlinear system, the deterministic filter has increasing error as ensemble size increases because all ensemble members but one become tightly clustered. In this case, the stochastic filter performs better for sufficiently large ensembles. A new method for computing ensemble increments in observation space is proposed that does not suffer from the pathological behavior of the deterministic filter while avoiding much of the sampling error of the stochastic filter. This filter uses the order statistics of the prior observation space ensemble to create an approximate continuous prior probability distribution in a fashion analogous to the use of rank histograms for ensemble forecast evaluation. This rank histogram filter can represent non-Gaussian observation space priors and posteriors and is shown to be competitive with existing filters for problems as large as global numerical weather prediction. The ability to represent non-Gaussian distributions is useful for a variety of applications such as convective-scale assimilation and assimilation of bounded quantities such as relative humidity.

1. Introduction

Many Monte Carlo methods for geophysical data assimilation and prediction have been developed. The most general methods are particle filters that can represent arbitrary analysis probability distributions (Van Leeuwen 2003). However, the number of particles required increases very rapidly as the size of the prediction model increases. At present, no variant of a particle filter that is practical for large geophysical models is known (Snyder et al. 2008).

Ensemble Kalman filter methods are not as general as particle filters but practical variants exist that work for very large problems like operational numerical weather prediction (Houtekamer and Mitchell 2005; Whitaker et al. 2008; Szunyogh et al. 2008). Ensemble methods use a set of model states to compute sample estimates of the variance and covariance of model state variables and observations. Ensemble methods that have been successfully applied to large geophysical problems share three simplifying assumptions. First, estimates of the probability distribution of an observation are approximated by a Gaussian during the assimilation. Second, the likelihood function for the observation is specified as a Gaussian. Third, the relation between the observation and all unobserved model state variables is approximated with a least squares fit.

Although many variants have been proposed, ensemble Kalman filters that work with large models can be categorized as either deterministic or stochastic. The time evolution of deterministic ensemble Kalman filters depends only on the initial ensemble and the observations.

Deterministic methods like the ensemble adjustment Kalman filter (EAKF; Anderson 2001) studied here are simply algorithms for computing the Kalman filter (Kalman 1960; Kalman and Bucy 1961) solution if the prediction model is linear, observations are a linear function of the state variables, and the observation likelihood is Gaussian. Stochastic filters make use of a random number generator when assimilating each observation. Stochastic filters like the ensemble Kalman filter (EnKF; Evensen 1994; Burgers et al. 1998) studied here approximate the solution to the Kalman filter, but include

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Jeffrey Anderson, NCAR, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: jla@ucar.edu

DOI: 10.1175/2010MWR3253.1

sampling error that is expected to decrease as ensemble size increases.

The Kalman filter is generally not the optimal solution to the data assimilation problem when the forecast model is not linear. Nevertheless, both the EnKF and EAKF produce successful analyses when applied in atmospheric general circulation models that are nonlinear. There have been several studies of how the different types of filters perform with nonlinear models (Lawson and Hansen 2004; Leeuwenburgh et al. 2005; Sakov and Oke 2008a,b; Sun et al. 2009; Mitchell and Houtekamer 2009). In fact, even a single variable model with a quadratic nonlinearity can prove challenging to ensemble filtering methods. Cycling of the nonlinear model with periodic assimilation can lead to prior ensemble distributions that are highly non-Gaussian. Assuming Gaussianity when assimilating an observation can result in poor filter performance. Ensemble filters also can have difficulty when assimilating observations that are not expected to have Gaussian distributions. The most obvious examples are observations of bounded quantities like precipitation, tracer concentration, or cloud fraction where Gaussianity is obviously inappropriate. Observation likelihood functions for instruments must also be approximated by Gaussians. For instance, instruments with skewed error distributions like some radiometers can be problematic.

An ensemble filter method that can represent arbitrary prior and posterior observation distributions and observation likelihoods is developed here. It is first shown to greatly reduce problems caused by weakly nonlinear forecast models that generate significantly skewed prior distributions. It is also shown to be competitive with both deterministic and stochastic ensemble filters for practical ensemble sizes in large numerical weather prediction applications. Section 2 describes the most common deterministic and stochastic ensemble Kalman filter algorithms in a Bayesian context. Section 3 illustrates the difficulties that occur when these ensemble filters are applied to problems with a nonlinear forecast model. Section 4 presents the new ensemble filter update algorithm, the rank histogram filter (RHF). Comparisons of the new filter with the deterministic and stochastic ensemble Kalman filters for both simple and complex models are presented in section 5 with conclusions presented in section 6.

2. Ensemble filters

Ensemble Kalman filter algorithms have two parts. First, a prediction model is applied to each member of a posterior ensemble at time t_1 to produce a prior ensemble at some later time t_2. Second, an update algorithm is used to fuse observations available at t_2 with the prior ensemble to produce a posterior ensemble at t_2. Only algorithms for the second part are examined in detail here.

Observational error distributions for each observation are assumed to be independent here so that the observations can be assimilated sequentially (Houtekamer and Mitchell 2001). Update algorithms are described for a single scalar observation without loss of generality. If observational error distributions for observations available at the same time are Gaussian but not independent, the update problem can be rotated to a space in which the error covariance matrix is diagonal and a scalar update algorithm can be applied (Anderson 2003). If the observational error distribution for observations available at the same time is not Gaussian and the observational errors are not independent, the update algorithms presented here are not sufficiently general. These algorithms are also insufficient to assimilate observations available at different times that have dependent observational error distributions (Evensen 2003).

It is also assumed that each observation is related to the model state variables such that linear regression is an accurate approximation of the relationship. It follows that the impact of an observation on each scalar state variable can be computed independently (Anderson 2003). Given these assumptions, an update algorithm consists of three steps. First, a prior ensemble estimate of the scalar observation y is computed by applying a forward observation operator h to each of the N prior ensemble estimates of the model state vector:

$$y_n^p = h(\mathbf{x}_n), \quad n = 1, \ldots, N, \tag{1}$$

where the superscript p indicates a prior.

Second, the observed value y^o and the observational error variance σ_o^2 are used to compute increments Δy_n for each prior ensemble estimate in (1). This report explores several variants for computing Δy_n in the second step of the update algorithm; the rest of the steps remain unchanged. In the EAKF algorithm described in section 2a, the ensemble mean of the Δy_n is the standard Kalman filter update increment at the measurement location.

Third, increments for each component of the prior state vector are computed by linear regression of the observation increments using the prior joint sample of the state variables and the observed variable:

$$\Delta\mathbf{x}_{m,n} = (\sigma_{p,m}/\sigma_p^2)\Delta y_n, \quad m = 1, \ldots, M, \quad n = 1, \ldots, N, \tag{2}$$

where $\sigma_{p,m}$ is the prior sample covariance of the observed variable and the mth element of the state vector,

σ_p^2 is the prior sample variance of the observed variable and M is the size of the model state vector.

