Implications of Stochastic and Deterministic Filters as Ensemble-Based Data Assimilation Methods in Varying Regimes of Error Growth

W. GREGORY LAWSON AND JAMES A. HANSEN

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 23 June 2003, in final form 6 January 2004)

ABSTRACT

Accurate numerical prediction of fluid flows requires accurate initial conditions. Monte Carlo methods have become a popular and realizable approach to estimating the initial conditions necessary for forecasting, and have generally been divided into two classes: stochastic filters and deterministic filters. Both filters strive to achieve the error statistics predicted by optimal linear estimation, but accomplish their goal in different fashions, the former by way of random number realizations and the latter via explicit mathematical transformations. Inspection of the update process of each filter in a one-dimensional example and in a two-dimensional dynamical system offers a geometric interpretation of how their behavior changes as nonlinearity becomes appreciable. This interpretation is linked to three ensemble assessment diagnostics: rms analysis error, ensemble rank histograms, and measures of ensemble skewness and kurtosis. Similar expressions of these diagnostics exist in a hierarchy of models. The geometric interpretation and the ensemble diagnostics suggest that both filters perform as expected in a linear regime, but that stochastic filters can better withstand regimes with nonlinear error growth.

1. Introduction

Accurate numerical prediction of fluid flows requires accurate initial conditions. In an effort to approximate optimal estimation methods for linear problems with Gaussian error statistics, Monte Carlo methods have become a popular and realizable approach to estimating the initial conditions necessary for forecasting (e.g., Evensen and van Leeuwen 1996; Houtekamer and Mitchell 2001). Ensemble methods based on linear theory have taken two general forms: stochastic filters and deterministic filters. Both filters strive for the same result, namely to achieve the error statistics predicted by optimal linear estimation, but accomplish their goal in different fashions, the former by way of random number realizations and the latter via explicit mathematical transformations. We seek to understand the implications of these different methodologies when applied to different regimes of error growth.

Through geometric considerations in a one-dimensional example and in a two-dimensional system's state space, we formulate an interpretation of the update process for both filter types. We examine how this interpretation expresses itself in three ensemble diagnostics, root-mean-square (rms) analysis error, ensemble rank histograms, and measures of ensemble skewness and kurtosis. We then monitor filter behavior with these diagnostics in a hierarchy of models. By running observation system simulation experiments in a perfect model scenario with relatively large ensembles, we test filter reliability and performance in regimes with well-contained linear error dynamics and regimes with appreciable nonlinear error growth.

We find that both filters perform well, as expected, in the linear regime, but that stochastic filters are more robust *for these diagnostics* as nonlinearity becomes important. We stress that issues of robustness are highly dependent on one's choice of measure and are application dependent. For instance, the formulators of one deterministic filter algorithm, Whitaker and Hamill (2002), show that realizing the random numbers necessary for stochastic filters adds another source of sampling error to data assimilation that can easily lead to underestimation of the desired analyzed uncertainties for small to moderately sized ensembles. Similarly, Anderson (2001) points out that deteriministic filters are better at retaining higher-order moments through the assimilation process and allow an easier time tracking individual ensemble member trajectories. Deterministic filters are likely to be more robust if any of these features are deemed important for one's given application. This paper does not seek to proclaim that one filter formulation is better than another; it only seeks to understand the differences between them and the implications of those differences in various regimes of linearity for systems' error growth dynamics.

In section 2 we give a small review of the ensemble

Corresponding author address: W. Gregory Lawson, 77 Massachusetts Avenue, Room 54-1724, Cambridge, MA 02139.
E-mail: wglawson@mit.edu

filter formulations. We begin section 3 by examining the filters' respective update kinematics for two hypothetical one-dimensional ensembles. We then link insight gained from this hypothetical case to three ensemble assessment methods applied to a filtering exercise in a two-dimensional chaotic system, the Ikeda system. Then, having seen how the updates in state space express themselves in these diagnostics, we turn our attention in section 4 to higher-dimensioned, spatially extended models, and find the same signatures in the diagnostics for those systems. Finally, we summarize our findings in section 5.

2. Filter formulations

No model estimate nor observation is ever perfect, and their uncertainties render data assimilation and forecasting exercises in probability. For systems following linear dynamics with errors from Gaussian probability density functions (PDFs), an optimal-state estimate, in both a minimum variance sense and a maximum likelihood sense, can be achieved through use of the Kalman filter. The Kalman filter is a recursive state estimation technique that can be derived from Bayes' rule seeking either a maximum likelihood estimate (e.g., Lorenc 1986) or a minimum variance estimate (e.g., within estimation theory, Cohn 1997), or it can be identified as a recursive least squares problem properly weighted by the inverses of the relevant error covariance matrices (e.g., Wunsch 1996). The filter consists of a forecast step and an update step for both the mean and the covariance of a state estimate's PDF. These two moments fully characterize a Gaussian PDF.

For nonlinear systems, where PDFs can in general develop important moments not captured by a Gaussian, no such optimality conditions can be asserted. However, through the use of various linearizations, the Kalman filter framework can be retained. One such implementation is the extended Kalman filter (EKF). This filter uses the full nonlinear model to advance the state estimate, that is, the PDF's mean, but uses a linearized version of the model to advance the state's uncertainty, that is, the PDF's covariance. The linearized version of the model, or the tangent linear model, is reevaluated at each time step about the latest state estimate.

Among the main challenges of implementing an EKF in geophysical problems is the computational demand. A state estimate with n elements requires a covariance matrix with n^2 elements. For $n \sim O(10^7)$, matrix storage and arthimetic are prohibitively expensive. This computational barrier has led scientists to explore different ways to approximate the EKF's analyses. Many reduced-state filters have been proposed, most relying on the notion that the actual number of dynamically important directions for error growth is much less than the full dimension of the state estimate. Examples include Verlaan and Heemink's (1997) reduced-rank square root filter (RRSQRT), Pham et al.'s (1998) singular evolutive extended Kalman filter (SEEK), Chin et al.'s (1999) reduced-order information filter (ROIF), and Farrell and Ioannou's (2001) balanced truncation Kalman filter.

Another class of reduced-state filters utilizes Monte Carlo methods. Evensen (1994) was the first to use an ensemble of forecasts to estimate the PDFs necessary in Kalman filtering, though the idea of using ensembles for probabilistic weather forecasting was posed at least 20 yr earlier (e.g., Leith 1974). In Evensen's so-called ensemble Kalman filter (EnKF), the key assumption is equating the covariance of an ensemble of forecasts, each evolved with the full nonlinear model, with the linearly evolved forecast error covariance matrix necessary for the EKF. Concomitantly, one must identify the ensemble mean as one's best-state estimate. This assumption allows efficient computation of the Kalman gain matrix necessary for the update step of the filters. Houtekamer and Mitchell (1998) and Burgers et al. (1998) introduced and justified the notion of using perturbed observations in an EnKF scheme. They found that in order to approximate the proper analysis error covariance, the observations must be treated as random variables. Hence, to arrive at the EnKF, the forecast step of the EKF is replaced by N ensemble forecasts, where N is the number of ensemble members, and the update step of the EKF is replaced by updating each ensemble member with the same estimated Kalman gain matrix and observations perturbed with random samples from the assumed observational uncertainty.

Equating the ensemble spread covariance with the error covariance necessary for the EKF should be a valid assumption if the ensemble is constructed in a fashion so it can sample from the important growing modes in the current state estimate. Experience has shown that the EnKF is an effective data assimilation scheme (e.g., Evensen and van Leeuwen 1996; Houtekamer and Mitchell 1998; Hansen and Smith 2000; Reichle et al. 2002). However, experience has also shown that for insufficiently large ensembles, problems may arise in implementation of these methods. Thus, compensations such as covariance localization, where one imposes a fixed correlation length scale on one's estimate of the ensemble covariance, and covariance inflation, where one artificially increases the ensemble spread about the mean, are often employed. By using relatively large ensemble sizes, we avoid implementing these practices in the work presented here.

The EnKF is not the only algorithm for a recursive ensemble filter to achieve the EKF-predicted analysis error covariance. Several independent parties have proposed equivalent variations of an ensemble filter that do not require generating perturbed observations. Because these filters do not require further random number realizations, they are termed *deterministic* filters—once the initial ensemble and the observations are known, the updated ensemble is immediately available. In contrast, the EnKF is termed a *stochastic* filter because the random numbers necessary preclude knowledge of the an-

alyzed ensemble until they are realized. Bishop et al.'s (2001) ensemble transform Kalman filter (ETKF), Anderson's (2001) ensemble adjustment Kalman filter (EAKF), and Whitaker and Hamill's (2002) ensemble square root filter (EnSRF) are all deterministic filters that arrive at the desired analysis error statistics, the analysis error covariance matrix that the EKF would have retrieved, by updating the ensemble mean and then linearly transforming the ensemble members about that mean into a distribution with exactly the desired error covariance. The transformations available to the filters are rotations, translations, and rescalings in various directions.

