Hidden Error Variance Theory. Part I: Exposition and Analytic Model

CRAIG H. BISHOP

Naval Research Laboratory, Monterey, California

ELIZABETH A. SATTERFIELD

National Research Council, Monterey, California

(Manuscript received 17 April 2012, in final form 19 September 2012)

ABSTRACT

A conundrum of predictability research is that while the prediction of flow-dependent error distributions is one of its main foci, chaos fundamentally hides flow-dependent forecast error distributions from empirical observation. Empirical estimation of such error distributions requires a large sample of error realizations *given the same flow-dependent conditions*. However, chaotic elements of the flow and the observing network make it impossible to collect a large enough conditioned error sample to empirically define such distributions and their variance. Such conditional variances are ''hidden.'' Here, an exposition of the problem is developed from an ensemble Kalman filter data assimilation system applied to a 10-variable nonlinear chaotic model and 25 000 replicate models. The 25 000 replicates *reveal* the error variances that would otherwise be hidden. It is found that the inverse-gamma distribution accurately approximates the *posterior* distribution of conditional error variances *given an imperfect ensemble variance* and provides a reasonable approximation to the *prior* climatological distribution of conditional error variances. A new analytical model shows how the properties of a *likelihood* distribution of ensemble variances given a true conditional error variance determine the *posterior* distribution of error variances given an ensemble variance. The analytically generated distributions are shown to satisfactorily fit empirically determined distributions. The theoretical analysis yields a rigorous interpretation and justification of hybrid error variance models that linearly combine static and flow-dependent estimates of forecast error variance; in doing so, it also helps justify and inform hybrid error covariance models.

1. Introduction

A *hidden error distribution* is a condition-dependent error variance that is formally unobservable because the condition of interest does not repeat itself enough times to enable the condition-dependent error variance to be accurately computed—even if perfectly precise observations were available to evaluate the errors. The properties of hidden error distributions/variances have received very little scientific attention to date. This is regrettable given the overwhelming prominence of the concept of condition-dependent error distributions in the design of advanced probabilistic data assimilation and forecasting schemes (Toth and Kalnay 1993; Molteni et al. 1996; Houtekamer et al. 1996; Bishop et al. 2001; Houtekamer and Mitchell 2001; Anderson 2001; Whitaker and Hamill 2002; Ott et al. 2004; Wang and Bishop 2005; Raftery et al. 2005; Hunt et al. 2007; McLay et al. 2008; van Leeuwen 2009).

Forecast error distributions depend on whether a front, cyclone, anticyclone, or other flow feature is nearby. They also depend on the quality and location of observations used to generate the forecast. Hence, condition-dependent forecast error distributions are contingent upon the state of the flow and on the observational network. The flow changes chaotically because the atmosphere exhibits nonperiodic chaos (Lorenz 1963; Pedlosky and Frenzen 1980). Aspects of the observational network such as cloud track winds, clear-air radiances, and adaptive observations are also subject to quasi-chaotic change.

An explicit aim of ensemble forecast design is to predict forecast error distributions conditioned on both

Corresponding author address: Craig H. Bishop, Marine Meteorology Division, Naval Research Laboratory, 7 Grace Hopper Ave., Stop 2, Bldg. 702, Room 212, Monterey, CA 93943-5502.
E-mail: bishop@nrlmry.navy.mil

DOI: 10.1175/MWR-D-12-00118.1

the flow and the observational network (Houtekamer et al. 1996). Here, we conceptually define the *true conditioned error variance* as the mean square of instantaneous forecast errors of replicate forecasting systems whose forecasts pertain to the same true state (condition 1) and observational network (condition 2) but to differing random realizations of observation and model error, which results in differing forecasts and forecast errors. The chaotic aspects of the flow and observational network make it practically impossible to observe and collect large samples of errors conditioned on specific flows and specific configurations of observational resources. Consequently, our ability to characterize and measure the accuracy of flow-dependent error distribution predictions is extremely limited.

Methods, such as those described in Majumdar et al. (2001), Wang and Bishop (2003), and Leutbecher and Palmer (2008) (among others) that involve binning differences between forecasts and observations according to the magnitude of the predicted error variance, can perceive the extent to which a condition-dependent error variance prediction scheme can distinguish large error variance from small error variances. However, such methods do not provide clear quantitative information about 1) the extent to which the true condition-dependent error variance fluctuates around predicted error variances and 2) the actual climatological range of true condition-dependent error variances associated with a particular forecasting system. Knowledge of these two quantities would provide a clearer measure of the extent to which the ensemble performance met the ensemble designers' aim of predicting condition-dependent forecast error distributions.

This is the first of two papers that seek to improve our understanding of hidden error variance distributions and then to use this understanding to improve data assimilation and ensemble forecasting techniques. This first paper presents a new method that can be used to quantify hidden error distributions in dynamical systems for which it is possible to run a very large number of statistically independent replicate data assimilation forecast cycles corresponding to the same true trajectory. The revealed flow-dependent error variances from this method are then used to develop an analytical model of the prior climatological distribution of true error variances associated with a forecasting system and also the posterior distribution of true error variances given an imperfect ensemble variance. In Bishop et al. (2013, hereafter Part II) the challenge of developing and thoroughly testing a reliable tool for deducing the parameters defining these analytic models of hidden error variance when the truth is not known and only a single ensemble data analysis and forecasting system is available is addressed. The overarching objectives of this first paper are to

(i) reveal otherwise hidden distributions of error variance and
(ii) develop an analytical model for hidden error variance that is consistent with the empirical findings of objective i.

Sections 2 and 3 achieve objectives i and ii, respectively, while conclusions follow in section 4.

2. Hidden error variances revealed

We begin this study with a thought experiment. Consider an infinite number of ''replicate earths,'' all having the same true atmospheric–oceanic state, the same forecast model, and the same observation types and observation locations but with differing random unbiased observation and model errors; in other words, the *conditions* on each of these earths are identical except for random observation and model errors. The differing realizations of observation and model error on each earth cause the forecasts on each earth to differ. From the perspective of an inhabitant of one of these earths, the error variance *given a particular truth* would be hidden. This is because the earth's ocean–atmosphere and observational network exhibits aperiodic chaos so that the system rarely returns to the same state and hence one is unlikely to have more than a few realizations of error. One realization of error does not define a variance or any other aspect of a probability distribution function (pdf) of errors. On the other hand, if one were omniscient and knew the (differing) forecasts on each of the replicate earths and the particular truth common to all the earths, then he or she could compute the forecast error on each of the replicate earths. By computing the variance of these errors, one could precisely determine the error variance *given a particular truth and a particular observational network* and hence the otherwise hidden error variances would be revealed.

We simulated this thought experiment with a system that employs a 10-variable version of the Lorenz '96 model [Lorenz (1996)—equivalent to model 1 of Lorenz (2005)] and an implementation of the ensemble transform Kalman filter (ETKF; Bishop et al. 2001); both of which are described in more detail in appendix A. A primary design objective for the system was to minimize the number of ensemble members required to ensure that the system would converge to a perfect linear extended Kalman filter for an *imperfect model* in the limit of infinitesimally small analysis and forecast errors. We set this objective based on the belief that if this objective was met *and* if the ensemble perturbations and

errors were sufficiently small to justify the linearity assumptions of the extended Kalman filter, then the resulting system would accurately predict condition-dependent error variances. Accurate error variance predictions would then enable us to systematically degrade the error variance prediction accuracy and thus reveal how variance prediction accuracy affects the distribution of true conditional error variances given an imperfect variance prediction. As can be seen from appendix A, these requirements led to a system featuring a *control* analysis–forecast, as well as 20 analysis–forecast ensemble members. The control forecast is subject to a random, quasi-flow-dependent model error that is uncorrelated from one forecast time to the next. Notable differences between this formulation and formulations typically seen in the literature include (i) the control forecast, rather than the ensemble mean, is used to compute the innovation; (ii) ensemble perturbations are computed about the control forecast rather than the ensemble mean; (iii) the analysis perturbations are not constrained to sum to zero; and (iv) no localization is used because the number of independent perturbations equals the number of independent model variables. These differences make this ETKF system similar to a full-rank square root Kalman filter. It is known that the Kalman filter is optimal for a linear system with Gaussian observation error statistics. Removing differences i–iii would have involved doubling the ensemble size to 40 members to obtain a symmetric ensemble with a mean of zero and hence doubling the cost of the already expensive calculations performed in this paper. The effect of differences i–iii is guaranteed to be negligible in the limit of vanishingly small forecast errors and ensemble perturbations.

We simulate the replicate earths described in our thought experiment by performing a large number of independent forecast–analysis cycles, each having the same truth but differing random observation and model errors. Hereafter, we will refer to these independent time series of forecasts and analyses as replicate systems. Each replicate system assimilates simulated observations that differ solely because each observation error is drawn independently from a normal distribution with zero mean and variance $R = 0.05$. This gives each individual system an independent time series of observation errors, which in turn leads to an independent time series of analyses and forecasts. The observation error variance (0.05) was chosen to be much smaller than the climatological variance of the true state of our model (13.1) in the hope that this would yield a data assimilation cycle in which the linear assumptions of an extended Kalman filter would be approximately satisfied.

