A Comparison of the Hybrid and EnSRF Analysis Schemes in the Presence of Model Errors due to Unresolved Scales

XUGUANG WANG

School of Meteorology, and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

THOMAS M. HAMILL AND JEFFREY S. WHITAKER

Physical Sciences Division, NOAA Earth System Research Laboratory, Boulder, Colorado

CRAIG H. BISHOP

Naval Research Laboratory, Monterey, California

(Manuscript received 5 January 2009, in final form 6 April 2009)

ABSTRACT

A hybrid analysis scheme is compared with an ensemble square root filter (EnSRF) analysis scheme in the presence of model errors as a follow-up to a previous perfect-model comparison. In the hybrid scheme, the ensemble perturbations are updated by the ensemble transform Kalman filter (ETKF) and the ensemble mean is updated with a hybrid ensemble and static background-error covariance. The experiments were conducted with a two-layer primitive equation model. The true state was a T127 simulation. Data assimilation experiments were conducted at T31 resolution (3168 complex spectral coefficients), assimilating imperfect observations drawn from the T127 nature run. By design, the magnitude of the truncation error was large, which provided a test on the ability of both schemes to deal with model error. Additive noise was used to parameterize model errors in the background ensemble for both schemes. In the first set of experiments, additive noise was drawn from a large inventory of historical forecast errors; in the second set of experiments, additive noise was drawn from a large inventory of differences between forecasts and analyses. The static covariance was computed correspondingly from the two inventories. The hybrid analysis was statistically significantly more accurate than the EnSRF analysis. The improvement of the hybrid over the EnSRF was smaller when differences of forecasts and analyses were used to form the random noise and the static covariance. The EnSRF analysis was more sensitive to the size of the ensemble than the hybrid. A series of tests was conducted to understand why the EnSRF performed worse than the hybrid. It was shown that the inferior performance of the EnSRF was likely due to the sampling error in the estimation of the model-error covariance in the mean update and the less-balanced EnSRF initial conditions resulting from the extra localizations used in the EnSRF.

1. Introduction

The ensemble Kalman filter (EnKF)–based data assimilation (DA) method has been explored extensively since it was described and tested by Evensen (1994) in the oceanographic application and by Houtekamer and Mitchell (1998) in the atmospheric application. Relative to the three-dimensional variational data assimilation method (3DVAR) that utilizes stationary background-error covariances (e.g., Parrish and Derber 1992; Courtier et al. 1998; Gauthier et al. 1998; Cohn et al. 1998), the presumed benefit of the EnKF method is its ability to provide flow-dependent estimates of the background-error covariances through ensemble covariances so that the observations and the background are more appropriately weighted during the assimilation. Encouraging results have been reported in both observing system simulation experiments (OSSEs) and experiments with real numerical weather prediction (NWP) models and observations; in the numerical predictions from global to convective scales; and for applications on the state

Corresponding author address: Dr. Xuguang Wang, School of Meteorology, University of Oklahoma, 120 David L. Boren Blvd., Suite 5341, Norman, OK 73072.
E-mail: xuguang.wang@ou.edu

DOI: 10.1175/2009MWR2923.1

© 2009 American Meteorological Society

estimates of the atmosphere, ocean, and land surface (Evensen 1994; Houtekamer and Mitchell 1998, 2001; Anderson 2001; Szunyogh et al. 2005, 2008; Torn et al. 2006; Houtekamer et al. 2005, 2009; Whitaker et al. 2004, 2008; Miyoshi et al. 2007; Snyder and Zhang 2003; Dowell et al. 2004; Tong and Xue 2005; Meng and Zhang 2008; Liu et al. 2008b; Dirren et al. 2007; Reichle et al. 2002; Keppenne and Rienecker 2002; Yang et al. 2008; for reviews, see Evensen 2003; Lorenc 2003; Hamill 2006; Ehrendorfer 2007).