Consider an ensemble $\mathbf{z}_j$, where $j = 1, N$ and each ensemble state has dimension n. Because the analysis is formed as a linear combination of the forecast best estimate and the observations, the update step for ensemble filters can be written in terms of the ensemble mean and perturbations about that mean, $\mathbf{z}_j = \overline{\mathbf{z}} + \mathbf{z}'_j$. Following Whitaker and Hamill (2002), the update step for a stochastic filter, for example, the EnKF, can be formulated as

$$\overline{\mathbf{z}}^a = \overline{\mathbf{z}}^f + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\overline{\mathbf{z}}^f), \tag{1}$$

$$\mathbf{z}'^a_j = \mathbf{z}'^f_j + \mathbf{K}(\mathbf{y}'^o - \mathbf{H}\mathbf{z}'^f_j), \qquad j = 1, N, \tag{2}$$

$$\mathbf{P}^a = \frac{1}{N-1}(\mathbf{Z}^a - \overline{\mathbf{Z}}^a)(\mathbf{Z}^a - \overline{\mathbf{Z}}^a)^{\mathrm{T}}, \tag{3}$$

where superscript a denotes the updated quantity, or analysis, superscript f denotes the forecasted quantity, $\mathbf{K}$ is the approximate Kalman gain matrix, or simply a weighting matrix, found from equating the forecast ensemble covariance with the forecast error covariance, $\mathbf{y}^o$ is an m vector of observations, $\mathbf{H}$ is an operator that transforms from state space to observation space, $\mathbf{Z}$ is an $n \times N$ matrix having the individual ensemble members as its columns, $\overline{\mathbf{Z}}$ is a similarly dimensioned matrix with the ensemble mean in each column, and $\mathbf{Z}'$ is a matrix whose columns contain the deviations from the ensemble mean. The vector $\mathbf{y}'^o$ emphasizes the need for perturbed observations. The update step for a deterministic filter is similar, except Eq. (2) is written

$$\mathbf{z}'^a_j = \mathbf{z}'^f_j - \tilde{\mathbf{K}}\mathbf{H}\mathbf{z}'^f_j, \qquad j = 1, N, \tag{4}$$

where $\tilde{\mathbf{K}}$ is a gain matrix reserved for updating the deviations from the ensemble mean. Deterministic filters do not implement perturbed observations and, therefore, are forced to have $\tilde{\mathbf{K}} \neq \mathbf{K}$.

Whitaker and Hamill (2002) cite a solution for $\tilde{\mathbf{K}}$ found in the control theory literature, and they go further to note that if the observations have uncorrelated errors, then processing the observations serially allows one to write

$$\tilde{\mathbf{K}} = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R}}}\right)^{-1}\mathbf{K}, \tag{5}$$

where $\mathbf{R}$ is the observational uncertainty expressed as an error covariance matrix and $\mathbf{P}^f$ is the forecast ensemble covariance matrix. Processing observations one at a time means that the matrices $\mathbf{R}$ and $\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}}$ are simply scalars. This is a particularly useful form for $\tilde{\mathbf{K}}$ as it renders their EnSRF algorithm computationally comparable to an EnKF algorithm. Bishop et al.'s (2001) and Anderson's (2001) algorithms can be shown equivalent to Whitaker and Hamill's EnSRF (Tippett et al. 2003), but they look different because those authors phrased their update processes slightly differently. Instead of solving for $\tilde{\mathbf{K}}$, Bishop et al. solve for $\mathbf{T}$ such that $\mathbf{P}^a = \mathbf{Z}'^f\mathbf{T}\mathbf{T}^{\mathrm{T}}\mathbf{Z}'^{f\mathrm{T}}$ and Anderson solves for $\mathbf{A}^{\mathrm{T}}$ such that $\mathbf{P}^a = \mathbf{A}^{\mathrm{T}}\mathbf{Z}'^f\mathbf{Z}'^{f\mathrm{T}}\mathbf{A}$. Regardless of the details, the important distinction is that these deterministic filters all achieve the proper EKF analysis error covariance statistics because a forecast ensemble is transformed to fit the expected analysis covariance exactly; stochastic filters attempt to achieve the proper analysis error covariance by blending the forecast ensemble spread with the observational uncertainty via explicitly perturbed observations. This paper seeks to understand the implications of this key difference.

3. Geometric interpretation of ensemble filter updating

As typical geophysical problems are hugely dimensioned, it is very difficult to visualize and intuit how data assimilation methods applied to them achieve their updates. Thus, as a starting point, we examine the update process in one and two dimensions where we can easily plot state space diagrams. We first construct one-dimensional contrived ensembles to examine the filters' updates in various linearity limits. We then apply the filters to a two-dimensional dynamic system, the Ikeda map, to examine how this behavior compounds through the assimilation process. We gauge the filters via three assessment tools: rms analysis error, ensemble rank histograms, and univariate measures of ensemble skewness and kurtosis. We will show that, consistent with the pictures from our kinematic example, both filters perform well in the linear regime; as nonlinearity becomes appreciable, deterministic filters break down earlier.

a. Update kinematics: One-dimensional example

To demonstrate how the update processes work for these two filter classes and how they behave as nonnormality/nonlinearity becomes appreciable, we examine the update processes in a contrived one-dimensional example. We generate two hypothetical ensembles, one perfectly represented by a Gaussian, a linear example, and another poorly represented by a Gaussian, a nonlinear example. We then generate a synthetic observation and update the ensembles via a stochastic filter, the EnKF, and a deterministic filter, in this case the EnSRF scheme.

We first generate an ensemble with definite higher

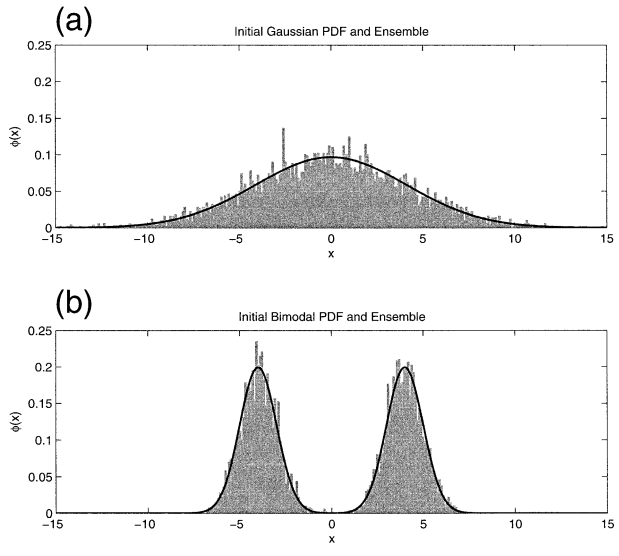

FIG. 1. (a) A Gaussian PDF with zero mean and variance $\sigma^2 = 17.0$, (black curve) and an ensemble distribution of 5000 members generated from the PDF (gray histogram). (b) A bimodal PDF with the same mean and variance as the Gaussian PDF in (a) (black curve) and a 5000-member ensemble generated from the bimodal PDF (gray histogram).

moment structure in its distribution. For this we choose a bimodal distribution comprised of two equally likely Gaussians:

$$\phi(x) = \frac{1}{2\sqrt{2\pi}}\left\{\exp\left[-\frac{1}{2}(x - \nu)^2\right] + \exp\left[-\frac{1}{2}(x + \nu)^2\right]\right\}. \quad (6)$$

Here, ϕ is the PDF and ν is the offset of each peak from 0. For the example shown here we use $\nu = 4$. This PDF has zero mean and a variance of 17, which, in turn, can define a Gaussian PDF, $N(0, 17)$. Having defined these two PDFs, we generate random numbers from them. For the example here, we generate 5000 ensemble members from each PDF.[1] Hence, we have two ensembles approximating PDFs, one Gaussian and one non-Gaussian, whose first two moments are identical. Figure 1 shows the two initial ensembles and the PDFs they represent. Clearly, by construction, only the top of the figure is sufficiently described by its first two moments.

In order to update our ensembles, observations are necessary along with an estimate of their accuracy. We suppose a hypothetical value for ''truth,'' 3.5, consistent with a state in the right peak of the bimodal PDF and also within one standard deviation of the Gaussian PDF's mean. As the updated ensemble depends on the chosen observational error size, we test three cases: one where the observational error is half the ensemble's standard deviation [$\mathbf{R} = (\sigma/2)^2$], one where the errors are equal ($\mathbf{R} = \sigma^2$), and one where the observational error is twice the ensemble's standard deviation [$\mathbf{R} = (2\sigma)^2$]. All three cases use the same observation, which for simplicity we set equal to truth; they differ only in their confidence of this observation.

Applying an EnKF and Whitaker and Hamill's EnSRF to both ensembles given the observation and a level of its presumed error yields a total of 12 updated ensembles. Figure 2 shows the six from the initially Gaussian ensemble, and Fig. 3 shows the six from the initially bimodal ensemble. The top row in each figure shows the observation with each level of its uncertainty, and the other two rows show the updates of the initial ensembles shown in Fig. 1, one graph for each observational error/filter combination. The gray histograms show the updated ensemble distributions, and the overlain solid black lines denote the distributions predicted by application of Bayes rule to each initial PDF/observational PDF pairing. These Bayes' rule distributions are ultimately what the filters are trying to approximate.

For the initially Gaussian ensemble, the respective updated ensembles in Fig. 2 from each filter are essentially indistinguable from one another and from the posterior distribution predicted by Bayes rule for all three observational error cases. Note the updated ensembles resemble the more accurate information source, that is, the initial ensemble or the observation, in each case. This is precisely what the update step of these filters is trying to accomplish, as the update should reflect the relative confidence in the contributing information sources.

For the initially bimodal ensemble, the respective updated ensembles in Fig. 3 are clearly different between filters, and each differ from the distributions predicted by Bayes rule. In the high-accuracy observation case, the observation is accurate enough to confidently choose between the initial ensemble's two peaks. The EnKF's update has erased the bimodality, and as the observational error decreases in accuracy, more of the initial bimodality is retained. This makes sense because in the low-accuracy observation case, the observational uncertainty is large enough that one cannot confidently say which of the two peaks truth is likely populating. The updated ensembles from the EnSRF, on the other hand, all show a marked bimodality, regardless of the observation's relative accuracy. This is the crux of deterministic filters, the initial ensemble is linearly transformed so that it exactly fits the EKF-predicted analysis error covariance. So while the first and second moments of the distributions sampled by the updated ensemble have the desired values, they have the added feature of retaining higher moments from the initial ensemble. While this may be useful behavior in some applications, it can also be undesirable in others. For instance, note that for

[1] Clearly, no operational implementation ever approaches ensemble sizes so large, particularly in relation to the size of the ''state vector.'' We are not trying to mimic an operational setting here, rather we are trying to understand how these filters behave unmired by implementation constraints.