The ''true state'' trajectory was obtained from a free run of the Lorenz model that was initialized at each grid point with random numbers (see appendix A for details). To allow for model spinup, we disregarded the initial 240 time steps. Beginning at the end of this spinup period, the data assimilation cycle was run for 400 time steps (~100 days) with data assimilation being performed every second time step (~12 h). Figure 1 shows how the mean of the squared errors of the 10 variables of the analysis for a single replicate system is reduced from a value of about 0.03 to roughly 0.005 after 10–20 data assimilation cycles. Subsequent data assimilation cycles maintain the error variance at this fairly low value, but note that this mean square error does exhibit considerable fluctuations from one data assimilation cycle to the next.

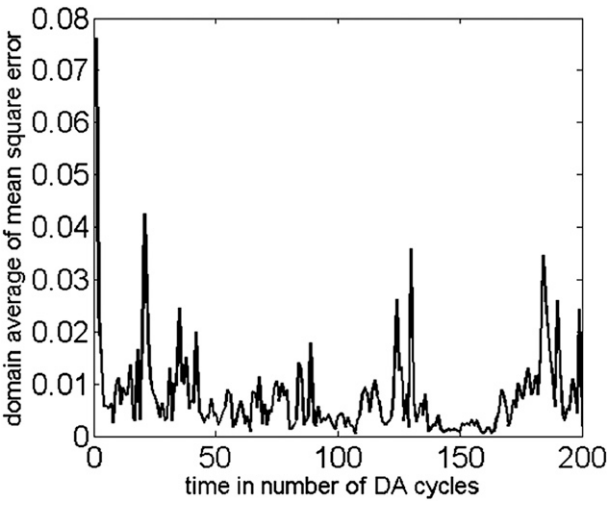

FIG. 1. The domain-averaged mean square of forecast error (MSE of analysis) shown for the first 200 data assimilation cycles. Data assimilation was performed every two time steps (~12 h). The first 25 data assimilation cycles (~12.5 days) are considered part of a spinup period and are excluded from our evaluation. The forecasts for this computation were obtained using 100% observation coverage, $R = 0.05$, and 4% ensemble covariance inflation. The model error variance parameter was $q = 0.0001$.

To avoid including the initial period of rapid error reduction, we remove all data from the first 25 data assimilation cycles (~12.5 days) and perform our computations on the data from the last 175 data assimilation cycles (~87.5 days). The variation of the domain-averaged mean square error over time for independent replicate systems was qualitatively similar to that shown in Fig. 1 but the timing and magnitude of increases and decreases of the mean squared error differed considerably from one replicate system to the next.

It is often claimed that ensemble Kalman filters are capable of predicting flow-dependent error variances. For the reasons discussed above, in chaotic systems it is often impossible to check whether this claim is valid.

However, with 25 000 replicate systems, we can accurately estimate the true forecast error variance given the current flow and the locations and error variances of the current and historical observations. An accurate estimate of the true condition-dependent variance of the error of the control forecast for a particular variable at a particular time can be obtained by subtracting the single true value of this variable from each of the 25 000 independent control forecasts of this variable to obtain a set of 25 000 independent control forecast errors. The variance of this set then gives an accurate estimate of the forecast error variance given a particular truth and a particular observational network. It is this error variance that ensemble variances attempt to predict. In Fig. 2, the coordinates of the ($1750 = 10 \times 175$) black circles are given by the estimated true error variance from the replicate systems (abscissa) and the corresponding ETKF ensemble variance (ordinate) from a single system (replicate system number 1, in this case). This scatterplot demonstrates that our implementation of the ETKF provides remarkably accurate predictions of the true flow-dependent forecast error variance. As far as the authors are aware, Fig. 2 gives the first empirical evidence that clearly supports the claim that deterministic ensemble Kalman filters (with a relatively small ensemble size of just 20 members) are capable of accurately predicting flow-dependent forecast error variances.

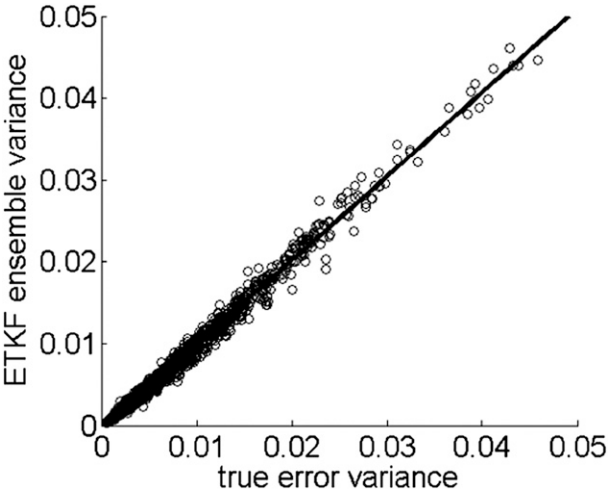

FIG. 2. Scatterplot of ETKF ensemble variance from a single replicate system as a function of true error variance. The true error variance is estimated from all 25 000 replicate systems. The linear fit to the points on the scatterplot is governed by the equation $\langle s^2 \mid \sigma^2 \rangle = 1.03\sigma^2 - 3.96 \times 10^{-4}$.

EnKF implementations for real systems cannot be expected to produce error variance predictions that will be as accurate as that shown in Fig. 2. The ETKF that produced Fig. 2 was constructed using perfect information about the distribution of stochastic model errors and did not require ad hoc error covariance localization. In real-world applications to high-dimensional systems, such as the ocean and atmosphere, the model error distribution is not known and ad hoc alterations such as ensemble covariance localization are employed. Hence, ensemble variances from such systems will not, in general, be able to predict the true error variance as accurately as that indicated by Fig. 2.

Because authors such as Wang and Bishop (2003) and Leutbecher and Palmer (2008) have shown that mean forecast error variance is a linearly increasing function of ensemble variance in some ensemble forecasting systems, it is of interest to consider ensemble forecasting schemes that are at least able to produce ensemble variance predictions s^2 that behave like random variables drawn from distributions with means that are monotonically increasing functions of the true conditional error variance σ^2. We model such ensemble forecasting schemes by assuming that s^2 is a random draw from a probability distribution whose mean is a linear function of the true conditional error variance σ^2. Specifically, we assume that the difference $s^2 - s^2_{\min}$ of the ensemble variance s^2 about a climatological minimum of ensemble variance $s^2_{\min}$ is a random draw from a gamma distribution whose mean is a linear function ($a\sigma^2 + b$) of σ^2 and whose relative variance k^{-1} (defined below) is linked to the accuracy of the ensemble-variance prediction of forecast error variance. To emphasize that $a\sigma^2 + b$ is semipositive definite, we let $b = -a\sigma^2_{\min}$ so that $a\sigma^2 + b = a(\sigma^2 - \sigma^2_{\min})$, where $\sigma^2_{\min}$ is the minimum possible value of the true forecast error variance in the chaotic system under consideration.

This approach allows us to control the level of accuracy of the ensemble forecast in a very precise way. To see this, note that gamma distributions are typically defined in terms of a shape and a scale parameter, k and θ, respectively. The properties of the gamma distribution are such that if x denotes a random draw from a gamma distribution, then the mean $\langle x \rangle = k\theta$ and the relative variance of x is defined by $[\langle (x - \langle x \rangle)^2 \rangle / \langle x \rangle^2] = k^{-1}$, where the angle brackets indicate the expectation operator. Thus, by setting $\theta = a(\sigma^2 - \sigma^2_{\min})/k$, we ensure that the mean $\langle s^2 \mid \sigma^2 \rangle - s^2_{\min}$ of the random distribution of $s^2 - s^2_{\min}$ given a fixed true error variance σ^2 is equal to $a(\sigma^2 - \sigma^2_{\min})$, while by increasing or decreasing k we can control the typical size of the stochastic variations of $s^2 - s^2_{\min}$ about its mean value of $a(\sigma^2 - \sigma^2_{\min})$. To be mathematically explicit, we assume that the likelihood pdf $L(s^2 \mid \sigma^2)$ of the ensemble variances given a true error variance is given by

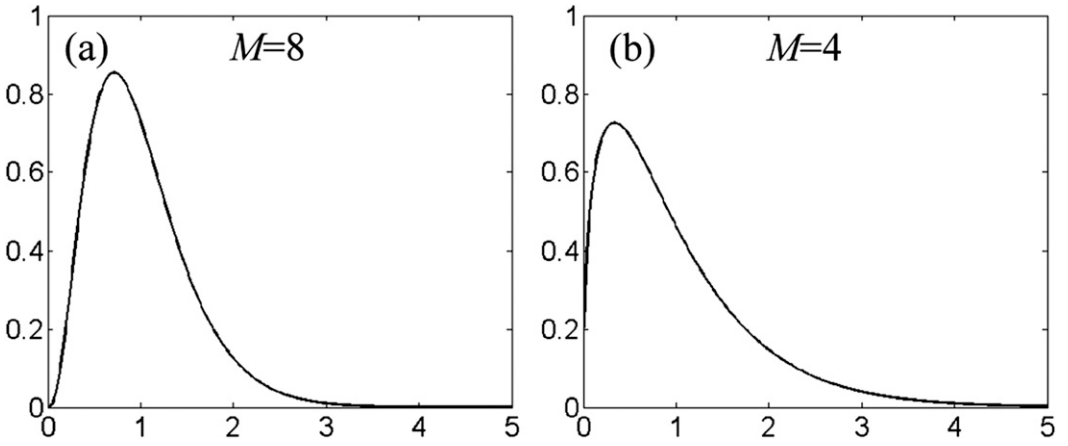

FIG. 3. Examples of assumed likelihood gamma pdf's of sample ensemble variances with a mean of unity. (a) Simulation for an effective ensemble size of $M = 8$, or equivalently, a relative variance of $2/7$. (b) As in (a), but for an effective ensemble size of 4, or equivalently, a relative variance of $2/3$.

