Hidden Error Variance Theory. Part II: An Instrument That Reveals Hidden Error Variance Distributions from Ensemble Forecasts and Observations

CRAIG H. BISHOP

Naval Research Laboratory, Monterey, California

ELIZABETH A. SATTERFIELD

National Research Council, Monterey, California

KEVIN T. SHANLEY

Department of Mechanical Engineering, Clarkson University, Potsdam, New York

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

ABSTRACT

In Part I of this study, a model of the distribution of true error variances given an ensemble variance is shown to be defined by six parameters that also determine the optimal weights for the static and flow-dependent parts of hybrid error variance models. Two of the six parameters (the climatological mean of forecast error variance and the climatological minimum of ensemble variance) are straightforward to estimate. The other four parameters are (i) the variance of the climatological distribution of the true conditional error variances, (ii) the climatological minimum of the true conditional error variance, (iii) the relative variance of the distribution of ensemble variances given a true conditional error variance, and (iv) the parameter that defines the mean response of the ensemble variances to changes in the true error variance. These parameters are *hidden* because they are defined in terms of *condition*-dependent forecast error variance, which is unobservable if the *condition* is not sufficiently repeatable. Here, a set of equations that enable these hidden parameters to be accurately estimated from a long time series of (observation minus forecast, ensemble variance) data pairs is presented. The accuracy of the equations is demonstrated in tests using data from long data assimilation cycles with differing model error variance parameters as well as synthetically generated data. This newfound ability to estimate these hidden parameters provides new tools for assessing the quality of ensemble forecasts, tuning hybrid error variance models, and postprocessing ensemble forecasts.

1. Introduction

Much work has been done in the last decade on the construction and use of probability forecasts (Hersbach 2000; Wilks 2001; Hamill 2001; Mason and Graham 2002; Roulston and Smith 2003; Wang and Bishop 2005; Raftery et al. 2005; Fortin et al. 2006; Vrugt et al. 2006; Wilks and Hamill 2007; Wilson et al. 2007; Gneiting et al. 2007; Casati et al. 2008). In part, this work was motivated by the introduction of routine ensemble forecasts at major operational weather forecasting centers in the 1990s (Toth and Kalnay 1993; Molteni et al. 1996; Houtekamer et al. 1996) from which it was relatively straightforward to generate forecasts of the probability of weather events occurring. Many of these studies began with the assumption/hope that ensemble forecasts were sampling flow-dependent error distributions or at least an approximation to the true flow-dependent error distribution. As discussed in Bishop and Satterfield (2013, hereafter Part I), aperiodic chaos makes it difficult to assess the accuracy of forecasts of flow-dependent error distributions.

How can one measure the accuracy with which imperfect ensemble variances predict the true error variance? Existing measures used to evaluate the performance of an ensemble such as the Brier score (Brier 1950), ignorance (Roulston and Smith 2002), entropy, relative entropy and mutual information (Kleeman 2002; DelSole

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-12-00119.1

2004) convolve distinct aspects of ensemble performance, such as the accuracy of the ensemble mean and the ability of the ensemble to distinguish fluctuations in error variance from the climatological mean of the error variance. The sensitivity of these diagnostics to flow-dependent error variance prediction accuracy is generally obfuscated. Atger (1999) found that flow-dependent ensemble variances added no more value to the Brier skill score of an ensemble forecast than did static empirically derived variances. Is this result indicative of an ensemble whose variance is a poor predictor of error variance or is it that the Brier skill score is insensitive to the accuracy of flow-dependent error variance predictions? It is difficult to say because the sensitivity of the Brier skill score (Brier 1950) to flow-dependent error variance prediction accuracy is indirect and would depend on, among other things, the true climatological range of error variances experienced by the forecasting system. If, for example, the true error variance showed little variance from one flow to the next, then Atger's result would be unsurprising. It would be helpful for many applications as well as to efforts to understand predictability results such as Atger's if we could deduce (i) the climatological distribution of the true conditional forecast error variances, (ii) the distribution of the error variance predictions given a true error variance, and (iii) the distribution of the true error variances given an error variance prediction. In Part I, we showed that a compelling model of these three distributions can be built from just six parameters. The goal of this paper is to show how these six parameters can be estimated from a long data record of (observation minus forecast, ensemble variance) pairs.

In section 2, we derive the six equations that define the six parameters. Section 3 tests these parameter recovery equations. Discussion and conclusions follow in section 4.

2. Estimation of hidden error variance parameters with data

a. Review of hidden error variance distributions and parameters

The first two distributions to be estimated from Part I are as follows.

(i) Inverse-gamma probability distribution function (pdf) describing the prior pdf of forecast error variances σ^2 given by

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

where $\sigma^2_{\min}$ is the climatological minimum of the true forecast error variance and α and β are parameters defining the climatological mean $\langle\sigma^2\rangle$ and variance $\text{var}(\sigma^2)$ of the climatological distribution of σ^2. As pointed out in Part I, α and β satisfy

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

where the angle bracket indicates the expectation operator. (ii) A gamma pdf describing the likelihood pdf of ensemble variances s^2 given a true error variance σ^2. Namely,

$$L(s^2 \mid \sigma^2) = \frac{1}{\Gamma(k)}\frac{1}{(s^2 - s^2_{\min})}\left\{\frac{k(s^2 - s^2_{\min})}{a(\sigma^2 - \sigma^2_{\min})}\right\}^k \times \exp\left\{\frac{-k(s^2 - s^2_{\min})}{a(\sigma^2 - \sigma^2_{\min})}\right\}, \tag{3}$$

where $s^2_{\min}$ is the climatological minimum of the ensemble variances and a and k determine the mean and relative error variances of the likelihood pdf via the equations

$$\langle(s^2 - s^2_{\min}) \mid \sigma^2\rangle = \langle s^2 - s^2_{\min}\rangle_{\sigma^2} = a(\sigma^2 - \sigma^2_{\min}) \tag{4}$$

and

$$\frac{\text{var}[(s^2 - s^2_{\min}) \mid \sigma^2]}{[a(\sigma^2 - \sigma^2_{\min})]^2} = \frac{1}{k}, \tag{5}$$

where $\langle(s^2 - s^2_{\min}) \mid \sigma^2\rangle = \langle s^2 - s^2_{\min}\rangle_{\sigma^2}$ and $\text{var}[(s^2 - s^2_{\min}) \mid \sigma^2]$, respectively, refer to the mean and variance of the distribution of $(s^2 - s^2_{\min})$ for a fixed σ^2. Note also that $\text{var}[(s^2 - s^2_{\min}) \mid \sigma^2] = \text{var}(s^2 \mid \sigma^2)$. We will later use $\text{var}(s^2)$ to refer to the variance of s^2 over the entire climatological sample (and hence all possible values of σ^2).

The third key distribution of the hidden variance model introduced in Part I is the posterior distribution of the true forecast error variances given an ensemble variance, but this distribution can be obtained directly from (1) and (3) via Bayes' theorem. Hence, the key challenge is to accurately estimate the six unknown parameters $[\sigma^2_{\min}, \alpha, \beta, s^2_{\min}, a, k]$, which define (1) and (3), from a long time series of output from an ensemble forecasting scheme. Since (2) shows how to derive α and β from $\langle\sigma^2\rangle, \mathrm{var}(\sigma^2)$ and $\sigma^2_{\min}$, it is sufficient to recover the true values of $[\langle\sigma^2\rangle, \mathrm{var}(\sigma^2), \sigma^2_{\min}, s^2_{\min}, a, k]$.

The four parameters $[\mathrm{var}(\sigma^2), \sigma^2_{\min}, a, k]$ would be trivial to estimate if one could directly observe the true flow-dependent forecast error variance σ^2. However, as noted in Part I, σ^2 is practically unobservable in systems exhibiting aperiodic chaos. For this reason, $[\mathrm{var}(\sigma^2), \sigma^2_{\min}, a, k]$ are the *hidden* parameters of Part I's analytical pdf's. In Part I, we used replicate systems to reveal their values. In the following, we present equations that allow them to be deduced from a single system.

b. Equations to estimate hidden parameters of pdf's of true error variances

Ensemble forecasting schemes usually provide a high-resolution control forecast and a lower-resolution ensemble mean forecast. Either one of these or a linear combination of them could be used to define a single "best" forecast x_i^f of an observation y_i with error variance R_i (R_i may differ from one observation to the next). Corresponding to each forecast of an observation is an ensemble variance s_i^2 and an *innovation* $v_i = y_i - x_i^f$. Thus, it is straightforward to use an ensemble forecasting system to create a long data record $(v_i, s_i^2), i = 1, 2, \ldots, n$ of (innovation, ensemble variance) data pairs. The climatological mean of v_i would be zero if all observations and forecasts were unbiased. If bias is present and $\langle v_i\rangle \neq 0$, it is a simple matter to create an unbiased set of innovations by subtracting $\langle v_i\rangle$ from each individual realization of v_i. Hence, we hereafter assume that the dataset $(v_i, s_i^2), i = 1, 2, \ldots, n$ has been bias corrected so that $\langle v_i\rangle = 0$. In the appendix, we derive equations that estimate the parameters $[\langle\sigma^2\rangle, \mathrm{var}(\sigma^2), \sigma^2_{\min}, s^2_{\min}, a, k]$ from $(v_i, s_i^2), i = 1, 2, \ldots, n$; namely,