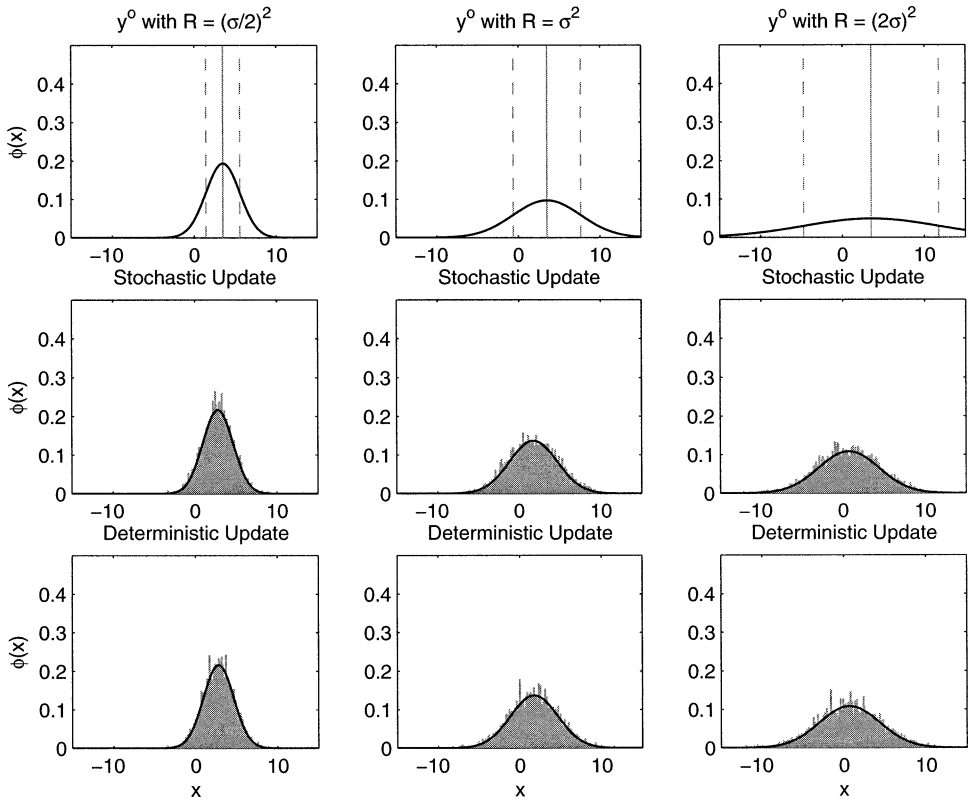

FIG. 2. (top row) The observation taken with its three different accuracies, the vertical lines showing the observation ± one observational standard deviation. (middle and bottom rows) The resulting ensemble updates from using either (middle) an EnKF or (bottom) an EnSRF with each of the different observation accuracies. The solid black lines in the bottom two rows show the updated PDFs from application of Bayes rule.

the high- and moderate-accuracy observations almost no ensemble members have the value of the observation (i.e., truth). By retaining bimodality and fitting the updated ensemble to rather accurate observations, the filter has consequently failed to sample from the most likely region for truth in the system's assumed ''state space.'' Conversely, the EnSRF does not change the initial ensemble significantly when used with the low-accuracy observation, and thus does not populate the initial probability gap at $x = 0$ as the EnKF's updated ensemble does.

These examples indicate that both filters should work very well in linear regimes. Also, inasmuch as non-Gaussian PDFs spring from nonlinear dynamics, these examples indicate that as nonlinearity becomes important, the filters can produce ensembles with members sampling regions in state space unlikely to contain truth, stochastic filters due to their random number realizations and deterministic filters due to their retaining higher-order moments. We next test these filters in a full dynamical system to confirm this suspected behavior and to gauge its consequences. We detect this behavior using ensemble assessment techniques.

b. Update dynamics: The Ikeda system

The Ikeda system is a two-dimensional, chaotic map originally posed as a model for the transmission of light by an optical ring cavity (Ikeda 1979). It has been used as a test bed for data assimilation applications before (e.g., Hansen and Smith 2001). The Ikeda system follows the equations

$$x_{i+1} = 1 + \mu[x_i \cos(t) - y_i \sin(t)], \tag{7}$$

$$y_{i+1} = \mu[x_i \sin(t) + y_i \cos(t)], \tag{8}$$

$$t = a - \frac{b}{x_i^2 + y_i^2 + 1}, \tag{9}$$

where $[x, y]^{\mathrm{T}}$ is the state vector, and a, b, and μ are parameters. The system behaves chaotically in certain parameter ranges. Figure 4 shows the attractor for the Ikeda system with $a = 0.4$, $b = 6.0$, and $\mu = 0.83$, compiled from collecting many iterated states.

Because the Ikeda system is chaotic, states initially close together in state space will tend to diverge from one another. Hence, a good, yet imperfect estimate of truth will have a limited range of useful prediction since

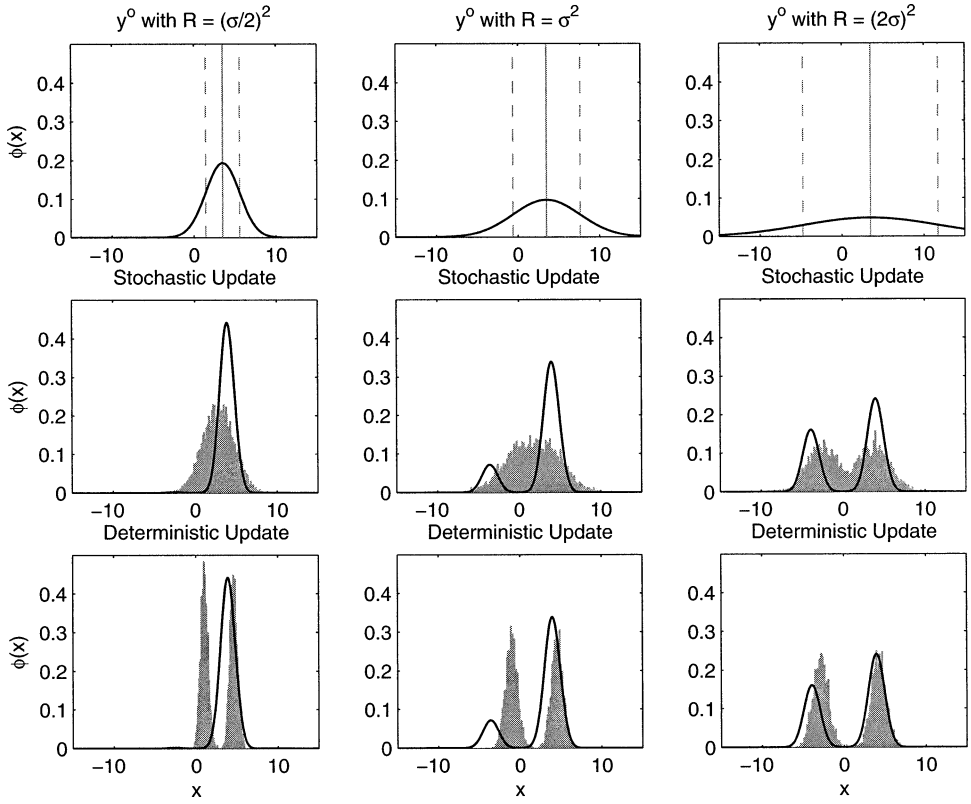

FIG. 3. This shows the same information as in Fig. 2 only for the bimodal ensemble case.

the initial error will grow. For small enough initial errors, the early stages of this growth generally follow linear dynamics, while the later stages begin to develop nonlinearly. Growth usually continues until the errors saturate at some climatological level, a magnitude typical of the attractor size in state space. Examining the states that comprise Fig. 4, we find a climatological mean and standard deviation of $(\overline{x}, \overline{y}) \pm (\sigma_x, \sigma_y) = (0.66, -0.28) \pm (0.42, 0.59)$, from which we estimate a typical climatological error size of 0.5. Note that the attractor is not Gaussian in appearance and that the climatological mean is not itself on the attractor.

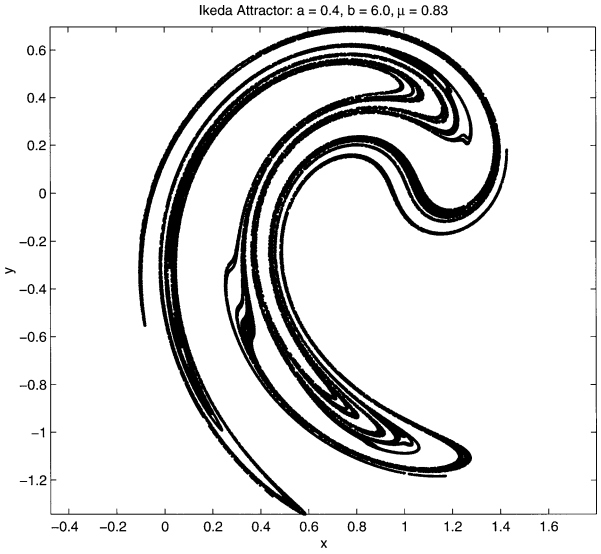

FIG. 4. The attractor for the Ikeda system for the parameters $a = 0.4$, $b = 6.0$, and $\mu = 0.83$. There are 3×2^{14} points plotted from many iterations.