$$L(s^2 \mid \sigma^2) = \frac{1}{\Gamma(k)} \frac{1}{(s^2 - s^2_{\min})} \left\{ \frac{k(s^2 - s^2_{\min})}{a(\sigma^2 - \sigma^2_{\min})} \right\}^k \exp\left\{ \frac{-k(s^2 - s^2_{\min})}{a(\sigma^2 - \sigma^2_{\min})} \right\}, \quad s^2 \geq s^2_{\min}$$

$$\text{so that} \quad \langle s^2 \mid \sigma^2 \rangle = a(\sigma^2 - \sigma^2_{\min}) + s^2_{\min}. \tag{1}$$

In appendix B it is shown that the relative variance k^{-1} of a gamma distribution is exactly the same as the relative variance of the distribution of sample variances obtained from M member random draws from a Gaussian distribution provided that $M = 2k + 1$. This makes the gamma distribution particularly appropriate for representing stochastic fluctuations in ensemble variances because there is an *effective random sample ensemble size* $M = 2k + 1$ associated with the relative variance k^{-1}. Figure 3 depicts two gamma pdf's with effective random sample ensemble sizes of $M = 8$ and $M = 4$. Both distributions have a mean of unity. Note how the variance of the $M = 4$ case is much larger than that for the $M = 8$ case.

The connection between the relative variance k^{-1} of (1) and the effective ensemble size enable the accuracy of ensemble variance predictions to be described in terms of random sample Gaussian ensembles of size $M = 2k + 1$ from distributions whose variances are linearly related to the true error variance. Because of unaccounted model errors and other deficiencies in ensemble design, in practice, a 100-member ensemble might have a connection to the true error variance like that of an 8-member random draw from a distribution whose variance is a linear function of the true error variance. In other words, a 100-member ensemble might have an effective ensemble size of just 8.

To investigate the distribution of true error variances given an imperfect ensemble variance, we generate degraded ensemble variances from the near-perfect ETKF variances. One way of doing this is to replace ETKF variances s^2_{ETKF} by

$$s^2 = (s^2_{\text{ETKF}} - s^2_{\min})\eta + s^2_{\min}, \tag{2}$$

where η is a random draw from a gamma distribution with a mean of unity and a relative error variance of k^{-1}.

We choose to estimate the climatological minimum of ensemble variances $s^2_{\min}$ in the above equation from the smallest realized value of s^2 returned by the ETKF ensemble. In Part II, we describe seven independent runs of length 400 000 time steps (200 000 data assimilation cycles). The smallest value of ETKF s^2 obtained in any of these runs was 1.6305×10^{-4}. This number is of the same order of magnitude as the minimum ETKF ensemble variance of the 1750 s^2_{ETKF} realizations that occurred on the first replicate system during the test period, which was 3.4×10^{-4}. For the sake of consistency with Part II, we chose to set $s^2_{\min} = 1.6305 \times 10^{-4}$.

To define the likelihood distribution given by (1) from the ensemble variances, generated from (2), we must estimate the parameters a and $\sigma^2_{\min}$. To estimate a, we note that the linear regression shown in Fig. 2 provides an approximation to the mean $\langle s^2 \mid \sigma^2 \rangle = a(\sigma^2 - \sigma^2_{\min}) + s^2_{\min}$. Hence, we can set a equal to the slope of this line. Doing so gives $a = 1.0262$. For the parameter $\sigma^2_{\min}$, we simply let it be equal to the smallest observed value of σ^2 within the set of points shown in Fig. 2. This approach yields

$\sigma^2_{\min} = 5.5021 \times 10^{-4}$, which is likely to be an overestimate of the true climatological minimum of the true forecast error variance because it is based on just 175 data assimilation cycles. Nevertheless, it is very close to the value 5.4463×10^{-4} one would obtain for $\sigma^2_{\min}$ by insisting that the ordinate axis intercept (-3.96×10^{-4}) in Fig. 2 be equal to $s^2_{\min} - a\sigma^2_{\min}$.

The computation that produced Fig. 2 took about 5 days to run on a personal computer and produced 1750 $(\sigma^2, s^2_{\text{ETKF}})$ pairs. This number of pairs would be too small for us to produce a large number of distinct bins of similar s^2 values. In addition, it does not allow the full range of possible deviations of s^2 about s^2_{ETKF} that are permitted by (2) to be explored. Consequently, we generated 306 250 pairs of (σ^2, s^2) from the 1750 pairs of $(\sigma^2, s^2_{\text{ETKF}})$ by creating 175 random values of s^2 for each $(\sigma^2, s^2_{\text{ETKF}})$ pair using (2). This creates 175 (σ^2, s^2) pairs, all having the same σ^2 value. This procedure was independently repeated for each of the 1750 $(\sigma^2, s^2_{\text{ETKF}})$ pairs, thus yielding 306 250 pairs of (σ^2, s^2). These pairs were then ordered from lowest s^2 value to highest s^2 value and then split into 35 bins of 8750 pairs having similar s^2 values.

Figure 4 depicts probability density histograms of true error variance for the 2nd, 18th, and 35th of these bins, for the case in which the s^2 values were generated using a relative error variance corresponding to an eight-member ensemble ($k = 3.5$). Figure 4's probability density histograms were obtained by (i) splitting the range of true error variances within each s^2 bin into 50 equally spaced intervals, (ii) computing the frequency with which the true error variance falls within each of these intervals, and (iii) dividing each frequency by the corresponding interval width to obtain a discrete approximation to the probability density function. The smooth solid lines shown in Fig. 4 give the fit of an inverse-gamma pdf approximation,

$$\rho(\sigma^2 \mid s^2) \approx \frac{\beta_{\text{post}}^{\alpha_{\text{post}}}}{\Gamma(\alpha_{\text{post}})} (\sigma^2)^{-\alpha_{\text{post}}-1} \exp\left\{-\frac{\alpha_{\text{post}}}{(\sigma^2)}\right\}, \quad (3)$$

to the distribution of true error variances found within each s^2 bin. The technique used to derive the parameters that define the fit of the inverse-gamma distribution is given in appendix C. Inspection of Fig. 4 indicates that the empirically derived discrete pdf appears to oscillate randomly around the inverse-gamma pdf fit to the data. The inverse-gamma pdf provides a qualitatively close match to the empirical pdf in that it correctly represents the skewness of the distribution and the empirical finding that the probability of very small error variances is itself very small. Figure 4 serves to (i) motivate the assumption that the distribution of true error

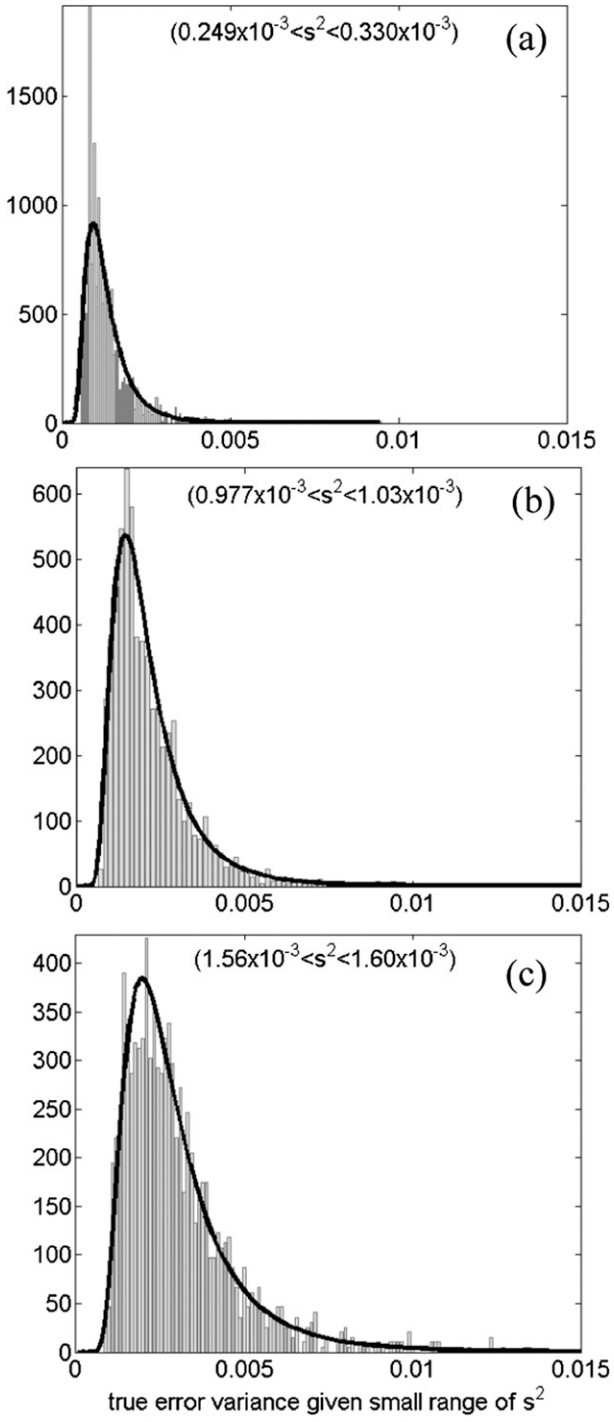

FIG. 4. (a)–(c) Histograms giving an empirical estimate of the pdf of the true error variances considering a constrained range of sample variances s^2 for an eight-member ensemble. The s^2 ranges are given in each figure; they correspond to the 2nd, 18th, and 34th bins, respectively, of the 35 bins of true error variance. The solid lines give the fit of an inverse-gamma function to the data in each bin.