Several ensemble Kalman filter update algorithms are described in the geophysics literature. They are commonly described in terms of linear algebra using a Kalman gain (Houtekamer and Mitchell 1998), but can also be presented in the Bayesian context described above (Anderson 2003). These algorithms can be divided into two classes: stochastic perturbed observation filters and deterministic square root filters. Perturbed observation filters (Burgers et al. 1998; Pham 2001) are commonly referred to as EnKFs. Several variants of deterministic square root ensemble Kalman filters appear in the literature (Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; Tippett et al. 2003). The ensemble adjustment Kalman filter is used here but the behavior described is generic. The relative performance of the perturbed observation and deterministic filter algorithms is application dependent (Thomas et al. 2009). All of these algorithms assume that the observational error variance is Gaussian so that the observational likelihood is Normal(y^o, σ_o^2).

a. *Ensemble adjustment Kalman filter update algorithm*

The EAKF update first computes the ensemble mean $\overline{y}_p$ and variance σ_p^2 of the prior observation ensemble and approximates the prior with Normal($\overline{y}_p, \sigma_p^2$). The product of this prior Gaussian with the observational likelihood Gaussian is a continuous posterior that is a constant times Normal($\overline{y}_u, \sigma_u^2$). The prior ensemble is then compressed [first term in (3)] and translated [last two terms in (3)] to give an updated ensemble with sample mean $\overline{y}_u$ and variance σ_u^2. The resulting increments for the observation ensemble are

$$\Delta y_n = (y_n^p - \overline{y}_p)(\sigma_u/\sigma_p) + \overline{y}_u - y_n^p, \quad n = 1, \ldots, N. \qquad (3)$$

For a linear prediction model, linear forward observation operators, and normally distributed observational error distributions, the EAKF is an exact method for computing the standard Kalman filter (KF) for all ensemble sizes $N > D_{\max}$. The KF represents the prior and posterior distributions at each time as Gaussians. The $D_{\max}$ is the largest number of nonzero singular values in a singular value decomposition of the prior and posterior covariance matrices from the KF application. If an N-member ensemble with $N > D_{\max}$ has prior sample mean $\overline{x}_t$ and covariance matrix $\mathbf{\Sigma}_t$ at time t and a KF at time t has prior distribution Normal($\overline{x}_t, \mathbf{\Sigma}_t$) then the EAKF sample mean and covariance will be identical to the KF mean and covariance at all subsequent times (Anderson 2009a).

If $N \le D_{\max}$, the EAKF sample covariance is rank deficient and no longer duplicates the KF results. If the linear prediction model operator has any eigenvalues greater than 1 then the EAKF estimate of the mean will diverge from the KF solution. For geophysical problems, it is usually not computationally affordable to have $N > D_{\max}$. To avoid filter divergence, localization (Houtekamer and Mitchell 2001; Hamill et al. 2001) can be applied to allow small ensembles to work for large problems. Errors associated with using localization for small ensembles increase as the ensemble size decreases, but there is no general analytic characterization of this sampling error.

b. *Perturbed observation EnKF update algorithm*

The EnKF (Burgers et al. 1998) update also computes the mean $\overline{y}_p$ and variance σ_p^2 of the prior observation ensemble. An N-member random sample of the observational error distribution is generated

$$\hat{y}_n^o = y^o + \varepsilon_n, \quad n = 1, \ldots, N, \qquad (4)$$

where ε_n is drawn from Normal(0, σ_o^2). The mean of this sample is replaced by the observed value y^o giving an ensemble with elements:

$$y_n^o = \hat{y}_n^o - \sum_{n=1}^{N} \hat{y}_n^o/N + y^o. \qquad (5)$$

The nth updated ensemble member is the mean of the product of Normal(y_n^p, σ_p^2) and Normal(y_n^o, σ_o^2) so that

$$y_{u,n} = \sigma_u^2(\overline{y}_p/\sigma_p^2 + y_n^o/\sigma_o^2), \qquad (6)$$

where

$$\sigma_u^2 = [(\sigma_p^2)^{-1} + (\sigma_o^2)^{-1}]^{-1}. \qquad (7)$$

Observation increments are computed as $\Delta y_n = y_{u,n} - y_{p,n}$.

Unlike the EAKF, the EnKF is a Monte Carlo approximation of the KF if $N > D_{\max}$ and the sampling error in estimates of the ensemble variance is expected to be proportional to $N^{-0.5}$. Like the EAKF, the EnKF becomes degenerate when $N \le D_{\max}$ and in this case filter divergence is likely for geophysical applications without the use of localization.

The EnKF algorithm applied here also pairs the sorted observation prior and sorted observation posterior ensemble members when computing observation increments to minimize the expected value of the increments. Let Z_k^p and Z_k^u be the kth-order statistics of the prior and updated observed ensemble, respectively. The order statistics of the increments are $\Delta Z_k = Z_k^u - Z_k^p$. The values of the updated ensemble members are unchanged, but prior and updated members are paired so that the expected value of the increment is minimized. In a linear Gaussian system, sorting the increments like this does not change the answers produced by the EnKF. However, if the EnKF is applied to a nonlinear system, sorting reduces the expected errors from assuming a linear relation when regressing the observation increments onto state variable increments (Anderson 2003, section 4c). The expected value of this regression error increases as the ensemble size decreases. Without this algorithm the EnKF would be seriously handicapped for the nonlinear applications in the next sections.

3. Application to single variable system

The EAKF and EnKF are applied to the prediction model:

$$x_{t+1} = x_t + 0.05(x_t + \alpha |x_t| x_t), \tag{8}$$

where x is a scalar, the subscript indexes time, and α is a constant. A perfect model assimilation experiment is performed with the truth assumed to be 0 at all times. This is the same as a situation where (8) represents the evolution of the state linearized around an arbitrary time-varying trajectory. The true state is observed once every time step with unit error variance so that observations are drawn from Normal(0, 1). Observations are generated for 51 000 time steps. An initial ensemble for x is drawn from Normal(0, 1). The first 1000 assimilation steps are discarded to avoid any initial transient behavior.

a. Linear system

When $\alpha = 0$ in (8) the system is linear and the distribution of the prior ensemble around its ensemble mean at time t_2 is a linear expansion of the distribution of the posterior at time t_1 around its mean. The KF is the optimal solution for this problem and the EAKF solution is identical to the KF solution for any ensemble size $N > 1$. Figure 1 shows the time-mean value of the absolute value of the ensemble mean as a function of ensemble size for $N = 10, 20, 40, 80, 160, 320$, and 640. The EAKF time-mean root-mean-square error (RMSE) is independent of ensemble size and is slightly less than 0.31. The time-mean RMSE for the EnKF is a decreasing function of ensemble size; for 640 ensemble members it has decreased to about 0.315.

For all ensemble sizes, the sample variance for the EAKF converges to the same value as for the KF and is constant in time after fewer than 100 time steps (with the 64-bit reals used in the computation). All higher-order moments of the EAKF are also constant in time after the transient spinup, but their values depend on the details of the initial ensemble. The sample variance and all other moments of the EnKF vary in time with the amount of variation decreasing as ensemble size increases.