In order to compare our ensemble data assimilation systems, we run observation system simulation experiments (OSSEs) in a perfect model scenario, where one freely evolving state called *truth* is used to provide synthetic observations so that an ensemble of states slightly differing from truth remains a reasonable estimate of truth's PDF via data assimilation. This is a commonly used experimental setup to test new data assimilation (DA) systems and targeted observation strategies. To avoid implementation concerns such as covariance boosting and covariance localization, we run our OSSEs with a very large ensemble. After a spinup period to allow the ensemble to equilibrate to the specific observing system, we employ ensemble assessment and verification techniques, all of which entail comparison to truth since it is available. Because we are comparing two data assimilation systems, our OSSEs have two ensemble estimates running in parallel. The estimates from both filters are updated using the same observations at each assimilation time.

In order to compare the filters' behavior in different

linearity regimes, we need to control the degree of linearity in the system's error growth, the dynamics of which are known to be state dependent. We have two main controls available, the initial size of errors and the length of the integration over which the errors evolve. Thus, experiments considered here will differ in both observation time (i.e., the amount of time between assimilations) and observation system design. We choose to observe both state elements directly ($\mathbf{H} = \mathbf{I}$) at each assimilation time; this essentially reduces varying observation system design to varying the relative accuracy of observations taken.

Since the Ikeda system is a map, the effective time step is fixed as a single map iteration. In our chosen parameter range, calculations show a Lyapunov number of about 1.6, meaning one time step is about four-fifths the average error doubling time, a relatively large growth per time step. Hence, the observation time is generally kept at a small number of map iterations, between 1 and 10 steps inclusive. We choose to always observe both state components with equal and uncorrelated accuracy, that is, observational uncertainty isopleths are circles. We vary the level of observational accuracy over several orders of magnitude, from 1.375×10^{-6} to 1.375×10^{-2}. In the results shown later, all OSSEs have been performed with an ensemble size of 511. Other ensemble sizes were tested, and the results do not change significantly for larger sizes.[2] Verifications and statistics were compiled over 10 000 successive assimilations for each of several different initial ensemble realizations.

1) Ensemble snapshots

As the Ikeda system has a two-dimensional state space, it can be easily visualized. Examining individual cases of initial ensembles versus updated ensembles reveals several general aspects of the filter updating process. The first is that ensembles are observed to disperse mainly along attractor branches. Dispersion along a primarily straight-line direction in state space leads to covariance isopleths that have high eccentricity (i.e., covariance matrices have high condition numbers). Covariance isopleths have less eccentricity if there is cross-attractor branch scatter or if there is curvature in the underlying attractor. The former case is typically a product of DA pulling ensemble states off the attractor with the observation time not being long enough for all states to find the attractor before undergoing DA again. The latter case occurs whenever ensemble dispersion is comparable to the curvature length scale of the local attractor structure, for example, when the ensemble has relatively large dispersion or when the attractor has relatively sharp curvature. There are a few areas of sharp curvature readily apparent in Fig. 4, and because the attractor is fractal, we expect there to be similarly proportioned areas with large local curvature at all scales. Though these areas are relatively sparse, truth and its ensemble estimates should still visit them intermittently.

Ensemble dispersion along a straight line in state space is well approximated by a Gaussian PDF. Ensemble dispersion with cross-attractor branch scatter can be well approximated by a Gaussian PDF, especially if the scatter is symmetric about the attractor branch. Ensemble dispersion along a curved branch of the attractor will in general not be well approximated by a Gaussian PDF. Large deviations from a Gaussian PDF violate the filters' assumptions, and we thus begin to suspect problems in these situations. Common measures of departure from a Gaussian distribution are skewness and kurtosis. Skewness is a nondimensional measure of a distribution's asymmetry, and kurtosis is a nondimensional measure of a distribution's peakedness, that is, how long its tails are compared to a Gaussian's. There are several common definitions for these quantities, but all entail normalizing some measure of a distribution's higher moments by an appropriate power of the distribution's standard deviation. See appendix A for details on the definitions used here. While multivariate extensions to these ideas exist (Mardia 1970), they can be quite expensive to calculate for large-dimensioned systems being run with large ensemble sizes. Also, these multivariate measures are difficult to interpret and to verify. Hence, we find it more illuminating to consider the skewness and kurtosis for a single state element within the ensemble. Such measures can easily be assessed within any model for most reasonably sized ensembles. Note that having the skewness and kurtosis of a single element be consistent with a Gaussian distribution is not sufficient to assert the ensemble distribution as a whole is Gaussian, but that a single element not being consistent with a Gaussian distribution necessarily means the ensemble distribution as a whole is not Gaussian.

By calculating the skewness and kurtosis for x over our ensembles, we find that the stochastic filter (EnKF) keeps its ensemble estimates nearly Gaussian for a wide range of parameter space, whereas the deterministic filter (EnSRF) maintains estimates well approximated by Gaussians only when the observation time is one step and the observational error is 1.375×10^{-5} or less. When this is the case, the observations are accurate enough to keep the estimate very near the attractor (no cross-attractor branch scatter) and the ensemble dispersion is so small that it hardly ever finds itself in a region with attractor curvature at a comparable scale. A comparison of the EnKF update and EnSRF update for such a case, with an observation time of one step and an observational error of 1.375×10^{-6}, is shown in Fig. 5. The top half of Fig. 5 shows the initial ensemble (dark gray dots) and the updated ensemble (light gray dots) for the EnKF, and the bottom half shows the same information for the EnSRF. The position of truth and the observations used to update the ensembles have been

[2] Again, the very large ensemble size is for demonstration purposes only, not an attempt to mimic an operational setting.

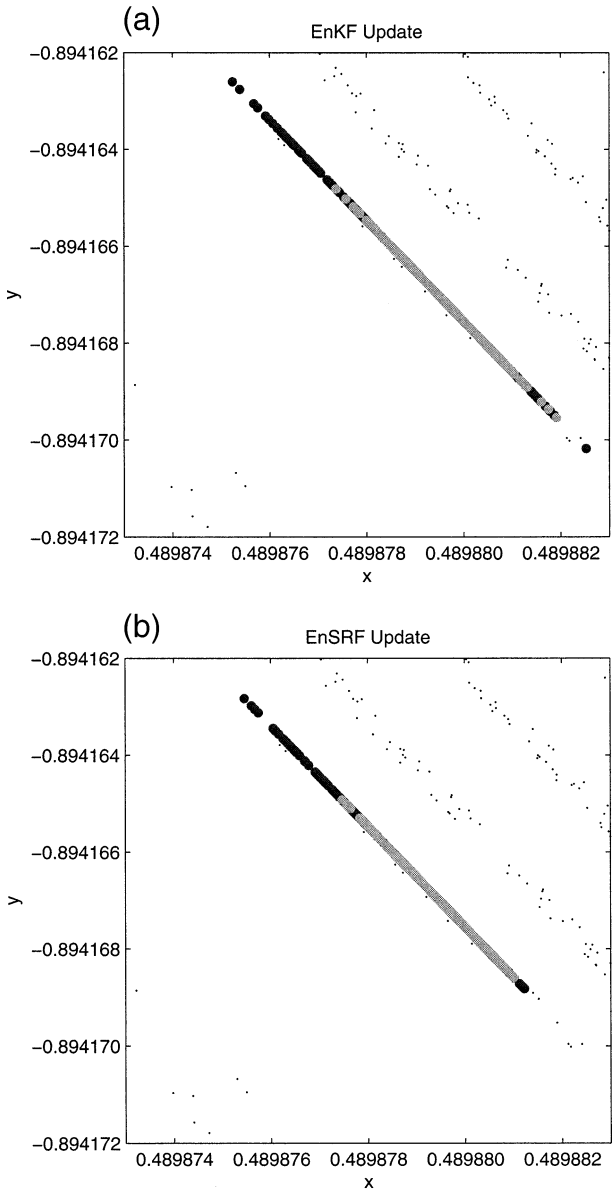

FIG. 5. Ikeda system in a linear regime: (a) a snapshot from an OSSE using the EnKF in a linear regime (observation time is one step and observational error is 1.375×10^{-6}). The small black dots are locations on the attractor, the dark gray dots the initial ensemble, and the light gray dots are the updated ensemble. (b) The same information only for the EnSRF's update.

omitted, but locations on the underlying attractor are plotted as small black dots. Note the ensemble spread is contained within a state space area of 1×10^{-5} by 1×10^{-5}. Though the Ikeda system has a two-dimensional state space, the ensemble spread and updates are essentially occurring along lines.

As we allow the observation time to lengthen and the observational error to become less accurate, we find the deterministic filter's ensemble estimates readily diverge from Gaussian PDF's. Figure 6 succinctly summarizes the absolute value of ensemble skewness over many different OSSEs. Because skewness is related to an odd central moment, it can easily be positive or negative depending on what side of the mean the farthest outliers are, thereby making the mean skewness over many realizations nearly zero. Hence, we consider the absolute value of the skewness. To characterize the values and overall distributions that ensemble skewness attains within each OSSE's time series, we employ box plots demarcating the maximum and minimum values, the median value, and the interquartile range. There are two box plots in each of the 12 Fig. 6 graphs, the left one is for the EnKF and the right for the EnSRF. Each graph shows the results from a different combination of observation time and observational error. The dotted line at the top of each graph shows the maximum possible value ensemble skewness can attain for the given ensemble size (for a discussion of how these maximum values are calculated, see appendix A). We omit a similar figure for kurtosis measures because they are largely redundant, that is, the sources of large skewness are also sources of large kurtosis.