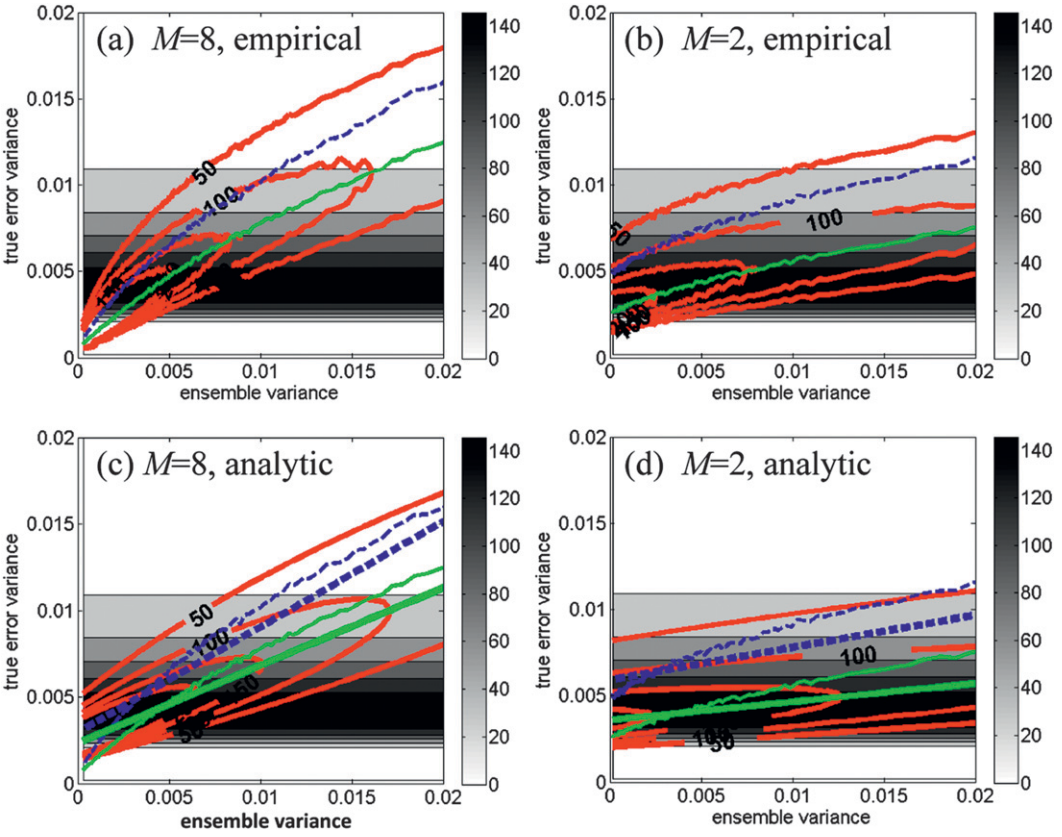

FIG. 5. Red lines depict estimates of the pdf's of the true error variance (ordinate axis) given fixed values of ensemble variance (abscissa axis). Thin green and blue lines give the mode and mean of the empirical estimates of the mode and mean of these estimates. Thick lines give the corresponding estimates for the analytic estimates of the pdf's. (a),(b) Empirical estimates for random sample ensembles of sizes $M = 8$ and $M = 2$, respectively. (c),(d) As in (a),(b), but for the corresponding analytic estimates of these pdf's. The gray shading gives an inverse-gamma pdf fit to the climatological pdf of the true error variances.

variance given an ensemble variance is an inverse-gamma distribution and (ii) indicate the limitations of such an assumption. Note how the range of true error variances markedly increases as the s^2 value associated with the bin increases.

To provide an overview of how the distribution of true error variances changes as s^2 changes, we computed the inverse-gamma pdf fit to the true error variances in each s^2 bin and then used contours to illustrate how the fitted inverse-gamma pdf's change as s^2 increases. Figures 5a and 5b give the result for effective ensemble sizes of eight and two, respectively. Note how the range of error variances given an ensemble variance is much more tightly constrained for an effective ensemble size of eight than it is for an effective ensemble size of two. As far as we are aware, Figs. 5a and 5b give the first empirically derived estimates of the distribution of the pdf of true conditional error variances with increasing ensemble variance. Figures 5c and 5d refer to an analytic model to be developed in the next section.

Our method of controlling the ensemble variance accuracy by simply degrading accurate ETKF variances that were actually used in the data assimilation cycle made it easy to vary the ensemble variance accuracy while keeping other factors the same. Although the motivation for doing this was primarily one of convenience, we emphasize that these degraded ensemble variances were not used in the data assimilation cycle. This aspect of our experimental setup is not unlike the current situation at many operational centers in which the ensemble covariances do not feed into the data assimilation cycle but are used to characterize the uncertainty of a high-resolution control forecast (e.g., Toth and Kalnay 1997; Molteni et al. 1996; Houtekamer et al. 1996; McLay et al. 2008).

3. Analytical model of hidden error variances

Figures 4 and 5 provide evidence that the inverse-gamma distribution is a good approximation to the distribution of hidden error variances given an imperfect ensemble

variance. Here, we develop an analytic model for hidden error variances that is consistent with this finding.

Apart from enabling us to perceive the distribution of error variances given an ensemble variance, our replicate system approach also enables us to see unconditioned, climatological distributions of error variances. Figure 6 shows the climatological distribution of error variances from our replicate system experiment together with an inverse-gamma approximation to this distribution. Although the fit of the inverse-gamma distribution to the empirically determined probability density histogram is not as good as those shown in Fig. 4, it correctly assigns very low probabilities to very low forecast error variances and correctly anticipates the low but finite probability of very large forecast error variances. It is possible that the fit would be better if we had continued our replicate system experiment beyond 400 time steps and thus provided a more complete sample of the error variances experienced by the forecasting system. Regardless of whether that speculation is true or not, in appendix D we demonstrate that a prior inverse-gamma pdf when combined with a gamma likelihood pdf results in a posterior pdf that is also an inverse-gamma pdf.

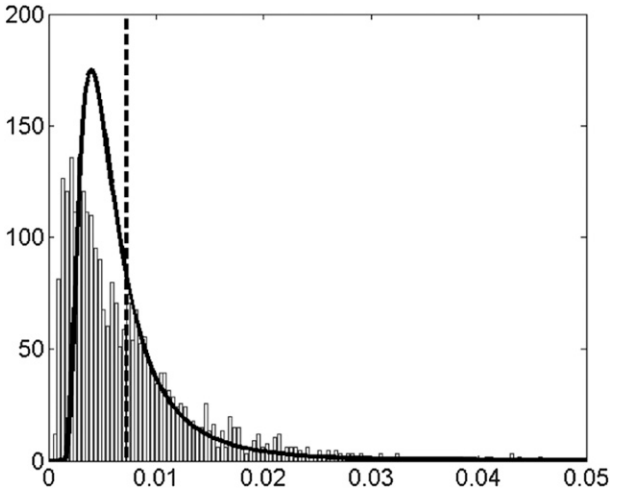

FIG. 6. Prior climatological distribution of true error variances. Bars show the probability density histogram of forecast error variances. Solid line shows the fit of the pdf in (4) to the data. The thick dashed line marks the mean of both the pdf and the data.

These facts lead us to propose an analytical model for hidden error variance that features a prior climatological inverse-gamma pdf of error variances, a likelihood gamma pdf of ensemble variances given a true error variance in (1), and a posterior inverse-gamma pdf of error variances *given an ensemble variance*.