$$\langle\sigma^2\rangle = \langle v^2 - R\rangle, \tag{6}$$

$$\begin{aligned}\mathrm{var}(\sigma^2) &= \frac{\langle v^4\rangle}{3} - (\langle\sigma^2\rangle + \langle R\rangle)^2 - \mathrm{var}(R)\\ &= (\langle\sigma^2\rangle + \langle R\rangle)^2\left[\frac{\mathrm{kurtosis}(v) - 3}{3}\right] - \mathrm{var}(R),\end{aligned} \tag{7}$$

$$\begin{aligned}&s^2_{\min}\text{- defined by ensemble designer} \quad \text{or}\\ &s^2_{\min} = \min(s_i^2) \text{ over all } i \text{ if sample is large},\end{aligned} \tag{8}$$

$$a = \frac{\mathrm{covar}(v^2, s^2)}{\mathrm{var}(\sigma^2)}, \tag{9}$$

$$\sigma^2_{\min} = \langle\sigma^2\rangle - \frac{\langle s^2\rangle - s^2_{\min}}{a}, \tag{10}$$

and the relative error variance

$$k^{-1} = \frac{\mathrm{var}(s^2) - a^2\mathrm{var}(\sigma^2)}{a^2[(\langle\sigma^2\rangle - \sigma^2_{\min})^2 + \mathrm{var}(\sigma^2)]} \tag{11}$$

empirically define the parameters of Part I's analytical model. Equation (7) assumes that the innovation v_i is a stochastic normally distributed random process given a true flow-dependent error variance σ_i^2 and a true observation error variance R_i, $\mathrm{covar}(v^2, s^2)$ gives the covariance between v^2 and s^2, and $\mathrm{var}(s^2)$ give the variance of the climatological distribution of ensemble variances.

In Part I, we point out that the minimum error variance estimate of the true error variance is the mean of the posterior distribution of the true error variances given an imperfect ensemble variance s^2. In addition, we show that this mean error variance is equal to a weighted combination of the ensemble variance s^2 and the climatological error variance $\langle\sigma^2\rangle$. We suggest that these weights might be of assistance to researchers trying to optimally tune hybrid error covariance models that linearly combine static covariances with flow-dependent covariances. Specifically, in Part I we give the equation

$$\begin{aligned}\mathbf{P}^f_{\mathrm{Hybrid}} = &\frac{k}{k + \alpha - 1}\\ &\times\left\{\left[\frac{\mathbf{P}^f_{\mathrm{Ensemble}}}{a}\right] + \left[\sigma^2_{\min} - \frac{s^2_{\min}}{a}\right]\tilde{\mathbf{Q}}^{\min}_{\mathrm{climatology}}\right\}\\ &+ \left[\frac{(\alpha - 1)}{k + \alpha - 1}\right]\mathbf{P}^f_{\mathrm{climatology}}\end{aligned} \tag{12}$$

as an ansatz for the optimal weights for hybrid error covariance models; although, strictly speaking, (12) is only valid for variances, not covariances. For simplicity and to conform to the hybrid covariance model suggested by Hamill and Snyder (2000), let us assume that

$$\left[\frac{\mathbf{P}^f_{\mathrm{Ensemble}}}{a}\right]^{-1}\left[\sigma^2_{\min} - \frac{s^2_{\min}}{a}\right]\tilde{\mathbf{Q}}^{\min}_{\mathrm{climatology}} \approx 0, \tag{13}$$

so that (12) can be approximated by

$$\begin{aligned}\mathbf{P}^f_{\mathrm{Hybrid}} &= \frac{1}{a}\left[\frac{k}{k + \alpha - 1}\right]\mathbf{P}^f_{\mathrm{Ensemble}} + \left[\frac{(\alpha - 1)}{k + \alpha - 1}\right]\mathbf{P}^f_{\mathrm{climatology}}\\ &= w_E\mathbf{P}^f_{\mathrm{Ensemble}} + w_c\mathbf{P}^f_{\mathrm{climatology}},\\ \text{where}\quad w_E &= \frac{1}{a}\left[\frac{k}{k + \alpha - 1}\right] \quad \text{and} \quad w_c = \left[\frac{(\alpha - 1)}{k + \alpha - 1}\right].\end{aligned} \tag{14}$$

To use (2) and (6)–(11) to estimate the weights w_E and w_c for the respective ensemble-based and climatological short-range forecast error covariances, one must apply (6)–(11) to innovations from a short-range forecast. One possible concern in applying (6)–(11) to such innovations is that one of the key tests of the quality of an observation is the degree by which its associated innovation magnitude exceeds the expected innovation standard deviation. Many observation quality control schemes deem observations associated with very large innovations to be ''poor'' and remove them from the observational dataset. Such systematic removal of large innovations will compromise the estimate of the innovation kurtosis used in (7). For this reason, we were initially concerned that estimates of the hybrid covariance model weights given by Part I's (10) might be adversely affected by this data quality control issue. However, the appendix shows that

$$w_E = \frac{1}{a}\frac{k}{(k+\alpha-1)} = \frac{\mathrm{covar}(\upsilon^2, s^2)}{\mathrm{var}(s^2)}. \quad (15)$$

Consequently, w_E is not directly affected by the kurtosis of the innovations and hence observation quality control schemes will affect w_E only to the extent that observation quality control would negatively influence estimation of the quantity $\mathrm{covar}(\upsilon^2, s^2)/\mathrm{var}(s^2)$. Once (15) has been used to define w_E, the requirement that the climatological average of the hybrid variance be equal to the observed climatological average of forecast error variance gives

$$\langle\sigma^2\rangle = w_E\langle s^2\rangle + w_c\langle\sigma^2\rangle \Rightarrow w_c = \frac{\langle\sigma^2\rangle - w_E\langle s^2\rangle}{\langle\sigma^2\rangle}. \quad (16)$$

Equation (16) assumes that the variance elements in $\mathbf{P}^f_{\mathrm{climatology}}$ are equal to the climatological average of the true error variance $\langle\sigma^2\rangle$ of their corresponding variables. In practice, the static covariance matrix used in place of $\mathbf{P}^f_{\mathrm{climatology}}$ may have inaccurate variance elements, which we will generically denote as $\langle\sigma^2\rangle_{\mathrm{guess}}$, where it is understood that, in general, $\langle\sigma^2\rangle_{\mathrm{guess}} \neq \langle\sigma^2\rangle = \langle\upsilon^2 - R\rangle$. In this case, the requirement (16) gives

$$\langle\sigma^2\rangle = w_E\langle s^2\rangle + w_c\langle\sigma^2\rangle_{\mathrm{guess}} \Rightarrow w_c = \frac{\langle\sigma^2\rangle - w_E\langle s^2\rangle}{\langle\sigma^2\rangle_{\mathrm{guess}}}. \quad (17)$$

Equations (16) and (17) show that the weight w_c is also largely independent of the kurtosis of innovations and hence likely to be largely insensitive to adjustments of data quality control schemes.

The above equations define the parameters of the hidden error variance pdf's defined in Part I. However, they are more general than this because (i) they define the mean and variance of the prior distribution of the true error variances irrespective of whether the assumption of an inverse-gamma prior is correct or not and (ii) provided that the likelihood pdf $L(s^2 \mid \sigma^2)$ of the ensemble variances given a true error variance is created by a stochastic process of the following form:

$$(s^2 - s^2_{\mathrm{min}}) = a(\sigma^2 - \sigma^2_{\mathrm{min}}) + \xi, \quad \text{where } \xi \text{ is random and } \langle\xi\rangle = \langle\xi\sigma^2\rangle = 0. \quad (18)$$

Equations (6)–(11) define the parameter a, which gives the change of the mean of the pdf of the variance predictions per unit change in the true variance σ^2. The rate value a obtained from these equations is valid irrespective of the precise form of the likelihood pdf. Hence, the above equations can be used to *measure* the accuracy of ensemble-based predictions of forecast error variance irrespective of the underlying forms of the prior and likelihood pdf's. We note that even if (18) was not precisely satisfied, (18) can nevertheless serve as a first-order statistical approximation to the relationship between ensemble variance and true error variance.

Our analysis provides a number of new measures of ensemble forecast accuracy. Provided that the likelihood pdf $L(s^2 \mid \sigma^2)$ of the ensemble variances given a true conditional error variance is created by a stochastic process of the form given by (18), then the relative variance of the likelihood pdf is given by

$$\frac{1}{k} = \frac{2}{M-1} = \frac{\mathrm{var}[(s^2 - s^2_{\mathrm{min}}) \mid \sigma^2]}{[a(\sigma^2 - \sigma^2_{\mathrm{min}})]^2} = \frac{\mathrm{var}(s^2 \mid \sigma^2)}{[a(\sigma^2 - \sigma^2_{\mathrm{min}})]^2} \quad (19)$$

while its mean is $a(\sigma^2 - \sigma^2_{\mathrm{min}}) + s^2_{\mathrm{min}}$ irrespective of whether the likelihood pdf takes the form of a gamma pdf or some other distribution. Equation (18) can be rearranged to obtain a debiased ensemble-based error variance prediction $\sigma^2_n \approx s^2/a - (s^2_{\mathrm{min}}/a - \sigma^2_{\mathrm{min}})$, which is unbiased in the sense that its average value for a fixed σ^2 is precisely equal to σ^2. If the relative variance of $L(s^2 \mid \sigma^2)$ is very small (large), then the approximation $\sigma^2_n \approx s^2/a - (s^2_{\mathrm{min}}/a - \sigma^2_{\mathrm{min}})$ is very accurate (inaccurate).