Aside from showing the departure from Gaussianity, these box plots show that in appreciable nonlinear regimes the deterministic filter's ensembles achieve skewness and kurtosis near the maximum values that ensembles their size can attain, sometimes quite frequently (e.g., the case with an observation time of three steps and an observational uncertainty of 1.375×10^{-4}). This means that the ensembles are approaching or attaining a configuration where nearly all members agree on a specific value for x with just a few outliers disagreeing. Examining individual cases confirms this, as seen in Fig. 7, a snapshot from an OSSE with an observation time of three steps and an observational uncertainty of 1.375×10^{-3}. The bottom part of Fig. 7 shows the initial and updated ensembles of the deterministic filter. The far-left dark gray dot and light gray dot each constitute 510 ensemble members (all but one) occupying such close proximity that they each appear as one dot. The sole far outliers allow the ensembles as wholes to maintain fairly accurate means and to match the EKF-predicted analysis error covariance.

Clearly, the deterministic filter's ensembles at the bottom of Fig. 7 have highly non-Gaussian structures, and monitoring the filter's behavior over time shows that the ensembles achieve these configurations often. But the ensembles are not always in such configurations. Indeed, they do sporadically approach Gaussian PDFs as well as many intermediate configurations. One telling example is shown in Fig. 8, a snapshot from an OSSE with an observation time of three steps and an observational error of 1.375×10^{-2}. Note the initial ensembles in the two parts of Fig. 8 are similarly dispersed but that the updated ensembles are quite different. This is very similar to the one-dimensional example earlier where the stochastic filter tended to fill out the desired analysis covariance structure and the deterministic filter

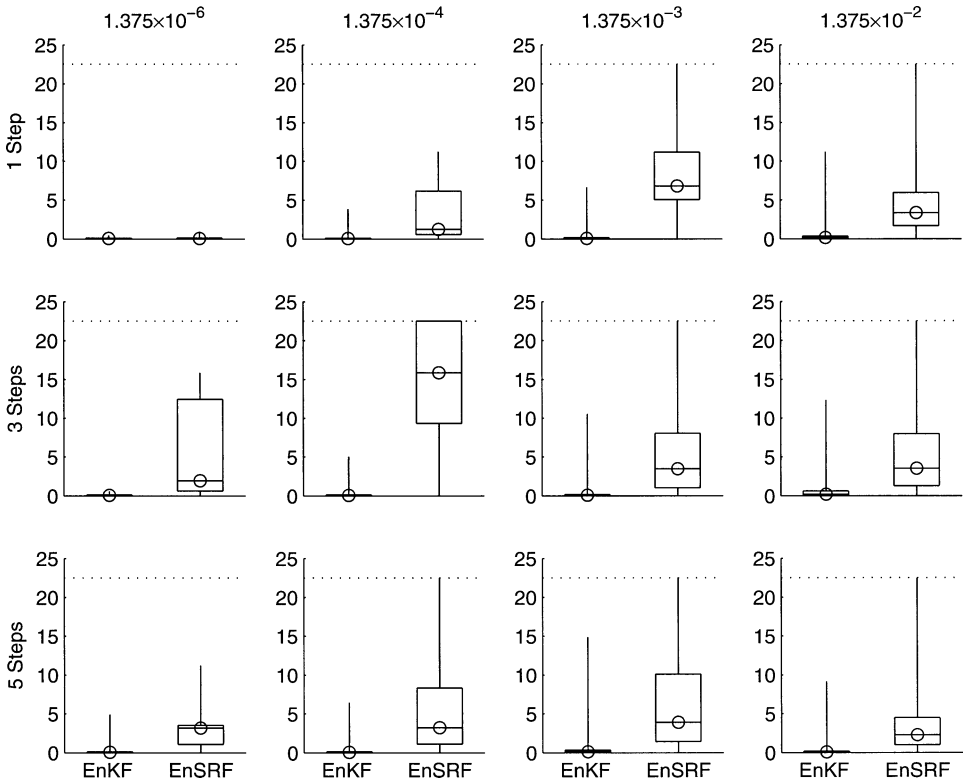

FIG. 6. Ikeda system skewness: box plots summarizing the time series for the absolute value of ensemble skewness for each filter type over 12 different OSSEs. Each graph corresponds to a given pairing of observation time (one value for each row) and observational error (one value for each column). The left-hand box plot in each graph refers to the EnKF and the right-hand plot is for the EnSRF. The centerline of each box plot shows the median value of the absolute value of ensemble skewness over the 10 000 assimilations and is further marked with a circle. The box's bounds show the interquartile range, and the whiskers outside the bound extend to the min and max values found over the assimilations. The horizontal dotted lines show the max possible value of ensemble skewness when using 511 ensemble members.

fit the ensemble exactly to that structure. In this case we see many ensemble members have been transformed well off the attractor, most especially the cluster of members already at the far-right edge of the deterministic filter's ensemble. Repeated updates such as this by the EnSRF lead quickly to small groups of far outliers which, when the ensemble is forced to have the desired error covariance, necessitates that the remaining members collapse onto one another so as to counterbalance the outliers.

2) ENSEMBLE ASSESSMENT

The three main tools utilized and presented here for assessing ensemble filter behavior are rms analysis errors, ensemble rank histograms, and univariate measures of ensemble skewness and kurtosis. Rms errors measure the accuracy of the ensemble mean as an estimate of truth. Ensemble rank histograms assess the reliability of the probabilistic estimation. And as stated earlier, skewness and kurtosis are standard measures for detecting departures from normality.

Rank histograms, also known as Talagrand diagrams, are used to assess whether truth and the ensemble all share a common PDF, that is, whether truth is statistically indistinguishable from our spanning ensemble estimate. If an ensemble filter is performing properly, its members ought to be unbiased and be sampling from a probability distribution of which truth is a member. If this is the case, then probabilistic analyses and forecasts are considered reliable, for example, events of 70% likelihood will verify 70% of the times predicted. To construct a rank histogram from an ensemble, one first chooses an index over which to sort the ensemble, typically a single element of the state vector. Over many successive assimilations, one then finds where the verification ranks in that sorted list, and keeps tally in the bins corresponding to those ranks. If there are N ensemble members, there are $N + 1$ bins to consider including the two end bins where verification is either greater than or less than the entire ensemble. Under the premise that the ensemble and truth are all members of the same PDF, then there should be no discernible preference to their respective ranks. Hence, reliable probabilistic data assimilation and forecasting systems should yield uniformly distributed rank histograms.

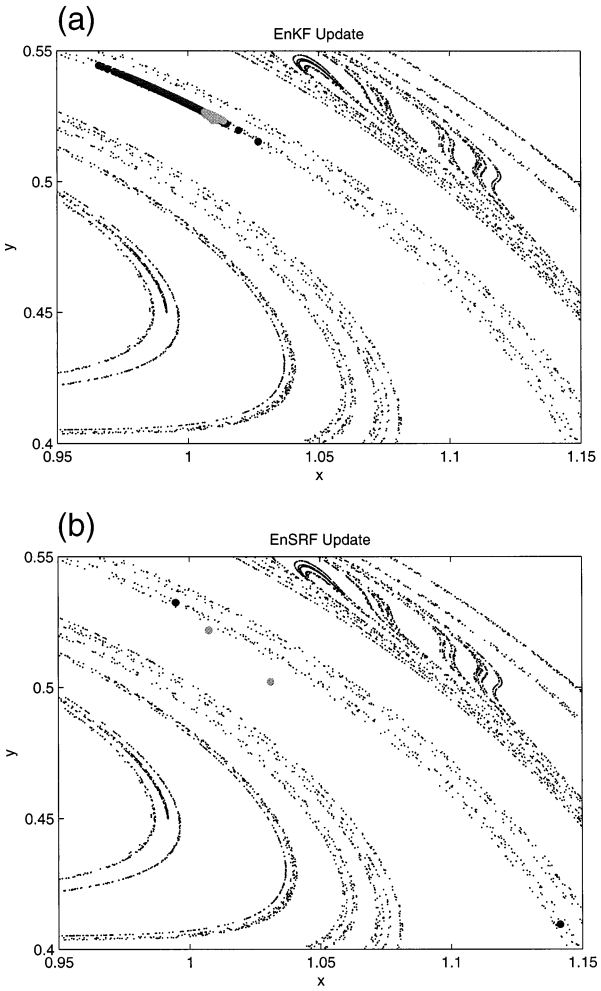

FIG. 7. Ikeda system in a nonlinear regime: (a) a snapshot from an OSSE using the EnKF in a nonlinear regime (observation time is three steps and observational error is 1.375×10^{-3}). The dots mean the same as in Fig. 5. (b) The same information only for the EnSRF's update. The far-left ensemble dots actually each contain 510 ensemble members (all but one).

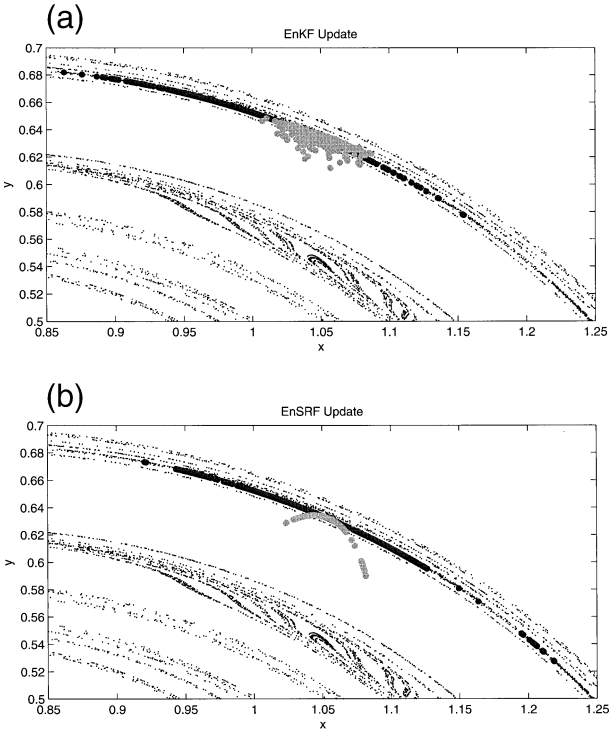

FIG. 8. Ikeda system in a nonlinear regime: (a) a snapshot from an OSSE using the EnKF in a nonlinear regime (observation time is three steps and observational error is 1.375×10^{-2}). The dots have mean the same as in Fig. 5. (b) shows the same information, only for the EnSRF's update.