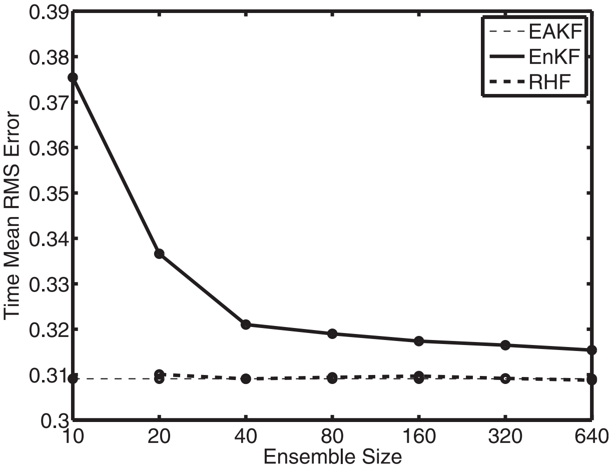

FIG. 1. Time-mean prior RMSE of the ensemble mean for the linear single variable model as a function of ensemble size. Results are shown for the ensemble adjustment Kalman filter (thin dashed line), perturbed observation ensemble Kalman filter (solid line), and the rank histogram filter (thick dashed line).

b. Nonlinear systems

When $\alpha > 0$ in the model in (8) an extra nonlinear growth in time is added to the linear expansion of the ensemble around its mean. The prior and posterior distributions for x are no longer Gaussian (Reichle et al. 2002) and the Kalman filter is no longer optimal. Figure 2 shows the time-mean RMSE for the EnKF and EAKF as a function of ensemble size for the case with $\alpha = 0.2$. The EnKF RMSE decreases with increasing ensemble size and has a value less than 0.33 for $N = 640$. The EAKF has time-mean RMSE of about 0.33 for $N = 10$, but the RMSE increases with increasing ensemble size to about 0.46 for $N = 640$. Similar problems for deterministic filters in nonlinear systems are documented in Sakov and Oke (2008b) and Mitchell and Houtekamer (2009, see their Fig. 4).

The initial evolution of an $N = 20$ EAKF for $\alpha = 0.8$ is shown in Fig. 3. The same behavior can be seen for $\alpha = 0.2$, but the increased nonlinearity in Fig. 3 accentuates the problem. The smallest ensemble member is farthest from the ensemble mean at the initial time. The other 19 ensemble members collapse to a single value during the first 250 assimilation steps, while the smallest member remains distinct and moves away from the rest. This

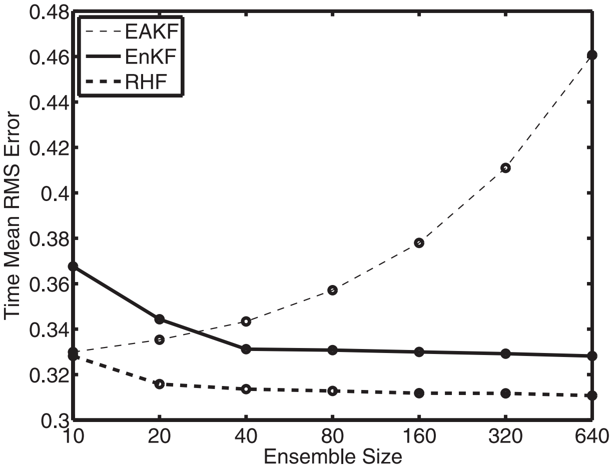

FIG. 2. As in Fig. 1, but for the nonlinear single variable model with nonlinearity parameter $\alpha = 0.2$.

behavior can be understood by comparing the EAKF update step and the model advance. In the update, ensemble members are linearly compacted around the ensemble mean. In the advance, members farther from the mean are moved out quadratically with the most remote member being moved out the most. To keep the ensemble variance constrained, the update uses a linear compaction that is sufficient to constrain the outermost ensemble member, but this compaction is larger than required for the other members. The result is that $N - 1$ ensemble members eventually collapse to a single value and one member remains distinct. As N increases, the expected distance between the outlier and the cluster must increase in order to maintain approximately the correct sample variance.

The sample kurtosis:

$$k = \frac{\sum_{n=1}^{N}(x_n - \overline{x})^4}{(N-1)\sigma^4}, \tag{9}$$

where $\overline{x}$ is the ensemble mean and σ^2 is the sample variance is a statistic that is especially sensitive to the presence of outliers (Lawson and Hansen 2004). For the $\alpha = 0.2$ case, the time-mean kurtosis for the EAKF is an increasing function of ensemble size and ranges from 8 for $N = 10$ to 98 for $N = 640$. Figure 4 plots the time-mean and maximum kurtosis of the ensemble for the EnKF as a function of ensemble size. The mean value increases relatively slowly while the maximum value does not demonstrate an obvious trend with values around 8 for all ensemble sizes.

Figure 5 shows the initial evolution of an $N = 20$ EnKF for $\alpha = 0.8$ for comparison with Fig. 3. The EnKF does not suffer from the obvious degenerate behavior of the EAKF, but it does have frequent instances where one ensemble member has become an outlier due to the nonlinear dynamics of the model advance. For instance, around $t = 60$ one ensemble member has a much smaller value than the rest and the kurtosis is close to 10. Although there is no analytic answer available for this nonlinear problem, the fact that the kurtosis is much larger than the value of 3 expected for a Gaussian distribution is some indication that the EnKF outlier is too extreme. Further evidence is presented in later sections. The outlier behavior does not persist in time for the EnKF because the update algorithm adds random noise to the ensemble members, essentially mixing them at each assimilation step.

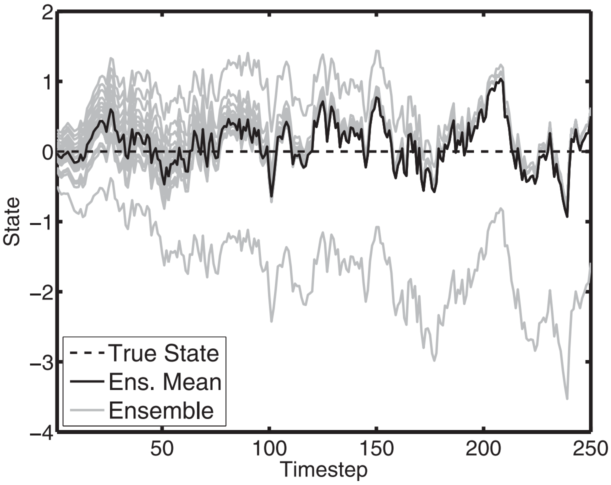

FIG. 3. Initial evolution of 20 ensemble members (light solid lines), the ensemble mean (dark solid line), and the truth (dashed line, constant value of 0) for an ensemble adjustment Kalman filter applied to the single variable model with the nonlinearity parameter $\alpha = 0.8$.

4. A non-Gaussian filter update for nonlinear systems

A deterministic filter update designed to work appropriately with non-Gaussian likelihoods and priors is described here. It avoids the EnKF sampling error and the EAKF outlier problems and is able to handle high-kurtosis priors like those seen in Fig. 3.

A viable ensemble update algorithm for geophysical applications must satisfy the following constraints. First, increments should be as small as possible in order to minimize errors from assuming linearity when computing state variable increments. Second, the algorithm should work nearly as well as the EAKF for problems that are Gaussian. Third, it should outperform the EAKF and EnKF for cases that are significantly non-Gaussian.