To be specific, we hypothesize that the prior climatological pdf of forecast error variances associated with a particular forecasting system can be described by the inverse gamma distribution:

$$\rho_{\text{prior}}(\sigma^2) = \begin{cases} \dfrac{\beta^{\alpha}}{\Gamma(\alpha)}(\sigma^2 - \sigma^2_{\text{min}})^{-\alpha-1} \exp\left[-\left\{\dfrac{\beta}{(\sigma^2 - \sigma^2_{\text{min}})}\right\}\right] & \text{for} \quad \sigma^2 > \sigma^2_{\text{min}} \\ 0 & \text{for} \quad \sigma^2 \le \sigma^2_{\text{min}} \end{cases}. \tag{4}$$

As in (1), the purpose of the σ^2_{min} in (4) is to account for the possibility that some forecasting systems may have a climatological minimum of forecast error variance. As previously discussed, analysis of the points in Fig. 2 suggests that $\sigma^2_{\text{min}} = 5.5021 \times 10^{-4}$ for our toy forecasting system. To determine the parameters α and β, we use the well-known properties of the inverse-gamma function (Gelman et al. 2003) to deduce that

$$(\langle\sigma^2\rangle - \sigma^2_{\text{min}}) = \frac{\beta}{\alpha - 1} \quad \text{and} \quad \frac{\text{var}(\sigma^2)}{(\langle\sigma^2\rangle - \sigma^2_{\text{min}})^2} = \frac{1}{(\alpha - 2)},$$

$$\text{so that} \quad \alpha = \frac{(\langle\sigma^2\rangle - \sigma^2_{\text{min}})^2}{\text{var}(\sigma^2)} + 2 \quad \text{and}$$

$$\beta = (\langle\sigma^2\rangle - \sigma^2_{\text{min}})\left[\frac{(\langle\sigma^2\rangle - \sigma^2_{\text{min}})^2 + \text{var}(\sigma^2)}{\text{var}(\sigma^2)}\right], \tag{5}$$

where $\langle\sigma^2\rangle$ and $\text{var}(\sigma^2)$ give the mean and variance of the prior climatological distribution of true forecast error variances. We estimate $\langle\sigma^2\rangle$ and $\text{var}(\sigma^2)$ from the sample mean and variance of the 1750 values of σ^2 obtained from our replicate system experiment. The values obtained by this approach are $\langle\sigma^2\rangle = 7.2364 \times 10^{-3}$ and $\text{var}(\sigma^2) = 4.0782 \times 10^{-5}$.

Figure 6 compares the probability density histogram for the 1750 values of σ^2 obtained from our experiment with the pdf obtained from (4) using the aforementioned parameter values. Figure 6 shows that with our estimated parameters the pdf given by (4) underestimates the probability of very small forecast error variances and overestimates the probability of moderately small error variances but gives close to the correct probability densities for forecast error variances larger than the mean $\langle\sigma^2\rangle = 7.2364 \times 10^{-3}$ of the distribution. It is possible that a more sophisticated method of

estimating $\sigma^2_{\min}$ that was less inclined to overestimate the actual value of $\sigma^2_{\min}$ would further improve the fit of (4) to the empirically derived probability density histogram.

Bayes' theorem states that the *posterior* $\rho_{\text{post}}(\sigma^2 \mid s^2)$ pdf of error variances given an ensemble variance s^2, a *prior* pdf $\rho_{\text{prior}}(\sigma^2)$ of innovation variances, and a *likelihood* pdf $L(s^2 \mid \sigma^2)$ for s^2 given a true error variance σ^2 is given by

$$\rho_{\text{post}}(\sigma^2 \mid s^2) = \frac{L(s^2 \mid \sigma^2)\, \rho_{\text{prior}}(\sigma^2)}{\int_0^\infty L(s^2 \mid \sigma^2)\, \rho_{\text{prior}}(\sigma^2)\, d(\sigma^2)}. \quad (6)$$

In appendix D, we show that substitution of (1) and (4) into (6) gives

$$\rho_{\text{post}}(\sigma^2 \mid s^2) = \begin{cases} \dfrac{\beta_{\text{post}}^{\alpha_{\text{post}}}}{\Gamma(k+\alpha)} (\sigma^2 - \sigma^2_{\min})^{-\alpha_{\text{post}}-1} \exp\left[-\dfrac{\beta_{\text{post}}}{(\sigma^2 - \sigma^2_{\min})}\right] & \text{for} \quad \sigma^2 \geq \sigma^2_{\min} \\ 0 & \text{for} \quad \sigma^2 < \sigma^2_{\min} \end{cases}, \quad (7a)$$

where

$$\alpha_{\text{post}} = \alpha + k \quad \text{and} \quad \beta_{\text{post}} = [(s^2 - s^2_{\min})k/a] + \beta. \quad (7b)$$

As anticipated, (7) shows that the posterior distribution is of the same inverse-gamma form as the prior inverse-gamma distribution but has parameters α_{post} and β_{post} that have been altered by s^2, a, k, and $s^2_{\min}$.

Figures 5c and 5d contour (7a), for the previously estimated parameter set of $[\langle\sigma^2\rangle, \text{var}(\sigma^2), \sigma^2_{\min}, a, s^2_{\min}] = [7.2 \times 10^{-3}, 4.1 \times 10^{-5}, 5.5 \times 10^{-4}, 1.0262, 1.6 \times 10^{-4}]$ and relative variance parameter $k = 2/(M-1)$ defined by $M = 8$ and $M = 2$ for Figs. 5c and 5d, respectively. The high degree of similarity between the contours of Figs. 5a and 5c ($M = 8$ case) and also between Figs. 5b and 5d ($M = 2$ case) demonstrate that the analytical model of true (hidden) error variances and its relation to imperfect ensemble variance is qualitatively correct. Both the prior and posterior pdf's underestimate the probability of extremely low true forecast error variances. As previously mentioned, a likely reason for this is that our time period of integration was not long enough to accurately estimate the climatological minimum $\sigma^2_{\min}$ of the forecast error variance. Additionally, the analytical model gives a strictly linear relationship between the ensemble variance and the true error variance (Figs. 5c and 5d), while Figs. 5a and 5b suggest a slight curvature in the empirical relationship between true error variance and ensemble variance.

We speculate that the slight curvature in the empirically derived contours is a spurious artifact of the method we used to increase our sample size of (σ^2, s^2) pairs. Recall that we created 175 random realizations of s^2 for each single realization of σ^2. This resulted in 306 520 distinct values of s^2 but only 1750 distinct values of σ^2. The chance of obtaining extreme values of s^2 among the 306 520 values is much greater than the chance of finding extreme values in the 1750 distinct values of σ^2. This behavior would lead to the mean value of σ^2 being too low in the bin corresponding to very large values of s^2, thus explaining curvature of the type shown in Figs. 5a and 5b. This speculation is consistent with Fig. 2, which shows a strictly linear relationship between σ^2 and s^2_{ETKF}.

In subsequent work, we will show how knowledge of the posterior distribution of true forecast error variances given an ensemble variance can be used to improve ensemble forecasting. For some purposes, such as error covariance modeling for data assimilation, one may want to extract a single "best estimate" of the error variance given an ensemble variance. The mean of the posterior distribution is "best" in the sense that it minimizes the mean square deviation of the estimate from the true value. In appendix D, we show that the posterior mean $\langle\sigma^2 \mid s^2\rangle$ over all realizations of the error variance given a fixed value of s^2 is given by

$$\langle\sigma^2 \mid s^2\rangle = \frac{k}{k + (\alpha - 1)} \sigma_n^2 + \frac{\alpha - 1}{k + (\alpha - 1)} \langle\sigma^2\rangle, \quad (8)$$

where $\sigma_n^2 = [s^2/a + (\sigma^2_{\min} - s^2_{\min}/a)]$ is a debiased ensemble-based estimate of the forecast error variance and $\langle\sigma^2\rangle$ is the mean of the prior climatological distribution of the error variances.

To interpret (8), recall that k^{-1} gives the relative variance $\{k^{-1} = [\langle(x - \langle x\rangle)^2\rangle / \langle x\rangle^2]\}$ of the gamma likelihood distribution of ensemble variances given a fixed true error variance. Second, from the properties of the mean and variance of an inverse-gamma distribution in the random variable x, one may deduce that $(\alpha - 1)^{-1} = [\langle(x - \langle x\rangle)^2\rangle / \langle x^2\rangle]$ and hence that $(\alpha - 1)^{-1}$ is a measure of the relative error variance of the guess that $\langle\sigma^2\rangle$ is the true flow-dependent error variance. If we denote the

relative variances k^{-1} and $(\alpha-1)^{-1}$ by R_r and P_r, respectively, and then substitute in (8), we obtain the expression

$$\langle \sigma^2 \mid s^2 \rangle = \left(\frac{R_r^{-1}}{R_r^{-1}+P_r^{-1}} \right) \sigma_n^2 + \left(\frac{P_r^{-1}}{R_r^{-1}+P_r^{-1}} \right) \langle \sigma^2 \rangle, \quad (9a)$$

which has the same form as the minimum error variance estimate of a scalar T from two scalar estimators T_1 and T_2 having normally distributed error distributions with error variances σ_1^2 and σ_2^2, respectively, given by Kalnay [(2003), Eqs. (5.3.5) and (5.3.10)]. To be specific,

$$T = \frac{1/\sigma_1^2}{(1/\sigma_1^2)+(1/\sigma_2^2)} T_1 + \frac{1/\sigma_2^2}{(1/\sigma_1^2)+(1/\sigma_2^2)} T_2. \quad (9b)$$

Comparing (9a) with (9b) shows that the error variances σ_1^2 and σ_2^2 are the counterparts of the *relative* error variances R_r and P_r and that T_1 and T_2 are the counterparts of $\langle \sigma^2 \rangle$ and σ_n^2 in (9a).