Since each of the $N + 1$ bins should be equally probable, then without loss of generality, one is free to group neighboring bins (with the caveat that each grouping must be comprised of an equal number of bins) or to subsample from the ensemble, for example, only rank a fraction of the full N members. Nonuniformity in rank histograms bespeaks a number of potential problems, usually concerned with ensemble biases and spread (e.g., Hamill 2001). A multidimensional extension to ensemble rank histograms exists, namely minimal spanning trees (e.g., Smith and Hansen 2004), but we find these univariate measures are adequate for demonstrating the necessary points. It should also be noted that in their respective papers, Anderson (2001) and Whitaker and Hamill (2002) both chose to examine ensemble reliability by comparing the rms error of the ensemble mean to the mean of the rms errors for each ensemble member. Anderson cites the expected value this ratio ought to have if the ensemble variance is accurately representing the error in the ensemble mean. We have evaluated this ratio for several of our OSSEs and find that it confirms the behavior we observe with our assessment techniques, but being only a number, does not lend the visual interpretation that the rank histograms give.

Figure 9 shows the rank histograms for both the EnKF and the EnSRF applied to both a linear/Gaussian case, the left column, and a ''slightly nonlinear''/non-Gaussian case, the right column. The linear case is for that shown in Fig. 5, where the observation time is one step and the observational uncertainty is 1.375×10^{-6}; the slightly nonlinear case uses an observation time of one step and an observational uncertainty of 1.375×10^{-4}. The results from all 10 000 assimilations are displayed. In each case we are comparing the true value of x to the analyzed ensemble's values, not to a forecast. Since we are using an ensemble size of 511, there are 512 bins to tally for the rank histograms, though we group neighboring bins such that the graphs in Fig. 9 only display 32 bins. The tallies have been converted to probabilities of ensemble rank. The top row shows the rank histograms for the EnKF. The solid horizontal line shows the expected mean value for each bin were the distribution actually uniform, and the dashed lines show the

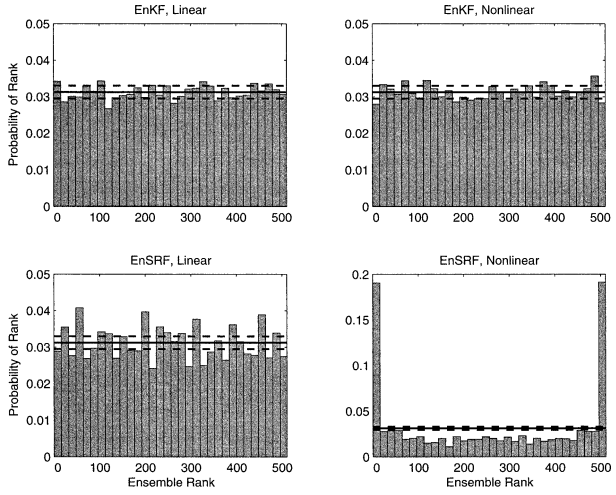

FIG. 9. Ikeda ensemble rank histograms: rank histograms are based on an ensemble size of 511, and hence each of the 32 bins shown are groupings of 16 underlying neighboring bins. Histograms are comprised of 10 000 members, and the bin counts have been converted to probabilities of rank. The solid horizontal lines show the expected value for each bin were the histograms actually uniform, and the dashed lines show the expected std dev from uniformity for the bin totals given that the histograms are comprised of a fixed number of random samples. (left column) The results from a linear OSSE (observation time is one step and observational error is 1.375×10^{-6}) and (right column) the results from a slightly nonlinear OSSE (observation time is one step and observational error is 1.375×10^{-4}). (top row) The results from the EnKF and (bottom row) from the EnSRF. Note that the std dev lines are not strictly applicable for the EnSRF results because the tallies are not independent ranks.

expected standard deviation from uniformity for the bin totals given that it is comprised of a fixed number of random samples.

Considering the linear case, the EnKF's rank histogram seems consistent with a uniform distribution as approximately two-thirds of the bin totals are within the expected standard deviation. Figure 9 (bottom left) shows the rank histogram for the EnSRF. Its bin totals show a larger variance in their totals. This is expected because, as cited in Anderson (2001), the tallies comprising the histogram are not independent due to the deterministic nature of the update imparting a rank correlation from one assimilation to the next. Therefore, the standard deviation lines plotted are incorrect and should reflect fewer independent degrees of freedom. Still, there are no discernible features to make us think the ensemble is biased or not capturing the proper variability. We conclude that it also seems consistent with a uniform distribution. Inspecting the time series for rms analysis errors (not shown), we see that both filters give essentially indistinguishable analyses for all 10 000 assimilations. Similarly, as indicated by the ensemble skewness box plots in the upper-right-hand portion of Fig. 6, the ensemble estimate distributions, at least for state element x, are never far from Gaussian. From these ensemble assessment measures we conclude that in this linear range, both ensemble filters are performing accurately and reliably.

The right column of Fig. 9 shows the rank histograms for the slightly nonlinear case. The EnKF rank histogram is still approximately uniform, but the EnSRF rank histogram is not. The end bins of the EnSRF rank histogram are heavily populated, meaning that truth is often ranked near the ensemble edge members or outside the ensemble completely. On its own, this result is difficult to interpret beyond asserting that it is not uniform and, therefore, not sampling from the same PDF that truth is following. Hamill (2001) notes that ''U-shaped'' rank histograms are typically a sign that the ensemble is undispersive; however, we know here that the ensemble is forced to have the desired analysis error covariance, so that is not the cause per se. A U-shaped rank histogram is consistent with the idea of periodically spawning far outliers. This is simply because a structured, non-Gaussian distribution can generally have far outliers from its mean, and hence, even if the ensemble mean is a very good estimate of truth, many of the ensemble members may systematically be to one given side of the verification. The box plots for this case, shown in the second panel in the top row of Fig. 6, indicate that the ensembles in this parameter range are often non-Gaussian, but the measures never approach their maximum values, hence we conclude that ensemble clumping and outlier generation are probably not so severe.

That an OSSE can give a U-shaped ensemble rank histogram yet give analyses with the expected rms error statistics indicates it is possible to form unreliable analyses that are still accurate. This behavior does not pervade much of parameter space; however, not properly sampling from truth's PDF will eventually degrade an estimate's accuracy. For instance, the nonlinear case shown in Fig. 8 has very similar ensemble rank histograms (not shown) to the ones for the slightly nonlinear case shown in the right-hand side of Fig. 9. However, the time series for rms analysis error, skewness, and kurtosis are quite different between these two OSSEs, with the fully nonlinear case averaging rms errors above the observational uncertainty level and skewness and kurtosis often approaching their maximum values (see Fig. 6). These indicate that even modest departures from normality can lead to U-shaped ensemble rank histograms. Conversely, it seems a filter can maintain appropriate levels of accuracy even when its ensembles have modestly departed from Gaussian distributions. By comparing the skewness and kurtosis measures from widely differing OSSEs using the EnKF to those using the EnSRF, we see the EnKF is able to keep its analyzed ensembles fairly close to Gaussian PDFs. Ensemble estimates from the EnSRF readily depart from Gaussian PDFs, and when they do their deviations from normality are kept through the filter's transformations and can easily derail our probabilistic interpretation of the ensemble via ensemble rank histograms. The stochastic filter is able to produce an expected uniform rank histogram

because its perturbed observations are acting sufficiently well as a Gaussian PDF repopulation device.

By plotting full-state space representations of our ensembles, we have gleaned a geometric interpretation of the ensemble filter update mechanisms. By running OSSEs with the Ikeda system, we were able to collect rms error statistics, compile rank histograms, and examine measures of skewness and kurtosis. From these we have argued that the geometric interpretation is consistent with these ensemble assessments. Now we turn to higher-dimensioned, spatially extended systems to see if this ensemble assessment behavior carries over when the filters are used in various linearity regimes.

4. Ensemble filter behavior in higher-dimensioned systems

We would like to know whether the insight gained from our simple 2D experiments extends to higher-dimensioned systems. Given that we cannot plot the state space for systems much over two dimensions, we look to show consistency with this insight by way of the ensemble assessment signatures discussed for the Ikeda system. We examine rms errors and ensemble rank histograms, again bolstered by skewness and kurtosis measures, for two models with differing complexity: Lorenz's one-latitude circle model and a two-dimensional barotropic form of the quasigeostrophic Harvard Open Ocean Model. By judiciously selecting the relevant parameters, we can place the filtering process in any desired regime along a linearity continuum. The results that follow generally compare linear cases to weakly nonlinear cases as they seem most relevant to real world applications, though we have explored some strongly nonlinear cases as well. The degree of nonlinear error growth can be sensibly compared between different systems, including the real world, by reference to typical error magnitudes and typical error doubling times. Ehrendorfer (1997), and references within, estimate the atmosphere's error doubling time as 2 ± 0.5 days. While we do not make strict numerical comparisons here, the success of running data assimilation cycles every 6 to 12 h for NWP models with the given levels of observational uncertainty present suggests a similar range in which to run the experiments here, a range we may term weakly nonlinear in error growth.

a. Lorenz latitude circle model

The first spatially extended model we consider is Lorenz's one-latitude circle model, often referred to as Lorenz 95 (or sometimes 96) (Lorenz 1995). It is a highly simplified model that governs the evolution of a variable around a latitude circle or a re-entrant channel. This model has been used by Lorenz and Emanuel (1998) and Hansen and Smith (2000) to explore strategies in targeted observations and more recently by Anderson (2001) and Whitaker and Hamill (2002) as test beds for their respective deterministic filter schemes. The model is well documented in these references, so we refrain from repeating many details or showing a figure of a typical state, though minimum detail is included in appendix B.