Note that this measure of the ability of the (unbiased) ensemble variance to predict true error variance is *independent* of (i) the accuracy of the forecast x^f and (ii) the ensemble variance bias $\langle s^2 - \sigma^2\rangle$. In contrast, as far as the authors are aware, existing measures of probabilistic forecasting accuracy such as the Brier score, ignorance, and relative operating characteristic methods are highly sensitive to both the accuracy of x^f and the degree of under- (over-) dispersion of the ensemble variance. Our new measurement of ensemble variance accuracy can be expressed either in terms of the relative variance k^{-1} or an effective ensemble size M [see (19)]. As shown in Part I, $M = 2k + 1$.

A second new quantity revealed by our analysis is the relative variance of the climatological distribution of true conditional error variances. From (2),

$$\frac{1}{(\alpha-1)}=\frac{\mathrm{var}(\sigma^2)}{(\langle\sigma^2\rangle-\sigma^2_{\min})^2+\mathrm{var}(\sigma^2)}\Rightarrow\frac{1}{(\alpha-2)}=\frac{\mathrm{var}(\sigma^2)}{(\langle\sigma^2\rangle-\sigma^2_{\min})^2}. \quad (20)$$

When this measure is large (small), the potential value of the imperfect flow-dependent variance prediction is large (small). As far as the authors are aware, this formula provides the first practical "instrument" for measuring this quantity.

3. Tests of estimation formulas

a. *Can the hidden parameters of Part I be recovered by (6)–(11)?*

In Part I, we revealed the posterior distribution of true error variances given an ensemble variance together with the associated hidden parameters [$\mathrm{var}(\sigma^2)$, $\sigma^2_{\min}, a, M$] by running 25 000 replicate systems all having the same true state but differing realizations of forecast and observation error. The setup allowed us to collect 25 000 independent realizations of forecast error for each flow, which, in turn, allowed us to accurately compute the true flow-dependent forecast error variance σ^2. From these realizations of σ^2, we were able to estimate the four hidden parameters [$\mathrm{var}(\sigma^2), \sigma^2_{\min}$, a, M]. Our claim is that (6)–(11) enable us to obtain these parameters, not from the impractical artifice of replicate systems, but from a long time series (v_i, s_i^2), $i = 1, 2, \ldots, n$ of (innovation, ensemble variance) pairs from a *single* ensemble forecasting system. To test this claim, we extended one of the data assimilation (DA)–forecast runs from Part I until we had a long time series of (innovation, ensemble variance) pairs on which to apply (6)–(11). We then compared the values of [$\langle\sigma^2\rangle$, $\mathrm{var}(\sigma^2), \sigma^2_{\min}, s^2_{\min}, a, M$] obtained using (6)–(11) on this long run with those obtained in Part I by applying the replicate system approach to a much shorter run.

Details of the long run and replicate system short run are as follow. The nonlinear model used is a 10-variable version of the Lorenz (1996) model [identical to model 1 of Lorenz (2005)]. Model imperfection is introduced by adding noise of variance q to the initial conditions before each forecast is made. The DA scheme is an adaptation of the ensemble transform Kalman filter (ETKF; Bishop et al. 2001) that accounts for model error (details are given in appendix A of Part I). The time step used in the model is analogous to a 6-h time step (see Part I for details). DA is performed every two time steps (~12 h). The short run that used replicate systems was composed of 400 time steps and 200 corresponding DA cycles. Results associated with the first 50 time steps were removed to allow for system "spinup." The long run was the same as a short run except it started from a different set of initial conditions and used differing random realizations of forecast and model error *and* it was run for 400 100 time steps, the first 100 of which were discarded (400 100 time steps corresponds to 100 025 pseudodays). This approach left 200 000 observation times for use in constructing the archive of (innovation, ensemble variance) pairs. Since 10 variables are observed at each DA time, the long run yields 2×10^6 (innovation, ensemble variance) pairs from which to attempt parameter recoveries using (6)–(11). To quantify the sensitivity of the parameter recoveries obtained by random processes, the long run and associated parameter recoveries were independently repeated seven times using differing random realizations of the initial, observation, and model errors.

Comparison of ETKF variance with the true error variance obtained from the replicate earth experiment showed the ETKF to be an extremely accurate predictor of error variance in our idealized system in which all sources of error were known and accurately accounted for (see Part I for details). Because such near-perfect variance predictions are unattainable in real systems, our primary interest here is in the distribution of the true error variances given an imperfect ensemble variance. To create imperfect ensemble variances, it was necessary for us to generate synthetic ensemble variances from the ETKF variances that were less accurate than the ETKF variance. For each of the seven long runs, we did this by first finding the minimum value of the ETKF ensemble forecast variance over the 400 000 time step test period over all seven runs and defined this value to be $s^2_{\min}$. This approach gave $s^2_{\min} = \min(s^2_{\mathrm{ETKF}}) = 1.6305 \times 10^{-4}$, where s^2_{ETKF} is a realization from the set of ETKF variances produced during the long run. Next, degraded ensemble variances were obtained using

$$s^2 = s^2_{\min} + \eta, \quad (21)$$

where η is a variance drawn from a gamma distribution with mean $(s^2_{\mathrm{ETKF}} - s^2_{\min})$ and relative variance $k^{-1} = 2/(M-1)$. This approach produces distributions of sample variances that would be identical to those given by (3) and (18) if the ETKF variance were precisely equal to the true flow-dependent error variance (as shown in Part I, there is a small variation in the ETKF variances about the true error variance). In the following experiments, we considered two distinct M values: $M = 2$ and $M = 8$, which give relatively inaccurate and accurate ensemble variances, respectively. For each of the seven 200 000 DA cycles, we used (21) to produce three sets of degraded ensemble variances for both the $M = 2$ and $M = 8$ cases. This approach gave 21 quasi-independent sets of (innovation, ensemble variance) pairs (7 strictly independent × 3 quasi independent).

Equation (8) gives two methods for estimating the climatological minimum $s^2_{\min}$ of s^2. The first is appropriate if the ensemble has been explicitly designed to prevent the ensemble variance from falling below some user-specified value $s^2_{\min}$. In our system, the ensemble variances have been designed to never fall below $\min(s^2_{\mathrm{ETKF}})$, so we simply set $s^2_{\min} = \min(s^2_{\mathrm{ETKF}})$ over the seven independent trials. This approach gave exactly the same value as was used to generate the ensemble variances. The second method given by (8) involves simply setting $s^2_{\min}$ equal to the minimum observed value of the stochastic s^2 over some long trial period. We performed experiments using this method as well and found that it had little impact on the values obtained for other variables. Presumably, this lack of sensitivity is associated with the fact that both methods of estimating $s^2_{\min}$ produced $s^2_{\min}$ values that were one to two orders of magnitude smaller than the climatological average of the ensemble variances $\langle s^2 \rangle$. For the remainder of the parameter recoveries to be discussed in this paper, the first method is used to determine $s^2_{\min}$.

Our idea of comparing the hidden parameters retrieved from a single long run to those "observed" in the replicate system experiments in Part I assumes that the climatological distribution of true error variances and the likelihood distribution of ensemble variances are the same for the "short" 400 time step period considered in Part I as the 400 000 time step period considered here. One crude indicator of the validity of this assumption is the degree of similarity between the estimate of the mean forecast error variance $\langle \sigma^2 \rangle$ from the initial 400 time step period and the longer 400 100 time step period. Figure 1 shows that the value of $\langle \sigma^2 \rangle$ obtained from the "short" 400 time step run is almost exactly the same as the mean of the values obtained from the seven "long" 400 000 time step runs.

The bars in Fig. 2 marked "observed" give the values of the hidden parameters $\mathrm{var}(\sigma^2), \sigma^2_{\min}, a$ and M revealed by Part I's short replicate system experiment. Figure 2 allows these observed values to be compared with the minimum, mean, and maximum values retrieved

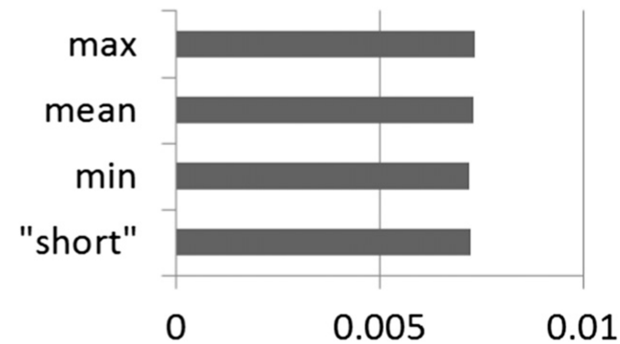

FIG. 1. Comparison of the estimate of the climatological mean forecast error variance $\langle \sigma^2 \rangle$ from the "short" 400 time step experiment with the minimum (min), mean, and maximum (max) values of $\langle \sigma^2 \rangle$ obtained from the seven independent "long" 400 000 time step experiments.

from the 7 × 3 long sets of (innovation, ensemble variance) pairs using (6)–(11). Figures 2a and 2c show that for the hidden parameters $\mathrm{var}(\sigma^2)$ and a, the mean of the values from the seven independent experiments is very close to the observed value (see Figs. 2a and 2c).

Figure 2b shows that the range of retrievals of $\sigma^2_{\min}$ obtained from individual runs is quite large. Equation (10) does not exclude the possibility of negative $\sigma^2_{\min}$ values and Fig. 2b shows that the minimum retrieved value is actually negative. Since $\sigma^2_{\min}$ is known to be a positive semidefinite quantity, one can imagine creating a more sophisticated Bayesian estimation scheme for $\sigma^2_{\min}$ that would combine the prior information about the sign of $\sigma^2_{\min}$ with the value given by (10). However, for the sake of simplicity, we do not explore that avenue in this paper.