Comparison of (9a) with Eq. (4) of Hamill and Snyder (2000) shows that it also has the same form as their hybrid model of forecast error covariance that linearly combines a static and ensemble-based estimate of the forecast error covariance matrix used in data assimilation. To the extent that variance estimation is an important part of the covariance models used in data assimilation, our theory and (9) provide a theoretical justification for the hybrid error covariance model and an interpretation of the weights in terms of the relative error variances of two independent estimates of the true error covariance. The simplicity of this interpretation provides a justification for the use of hybrid error covariance models in systems in which both the static and flow-dependent estimates of the forecast error covariance are imperfect.

Closer inspection of (9) shows that it is actually suggesting something slightly different from the hybrid form suggested by Hamill and Snyder (2000). To see this, note that the σ_n^2 is the sum of a term directly proportional to the ensemble variance s^2/a (which is in the Hamill and Snyder hybrid) and a constant term $(\sigma_{\min}^2 - s_{\min}^2/a)$ (which is not in the Hamill and Snyder hybrid). Recall that $\sigma_{\min}^2$ represents the climatological minimum of forecast error variance. In many ensemble systems, $s_{\min}^2$ would be precisely equal to zero. However, it is conceivable that in an effort to account for unknown model error, an ensemble designer would either intentionally or unintentionally produce an ensemble with a strict lower bound ($s_{\min}^2$) on the ensemble variance. In either case, $s_{\min}^2$ is the climatological minimum of the ensemble variances and $s_{\min}^2/a$ is a rescaling of this minimum. Thus, if $(\sigma_{\min}^2 - s_{\min}^2/a) > 0$, then it pertains to a part of the forecast error variance that is *always there* regardless of the stability of the flow and that was not accounted for by $s_{\min}^2$. It is conceivable that such immutable forecast error variances would arise due to the inability of discrete models to resolve coastlines and topography. One might estimate such immutable errors from the differences between two forecasts that are identical in every respect except their spatiotemporal resolution. If $(\sigma_{\min}^2 - s_{\min}^2/a) < 0$, then the term serves to counteract an overestimation of $\sigma_{\min}^2$ by $s_{\min}^2/a$. In summary, (7b) suggests a hybrid of the form

$$\mathbf{P}^f_{\text{Hybrid}} = \frac{k}{k+\alpha-1} \times \left\{ \left[\frac{\mathbf{P}^f_{\text{Ensemble}}}{a} \right] + \left[\sigma_{\min}^2 - \frac{s_{\min}^2}{a} \right] \tilde{\mathbf{Q}}^{\min}_{\text{climatology}} \right\} + \left[\frac{(\alpha-1)}{k+\alpha-1} \right] \mathbf{P}^f_{\text{climatology}}, \quad (10)$$

where $\mathbf{P}^f_{\text{Ensemble}}$ is a localized ensemble covariance matrix, $\mathbf{P}^f_{\text{climatology}}$ is an estimate of the climatological error covariance matrix, and $\tilde{\mathbf{Q}}^{\min}_{\text{climatology}}$ would be an estimate of the covariance of immutable errors. We emphasize that our current theory only pertains to variances and hence (11) represents an ansatz of the extension of our theory for variances to covariances.

The mode $(\sigma^2 \mid s^2)^{\text{mode}}_{\text{post}}$ of the posterior distribution is also of interest. As shown in appendix D,

$$(\sigma^2 \mid s^2)^{\text{mode}}_{\text{post}} = \frac{k}{k+(\alpha+1)} \sigma_n^2 + \frac{(\alpha+1)}{k+(\alpha+1)} [(\sigma^2)^{\text{mode}}_{\text{prior}}]. \quad (11)$$

Equation (11) shows that the posterior mode is a linear combination of the debiased ensemble variance σ_n^2 and the mode $(\sigma^2)^{\text{mode}}_{\text{prior}}$ of the prior distribution of the forecast error variances. Note the similarity of this expression to that for the mean in (8). The green lines in Fig. 5 show that the analytic model's prediction of the location of the modes of the posterior distribution is also close to those of the empirically determined posterior distribution.

4. Summary and conclusions

In the last two decades there has been an explosion of interest in predicting flow-dependent error distributions using ensemble methods. Here, we have pointed out that such error distributions are usually unobservable in

chaotic systems. This paper is one of the first to attempt to understand and explore the consequences of this fact.

By creating 25 000 parallel ETKF ensemble data assimilation and forecasting systems for a 10-variable nonlinear model, we were able to accurately estimate the true value of flow-dependent error variances that would be otherwise hidden. This enabled us to confirm that an ETKF that does not use error covariance localization and accurately accounts for model error covariance is able to predict the flow-dependent forecast error variance with a high degree of accuracy. The same result would be expected to hold for other forms of the deterministic ensemble square root Kalman filter because they solve the same set of equations (Tippett et al. 2003).

Because of unknown sources of model error and ad hoc alterations of ensemble square root filters such as ensemble covariance localization, it is inevitable that ensemble variances will be imperfect predictors of hidden error variance in real-world systems such as the ocean and atmosphere. We hypothesized that such inaccuracies result in gamma distributions of ensemble variances whose mean is a linear function of the true error variance. We also pointed out that the relative variance of such gamma distributions can be measured in terms of an *effective ensemble size M*. We then empirically determined the distribution of the conditional error variances of the ETKF forecasting system given a degraded ensemble variance. It was found that the distribution of error variances given an imperfect ensemble variance was satisfactorily approximated by inverse-gamma distributions. In addition, when these well-fitted inverse-gamma distributions were displayed as a function of the ensemble variance, it was found that the mode and mean of the distributions of true error variances were approximately linear functions of the true error variance.

A new analytical model of hidden error variances was introduced. It features a prior climatological inverse-gamma pdf of true error variances. This pdf represents the complete range of conditional forecast error variances experienced by a forecasting system. It also features a likelihood pdf of ensemble variances given a true error variance. The model employs Bayes' theorem to map the prior and likelihood pdf's into a posterior pdf of the distribution of true error variances given an ensemble variance. It was shown that this new analytical model was capable of accurately fitting the empirically derived pdf's of the true conditional forecast error variance given ensemble variance.

The analytical model yields simple expressions for the mode and mean of the posterior pdf. The equation for the mean is of essentially the same form as Hamill and Snyder's (2000) hybrid error covariance matrix. This equation demonstrates that if ensemble-based estimates of variances are imperfect in any way, then a superior estimate can always be obtained by combining the ensemble-based estimate with a climatological estimate. The analysis suggests an interpretation of the weights of the hybrid error covariance model in terms of the relative error variance of two independent estimates of the true error covariance. The simplicity of this interpretation provides a justification for the use of hybrid error covariance models in systems in which both the static and flow-dependent estimates of the forecast error covariance are imperfect. We caution, however, that our work has only made the veracity of this suggestion clear for variances and systems for which our analytical model is valid.

Careful inspection of the analytic expression for the posterior mean suggests that Hamill and Snyder's model might be improved by including an additional (small) static term that serves to account for a hypothesized climatological minimum of the true conditional error variance.

Acknowledgments. CHB gratefully acknowledges support from the U.S. Office of Naval Research Grant 4304-D-0-5. This research was performed while ES held a National Research Council Research Associateship Award.

APPENDIX A

Description of the ETKF System on the Lorenz '96 System

The following procedure outlines the analysis–forecast system used to create 25 000 replicate systems. A complete derivation of the ETKF is provided in Bishop et al. (2001).

a. Initialization of the control forecast and ensemble perturbations

1) The "true state" is taken from a free integration of the model described in point 6, below, initialized at each grid point by a random number drawn from a Gaussian distribution with a standard deviation of 0.008 and a mean of 8. (Here, 8 is the value of the applied forcing term F; see point 6). The first 240 time steps are disregarded to allow for model spinup and to render our results insensitive to our choice of initial conditions.
2) At the end of the spinup period, an analysis at the initial time $\mathbf{x}^a(0)$ is created for each 10-variable "replicate system" by adding a random number with variance $R = 0.05$ to the true state at each grid point; in other words,

$$\mathbf{x}^a(0) = \mathbf{x}^t(0) + \varepsilon^o, \quad \varepsilon^o \sim N(0, R\mathbf{I}). \tag{A1}$$

This ensures that the initial analysis error covariance matrix is diagonal with all elements equal to R. Because we employ an observation error variance $R = 0.05$, this initialization approach is equivalent to initializing by simply observing every variable and setting the initial value of the variables to their observed values.

3) The error of a forecast grows partly because the model has a different initial state with regard to reality and partly because the model misrepresents reality. Our idealization of forecast error growth in the presence of model imperfections is to add an additional "model error" random perturbation to the analysis and then propagate this through time with an otherwise perfect dynamical model. For simplicity, we assume that this noise has a covariance matrix $q\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Under these assumptions, the initial conditions for each control forecast are

$$\mathbf{x}_c^q(0) = \mathbf{x}_c^a(0) + \varepsilon^q, \quad \text{where} \quad \varepsilon^q \sim N(0, q\mathbf{I}), \tag{A2}$$

where $\mathbf{x}_c^a(0)$ is the control analysis. For the first forecast $\mathbf{x}_c^a(0)$ is obtained from (A1). For subsequent forecasts, it is the control analysis.