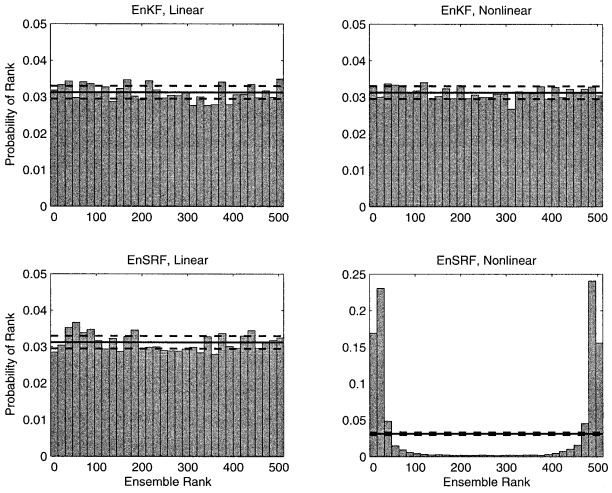

FIG. 10. Lorenz 95 ensemble rank histograms: The same as Fig. 9 except applied to the Lorenz 95 system. The linear case has an observation time of 12 h and an observational uncertainty of 2.0×10^{-4}, and the nonlinear case has an observation time of 12 h and an observational uncertainty of 0.2.

For our OSSEs we use an ensemble size of $N = 511$, and, as with the Ikeda system, we choose to observe all state elements at each observation time. We again choose to control the degree of linearity in the problem by varying both the observation time length and the observational error size. We have performed experiments with observation times ranging from 6 to 48 h (using the conversion from model time to ''atmosphere time'' suggested by Lorenz and Emanuel (1998) of 0.05 model time = 6 h) and observational error scales ranging from 2.0×10^{-4} to 2.0. As with the Ikeda system, we perform 10 000 assimilations per OSSE. We compile rank histograms and skewness and kurtosis data based on the analyzed values for state element X_1.

Figure 10 shows the same rank histograms as Fig. 9, only for the Lorenz 95 system. Here the linear/Gaussian case has an observation time of 12 h and an observational uncertainty of 2.0×10^{-4}. As with the Ikeda system, both filters give essentially uniformly distributed rank histograms. The skewness and kurtosis measures (not shown) confirm that neither filter is producing an ensemble that differs significantly from a Gaussian distribution. Further examination shows that both filters produce essentially indistinguishable analysis rms error time series (not shown). Based on these measures, both filters seem to be giving accurate and reliable probabilistic analyses. The nonlinear/non-Gaussian case in Fig. 10 also has an observation time of 12 h but now the observational error is 0.2. We see the deterministic filter's rank histogram shows a similar U-shaped struc-

ture to the one we saw for the nonlinear case in the Ikeda system (lower-right-hand corner of Fig. 9). Note that bins 2 and 31 are the most populated bins, as opposed to the edge bins for Ikeda. This is not too surprising as it indicates that our ensemble is still generally bounding truth, just preferentially within either end. It stands to reason that our ensembles must be bounding truth a majority of the time as the rms analysis errors hardly differ from those of the EnKF's estimates (not shown). However, seeing that these ensembles attain rather high values of skewness and kurtosis (not shown), then by our geometric reasoning discussed earlier and based on outliers, it makes sense that truth often falls near the ensemble ends.

b. Double-gyre barotropic ocean model

To test consistency in a larger, two-dimensional, spatially extended geophysical model, we examine the filters' behavior over OSSEs with a one-layer version of the Harvard Open Ocean Model (Haidvogel et al. 1980; Miller et al. 1983), a quasigeostrophic model. We run the model with zero buoyancy at the top and bottom boundaries, making the dynamics de facto barotropic. Potential vorticity is lost through bottom and lateral friction as well as from numerical filtering. The model is forced by a surface Ekman pumping to give the classic basin double-gyre western boundary current flow field. The version of the model we used is solved with finite elements and a second-order Adams–Bashforth time stepping routine. The model is well documented in its original references as well as more recently in an appendix of Moore et al. (2002; interested readers should seek further details there).

The system's state dimension is 2145 (=one prognostic variable at 33×65 grid points). For our OSSEs, we use $N = 511$ ensemble members, but we only observe at 29 locations, all fixed locations chosen to be near dynamically interesting places, either along the western boundary current or at the separation between the gyres. Though it is unphysical, we choose to observe potential vorticity directly so as to minimize complications in the observation operator. Again, we choose to vary observational uncertainty and observation time length to control the degree of linearity. We have swept quickly through the linearity range to find suitable cases to study. We find that errors seem to grow quite linearly using an observational uncertainty that is 5% of a climatological error level and an observation time of 2 days. The slightly nonlinear case we consider here uses 3 times this observational uncertainty and an observation time of 10 days. The results below consider 10 000 assimilations per OSSE.

A figure of the retrieved rank histograms is omitted because it is largely redundant—it looks indistinguishable from the analogous figure compiled for the Ikeda system, Fig. 9, with roughly uniformly distributed rank histograms for both filters in the linear case and for the stochastic filter in the nonlinear case, and with a U-shaped histogram (peaks in the end bins) for the deterministic filter in the nonlinear case. Whereas earlier with the simpler models, where we observed every state component at each assimilation time, we saw that the two filters produced essentially indistinguishable ensemble means from one another within the systems' respective linear regimes, we do not find this here with the ocean model. However, it is interesting to note that the filters' average rms analysis error statistics are identical. So, as earlier, we find that both filters are producing equally accurate updated ensembles for the linear regime. Skewness and kurtosis measures complement these indicators in the expected manner showing that both filters are producing ensembles closely resembling Gaussian PDFs.

In the nonlinear/non-Gaussian case, the rms analysis error statistics do not differ that significantly, the stochastic filter is about 10% more accurate on average over the 10 000 assimilations considered. The skewness and kurtosis time series show that the stochastic filter maintains a very Gaussian ensemble estimate whereas the deterministic filter's ensemble estimates are frequently non-Gaussian, though the measures never approach their maximum values.

The behavior we observe over the range of linearity for this barotropic ocean model is similar to that which we observed with the Ikeda system and the Lorenz 95 system. We again conclude from our ensemble assessment tools that the filters' behavior in this geophysically relevant model is consistent with the geometric interpretations devised from the simple systems.

5. Conclusions

We have sought to understand how and why stochastic and deterministic ensemble-based data assimilation systems can differ in their behavior. From geometric considerations in one- and two-dimensional state spaces we have confirmed that both filters behave as expected and perform well in regimes with linear error growth dynamics. In the linear range, the filters are simply combining two Gaussian PDFs, the ensemble distribution and the observations, in the near-optimal fashion theory predicts the EKF will accomplish. In these small state spaces we can visually confirm that the ensemble distributions are in fact Gaussian in shape. Measures of the skewness and kurtosis of the ensemble of course confirm this. By performing OSSEs with the Ikeda system we were able to show that both filters produce updated ensembles that are consistent with the PDF of which truth is a member.

Visualizing ensemble distributions in these small state spaces has also shown how ensemble estimates of truth can be affected by nonlinear error growth dynamics. By allowing errors to grow to moderate amplitudes, initial Gaussian ensembles can become non-Gaussian. When used with sufficiently accurate observations, the random

sampling necessary in stochastic filters allows the updated ensemble to assume much of the Gaussian form of the observations, thereby tending to erase the non-Gaussian higher moments nonlinear error growth has generated. These updated ensembles look Gaussian when plotted, and skewness and kurtosis measures confirm they have only small departures from normality. OSSEs run with the Ikeda system show that this tendency for the updated ensembles to repopulate Gaussians is effective at keeping its estimates consistent with truth's PDF, as told by their uniformly distributed ensemble rank histograms.

Deterministic filters, on the other hand, readily maintain non-Gaussian moments in their ensemble distributions by virtue of their direct transformations. We have observed that successive updates of non-Gaussian ensembles can shear the edge ensemble members away from the mean, particularly as they are transformed well off the system attractor. These edge members can easily become far outliers, which, under enforcement of having the EKF-predicted analysis error covariance, can lead the main ensemble cluster to tighten. For cases with fully nonlinear error growth, we observe the updated ensembles to often assume configurations with almost the entire ensemble having collapsed onto one state so as to counterbalance the very few (or even sole) distant outliers. This distribution approaches the limit of maximum ensemble skewness and kurtosis, that is, maximum non-Gaussianity. The outliers give the data assimilation system the degrees of freedom it needs to match the desired error covariance. Because the ensemble distributions are often skewed and leptokurtic (having positive kurtosis), truth often ranks near the edge members, between the outliers and the main cluster. This is seen to generate U-shaped ensemble rank histograms, indicating that deterministic filter ensemble estimates in nonlinear regimes are not sampling from PDFs that are consistent with truth.

While this geometric interpretation is satisfying, it is based on a contrived one-dimensional example and a two-dimensional chaotic map, and does not necessarily apply to the large and spatially extended models forecasters and geophysicists often care about. Hence, we have extended this interpretation to two higher-dimensioned models through the ensemble assessment measures of rms analysis error statistics, ensemble rank histograms, and univariate measures of ensemble skewness and kurtosis. We have found that the retrieved measures in these more complex models are in accordance with the expectations formed from the simple model. These indicate an ability on the stochastic filter's part to handle nonlinearity better than the deterministic filter, or at least over a larger parameter range.