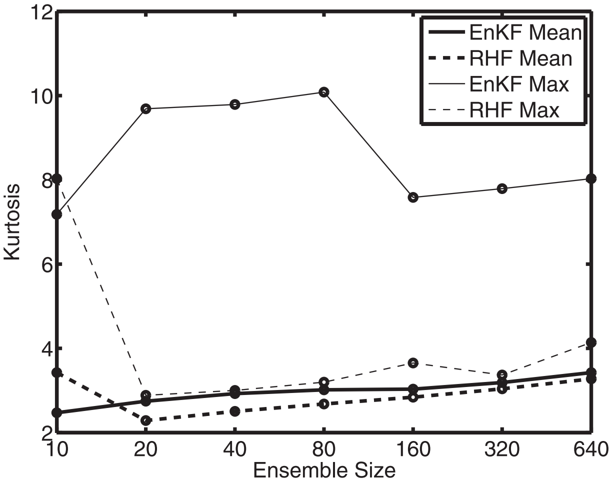

FIG. 4. Time-mean (thick line) and maximum (thin line) value of ensemble kurtosis as a function of ensemble size for the single variable model with the nonlinearity parameter $\alpha = 0.2$. Results are plotted for the ensemble Kalman filter (solid line) and the rank histogram filter (dashed line).

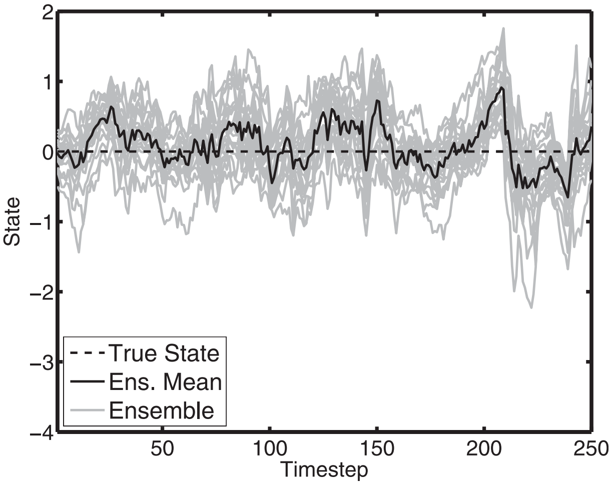

FIG. 5. As in Fig. 3, but for an assimilation with a perturbed observation ensemble Kalman filter.

Finally, it should not be prohibitively costly for large problems.

An ensemble sample of the prior and some continuous (usually Gaussian) representation of the likelihood are the inputs to an ensemble filter update algorithm. The output must be an ensemble representation of the updated distribution. One way to proceed is to compute a continuous approximation to the ensemble prior and multiply it by the likelihood to get a continuous posterior from which a posterior ensemble is constructed. For example, the EAKF approximates the prior with a Gaussian. In the kernel filters described in Anderson and Anderson (1999), the prior is represented as a sum of N Gaussians. The kernel filter satisfies the second and third criteria for a filter listed above, but not the first and it is computationally expensive.

Figure 6 is a schematic of how a continuous prior (shaded in the figure) is constructed from the ensemble in the filter update developed here. The ensemble members are assumed to partition the real line into $N + 1$ regions each of which contains $1/(N + 1)$ of the probability. This is the assumption made in using the rank histogram (Anderson 1996; Hamill 2001) for verification of ensemble predictions, so the method is referred to as a RHF here. The continuous prior probability density function is assumed to be constant in regions bounded on both sides by an ensemble member. The probability density in the unbounded regions on the tails of the ensemble is represented by a portion of a Gaussian distribution. If the smallest ensemble member is Z_1, then the probability density for $x < Z_1$ is Normal(μ, σ_p^2), where σ_p^2 is the prior ensemble sample variance and μ is selected so that the cumulative density at Z_1 is $1/(N + 1)$. The density for $x > Z_N$ is approximated analogously. Other representations of the density in the unbounded regions are possible and can lead to differences in the behavior of the filter that are discussed in section 6.

The RHF computes a continuous posterior distribution by taking the product of the prior and the likelihood at each point, and then normalizing the product so that it is a probability distribution. Given a continuous posterior, the RHF computes updated ensemble members with rank statistics Z_n^u that partition the real line into $N + 1$ regions, each of which contains $1/(N + 1)$ of the posterior probability. Computing this updated ensemble for a Gaussian likelihood is computationally intensive.

To reduce the cost of computing the updated ensemble members, the continuous Gaussian likelihood can be approximated in the interior regions by a piecewise linear approximation as shown by the dashed line in Fig. 7. The likelihood is unchanged in the unbounded regions on the tails of the prior ensemble. The continuous posterior distribution, shown by the thick solid lines in Fig. 7, is piecewise linear in the interior regions and is a weighted Gaussian in the two outer regions (Anderson and Anderson 1999).

To find the updated ensemble members, the cumulative probability of the posterior at the position of each prior ensemble member is computed by integrating the posterior PDF over the regions between each prior ensemble member. The value of the cumulative posterior density at the updated ensemble members is

$$C_n = n/(N + 1), \quad n = 1, \ldots, N.$$

Finding the location of an updated ensemble member that lies in one of the interior regions requires integrating

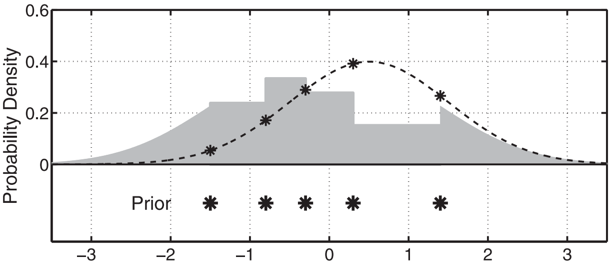

FIG. 6. Schematic of first phase of rank histogram filter algorithm. The locations of five prior ensemble members are indicated by large asterisks at the bottom. The continuous approximation to the prior probability density is indicated by the four shaded boxes and the shaded portions of Gaussians on the tails. The continuous likelihood is the dashed line with the values at the ensemble members marked by small asterisks.

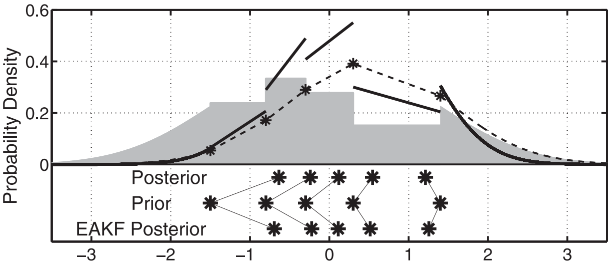

FIG. 7. Schematic of the rank histogram filter algorithm. The prior and posterior ensembles along with the posterior from an ensemble adjustment Kalman filter are marked by asterisks at the bottom. The continuous approximation to the prior probability density is shaded. The dashed line is a piecewise linear interior approximation to the likelihood. The continuous posterior probability distribution is the thick solid line.

the linear probability distribution in the region and solving the resulting quadratic equation to find where the cumulative posterior density equals C_n. Finding the location of an updated ensemble member that lies in the unbounded regions requires inverting the normal cumulative distribution function. This can be done efficiently by using an accurate polynomial approximation.