An emerging alternative to using ensemble forecasts in data assimilation is called the hybrid data assimilation scheme. In the hybrid method, ensemble forecasts are incorporated into the variational update of the background forecast. Like the EnKF, the hybrid runs short-term ensemble-forecast cycles. Unlike the EnKF, the hybrid can be implemented with minor changes to the existing variational codes that have been used operationally. The hybrid method, including a hybrid ensemble with both 3DVAR and 4DVAR, has been a subject of a number of recent papers (e.g., Hamill and Snyder 2000; Lorenc 2003; Etherton and Bishop 2004; Buehner 2005, Zupanski 2005; Wang et al. 2007a,b, 2008a,b; Liu et al. 2008a; Zhang et al. 2009). The focus of the discussion of the current paper is on the hybrid ensemble 3DVAR. So far, both simple model tests (e.g., Hamill and Snyder 2000; Etherton and Bishop 2004; Wang et al. 2007a) and real NWP model tests (e.g., Buehner 2005; Wang et al. 2008a,b) have demonstrated superior performance of the hybrid ensemble-3DVAR method (in this paper, simply called ''the hybrid'') relative to 3DVAR. Note that Etherton and Bishop 2004 and Wang et al. 2007a did not use the 3DVAR framework directly. They tested the performance of the hybrid by incorporating the ensemble covariance in the classic optimum interpolation (Schlatter 1975) framework. However, under their experiment design, it will provide the same solution as if they had adopted the 3DVAR framework (Daley 1991). The same approach will be adopted in this paper because we focus on the quality of the analysis rather than the computational expense and we want to save labors from developing 3DVAR for the model to be used.

Both the EnKF and the hybrid take advantage of the flow-dependent ensemble-estimated background-error covariance. How does the hybrid compare to the EnKF? There have not been many published studies focusing on directly comparing the performance of the hybrid and the EnKF-based approaches and on understanding their underlying differences. Recent work by Wang et al. (2007a) directly compared the hybrid where the ensemble was generated by the ensemble transform Kalman filter (ETKF; Wang and Bishop 2003; Wang et al. 2004, 2007a) and the ensemble square root filter (EnSRF; Whitaker and Hamill 2002) by using a primitive equation two-layer model with a perfect-model assumption. The EnSRF was chosen because it was one of the well- and extensively tested EnKF-based approaches and was demonstrated with improved performance relative to classical EnKF (Whitaker and Hamill 2002). Wang et al. (2007a) found that the analysis generated by the hybrid scheme was more accurate than the EnSRF when the ensemble size was small. However, to better simulate realistic NWP applications, the effect of model errors must be considered. How then does the hybrid compare to the EnSRF in the presence of model error? What are the underlying reasons for differences in their performances? These are the questions we seek to answer in this paper.

As a follow-up to Wang et al. (2007a), we compare the two schemes by using the same global primitive equation two-layer model but with model errors. As in Hamill and Whitaker (2005), we examine a relatively simple source of model error, the errors introduced by the truncation of the forecast model. With this relatively simple experiment setting, it will be easier to discern fundamental differences of the two schemes in the presence of one type of model error, which will be beneficial to future work of comparing the two schemes with real NWP models and observations.

Treatment of model errors has been a subject of many data assimilation studies. For ensemble-based methods, model errors were treated by using ensemble covariance inflation (e.g., Whitaker and Hamill 2002), additive noise (e.g., Whitaker et al. 2008; Houtekamer et al. 2005), ensemble covariance relaxation (e.g., Zhang et al. 2004), the multiphysics–multimodel method (e.g., Meng and Zhang 2008), and stochastic perturbations to physical tendencies and stochastic kinetic energy (KE) backscatter (e.g., Houtekamer et al. 2009). In 4DVAR, model error was treated by using a weak constraint in a 4DVAR (e.g., Zupanski 1997). Dee and da Silva (1998) described and applied a model-bias correction method for the sequential data assimilation system. For the simple model error resulting from truncation considered here, we follow Hamill and Whitaker (2005) and adopt the additive noise method, which is easy to implement and was shown to provide the most accurate analysis among all the model-error treatment methods considered by Hamill and Whitaker (2005).