We stress that this is not judgment on a ''better'' filter, it is simply a consequence of the formulations that comprise the filters. Each individual implementation of an ensemble filter will in a sense determine which aspects of its behavior are important. Also, armed with this insight, something might be able to be done to address these problems. Given that deterministic filters having trouble with nonlinearity seems to spawn from generating non-Gaussian ensemble distributions, one can imagine alleviating these problems in various ways like, for example, periodically resampling the ensemble in the ''bootstrapping'' fashion suggested by Anderson and Anderson (1999), essentially using a kernel density approach borrowed from particle filtering.

We have purposefully considered only large ensemble sizes so as to avoid some of the ''messy'' considerations of operational implementation such as ensemble covariance localization and covariance inflation. Even so, we feel the preceding analysis has begun to shed light on such an understanding, and that the issues raised may become direct concerns for numerical weather prediction in the years to come as computing power allows implementation of very large ensembles. Similarly, we have also restricted ourselves to consideration of the perfect model scenario to hone our focus on the pure implications of the filter formulations. Addressing model error is a burgeoning area of research and remains a difficult complication to treat. Clearly, model error is too large of a consideration to generalize over except to say that the results depend on the type of error including how it is compensated for. We have found the inclusion of model error could possibly change the earlier results and interpretations, depending on how the error is handled. If model error were to be addressed by the inclusion of a model error covariance term, typically written as $\mathbf{Q}$, its stochastic nature would tend to erase higher-order moments and render subsequent analyses more Gaussian. If $\mathbf{Q}$ is sufficient to compensate for the model error present and if it can keep the filters' estimates from diverging, then the deterministic filter's ensembles are found to be more reliable than in the earlier perfect model scenarios. However, if one were to address model error by simply including a large ensemble covariance inflation factor, then the geometric argument is generally found to hold since covariance inflation does not change the structure of ensemble distribution, only the members' distances from the ensemble mean.

The key to handling or avoiding undesirable behavior in these filters is simply understanding them and their limitations. To this, we note that even though we have studied a model hierarchy, we have not approached the dimensionality and complexity of modern-day numerical weather prediction models. That these interpretations extend to those systems is likely, but impossible at this point to assert.

Acknowledgments. W.G.L. wishes to thank Jake Gebbie for his help with the numerics to confirm the results in appendix A and for his general encouragement along the way. W.G.L. also wishes to thank Andrew Moore for help with the double-gyre ocean model and Tom Hamill for helpful comments on this manuscript. The

authors gratefully acknowledge the financial support of the National Science Foundation ITR Grants ACI-0121182, and ATM-0216866, and ONR YIP Grant N00014-02-1-0473.

APPENDIX A

Skewness and Kurtosis

a. Definitions

There are several common definitions for skewness and kurtosis, but all entail normalizing some measure of a distribution's higher moments by an appropriate power of the distribution's standard deviation. Here, we define skewness as the unbiased estimate of a distribution's third central moment divided by the cube of its standard deviation. Kurtosis is defined as the unbiased estimate of a distribution's fourth central moment divided by the fourth power of its standard deviation less three. Consider a univariate ensemble z_j where $j = 1, N$. The mean, variance, and the third and fourth central moments of this ensemble are

$$\bar{z} = \frac{1}{N}\sum_{j=1}^{N} z_j,$$

$$\sigma^2 = \frac{1}{N-1}\sum_{j=1}^{N} (z_j - \bar{z})^2,$$

$$\mu_3 = \frac{1}{N-1}\sum_{j=1}^{N} (z_j - \bar{z})^3,$$

$$\mu_4 = \frac{1}{N-1}\sum_{j=1}^{N} (z_j - \bar{z})^4.$$

These give the following expressions for skewness and kurtosis:

$$\text{skewness} = \frac{\mu_3}{\sigma^3},$$

$$\text{kurtosis} = \frac{\mu_4}{\sigma^4} - 3.$$

As a Gaussian distribution is symmetric, all odd central moments are zero, and hence the expected skewness for a Gaussian is zero. All even central moments for a Gaussian are completely determined by the value of its variance. A univariate Gaussian's fourth central moment is $3\sigma^4$. Hence, our definition of kurtosis subtracts three so that a Gaussian has zero kurtosis.

b. Maximum values

As skewness and kurtosis are nondimensional quantities, it turns out that there are maximum values they can attain for a finite ensemble size N. The distribution that leads to these maximum values may be devised intuitively. Proving this, however, is a nonlinear optimization problem most easily solved by numerical methods. We offer the following as the distribution with maximal skewness and kurtosis:

$$z_j = \begin{cases} z_0 & \text{if } 1 \le j \le N-1 \\ z_1 & \text{if } j = N. \end{cases}$$

This distribution gives the following central moments:

$$\bar{z} = \frac{(N-1)z_0 + z_1}{N},$$

$$\sigma^2 = \frac{1}{N}(z_1 - z_0)^2,$$

$$\mu_3 = \frac{N-2}{N^2}(z_1 - z_0)^3,$$

$$\mu_4 = \frac{N^2 - 3N + 3}{N^3}(z_1 - z_0)^4.$$

Using these with our earlier definitions of skewness and kurtosis produces the following expressions:

$$\text{skewness} = \frac{N-2}{\sqrt{N}},$$

$$\text{kurtosis} = \frac{N^2 - 6N + 3}{N}.$$

Note that these values do not depend on the values of z_0 and z_1, only on the ensemble size. Our claim is that these are the maximum values of skewness and kurtosis that a univariate distribution of N members can attain (as skewness includes an odd moment, it can just as easily be negative, in which case the preceding is also the minimum possible value for skewness). This has been confirmed through exhaustive searching for relatively small distribution sizes (five members for skewness only and six members for both measures). Though this does not constitute a proof, the retrieved cost functions appear rather smooth, and we suspect it is unlikely that adding further degrees of freedom will create significant local minima. Also, it should be noted that the ensembles considered earlier do often assume such configurations, and that these configurations have the highest skewness and kurtosis values we find.

APPENDIX B

Lorenz Latitude Circle Model

Strictly, the model is a set of J coupled ordinary differential equations:

$$\frac{dX_j}{dt} = (X_{j+1} - X_{j-2})X_{j-1} - X_j + F_j,$$

where $j = 1, J$, and J is commonly taken as 40. As it is a latitude circle, the boundary conditions are cyclic (e.g., $X_0 = X_J$). These equations exhibit the common characteristics of forced–dissipative flow: nonlinear terms for advection, a linear term for dissipation, and

a constant term for forcing. Accordingly, they can be made to exhibit chaotic behavior for adequate choice of F_j. The flow can be made unstable in a certain parameter range (here the only parameter is the constant forcing), and the instabilities are wavelike disturbances with westward (from high j to low j) phase speed and westerly (from low j to high j) group speed, thus mimicking traveling atmospheric disturbances.

The model is commonly run with $F_j = F = 8.0$ as this value places it well within the unstable range; we use this value as well. We also use $J = 40$ for our state dimension. Using the dissipative time scale of the model, Lorenz and Emanuel (1998) claim that a Δt of 0.05 is something like 6 h in the atmosphere; we use a time step of 0.005 for a fourth-order Runga–Kutta integration.

REFERENCES

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

——, 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.

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

Chin, T. M., A. J. Mariano, and E. P. Chassignet, 1999: Spatial regression and multiscale approximations for sequential data assimilation in ocean models. *J. Geophys. Res.,* **104C,** 7991–8014.

Cohn, S., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan,* **75,** 257–288.

Ehrendorfer, M., 1997: Predicting the uncertainty of numerical weather forecasts: A review. *Meteor. Z.,* **6,** 147–183.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99,** 10 143–10 162.

——, and P. van Leeuwen, 1996: Assimilation of *Geosat* altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasi-geostrophic model. *Mon. Wea. Rev.,* **124,** 85–96.

Farrell, B., and P. Ioannou, 2001: State estimation using a reduced-order Kalman filter. *J. Atmos. Sci.,* **58,** 3666–3680.

Haidvogel, D. B., A. R. Robinson, and E. E. Schulman, 1980: Accuracy, efficiency, and stability of 3 numerical-models with application to open ocean problems. *J. Comput. Phys.,* **34,** 1–53.

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

Hansen, J. A., and L. A. Smith, 2000: The role of operational constraints in selecting supplementary observations. *J. Atmos. Sci.,* **57,** 2859–2871.

——, and ——, 2001: Probabilistic noise reduction. *Tellus,* **53A,** 587–598.

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.

Ikeda, K., 1979: Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system. *Opt. Commun.,* **30,** 257–261.

Leith, C., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.,* **102,** 409–418.

Lorenc, A., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.,* **112,** 1177–1194.

Lorenz, E. N., 1995: Predictability—A problem partly solved. *Proc. Seminar on Predictability,* Reading, United Kingdom, ECMWF, 1–18.

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

Mardia, K. V., 1970: Measures of multivariate skewness and kurtosis with applications. *Biometrika,* **57,** 519–530.

Miller, R. N., A. R. Robinson, and D. B. Haidvogel, 1983: A baroclinic quasigeostrophic open ocean model. *J. Comput. Phys.,* **50,** 38–70.

Moore, A. M., C. L. Perez, and J. Zavala-Garay, 2002: A non-normal view of the wind-driven circulation. *J. Phys. Oceanogr.,* **32,** 2681–2705.

Pham, D., J. Verron, and M. Roubaud, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Sci.,* **16,** 323–340.

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

Smith, L. A., and J. A. Hansen, 2004: Extending the limits of forecast verification with the minimum spanning tree. *Mon. Wea. Rev.,* **132,** 1522–1528.

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.

Verlaan, M., and A. W. Heemink, 1997: Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrol. Hydraul.,* **11,** 349–368.

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

Wunsch, C., 1996: *The Ocean Circulation Inverse Problem.* Cambridge University Press, 442 pp.