4) To accurately model the error covariance associated with a forecast initialized from (A1) and (A2), we create an initial $K = 20$ member ensemble $\mathbf{X}^{a\prime}(0)$, where $\mathbf{X}^{a\prime}(0)$ is a 10×20 matrix with each column representing an initial ensemble perturbation using

$$\begin{aligned}\mathbf{X}^{a\prime}(0) &= [\mathbf{x}_1^{a\prime}, \mathbf{x}_2^{a\prime}, \ldots, \mathbf{x}_K^{a\prime}] \\ &= [(\sqrt{R} + \sqrt{q})\mathbf{I}, (-\sqrt{R} + \sqrt{q})\mathbf{I}]\sqrt{\frac{K}{2}},\end{aligned}$$

$$\text{so that} \quad \frac{\mathbf{X}^{a\prime}(0)\mathbf{X}^{a\prime}(0)^{\mathrm{T}}}{K} = R\mathbf{I} + q\mathbf{I}. \tag{A3}$$

5) Consistent with (A3), the initial $K = 20$ member ensemble was then created using

$$\mathbf{x}_i^q(0) = \mathbf{x}_c^q(0) + \mathbf{x}_i^{a\prime}\sqrt{\rho}, \quad i = 1, 2, \ldots, K, \tag{A4}$$

where $\mathbf{x}_c^q(0)$ is obtained from (A2) and ρ is a scalar constant covariance inflation factor. For all experiments reported on in this two-part paper, we set $\rho = 1.04$.

b. *Propagation using Lorenz (1996) [model 1 of Lorenz (2005)]*

6) The ensemble members and control analysis (21 members in all) were then integrated forward in time using the fourth-order Runga–Kutta time-stepping scheme with the equation

$$\frac{dX_n}{dt} = -X_{n-2}X_{n-1} + X_{n-1}X_{n+1} - X_n + F, \tag{A5}$$

with $F = 8$ and a nondimensional time step of 0.05. Lorenz (2005) points out that an analogy with the waves supported by (A5) and atmospheric Rossby waves suggests that a time step of 0.05 is "like" a time step of about 6 h.

c. *ETKF with simple ensemble-based model error representation*

7) Data assimilation is performed every second time step (~12 h). The symbols $\mathbf{x}_c^f$ and $\mathbf{x}_i^f, i = 1, 2, \ldots, K$ are used to, respectively, denote the control forecast and ensemble forecasts obtained by integrating $\mathbf{x}_c^q$ and $\mathbf{x}_i^q, i = 1, 2, \ldots, K$ forward in time, two time steps. The square root of the forecast error covariance matrix is then estimated using

$$\mathbf{Z}^f = \frac{[\mathbf{x}_1^f - \mathbf{x}_c^f, \mathbf{x}_2^f - \mathbf{x}_c^f, \cdots, \mathbf{x}_K^f - \mathbf{x}_c^f]}{\sqrt{K}}. \tag{A6}$$

8) The observations $\mathbf{y}$ (obtained by adding random noise with covariance $R\mathbf{I}$ to the truth) are then assimilated by mapping the control forecast to the observation space to form the innovation:

$$\mathbf{v} = \mathbf{y} - \mathbf{H}\mathbf{x}_c^f, \tag{A7}$$

where $\mathbf{H}$ denotes the observation operator. For all of our experiments, we assumed that every grid point was observed but only at every second time step.

9) Following Bishop et al. (2001), we then compute the eigenvector decomposition

$$\mathbf{Z}^{f\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{Z}^f = \mathbf{C}\boldsymbol{\Gamma}\mathbf{C}^{\mathrm{T}},$$

where $\mathbf{C}$'s columns are the orthonormal eigenvectors of $\mathbf{Z}^{f\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{Z}^f$ and $\boldsymbol{\Gamma}$ is a diagonal matrix of eigenvalues. Since there are 10 observations, 10 model variables, and 20 ensemble members, 10 of the eigenvalues in the diagonal matrix are precisely equal to zero. Since $\mathbf{C}^{\mathrm{T}}\mathbf{Z}^{f\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{Z}^f\mathbf{C} = \boldsymbol{\Gamma}$, it follows that the columns of $\mathbf{Z}^f\mathbf{C}$ corresponding to the zero eigenvalues are precisely equal to zero. Hence, a precise nonsymmetric left square root of the analysis error covariance matrix is given by

$$(\mathbf{P}^a)^{1/2} = \mathbf{Z}\mathbf{C}_{20\times 10}(\boldsymbol{\Gamma}_{10\times 10} + \mathbf{I})^{-1/2},$$

where $\mathbf{\Gamma}_{10\times10}$ is the 10×10 diagonal matrix listing the 10 leading eigenvalues of $\mathbf{\Gamma}$ and the 20×10 matrix $\mathbf{C}_{20\times10}$ lists the 10 eigenvectors corresponding to the eigenvalues in $\mathbf{\Gamma}_{10\times10}$. The unperturbed control analysis is then given by

$$\mathbf{x}_c^a = \mathbf{x}_c^f + \mathbf{P}^a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}_c^f) \quad \text{where}$$
$$\mathbf{P}^a = (\mathbf{P}^a)^{1/2}\left[(\mathbf{P}^a)^{1/2}\right]^{\mathrm{T}}. \tag{A8}$$

As previously discussed, model error is introduced by adding noise to the analysis given by (A8) with covariance $q\mathbf{I}$; hence, the control initial condition for the subsequent forecast is given by

$$\mathbf{x}_c^q = \varepsilon^q + \mathbf{x}_c^a, \quad \text{where} \quad \varepsilon^q \sim N(\mathbf{0}, q\mathbf{I}). \tag{A9}$$

To create initial perturbations that would accurately represent the degradation to the forecast accuracy associated with this simulated model error, we followed (A3) and obtained initial conditions for our 20-member ensemble by adding perturbations of the form

$$\mathbf{X}^{a\prime}(0) = \left[(\mathbf{P}^a)^{1/2} + \sqrt{q}\mathbf{I},\ -(\mathbf{P}^a)^{1/2} + \sqrt{q}\mathbf{I}\right]\sqrt{\frac{K}{2}} \tag{A10}$$

to the initial condition given by (A9) in the manner used in (A4) ($\rho = 1.04$). These initial conditions were then integrated forward in time two time steps to create forecast ensemble perturbations for the next data assimilation time.

10) Steps 6–9 are then iteratively repeated. Two hundred independent data assimilation cycles were performed for each of the 25 000 replicate earths.

APPENDIX B

Relationship between k and Effective Ensemble Size M

Setting $k^* = (M-1)/2$ and $\theta^* = 2$ in the Gamma distribution gives the Chi-square distribution

$$\chi^2(x) = \frac{1}{\Gamma[(M-1)/2]}\left(\frac{x}{2}\right)^{(M-1)/2}\frac{1}{x}\exp\left(-\frac{x}{2}\right). \tag{B1}$$

The Chi-square distribution has a mean of $(M-1)$. It precisely describes the distribution of the sum of the squares of deviations of M random draws from a normal distribution about their sample mean. If the variable x denotes the sum of the squared deviations about the mean, then $y = \mu x/(M-1)$ denotes a sample variance from an M sample of a Gaussian distribution with mean variance μ. We can deduce the pdf of y by substituting $x = [(M-1)\mu/y] = \upsilon\mu/y$, where $\upsilon = (M-1)$ in (B1), to obtain

$$\chi^2(x)dx = \chi^2(\upsilon y/\mu)\frac{1}{\frac{dy}{dx}}dy = \frac{1}{\Gamma(\upsilon/2)}\left(\frac{\upsilon y}{2\mu}\right)^{\upsilon/2}\frac{\mu}{\upsilon y}\exp\left(-\frac{\upsilon y}{2\mu}\right)\frac{1}{\frac{dy}{dx}}dy = \frac{1}{\Gamma(\upsilon/2)}\left(\frac{\upsilon y}{2\mu}\right)^{\upsilon/2}\frac{\upsilon\mu}{\mu\upsilon y}\exp\left(-\frac{\upsilon y}{2\mu}\right)dy$$
$$= \frac{1}{\Gamma(\upsilon/2)}\left(\frac{\upsilon y}{2\mu}\right)^{\upsilon/2}\frac{1}{y}\exp\left(-\frac{\upsilon y}{2\mu}\right)dy = g(y)dy, \quad \text{where} \quad g(y) = \frac{1}{\Gamma(\upsilon/2)}\left(\frac{\upsilon y}{2\mu}\right)^{\upsilon/2}\frac{1}{y}\exp\left(-\frac{\upsilon y}{2\mu}\right). \tag{B2}$$

The last line of (B2) implies that $g(y)$ is simply a gamma distribution with $k^* = \upsilon/2 = (M-1)/2$ and $\theta^* = \mu/k^* = 2\mu/(M-1)$, with the mean and relative variance given by

$$\mu = k^*\theta^* \quad \text{and} \quad \frac{\sigma^2}{\mu^2} = \frac{k^*\theta^{*2}}{(k^*\theta^*)^2} = \frac{1}{k^*} = \frac{2}{(M-1)}, \tag{B3}$$

respectively. Consequently, the relative variance $(k^*)^{-1}$ of a gamma distribution is exactly the same as the relative variance of the sample variance of an M-member random draw from a Gaussian distribution provided $M = 2k^* + 1$.