Figure 7 shows that the posterior ensembles for the RHF and the EAKF are similar for this simple five-member example. In general, for approximately Gaussian priors the RHF and EAKF perform similarly even for very small ensembles and are increasingly similar as ensemble size increases. Figure 8 illustrates the RHF continuous prior and posterior and the updated ensembles for the RHF and EAKF for a prior ensemble with an outlier. This prior is analogous to priors that arise during the initial stages of the nonlinear EAKF assimilation shown in Fig. 3. The RHF posterior eliminates the outlier since it occurs in a region where there is very small likelihood. The EAKF can only shift and compact the prior, so a significant outlier remains and the other ensemble members are shifted too far in the opposite direction. An ensemble Kalman filter with a mean-preserving random rotation (Sakov and Oke 2008b) is another mechanism for removing outliers (Evensen 2009) but discards any non-Gaussian information that might exist in the ensemble.

The key to the Kalman filter algorithm is the assumption that the prior and the observation likelihood are both Gaussian distributions. Since the RHF no longer makes this assumption, it is more appropriately categorized as an ensemble filter as opposed to an ensemble Kalman filter.

5. Results

a. Single variable model

Figure 1 shows the time-mean RMSE as a function of ensemble size for the linear model with the RHF. The RHF undergoes filter divergence for $N < 12$. For larger ensemble sizes, the RHF time-mean RMSE is within 0.001 of the exact KF value and is significantly smaller than the EnKF RMSE, while the spread is within 0.001 of the KF value.

Figure 2 shows the RHF time-mean RMSE for the nonlinear problem with $\alpha = 0.2$. In this case, the RHF shows a small decrease in RMSE with increasing ensemble size and has smaller time-mean RMSE than both the EAKF and EnKF for all ensemble sizes. Figure 4 shows the time-mean and maximum kurtosis as a function of ensemble size for the EnKF and the RHF. The time-mean values are similar with the EnKF being slightly larger for all ensemble sizes except $N = 10$. However, the maximum kurtosis over 50 000 assimilation steps for the RHF increases from approximately 3 to just greater than 4 as N increases from 20 to 640. The maximum kurtosis for the EnKF is approximately 8 for all ensemble sizes. This is additional circumstantial evidence that the EnKF is producing sporadic outliers.

b. Lorenz63 model

The three-variable model of Lorenz (1963) is used as the second in a series of increasingly complex tests of the three ensemble filter updates. Each of the three state variables is observed every 12th time step (with a

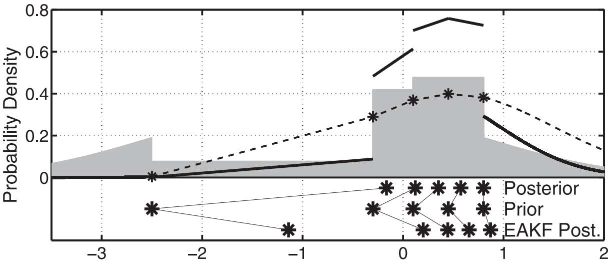

FIG. 8. As in Fig. 7, but for a prior ensemble with an outlier.

standard 0.01 nondimensional time step) with an observational error simulated by a draw from a normal with mean 0 and variance 8. Ensemble initial conditions are random draws from a very long free integration of the model. A total of 101 000 assimilation steps are performed and summary results are averaged over the last 100 000 steps of the assimilation to eliminate the impact of transient behavior. The prior covariance inflation value from the set $\{1, 1.01, 1.02, \ldots\}$ that minimizes the time-mean ensemble-mean RMSE over the last 100 000 assimilation steps is selected for each ensemble size and filter type.

Model trajectories in the Lorenz63 system rotate around two ellipsoidal lobes that are joined at one edge with transitions between the lobes occurring sporadically. The space between the two lobes is not part of the attractor. This attractor can lead to situations where some members of a prior ensemble have transitioned into the other lobe while other members do not, resulting in large kurtosis for prior distributions.

Figure 9 shows the evolution of the x variable from an 80-member EAKF ensemble and the truth during 40 assimilation steps. There is evidence of an outlier in the ensemble that is below the rest of the ensemble at times. Around assimilation 2795, the truth and 79 ensemble members transition to the upper lobe of the attractor, but the remaining ensemble member initially starts to return to the lower lobe. After that, this outlier becomes detached from the rest of the ensemble and is unrelated to the truth. Eventually, the outlier rejoins the rest of the ensemble by chance at about assimilation step 2900 and the assimilation begins to function properly again. The EAKF assimilation has repeated episodes of this type of behavior.

Figure 10 shows the time mean of the RMSE of the prior ensemble mean for the three filters as a function of ensemble size. The EAKF has smallest RMSE of about 1.18 for $N = 10$ and this increases rapidly as a function of N. The EnKF has RMSE of about 1.3 for $N = 10$, a minimum RMSE for $N = 40$, and a slight increase as ensemble size gets larger. The RHF has RMSE of about 1.5 for $N = 10$ and this decreases uniformly as a function of N. The RHF RMSE is less than that for the ENKF for $N = 80$ and larger. Even the best results here are not as good as those produced with the ESRF with mean-preserving random rotation by P. Sakov (2010, personal communication).

Figure 11 shows the time-mean and maximum kurtosis as a function of ensemble size for each filter. The time-mean values for all three filters increase with increasing ensemble size with the RHF being smallest, the EnKF slightly larger, and the EAKF much larger. The maximum kurtosis is a measure of the occurrence of occasional large outliers. For the EAKF, the maximum value is too large to display in the figure because of events like that in Fig. 9. The EnKF maximum also grows rapidly and is associated with less frequent and less severe outlier episodes while the RHF maximum grows to about 50 for $N = 640$. This suggests that the EnKF is subject to outlier problems, but that the mixing from the random part of the perturbed observation algorithm and the nonlinear aspects of the model dynamics eliminates outliers fairly quickly.

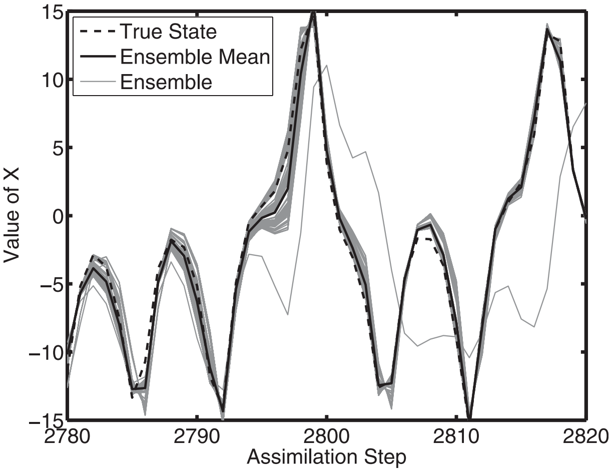

FIG. 9. A portion of the evolution of 80 ensemble members (light solid lines), the ensemble mean (dark solid line), and the truth (dashed line) for the x variable of an ensemble adjustment Kalman filter assimilation for the Lorenz63 model.