Figure 2b also shows that the mean value from the seven independent trials is significantly smaller than the observed value. However, recall from Part I that the observed value was obtained by setting $\sigma^2_{\min}$ equal to the smallest of the 1750 realizations of σ^2 obtained in the replicate system experiment. Since it seems likely that even smaller values of σ^2 would be obtained if this experiment were run for a longer period of time, it is highly probable that the value of $\sigma^2_{\min}$ obtained by this method is an *overestimate* of the true value of forecast error variance. Hence, the retrieved $\sigma^2_{\min}$ value may be more accurate than that suggested by Fig. 2b. Examination of (1) and (2) shows that, provided $\sigma^2_{\min}$ is significantly smaller than $\langle \sigma^2 \rangle$, it will have very little effect on the form of the prior pdf of the true conditional forecast error variances. For the case considered in Fig. 2, $\langle \sigma^2 \rangle = 7.3 \times 10^{-3}$, while the retrieved range of $\sigma^2_{\min}$ is $[-4.8 \times 10^{-4}, 1.2 \times 10^{-3}]$ for $M = 2$ and $[-4.8 \times 10^{-4}, 1.2 \times 10^{-3}]$ for $M = 8$. Hence, the retrieved $\sigma^2_{\min}$ were

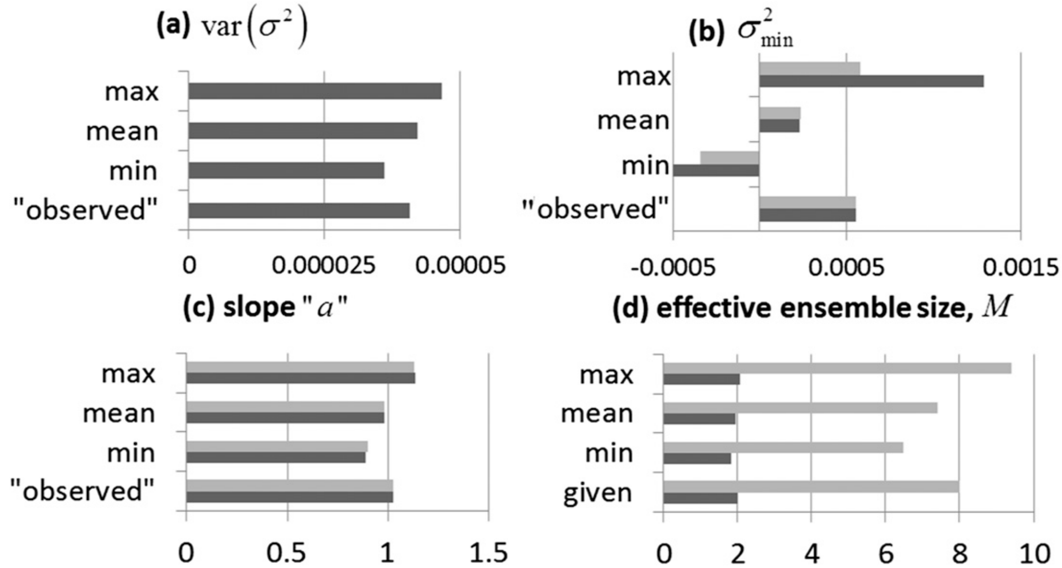

FIG. 2. Retrieved hidden parameters: (a) var(σ^2), (b) $\sigma^2_{\min}$, (c) a, and (d) M. Each plot summarizes information from 21 attempts to retrieve the hidden parameters from 200 000 DA–forecast cycles with the Lorenz model and model error parameter $q = 0.0001$. The values marked as ''observed'' are the values obtained from the 175 DA cycles of the 25 000 ''replicate systems'' described in Part I. The ''given'' ensemble sizes in (d) are the random sample ensemble sizes used to degrade the quality of the ETKF ensemble variance. The values marked as min, mean, and max are the minimum, mean, and maximum of the values retrieved from 21 sets of 200 000 DA cycles. In (b)–(d) the black and gray bars correspond to given values of $M = 2$ and 8 cases, respectively.

all at least an order of magnitude smaller than $\langle\sigma^2\rangle = 7.3\times10^{-3}$. For this reason, the uncertainty in the recovered $\sigma^2_{\min}$ only has a minor effect on the prior pdf in this case.

Figure 2d shows that (6)–(11) yield an effective ensemble size $M = 2k - 1$ that is very close to the ''given'' values of M that were used to generate the ensemble variances. In both the $M = 2$ and 8 cases, the retrieved values of M are slightly smaller than the given values. A possible explanation for this is that we saw in Part I that the ETKF variances exhibit (slight) stochastic variations around the true error variance. Hence, fluctuations of s^2 about the true flow-dependent error variance are not only caused by the size of the random sample used to generate s^2 from the ETKF variance but also from the inherent fluctuations of the ETKF variances about the true conditional error variance. From this perspective, one would expect the recovered effective ensemble sizes to be smaller than the given ensemble sizes. Figure 2d is consistent with this expectation.

In summary, Fig. 2 shows that mean of the retrievals of the hidden parameters var$(\sigma^2), \sigma^2_{\min}, a$ and M from long time series of (innovation, ensemble variance) pairs are in reasonable agreement with the replicate system approach. As shown in Part I, these parameters define the inverse-gamma posterior pdf of true error variances given an ensemble variance.

b. *Model error dependence*

Figure 2b showed that the relative error in individual recoveries of the parameter $\sigma^2_{\min}$ was quite large while the actual value of $\sigma^2_{\min}$ was very small compared with the climatological average of the forecast error variance. In addition, Fig. 1 shows that the mean forecast error variance is about 10 times smaller than the observation error variance ($R = 0.05$) used for these experiments. In atmospheric DA, observation error variance and forecast error variances are generally estimated to be of the same order of magnitude. Increasing model error will increase forecast error variance and bring the ratio of forecast error variance to observation error variance closer to that of the atmosphere. How would the parameters of the hidden distributions of true error variance change if the model error variance parameter was increased?

This question motivated us to explore how the retrieved parameters would change if we were to increase the model error variance. In our setup, as described in detail in Part I, model error is simulated by adding random, uncorrelated noise of variance q and mean zero to each of the initial conditions before the model is integrated.

Figure 3a shows how the ratio of the mean forecast error variance divided by observation error variance

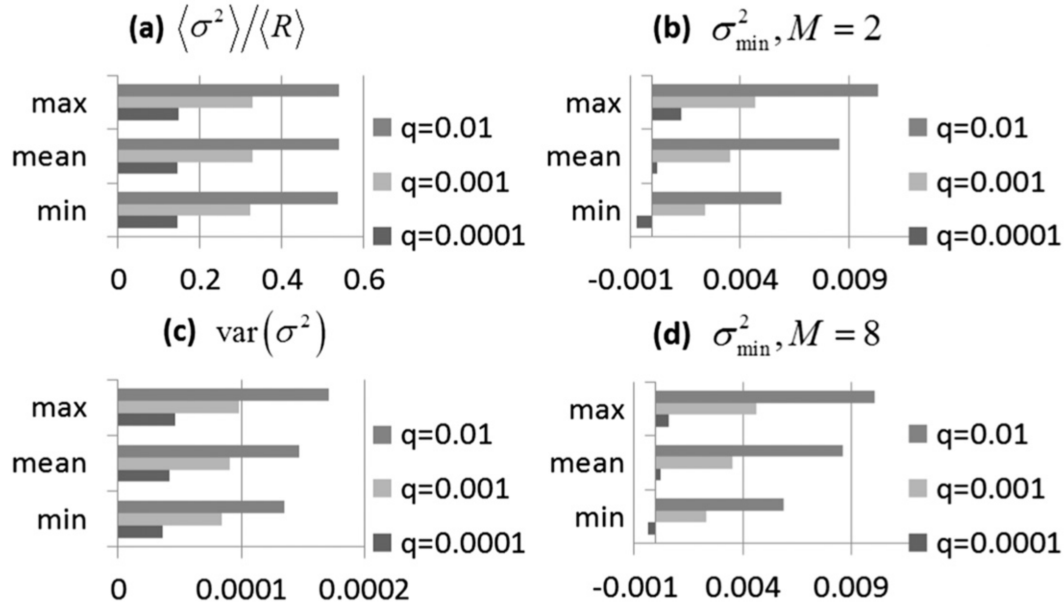

FIG. 3. Variation with model error parameter q of the parameters defining the climatological prior distribution of forecast error variances. Values marked min, mean, and max are the minimum, mean, and maximum values obtained from 21 trials. (a) Ratio of $\langle\sigma^2\rangle/\langle R\rangle$. (b),(d) The value of $\sigma^2_{\min}$ for $M = 2$ and 8, respectively. (c) The value of $\mathrm{var}(\sigma^2)$.

increased by about a factor of 3.7 as the model error variance parameter was increased by a factor of 100 from $q = 0.0001$ to $q = 0.01$. Using the same method as for the $q = 0.0001$ case, the estimated $s^2_{\min}$ for the $q = 0.001$ and $q = 0.01$ cases were estimated to be 1.29×10^{-3} and 3.37×10^{-3}, respectively. (Recall that $s^2_{\min} = 1.63 \times 10^{-4}$ in the $q = 0.0001$ case.) We expected $s^2_{\min}$ to increase with model error variance because the ensemble-based estimation of model error variance was designed to perfectly match a *constant* model error variance q associated with the first time step of each nonlinear integration of the model (see appendix A in Part I). Consequently, as the model error variance is increased, $s^2_{\min}$ ought to increase as well.