APPENDIX C

Fit of the Inverse Gamma Distribution to Error Variances

If x is distributed according to an inverse-gamma distribution, then it may be shown that

$$\langle x\rangle = \frac{\beta}{\alpha - 1} \quad \text{and} \quad \left\langle\frac{1}{x}\right\rangle = \frac{\alpha}{\beta}. \tag{C1}$$

Consequently,

$$\alpha = \frac{\beta}{\langle x\rangle} + 1 \quad \text{and} \quad \beta = \langle x\rangle\left(\beta\left\langle\frac{1}{x}\right\rangle - 1\right) \Rightarrow \beta = \frac{\langle x\rangle}{\langle x\rangle\left\langle\frac{1}{x}\right\rangle - 1} = \frac{1}{\left\langle\frac{1}{x}\right\rangle - \frac{1}{\langle x\rangle}} \Rightarrow \alpha = \frac{1}{\langle x\rangle\left\langle\frac{1}{x}\right\rangle - 1} + \frac{\langle x\rangle\left\langle\frac{1}{x}\right\rangle - 1}{\langle x\rangle\left\langle\frac{1}{x}\right\rangle - 1}$$

$$= \frac{\langle x\rangle\left\langle\frac{1}{x}\right\rangle}{\langle x\rangle\left\langle\frac{1}{x}\right\rangle - 1} \Rightarrow \alpha - 1 = \frac{1}{\langle x\rangle\left\langle\frac{1}{x}\right\rangle - 1} \Rightarrow \frac{1}{\alpha - 1} = \langle x\rangle\left\langle\frac{1}{x}\right\rangle - 1. \tag{C2}$$

Hence, the parameters α and β defining the inverse gamma pdf of x are uniquely defined by the expected values of x and x^{-1}.

APPENDIX D

Posterior Distribution of Error Variances

Readers may find it helpful to refer to wikipedia (http://en.wikipedia.org/wiki/) or Gelman et al. (2003) in order to recall or learn about the basic properties of gamma and inverse-gamma pdf's. Letting $\xi = s^2 - s^2_{\min}$, (1) can be rewritten as

$$L(s^2 \mid \sigma^2) = \frac{1}{\Gamma(k)}\frac{(\xi k/a)^k}{\xi}(\sigma^2 - \sigma^2_{\min})^{-k}\exp\left\{-\frac{\xi k/a}{(\sigma^2 - \sigma^2_{\min})}\right\}. \tag{D1}$$

Note that the denominator of Bayes' equation (6) is just the integral of the numerator over σ^2. Consequently, all terms in the numerator of (6) that are not functions of σ^2 will be canceled out by their appearance in the denominator. Hence, we only need concern ourselves with terms in the numerator that are functions of σ^2. Furthermore, we are free to multiply these numerator terms by any constant that is not a function of σ^2 provided we perform the exact same multiplication in the denominator. The idea is to choose a multiplication factor that causes the integral over σ^2 to be equal to unity. This then makes the denominator in (6) vanish. Using (D1) and (4), the terms in the numerator of (6) for $\sigma^2 \geq \sigma^2_{\min}$ that are proportional to σ^2 are given by

$$\begin{aligned} L(\xi \mid \sigma^2)\rho_{\text{prior}}(\sigma^2) &\propto (\sigma^2 - \sigma^2_{\min})^{-k}(\sigma^2 - R)^{-\alpha-1}\exp\left\{-\frac{(\xi k/a) + \beta}{(\sigma^2 - \sigma^2_{\min})}\right\} \\ &\propto (\sigma^2 - \sigma^2_{\min})^{-(k+\alpha)-1}\exp\left\{-\frac{(\xi k/a) + \beta}{(\sigma^2 - \sigma^2_{\min})}\right\} \\ &\propto \frac{\beta^{(k+\alpha)}_{\text{post}}}{\Gamma(k+\alpha)}(\sigma^2 - \sigma^2_{\min})^{-(k+\alpha)-1}\exp\left\{-\frac{\beta_{\text{post}}}{(\sigma^2 - \sigma^2_{\min})}\right\} \end{aligned} \tag{D2}$$

where $\beta_{\text{post}} = (\xi k/a) + \beta$ while $L(\xi \mid \sigma^2)\rho_{\text{prior}}(\sigma^2) = 0$ for $\sigma^2 < \sigma^2_{\min}$. We obtained the last line of (D2) by multiplying the preceding line by $[\beta^{(k+\alpha)}_{\text{post}}/\Gamma(k+\alpha)]$. This is permissible provided we remember to perform the same multiplication in the denominator of (6) as well. We choose this multiplicative factor because, by inspection, (D2) defines an inverse-gamma function with parameters $\alpha_{\text{post}} = k + \alpha$ and β_{post}. Equation (7) follows from (D2) because the integral of (D2) over σ^2 is unity, and so when it is used in the denominator of (6), it vanishes.

Since the mean of an inverse-gamma pdf is given by $\beta/(\alpha - 1)$, it follows from (7) that the mean value $\langle \sigma^2 \mid s^2\rangle$ of the posterior is given by

$$\begin{aligned} \langle \sigma^2 \mid s^2\rangle &= \sigma^2_{\min} + \frac{\beta_{\text{post}}}{k+\alpha-1} = \sigma^2_{\min} + \frac{(s^2 - s^2_{\min})k/a + \beta}{k+\alpha-1} = \sigma^2_{\min} + \frac{(s^2 - s^2_{\min})k/a + (\alpha-1)(\langle\sigma^2\rangle - \sigma^2_{\min})}{k+\alpha-1} \\ &= \frac{k\sigma^2_{\min} + (\alpha-1)\sigma^2_{\min}}{k+\alpha-1} + \frac{(s^2 - s^2_{\min})k/a + (\alpha-1)(\langle\sigma^2\rangle - \sigma^2_{\min})}{k+\alpha-1} = \frac{k}{k+(\alpha-1)}\left[\frac{s^2}{a} + \left(\sigma^2_{\min} - \frac{s^2_{\min}}{a}\right)\right] \\ &\quad + \frac{(\alpha-1)\langle\sigma^2\rangle}{k+\alpha-1}. \end{aligned} \tag{D3}$$

Equation (8) follows directly from (D3).

The mode of the posterior inverse-gamma distribution is given by

$$(\sigma^2 \mid s^2)^{\text{mode}}_{\text{post}} = \sigma^2_{\min} + \frac{\beta_{\text{post}}}{\alpha_{\text{post}} + 1} = \sigma^2_{\min} + \frac{\beta_{\text{post}}}{(k+\alpha)+1} = \sigma^2_{\min} + \frac{\xi k/a + \beta_{\text{prior}}}{k+\alpha+1} = \sigma^2_{\min} + \frac{k\dfrac{\xi}{a} + (\alpha+1)[(\sigma^2)^{\text{mode}}_{\text{prior}} - \sigma^2_{\min}]}{k+(\alpha+1)}$$

$$= \frac{k\sigma^2_{\min} + (\alpha+1)\sigma^2_{\min}}{k+(\alpha+1)} + \frac{k\dfrac{s^2 - s^2_{\min}}{a} + (\alpha+1)[(\sigma^2)^{\text{mode}}_{\text{prior}} - \sigma^2_{\min}]}{k+(\alpha+1)} = \frac{k}{k+(\alpha+1)}\left[\frac{s^2}{a} + \left(\sigma^2_{\min} - \frac{s^2_{\min}}{a}\right)\right]$$

$$+ \frac{(\alpha+1)}{k+(\alpha+1)}[(\sigma^2)^{\text{mode}}_{\text{prior}}]. \tag{D4}$$

Using the definition of σ^2_n given just after (8) in (D4) then yields (11).

REFERENCES

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

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.

——, E. A. Satterfield, and K. T. Shanley, 2013: Hidden error variance theory. Part II: An instrument that reveals hidden error variance distributions from ensemble forecasts and observations. *Mon. Wea. Rev.,* **141,** 1469–1483.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2003: *Bayesian Data Analysis.* 2nd ed. CRC Press, 696 pp.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.,* **128,** 2905–2919.

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

——, L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.,* **124,** 1225–1242.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D,* **230,** 112–126.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, 341 pp.

Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.,* **227,** 3515–3539.

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

——, 1996: Predictability—A problem solved. *Proc. Predictability,* Reading, United Kingdom, ECMWF.

——, 2005: Designing chaotic models. *J. Atmos. Sci.,* **62,** 1574–1587.

Majumdar, S. J., C. H. Bishop, I. Szunyogh, and Z. Toth, 2001: Can an ensemble transform Kalman filter predict the reduction in forecast error variance produced by targeted observations? *Quart. J. Roy. Meteor. Soc.,* **127,** 2803–2820.

McLay, J. G., C. H. Bishop, and C. A. Reynolds, 2008: Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.,* **136,** 1093–1108.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.,* **122,** 73–119.

Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus,* **56A,** 415–428.

Pedlosky, J., and C. Frenzen, 1980: Chaotic and periodic behavior of finite-amplitude baroclinic waves. *J. Atmos. Sci.,* **37,** 1177–1196.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.,* **133,** 1155–1174.

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.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.,* **74,** 2317–2330.

——, and ——, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.,* **125,** 3297–3319.

van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.,* **137,** 4089–4114.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.,* **60,** 1140–1158.

——, and ——, 2005: Improvement of ensemble reliability with a new dressing kernel. *Quart. J. Roy. Meteor. Soc.,* **131,** 965–986.

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