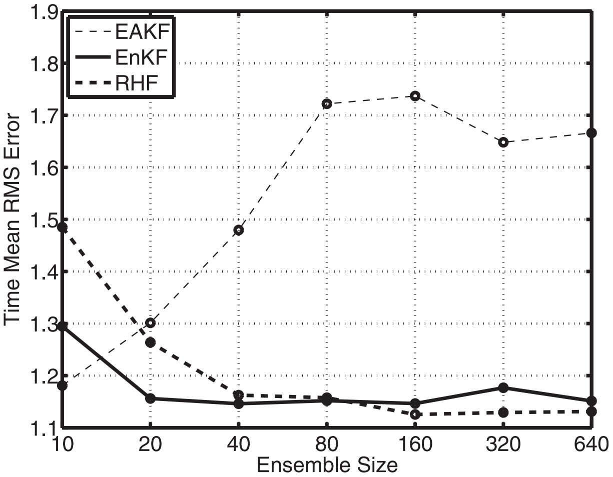

FIG. 10. Time-mean prior RMSE of the ensemble mean for an assimilation with the Lorenz63 model as a function of ensemble size. Results are shown for the ensemble adjustment Kalman filter (thin dashed line), perturbed observation ensemble Kalman filter (solid line), and the rank histogram filter (thick dashed line).

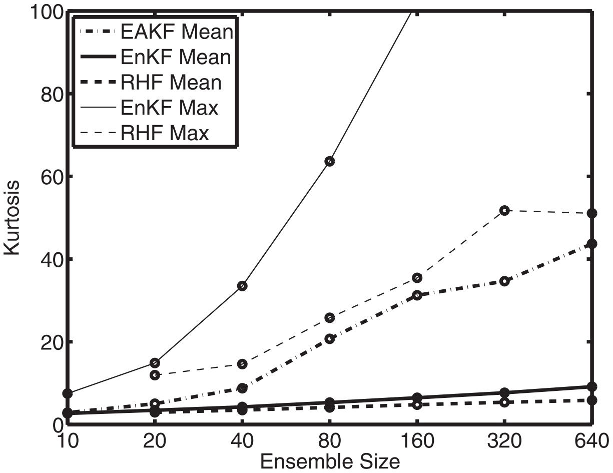

FIG. 11. Time-mean (thick line) and maximum (thin line) value of ensemble kurtosis as a function of ensemble size for the *x* variable of the Lorenz63 model. Results are plotted for the ensemble adjustment Kalman filter (dashed line, maxima too large to display on plot), the perturbed observation ensemble Kalman filter (solid line), and the rank histogram filter (dashed line).

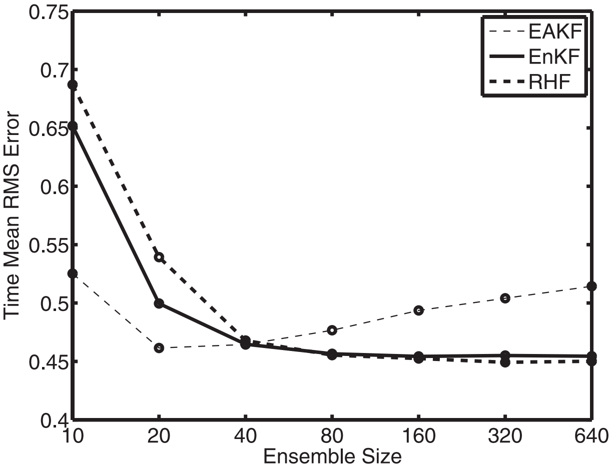

FIG. 12. Time-mean prior RMSE of the ensemble mean for an assimilation with the Lorenz96 40-variable model as a function of ensemble size. Results are shown for the ensemble adjustment Kalman filter (thin dashed line), perturbed observation ensemble Kalman filter (solid line), and the rank histogram filter (thick dashed line).

c. Lorenz96 model

The 40-variable configuration of the Lorenz96 model is frequently used for evaluating geophysical ensemble filters (Lorenz and Emanuel 1998). A truth integration of 100 000 time steps is observed every time step. Forty observations of the following form:

$$y_j = (x_j + x_{j+1})/2 + \text{Normal}(0, 4), \quad j = 1, \ldots, 40, \tag{10}$$

where *j* is a cyclic index are taken every time step. The impact of observations was localized (Hamill et al. 2001) with a Gaspari–Cohn (Gaspari and Cohn 1999) function with half-width equal to 30% of the domain width and an adaptive spatially varying inflation (Anderson 2009b) was applied for all three filters.

Figure 12 shows the RMSE of the ensemble mean for the three filters as a function of ensemble size. All three demonstrate their worst performance for $N = 10$. The EAKF has its smallest RMSE for $N = 20$ and has increasing RMSE for larger *N*. The RHF and EnKF have RMSE that decreases with *N* for all ensemble sizes. The EnKF has smaller RMSE for *N* up to 40 while the RHF has smaller RMSE for larger *N*. Figure 13 shows the time-mean and maximum kurtosis and provides evidence that the EAKF is still suffering from sporadic outlier problems; the maximum kurtosis for the EAKF is off the plot for all ensemble sizes. The time-mean kurtosis for the RHF and EnKF are very similar while the EnKF maxima are larger for large ensemble sizes suggesting that the EnKF is mostly immune to outlier issues for this model and observation set.

Sakov and Oke (2008a,b) have recently studied the performance of several other ensemble filter variants including a random rotation that removes outliers from ensemble distributions. The Lorenz96 case they document is significantly more linear and only uses observations of state variables, which reduces the challenges of non-Gaussianity. The appendix provides results from the filters used here for direct comparison with their results.

d. B-grid dynamical core

The three filters were used with a low-resolution version of the B-grid dynamical core that was developed for the Geophysical Fluid Dynamics Laboratory (GFDL) atmospheric model version 2 (AM2). The model grid has 60 longitudes and 30 latitudes with 5 levels in the vertical and a total state vector size of 28 200. The model has no orography and is forced following the Held–Suarez framework (Held and Suarez 1994). This is close to the minimum size at which the model produces baroclinic waves. The model was integrated from a state of rest for 1000 days to ‘‘spin up’’ the model climate. The integration was continued for 100 days and synthetic observations were generated by simulating 300 radiosonde locations randomly located on the surface of the sphere and fixed in time every 12 h. Each radiosonde observes the temperature and wind components at each model level as well as the surface pressure. The observational error variance is 1 K for the temperatures, 1 m s^{-1} for the wind components, and 1 hPa for the

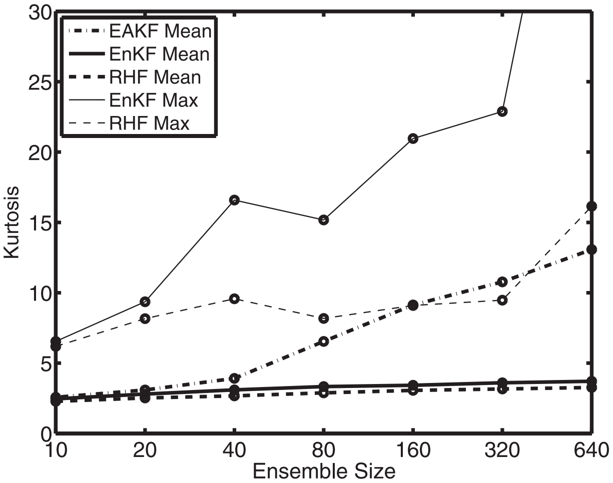

FIG. 13. As in Fig. 11, but for the first variable of the Lorenz96 40-variable model.