Figures 3b and 3d show how the 100-fold increase in q markedly increased the retrieved values of $\sigma^2_{\min}$ regardless of the given ensemble size M. To be precise, the mean of the recovered $\sigma^2_{\min}$ values increased by factors of 32 and 37 for the $M = 2$ (Fig. 3b) and $M = 8$ (Fig. 3d) cases as q was increased from 0.0001 to 0.01. With the low $q = 0.0001$ value, a spurious negative value for $\sigma^2_{\min}$ was obtained; however, with the $q = 0.001$ and $q = 0.01$ cases, no negative $\sigma^2_{\min}$ value were recovered. In principle, the climatological minimum of true forecast error variance $\sigma^2_{\min}$ should be independent of the "effective ensemble size" M associated with the ensemble that is used to retrieve the hidden parameter $\sigma^2_{\min}$. The similarity of the $\sigma^2_{\min}$ retrieved using the $M = 2$ and 8 ensembles is in accord with this principle.

Figure 3c shows that the retrieved variance of the climatological distribution of true conditional error variances is also an increasing function of q. To be precise, as q was increased from 0.01 to 0.0001, the mean of the retrieved values of $\mathrm{var}(\sigma^2)$ increased by a factor of 3.4 (this quantity is entirely independent of the given ensemble size M).

The relative variance of the prior distribution of error variances is tied to the α parameter of the inverse-gamma distribution via the equation $(\alpha - 2)^{-1} = \mathrm{var}(\sigma^2)/\langle\sigma^2\rangle^2$. When this parameter is large (small), the third moment of the climatological distribution of the true conditional error variances is large (small). Figure 4 shows that, for our model error representation, this parameter was a decreasing function of model error.

The slope parameter a determines the rate of increase of the mean of the likelihood distribution of the ensemble variances with the true error variance. In theory, this parameter is independent of the effective ensemble size of the ensemble used to retrieve it. Comparison of Figs. 5a and 5c demonstrates that this theoretical expectation is met by the retrieved a value. Figures 5a and 5c suggest that a decreases (slightly) as q increases.

Figures 5b and 5d show that the retrieved effective ensemble size is largely independent of q for given ensemble sizes of 2 and 8. Furthermore, it shows that the retrieved ensemble sizes are very close to the given ensemble sizes. It also shows that the range of retrieved effective ensemble sizes over the 7×3 trials decreased

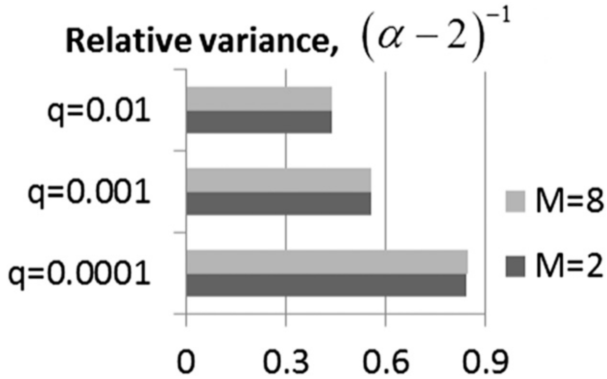

FIG. 4. Relative variance of the climatological distribution of the true forecast error variances as a function of model error.

as the model error was increased. Does this mean that the error of retrieved effective ensemble sizes is smaller in the parameter regime associated with the high model error $q = 0.01$? To investigate this question, in the next subsection, we generate synthetic data that are precisely consistent with the mean retrieved parameters obtained in this section.

c. Tests based on synthetic data

To test whether (6)–(11) can infer the correct prior, likelihood, and posterior distributions of forecast error variance, in this section, we apply the equations to synthetic data. This demonstration has the following steps:

1) Specify the six parameters $[\langle\sigma^2\rangle, \mathrm{var}(\sigma^2), \sigma^2_{\min}, s^2_{\min}, a, M]$. From this set, use (2) and (5) to derive $[\alpha, \beta, \sigma^2_{\min}, s^2_{\min}, a, k]$.
2) Generate n random samples $\sigma^2_i, i = 1, 2, \ldots, n$ from the inverse-gamma prior pdf of forecast error variances given by (1). This was done using built-in Matlab functions to generate gamma random numbers and using the fact that if $\sigma^2 \sim \Gamma^{-1}(\alpha, \beta)$, then $1/\sigma^2 \sim \Gamma(\alpha, 1/\beta)$.
3) Generate n corresponding innovations $v_i = \varepsilon^f_i - \varepsilon^o_i$ for $i = 1, 2, .., n$ events, where the forecast errors ε^f_i and observation errors ε^o_i are drawn from zero mean normal distributions with variances of σ^2_i and R_i, respectively.
4) Generate n corresponding random samples $s^2_i, i = 1, 2, .., n$ from the likelihood gamma pdf of ensemble variances corresponding to an ensemble size of M given by (3).
5) Test whether the parameters used to generate (v_i, s_i), $i = 1, 2, .., n$ can be recovered by applying the hidden error variance from (6)–(11) to the data pairs (v_i, s_i), $i = 1, 2, .., n$.

Each of the parameter recoveries from the previous subsections were based on 2×10^6 (innovation, ensemble variance) pairs. We performed 60 completely independent parameter recoveries using 2×10^6 synthetically generated (innovation, ensemble variance) pairs. By computing the minimum, mean, maximum, and standard

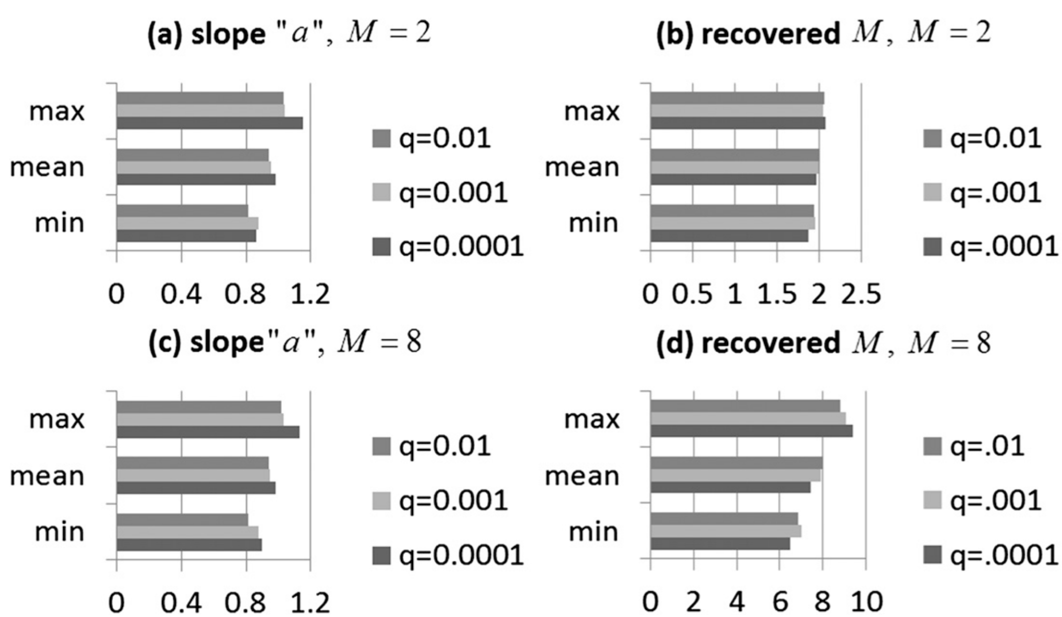

FIG. 5. Variation with model error parameter q of the parameters defining the likelihood distribution of ensemble variances given a true error variance. Values marked min, mean, and max are the minimum, mean, and maximum values obtained from 21 trials. (a),(c) The recovered slope parameters a for $M = 2$ and 8, respectively. (b),(d) The effective ensemble sizes M, retrieved from the $M = 2$ and 8 experiments, respectively.

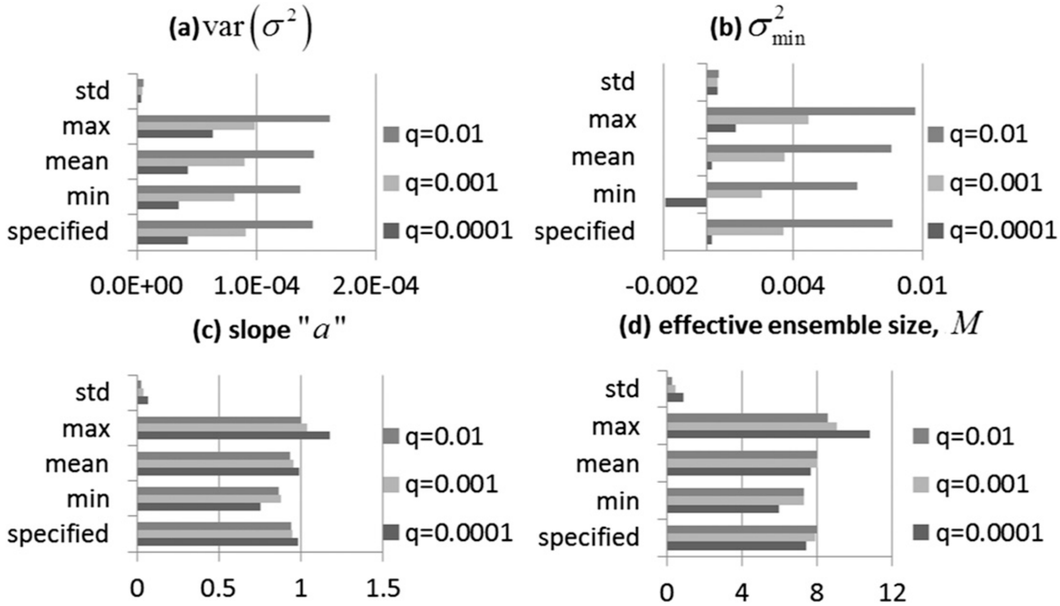

FIG. 6. Retrieved hidden parameters: (a) var(σ^2), (b) min(σ^2), (c) a, and (d) M. Each plot summarizes information from 60 independent attempts to retrieve the hidden parameters from 2 000 000 (innovation, ensemble variance) pairs synthetically generated from specified distributions. The values marked "specified" are equal to values retrieved from Lorenz model experiments with a given $M = 8$ and differing values of the model error q. Black, light gray, and gray bars correspond to q values of 0.0001, 0.001, and 0.01, respectively. The values marked as min, mean, max, and std are the minimum, mean, maximum, and standard deviation of the values retrieved from 60 completely independent synthetically generated datasets.