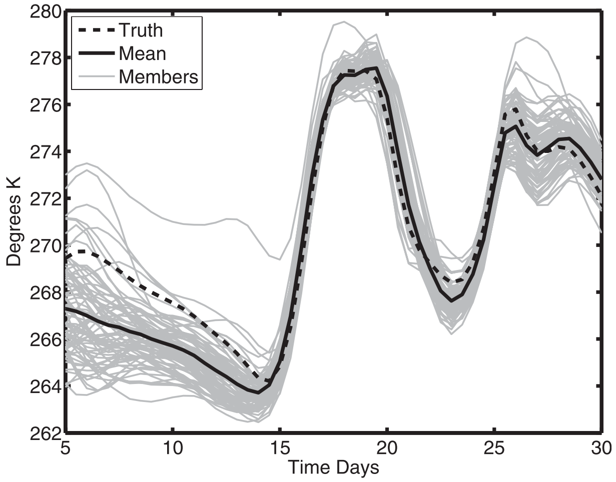

FIG. 14. A portion of the evolution of 80 ensemble members (light solid lines), the ensemble mean (dark solid line), and the truth (dashed line) for a temperature variable in the midlatitudes on the top model level from an assimilation with the dynamical core of the B-grid general circulation model.

surface pressure. Observation impact was localized with a Gaspari–Cohn function with a 0.2 rad half-width and no inflation was used as in Anderson et al. (2005).

For the B-grid assimilation, the RHF gave the lowest time-mean globally averaged RMSE for all state variables on each model level. The EAKF and EnKF produced somewhat larger time-mean RMSE. Even in this large model application where there is lots of ‘‘noise’’ from sampling error, localization, and a nonlinear model, there is evidence of outlier issues with the EAKF. Figure 14 shows a time series of the evolution of an 80-member EAKF and the truth for a temperature grid point at the top model level in the midlatitudes (50°N, 30°W). There is a clear outlier between days 6 and 17. The kurtosis for the EAKF ensembles continues to have very large maximum values for all variables; however, the time-mean kurtosis is similar to that for the EnKF. This suggests that outliers are occurring less frequently than was the case in low-order models or are being more rapidly eliminated by nonlinear model dynamics and the use of many observations that impact each state variable.

6. Discussion and conclusions

Many publications have discussed the relative performance of various flavors of ensemble Kalman filters in low-order models (Whitaker and Hamill 2002; Lawson and Hansen 2004) and there is ample evidence that different types of filters are better for different applications. Often, however, there is little understanding of why this is the case nor is there an ability to generalize results to other applications. The RHF algorithm described here is specifically designed to deal with significantly non-Gaussian priors. It is expected to do well compared with the traditional Gaussian filters when priors are bimodal, skewed, or bounded. As an example, it may have advantages in applications to radar data assimilation for convective scale storms. Priors in such cases may be bimodal for important variables; either convection has initiated at a grid point or it has not. The Gaussian filters tend to produce a posterior that is an average between convecting and not convecting, a situation that is unrealistic. With compelling observations, the RHF should produce a posterior that is convecting or not convecting as appropriate.

The RHF has been compared here to results from an EAKF and an EnKF with sorting of observation increments. There are many other ensemble Kalman filter variants described in the literature that could also have been used as controls. The appendix compares the EAKF and EnKF to a variety of ensemble Kalman filters tested by Sakov and Oke (2008a,b).

All three filter methods discussed have been applied to a global numerical weather prediction experiment with the Community Atmosphere Model (CAM) version 3.5 (Collins et al. 2006) with a 2° grid resolution and 26 vertical levels. All temperature and wind observations from radiosondes, the Aircraft Communication, Addressing, and Reporting System (ACARS), and other aircraft and satellite motion vector winds from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis input files for 1 September 2006–30 August 2007 were assimilated with an 80-member ensemble. A spatially varying adaptive inflation (Anderson 2009b) and localization with a 0.2 rad Gaspari–Cohn function were used.

The results were evaluated by computing the difference between the ensemble mean 6-h forecast predictions of the observations and the observed values. All three filters produce results that are comparable to or better than those produced by the NCEP operational system from this period with the EAKF and RHF being consistently better than the EnKF for this application. This demonstrates that the RHF can be competitive with the Gaussian filters even for very large applications with model error and real observations. Much additional study is needed to better understand the performance of the different filter variants in realistic atmospheric applications.

Several modifications to the RHF algorithm can reduce computational cost or lead to interesting new capabilities. The approximate representation of the likelihood can be further simplified so that its value between each pair of ensemble members is a constant equal to the average of the likelihood evaluated at the bounding ensemble members. In this case, computing the location of posterior ensemble members in the interior regions requires solving a linear equation, not a quadratic. This piecewise constant form for the likelihood also tends to give posteriors with greater spread. Loss of spread is a persistent problem in ensemble filters (Houtekamer and Mitchell 1998; Mitchell and Houtekamer 2009). The piecewise constant variant generally required less inflation (Anderson and Anderson 1999) and produced slightly smaller RMSE in cases where inflation was needed for good filter performance.

A second modification to the approximate representation of the likelihood involves replacing the Gaussians in the unbounded regions with a constant that is just the value of the likelihood evaluated at the outermost ensemble member. At first, this seems like a drastic approximation. However, in general it has only a small impact on the posterior since the product in the tails usually has small probability. Since the likelihood on the tails is increased with this approximation, posterior spread is increased again resulting in a reduced need for inflation. Applying this approximation to the low-order model examples above led to very small increases in time-mean RMSE in all cases.

Assuming a flat tail facilitates the use of arbitrary observation likelihoods with the RHF. The only information needed about the likelihood is its value for each of the prior ensemble members. There are a number of observations that have non-Gaussian likelihoods. The most obvious examples are observations of bounded quantities like precipitation, tracer concentration, or cloud fraction. Observations with lognormal likelihoods can be assimilated directly without the need to convert state variables to logarithmic forms and back. Some instruments like certain radiometers are known to have skewed likelihood functions. Initial applications of the RHF to observations of bounded quantities suggests that it is much more effective than Gaussian filters in keeping posterior ensemble members appropriately bounded. All of these variants of the RHF are available as part of the standard release of the Data Assimilation Research Test bed from NCAR (Anderson et al. 2009).

For most geophysical applications, the cost of the observation increment computation in an ensemble filter [computing Δy_n for use in (2)] is a small fraction of the computation. The observation increments for the three filters discussed here all have cost $O(KN)$, where K is the number of observations and N is the ensemble size. The constant factor for the RHF observation increment computation is about 4 times greater than that for the EnKF and EAKF. However, the cost of updating the state variables using (3) given the observation increments is $O(KM_cN)$, where M_c is the average number of state variables that are impacted by a given observation. For geophysical problems, M_c is generally much larger than 4 so the RHF cost is negligible. For instance, the total computational time for the RHF is less than 10% more than the cost of the EAKF for all multivariate experiments described here.