deviation of the recovered parameters and comparing these with the actual parameters that were used to generate the (innovation, ensemble variance) pairs, we were able to robustly describe the accuracy of the parameter recoveries obtained from (6)–(11). We test six distinct parameter sets, each precisely corresponding to the parameters previously retrieved from the Lorenz model experiments for the (q = 0.0001, M = 8), (q = 0.001, M = 8), (q = 0.01, M = 8), (q = 0.0001, M = 2), (q = 0.001, M = 2), and (q = 0.01, M = 2) cases.

Figure 6 gives the hidden parameter retrieval results from the (q = 0.0001, M = 8), (q = 0.001, M = 8), and (q = 0.01, M = 8) parameter sets. Comparison of the specified values of var(σ^2), $\sigma^2_{\min}$, a, and M with the mean of the 60 retrieved values shows that the mean value is very close to the specified value. Figure 7 gives the corresponding retrievals for the (q = 0.0001, M = 2), (q = 0.001, M = 2), and (q = 0.01, M = 2) cases. We do not show the retrieval of var(σ^2) for these cases because they are identical to those shown in Fig. 6a. Comparison of Figs. 7a and 7b with Figs. 6b and 6c shows that the degrees of accuracy of the retrievals of $\sigma^2_{\min}$ and the slope parameter a are similar for both the M = 2 and 8 cases. Comparison of Fig. 7c with Fig. 6d shows that the recovery of the effective ensemble size is even more accurate for the M = 2 case than it was for the M = 8 case.

This result demonstrates that (6)–(11) recover the true hidden parameters when the number of independent samples is large enough. Comparison of the specified values with the minimum, maximum, and standard deviation of the retrieved values over the 60 trials shows that the retrievals from a single set of (innovation, ensemble variance) pairs becomes more accurate as the model error parameter q increases. In addition, the accuracy of the retrievals from single sets of 2×10^6 (innovation, ensemble variance) pairs using (6)–(11) is usefully accurate for all hidden variables except for $\sigma^2_{\min}$ in the q = 0.0001 case (Figs. 6b and 7a). In this case, a larger sample size than 2×10^6 pairs would be required in order to make the noise in the recovered value small in comparison with the already very small $\sigma^2_{\min}$ value.

4. Conclusions

Modern DA and ensemble forecasting research is based on the assumption that the true forecast error variance is flow dependent. However, flow-dependent forecast error variance is difficult to observe in chaotic systems. This fact makes it difficult to answer basic questions like: What is the climatological variance of true conditional error variances? How do ensemble

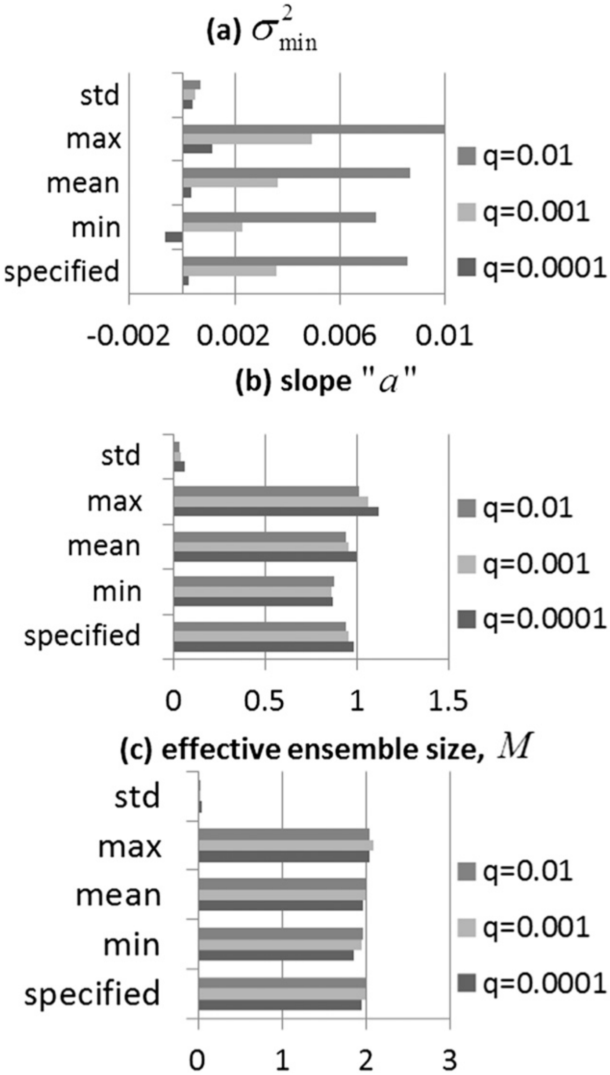

FIG. 7. As in Fig. 6, but for the given $M = 2$ case. Also, the var(σ^2) case is omitted because it is identical to that shown in Fig. 6a.

variances vary about the true conditional error variance? What is the relative variance of true conditional error variances given an ensemble variance?

The theory of hidden error variances developed in Part I of this paper and the current work provides compelling answers to these questions. Part I showed that the inverse-gamma distribution provided a plausible model of the prior climatological distribution of true conditional error variances and the posterior distribution of true conditional error variances given an ensemble variance. It also showed that the gamma distribution provided a plausible model of the likelihood distribution of ensemble variances given a true conditional error variance. These three distributions are uniquely defined by six parameters. Two of these parameters, the climatological mean of forecast error variance $\langle\sigma^2\rangle$ and the climatological minimum of ensemble variance $s^2_{\min}$, are trivially observable from a large set of (innovation, ensemble variance) pairs. The other four parameters are hidden. They are (i) var(σ^2), the variance of the climatological distribution of the true conditional forecast error variances; (ii) $\sigma^2_{\min}$, the climatological minimum of the true conditional forecast error variances; (iii) the slope a of the line mapping true conditional error variance to the mean of the likelihood distribution of ensemble variances; and (iv) the relative variance k^{-1} of the likelihood distribution of ensemble variances, which can also be expressed in terms of an effective ensemble size $M = 2k + 1$.

In Part I, these hidden parameters were estimated using the artifice of 25 000 replicate systems that produced 25 000 independent realizations of forecast error for each forecast and thus enabled accurate observations of the true flow-dependent forecast error variance σ^2. With observations of σ^2, it was a trivial matter to estimate all of the hidden parameters, except for $\sigma^2_{\min}$, because the length of the replicate system run was too short to accurately estimate the climatological minimum of error variance.

Here in Part II, we derived equations to retrieve the hidden parameters from a single run of an ensemble forecasting system. These equations accurately recovered the hidden parameter values of var(σ^2), a, and M revealed by Part I's replicate system experiment. Increasing the model error variance increased the retrieved values of $(\langle\sigma^2\rangle, \mathrm{var}(\sigma^2), \sigma^2_{\min})$. The retrieved slope parameter a and effective ensemble size parameter M of the likelihood distribution were insensitive to increases in model error variance. The experiments showed that the relative variance of the prior climatological distribution of error variances was a decreasing function of our representation of the model error variance.

This last finding has implications for hybrid models of error variance (Hamill and Snyder 2000) that linearly combine static and flow-dependent forecast error variances because, as was shown in Part I, the optimal weight for the static part increases as the relative variance of the prior climatological distribution decreases. Figure 8 shows how these ''optimal hybrid weights'' change as a function of model error and ensemble size. Figure 8 depicts how the weights on the ensemble-based variance (black bars) decrease with increasing model error but increase with an increased effective ensemble size M. Figure 8 quantitatively demonstrates how the value of ensemble variances to hybrid error variance prediction increases (decreases) as the relative variance of the climatological distribution of true conditional error variances increases (decreases).

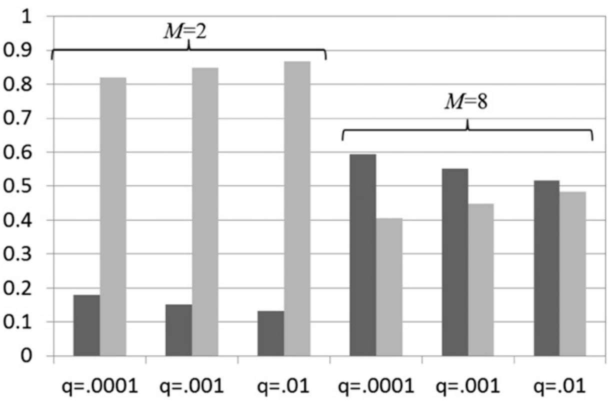

FIG. 8. Variation of weights for the mean of the posterior distribution of the true error variances with model error q and given effective ensemble size M. Black bars give the weights for the debiased *flow-dependent* ensemble variance while gray bars give the corresponding weights for the *static* mean of the climatological error variances.