The flat tail variant for the likelihood presents interesting possibilities for future ensemble filter development. The observation update only requires the values of the likelihood for each ensemble member. This is identical to the weights that are associated with each ensemble member in a particle filter. One could use this RHF algorithm to update state variables directly, given the likelihood, without the intermediate steps of computing observation increments and regressing them on the state variables. A filter constructed in this fashion would no longer assume linearity by using regression and could have a computational cost that is much smaller than the ensemble filters described above. However, there are challenges related to dealing with sampling noise, localization, and inflation in an ensemble particle filter of this type. It is not yet clear if there will be applications where such a filter will be competitive with traditional ensemble filters.

Acknowledgments. The author thanks Andy Majda, Chris Snyder, Stephen Anderson, Doug Nychka, Thomas Bengtsson, and Peter Bickell for their insights on ensemble and particle filters. Thanks to Tim Hoar, Kevin Raeder, Nancy Collins, Glen Romine, and Hui Liu for helping to build and maintain the Data Assimilation Research Testbed. Pavel Sakov, Herschel Mitchell, and two anonymous reviewers were a great help in improving this manuscript. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation.

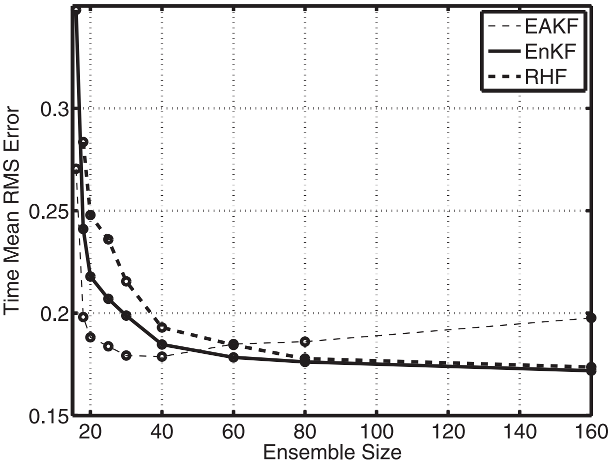

FIG. A1. Time-mean posterior (analysis) RMSE of the ensemble mean for an assimilation with the Lorenz96 40-variable model with observations of each state variable. Results are shown for the ensemble adjustment Kalman filter (thin dashed line), perturbed observation ensemble Kalman filter (solid line), and the rank histogram filter (thick dashed line).

APPENDIX

Additional Lorenz96 Results

A 1.1 million step control integration of the Lorenz96 model is observed every time step with 40 observations of the following form:

$$y_j = x_j + \text{Normal}(0,1), \quad j = 1, \ldots, 40. \qquad \text{(A1)}$$

Assimilations with ensemble sizes of 16, 18, 20, 25, 30, 40, 60, 80, and 160 are performed with the EAKF, the EnKF, and the RHF with no localization. The prior covariance inflation value from the set $\{1, 1.01, 1.02, \ldots\}$ that minimizes the time-mean ensemble-mean RMSE over the last 1 million steps is selected for each ensemble size and filter type. Initial ensemble members are random draws from an extended free run of the model. Results in Fig. A1 can be compared to Fig. 5 in Sakov and Oke (2008a) and Fig. 4 in Sakov and Oke (2008b). This system is very linear compared to the Lorenz96 case in section 5c and does not display much non-Gaussianity that would challenge the EAKF or give an advantage to the RHF. The EAKF is distinctly better than the best results from Sakov and Oke for ensembles of size 16 and 20, while all three filters are competitive for ensemble sizes larger than 20. The EnKF results here are also better than those from Sakov and Oke for small ensemble sizes. The EnKF used here sorts the observation increments (see section 2b) and also adjusts the mean of a perturbed observation to be the same as the original observed value; neither of these operations were done in the Sakov and Oke filter (P. Sakov 2010, personal communication) and may account for the differences in results.

REFERENCES

Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate,* **9,** 1518–1530.

——, 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.,* **129,** 2884–2903.

——, 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.,* **131,** 634–642.

——, 2009a: Ensemble Kalman filters for large geophysical applications. *IEEE Contr. Syst.,* **29,** 66–82.

——, 2009b: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus,* **61A,** 72–83.

——, 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.

——, B. Wyman, S. Zhang, and T. Hoar, 2005: Assimilation of surface pressure observations using an ensemble filter in an idealized global atmospheric prediction system. *J. Atmos. Sci.,* **62,** 2925–2938.

——, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009: The data assimilation research test bed: A community facility. *Bull. Amer. Meteor. Soc.,* **90,** 1283–1296.

Bishop, C. H., B. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.,* **129,** 420–436.

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

Collins, W. D., and Coauthors, 2006: The formulation and atmospheric simulation of the Community Atmosphere Model Version 3 (CAM3). *J. Climate,* **19,** 2144–2161.

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.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.,* **53,** 343–367.

——, 2009: The ensemble Kalman filter for combined state and parameter estimation. *IEEE Contr. Syst. Mag.,* **29,** 83–104.

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

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.,* **129,** 550–560.

——, 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.

Held, I. M., and M. J. Suarez, 1994: A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Amer. Meteor. Soc.,* **75,** 1825–1830.

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

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **129,** 123–137.

——, and ——, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.,* **131,** 3269–3289.

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

——, and R. S. Bucy, 1961: New results in liner filtering and prediction theory. *Trans. Amer. Soc. Mech. Eng. J. Basic Eng.,* **83D,** 95–108.

Lawson, W. G., and J. A. Hansen, 2004: Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Wea. Rev.,* **132,** 1966–1981.

Leeuwenburgh, O., G. Evensen, and L. Bertino, 2005: The impact of ensemble filter definition on the assimilation of temperature profiles in the Tropical Pacific. *Quart. J. Roy. Meteor. Soc.,* **131,** 3291–3300.

Lorenz, E. N., 1963: Deterministic non-periodic flow. *J. Atmos. Sci.,* **20,** 130–141.

——, and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.,* **55,** 399–414.

Mitchell, H. L., and P. L. Houtekamer, 2009: Ensemble Kalman filter configurations and their performance with the logistic map. *Mon. Wea. Rev.,* **137,** 4324–4343.

Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.,* **129,** 1194–1207.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, 2002: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.,* **130,** 103–114.

Sakov, P., and P. R. Oke, 2008a: A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters. *Tellus,* **60A,** 361–371.

——, and ——, 2008b: Implications of the form of the ensemble transformation in the ensemble square root filters. *Mon. Wea. Rev.,* **136,** 1042–1053.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.,* **136,** 4629–4640.

Sun, A. Y., A. Morris, and S. Mohanty, 2009: Comparison of deterministic ensemble Kalman filters for assimilating hydrogeological data. *Adv. Water Resour.,* **32,** 280–292.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus,* **60A,** 113–130.

Thomas, S. J., J. P. Hacker, and J. L. Anderson, 2009: A robust formulation of the ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.,* **135,** 507–521.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.,* **131,** 1485–1490.

Van Leeuwen, P. J., 2003: A variance-minimizing filter for large-scale applications. *Mon. Wea. Rev.,* **131,** 2071–2084.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.,* **130,** 1913–1924.

——, ——, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.,* **136,** 463–482.