The possibility that the accuracy of recovered parameters might depend on parameter values motivated us to develop a method for synthetically generating (innovation, ensemble variance) pairs that were consistent with specified parameters and Part I's hidden error variance model. This synthetic data method enables one to check the potential accuracy of recovered parameters for any given set of parameters and any given sample size of (innovation, ensemble variance) pairs. To illustrate the technique, we applied it to sets of 2×10^6 (innovation, ensemble variance) pairs synthetically generated from parameters identical to the parameters retrieved in our experiments using differing model errors. (The sample size of 2×10^6 was chosen because this was exactly the same as the sample size used in the long nonlinear DA experiments.) It was found that all six of the parameters defining the hidden error variance model could be accurately retrieved from a single set of 2×10^6 (innovation, ensemble variance) pairs except for the very small $\sigma^2_{\min}$ value associated with the $q = 0.0001$ experiment. Nevertheless, the mean of 60 independent retrievals did provide an accurate estimate of $\sigma^2_{\min}$ for this case. The accuracy of all retrieved parameter values was greater for higher model error variance than for lower model error variance.

The ability to estimate the parameters of Part I's hidden error variance model may be of assistance to designers of hybrid DA strategies because these parameters determine the optimal weights for the static and flow-dependent estimates of error variance in hybrid error variance models. The approach presented here gives the first framework for estimating the weights for hybrid error variance models, not from computationally expensive trial and error, but by a direct analysis of the (innovation, ensemble variance) pairs produced by the DA scheme of interest.

Our newly developed effective ensemble size measure of ensemble performance directly measures the extent to which fluctuations in ensemble variance are tied to fluctuations in true error variance. We are unaware of any existing metric that measures this quantity. This measure, unlike most probabilistic measures, does not depend on the accuracy of the ensemble mean. It is also directly connected to the weight that ensemble variances should receive in hybrid error variance models. We hope that it will prove useful to those attempting to quantify ensemble performance.

A close relation of ensemble performance quantification is ensemble postprocessing. The method, developed in this two-part paper, of estimating the distribution of true error variances given an imperfect ensemble variance may also be useful to ensemble postprocessing efforts that attempt to statistically correct the imperfect probabilities associated with a raw ensemble forecast. The potential value this method adds to ensemble postprocessing techniques is the topic of a subsequent paper.

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. KTS acknowledges support from NOAA THORPEX Grant NA04AANRG0233.

APPENDIX

Derivation of (6)–(11) and (15)

a. Derivation of (6)–(11)

First, note that the innovation $v_i = (y_i - x_i^t) - (x_i^f - x_i^t) = \varepsilon_i^o - \varepsilon_i^f$, where x_i^t is the true value of the observed variable and ε_i^o and ε_i^f are the observation and forecast errors, respectively. We assume that observation error is uncorrelated with forecast error; that is, $\langle \varepsilon^o \varepsilon^f \rangle = 0$. It follows that

$$\begin{aligned} \langle v_i^2 - R_i \rangle &= \langle (\varepsilon_i^f)^2 \rangle + \langle (\varepsilon_i^o)^2 \rangle - \langle R_i \rangle \\ &= \langle \sigma^2 \rangle + \langle R_i \rangle - \langle R_i \rangle = \langle \sigma^2 \rangle. \end{aligned} \quad \text{(A1)}$$

Note that in the case of normally distributed forecast and observation errors innovations may be viewed as a random draw from a Gaussian distribution with mean zero and variance $\sigma_i^2 + R_i$; in other words, $v_i = (R_i + \sigma_i^2)^{1/2} \zeta_i$ where $\zeta_i \sim N(0, 1)$. The notation $\zeta_i \sim N(0, 1)$ means that ζ_i is a random draw from a normal distribution with mean zero and a variance of unity. It turns out that the variance of the prior climatological distribution of true (hidden) error variances can be deduced from the expected value of the fourth power of v_i. To see this, note that

$$\begin{aligned}\langle \upsilon_i^4\rangle &= \langle \zeta^4(\langle\sigma_i^2\rangle + (\sigma_i^2 - \langle\sigma_i^2\rangle) + \langle R_i\rangle + R_i')^2\rangle, \quad \text{where} \quad R_i' = R_i - \langle R_i\rangle \\ &= \langle\zeta^4\rangle\{[\langle\sigma_i^2\rangle^2 + \langle(\sigma_i^2 - \langle\sigma_i^2\rangle)^2\rangle] + [\langle R_i\rangle^2 + \langle R_i'^2\rangle] + 2\langle R_i\rangle\langle\sigma_i^2\rangle\} \\ &= \langle\zeta^4\rangle\{(\langle\sigma_i^2\rangle + \langle R_i\rangle)^2 + \mathrm{var}(\sigma_i^2) + \mathrm{var}(R_i)\} = \langle\zeta^4\rangle\{\langle\upsilon_i^2\rangle^2 + \mathrm{var}(\sigma_i^2) + \mathrm{var}(R_i)\}. \end{aligned} \qquad \text{(A2)}$$

Since $\langle\zeta^4\rangle = 3$ for a normal distribution, (A2) implies that

$$\begin{aligned}\mathrm{var}(\sigma^2) &= \frac{\langle\upsilon^4\rangle}{3} - \langle\upsilon^2\rangle^2 - \mathrm{var}(R) \\ &= \frac{\langle\upsilon^4\rangle}{3} - (\langle\sigma^2\rangle + \langle R\rangle)^2 - \mathrm{var}(R). \end{aligned} \qquad \text{(A3)}$$

Equations (A1) and (A3) imply the estimators

$$\begin{aligned}\langle\sigma^2\rangle &\approx \frac{1}{n}\sum_{i=1}^{n}(\upsilon_i^2 - R_i), \\ \mathrm{var}(\sigma^2) &\approx \frac{\frac{1}{n}\sum_{i=1}^{n}\upsilon_i^4}{3} - \left[\frac{1}{n}\sum_{i=1}^{n}(\upsilon_i^2)\right]^2 \\ &\quad - \frac{1}{n-1}\sum_{i=1}^{n}\left(R_i - \frac{1}{n}\sum_{j=1}^{n}R_j\right)^2. \end{aligned} \qquad \text{(A4)}$$

The climatological minimum of ensemble variance may be predefined by the ensemble designer or, provided that a large sample is available, a reasonable estimator of $s^2_{\min}$ is simply the minimum of all realizations of s_i^2; in other words,

$$s^2_{\min} = \min(s_i^2) \quad \text{over all } i. \qquad \text{(A5)}$$

To estimate the parameter a in the likelihood distribution (3), note that (3) and (4) imply that s^2 is a stochastic process of the form

$$\begin{aligned}(s^2 - s^2_{\min}) &= a(\sigma^2 - \sigma^2_{\min}) + \xi, \quad \text{where } \xi_i \text{ is random}, \\ \langle\xi\rangle &= \langle\xi\sigma^2\rangle = 0 \quad \text{but} \quad \langle\xi^3\rangle \neq 0. \end{aligned} \qquad \text{(A6)}$$

Taking the mean of (A6) over all realizations gives $\langle s^2\rangle - s^2_{\min} = a(\langle\sigma^2\rangle - \sigma^2_{\min})$. Subtracting this from both sides of (A6) then gives

$$\begin{aligned}&s^2 - \langle s^2\rangle = a(\sigma^2 - \langle\sigma^2\rangle) + \xi \Rightarrow s^{2\prime} = a\sigma^{2\prime} + \xi, \\ &\text{where} \quad s^{2\prime} = s^2 - \langle s^2\rangle \quad \text{and} \quad \sigma^{2\prime} = \sigma^2 - \langle\sigma^2\rangle. \end{aligned} \qquad \text{(A7)}$$

We can then estimate a from the covariance

$$\begin{aligned}\langle\upsilon^2 s^{2\prime}\rangle &= \left\langle\left(\zeta\sqrt{(\sigma^2 + R)}\right)^2(s^{2\prime})\right\rangle = \langle[\zeta^2[\langle\sigma^2\rangle + (\sigma^{2\prime})][a\sigma^{2\prime} + \xi]]\rangle = a\langle\zeta^2(\sigma^{2\prime})^2\rangle \quad \{\langle\xi\rangle = \langle\xi\sigma^{2\prime}\rangle = \langle\xi\zeta\rangle = \langle\sigma^{2\prime}\langle\sigma_i^2\rangle\rangle = 0\} \\ &= a\langle\zeta^2\rangle\langle(\sigma^{2\prime})^2\rangle = a\,\mathrm{var}(\sigma^2) \quad \{\langle\zeta^2\rangle = 1 \quad \text{and} \quad \langle\zeta\sigma^{2\prime}\rangle = 0\}; \end{aligned} \qquad \text{(A8)}$$

since $\langle\upsilon^2 s^{2\prime}\rangle = \langle(\upsilon^2 - \langle\upsilon^2\rangle)s^{2\prime}\rangle = \mathrm{covar}(\upsilon^2, s^2)$, where $\mathrm{covar}(\upsilon^2, s^2)$ gives the covariance between υ^2 and s^2 because $\langle\langle\upsilon^2\rangle s^{2\prime}\rangle = \langle\upsilon^2\rangle\langle s^{2\prime}\rangle = 0$. Consequently,

$$\begin{aligned}a &= \frac{\mathrm{covar}(\upsilon^2, s^2)}{\mathrm{var}(\sigma^2)} \approx \frac{\left[\frac{1}{n-1}\sum_{i=1}^{n}\upsilon_i^2(s_i^2 - \langle s^2\rangle)\right]}{\mathrm{var}(\sigma^2)} \\ &= \frac{\left[\frac{1}{n-1}\sum_{i=1}^{n}(\upsilon_i^2 - \langle\upsilon^2\rangle)(s_i^2 - \langle s^2\rangle)\right]}{\mathrm{var}(\sigma^2)} \\ &\text{where} \quad \langle s^2\rangle = \frac{1}{n}\sum_{i=1}^{n}s_i^2 \quad \text{and} \quad \langle\upsilon^2\rangle = \frac{1}{n}\sum_{i=1}^{n}\upsilon_i^2 \end{aligned} \qquad \text{(A9)}$$

and $\mathrm{var}(\sigma^2)$ is obtained from (A3). To obtain $\sigma^2_{\min}$, we rearrange the mean $\langle s^2\rangle - s^2_{\min} = a(\langle\sigma^2\rangle - \sigma^2_{\min})$ of (A6) to obtain

$$\sigma^2_{\min} = \langle\sigma^2\rangle - \frac{\langle s^2\rangle - s^2_{\min}}{a}. \qquad \text{(A10)}$$

To define the likelihood pdf given by (3), we must be able to deduce the parameter k from the observational dataset. Equation (5) defines k for a single value of σ^2. Rearranging this equation gives

$$\begin{aligned}k\,\mathrm{var}[(s^2 - s^2_{\min})\,|\,\sigma^2] &= k\langle[(s^2 - s^2_{\min}) - \langle s^2 - s^2_{\min}\rangle_{\sigma^2}]^2\rangle_{\sigma^2} \\ &= a^2(\sigma^2 - \sigma^2_{\min})^2, \end{aligned} \qquad \text{(A11)}$$

where $\langle x\rangle_{\sigma^2}$ indicates that the variable x is averaged while holding the true error variance σ^2 constant. Using the fact that $\langle s^2 - s^2_{\min}\rangle_{\sigma^2} = a(\sigma^2 - \sigma^2_{\min})$ in (A11) and taking the expectation of (A11) over all realizations of σ^2 then gives

$$\begin{aligned}&k\langle\langle[(s^2 - s^2_{\min}) - a(\sigma^2 - \sigma^2_{\min})]^2\rangle_{\sigma^2}\rangle = \langle[a(\sigma^2 - \sigma^2_{\min})]^2\rangle \\ &\Rightarrow k\langle[(\langle s^2\rangle - s^2_{\min} + s^{2\prime}) - a[(\langle\sigma^2\rangle - \sigma^2_{\min}) + \sigma^{2\prime}]]^2\rangle \\ &= a^2[(\langle\sigma^2\rangle - \sigma^2_{\min})^2 + \langle(\sigma^{2\prime})^2\rangle] \end{aligned} \qquad \text{(A12)}$$

Using $\langle s^2\rangle - s^2_{\min} = a(\langle \sigma^2\rangle - \sigma^2_{\min})$ in (A12) gives

$$k\langle (s^{2\prime} - a\sigma^{2\prime})^2\rangle = a^2[(\langle \sigma^2\rangle - \sigma^2_{\min})^2 + \langle (\sigma^{2\prime})^2\rangle] \Rightarrow k\langle (s^{2\prime})^2 - 2a\sigma^{2\prime}s^{2\prime} + a^2(\sigma^{2\prime})^2\rangle = a^2[(\langle \sigma^2\rangle - \sigma^2_{\min})^2 + \langle (\sigma^{2\prime})^2\rangle]. \tag{A13}$$

Using (A7)'s relation $s^{2\prime} = a\sigma^{2\prime} + \xi$ to substitute into the $2a\sigma^{2\prime}s^{2\prime}$ term gives

$$\Rightarrow k\langle (s^{2\prime})^2 - 2a\sigma^{2\prime}(a\sigma^{2\prime} + \xi) + a^2(\sigma^{2\prime})^2\rangle = a^2[(\langle \sigma^2\rangle - \sigma^2_{\min})^2 + \langle (\sigma^{2\prime})^2\rangle]$$
$$\Rightarrow k\langle (s^{2\prime})^2 - a^2(\sigma^{2\prime})^2\rangle = a^2[(\langle \sigma^2\rangle - \sigma^2_{\min})^2 + \langle (\sigma^{2\prime})^2\rangle] \quad \{\text{used } \langle \sigma^{2\prime}\xi\rangle = 0\}$$
$$\Rightarrow k = \frac{a^2[(\langle \sigma^2\rangle - \sigma^2_{\min})^2 + \langle (\sigma^{2\prime})^2\rangle]}{\langle (s^{2\prime})^2\rangle - a^2\langle (\sigma^{2\prime})^2\rangle} = \frac{a^2[(\langle \sigma^2\rangle - \sigma^2_{\min})^2 + \mathrm{var}(\sigma^2)]}{\mathrm{var}(s^2) - a^2\,\mathrm{var}(\sigma^2)}, \tag{A14}$$

where one can use the estimator $\mathrm{var}(s^2) \approx 1/(n-1) \times \sum_{i=1}^{n}(s^2 - \langle s^2\rangle)^2$.

b. Derivation of (15)

To prove the veracity of (15), begin by noting that (2) implies that

$$\alpha - 1 = \frac{(\langle \sigma^2\rangle - \sigma^2_{\min})^2 + \mathrm{var}(\sigma^2)}{\mathrm{var}(\sigma^2)}. \tag{A15}$$

Using (A15) and (11) in (15) then gives

$$w_E = \frac{1}{a}\frac{k}{(k + \alpha - 1)} = \frac{1}{a}\frac{\dfrac{a^2[\mathrm{var}(\sigma^2)(\alpha-1)]}{\mathrm{var}(s^2) - a^2\,\mathrm{var}(\sigma^2)}}{\dfrac{a^2[\mathrm{var}(\sigma^2)(\alpha-1)]}{\mathrm{var}(s^2) - a^2\,\mathrm{var}(\sigma^2)} + (\alpha - 1)} = \frac{1}{a}\frac{\dfrac{a^2[\mathrm{var}(\sigma^2)(\alpha-1)]}{\mathrm{var}(s^2) - a^2\,\mathrm{var}(\sigma^2)}}{\dfrac{a^2[\mathrm{var}(\sigma^2)(\alpha-1)] + (\alpha-1)[\mathrm{var}(s^2) - a^2\,\mathrm{var}(\sigma^2)]}{\mathrm{var}(s^2) - a^2\,\mathrm{var}(\sigma^2)}}$$
$$= \frac{1}{a}\frac{a^2[\mathrm{var}(\sigma^2)(\alpha-1)]}{a^2[\mathrm{var}(\sigma^2)(\alpha-1)] + (\alpha-1)[\mathrm{var}(s^2) - a^2\,\mathrm{var}(\sigma^2)]} = \frac{1}{a}\frac{a^2[\mathrm{var}(\sigma^2)(\alpha-1)]}{(\alpha-1)[\mathrm{var}(s^2)]} = \frac{a[\mathrm{var}(\sigma^2)]}{[\mathrm{var}(s^2)]}$$
$$= \frac{\mathrm{covar}(v^2, s^2)}{\mathrm{var}(s^2)}, \quad [\text{from (6)}] \tag{A16}$$

as was required by (15).

REFERENCES

Atger, F., 1999: The skill of ensemble prediction systems. *Mon. Wea. Rev.,* **127,** 1941–1953.

Bishop, C. H., and E. A. Satterfield, 2013: Hidden error variance theory. Part I: Exposition and analytic model. *Mon. Wea. Rev.,* **141,** 1454–1468.

——, 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.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.,* **78,** 1–3.

Casati, B., and Coauthors, 2008: Forecast verification: Current status and future directions. *Meteor. Appl.,* **15,** 3–18.

DelSole, T., 2004: Predictability and information theory. Part I: Measures of predictability. *J. Atmos. Sci.,* **61,** 2425–2440.

Fortin, V., A.-C. Favre, and M. Saïd, 2006: Probabilistic forecasting from ensemble prediction systems: Improving upon the best-member method by using a different weight and dressing kernel for each member. *Quart. J. Roy. Meteor. Soc.,* **132,** 1349–1369.

Gneiting, T., F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *J. Roy. Stat. Soc.,* **69B,** 243–268.

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

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

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting,* **15,** 559–570.

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

Kleeman, R., 2002: Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.,* **59,** 2057–2072.

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

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

Mason, S. J., and N. E. Graham, 2002: Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quart. J. Roy. Meteor. Soc.,* **128,** 2145–2166.

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.

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.

Roulston, M. S., and L. A. Smith, 2002: Evaluating probabilistic forecasts using information theory. *Mon. Wea. Rev.*, **130,** 1653–1660.

——, and ——, 2003: Combining dynamical and statistical ensembles. *Tellus,* **55A,** 16–30.

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

Vrugt, J. A., M. P. Clark, C. G. H. Diks, Q. Duan, and B. A. Robinson, 2006: Multi-objective calibration of forecast ensembles using Bayesian model averaging. *Geophys. Res. Lett.,* **33,** L19817, doi:10.1029/2006GL027126.

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

Wilks, D. S., 2001: A skill score based on economic value for probability forecasts. *Meteor. Appl.,* **8,** 209–219.

——, and T. M. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev.,* **135,** 2379–2390.

Wilson, L. J., S. Beauregard, A. E. Raftery, and R. Verret, 2007: Calibrated surface temperature forecasts from the Canadian Ensemble Prediction System using Bayesian model averaging. *Mon. Wea. Rev.,* **135,** 1364–1385.