The hybrid and the EnSRF schemes will be described in section 2. Section 3 provides a description on the experiment design. Sections 4, 5, and 6 describe the results and tests conducted to understand the difference between the two schemes. Section 7 offers conclusions and conjectures of the results presented in the prior sections.

2. The hybrid and the EnSRF data assimilation methods

a. The hybrid scheme

Figure 1 of Wang et al. (2007a) describes, in general, how the hybrid data assimilation cycle works. Compared to the perfect-model experiment from Wang et al. (2007a), the main difference in the current application of the hybrid scheme is the representation of model error in the ensemble update, which will be specified later in this section.

We first consider the update of the mean in the hybrid method. The ensemble-mean forecast $\overline{\mathbf{x}}^b$ is updated by observations $\mathbf{y}$ to obtain the ensemble-mean analysis $\overline{\mathbf{x}}^a$ by using

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{P}^b\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^b), \qquad (1)$$

where $\mathbf{H}$ is the observation operator mapping from the model state variables to the observed variables, which is presumed linear here; $\mathbf{R}$ is the observation-error covariance matrix; and $\mathbf{P}^b$ is the background-error covariance. As in Wang et al. (2007a), $\mathbf{P}^b\mathbf{H}^{\mathrm{T}}$ and $\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}}$ are formed by

$$\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} = (1-\alpha)(\boldsymbol{\rho}_s^{p\times p} \circ \mathbf{H}\mathbf{P}^e\mathbf{H}^{\mathrm{T}}) + \alpha(f\mathbf{H}\mathbf{B}\mathbf{H})^{\mathrm{T}} \quad \text{and} \qquad (2)$$

$$\mathbf{P}^b\mathbf{H}^{\mathrm{T}} = (1-\alpha)(\boldsymbol{\rho}_s^{n\times p} \circ \mathbf{P}^e\mathbf{H}^{\mathrm{T}}) + \alpha(f\mathbf{B}\mathbf{H})^{\mathrm{T}}, \qquad (3)$$

where $\mathbf{P}^e\mathbf{H}^{\mathrm{T}}$ and $\mathbf{H}\mathbf{P}^e\mathbf{H}^{\mathrm{T}}$ are calculated from the K ETKF (Wang et al. 2007a) ensemble-forecast perturbations ($\mathbf{x}_k^{e'}, k = 1, \ldots, K$). Note in the second terms of Eqs. (2) and (3) that covariance localization was applied through Schur product $\circ$ between a correlation matrix and the raw ensemble-covariance matrix (Hamill et al. 2001; Houtekamer and Mitchell 2001). Horizontal localization by using Gaspari and Cohn's (1999) locally supported, approximately Gaussian-shaped function is used to form the correlation matrices. As in Wang et al. (2007a), the static covariance $\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}$ and $\mathbf{B}\mathbf{H}^{\mathrm{T}}$ are formed from a large inventory of historical forecast errors over many separate times (see section 3b). Following Etherton and Bishop (2004) and Wang et al. (2007a), a rescaling factor f was used to rescale the static covariance matrix so that the total variance of the rescaled covariance matrix was equal to the total forecast-error variance in the observation space [under the norm of trace($\mathbf{R}^{-1/2}\mathbf{H}\mathbf{P}^t\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1/2}$)]. As in Eq. (21) of Wang et al. (2007a), the rescaling factor f was determined dynamically. The user-tunable factor α, where $0 \le \alpha \le 1$, determines the relative weights placed on the static and the ensemble covariances. As discussed in Wang et al. (2007a), an inflation factor was applied so that the ETKF ensemble-forecast variance was equal to the total forecast-error variance in the observation space as well. Designed this way, the weighting factor α preserves the total variance (e.g., Etherton and Bishop 2004). As noted in the previous section, although we updated the mean by using the classic optimum interpolation (OI) formula (Schlatter 1975), it will provide the same solution as if we had adopted the 3DVAR framework (Daley 1991; Wang et al. 2007b).

We now consider the method for updating perturbations around the mean state. The ensemble perturbations are updated by the ETKF (Wang and Bishop 2003; Wang et al. 2004, 2007a). The ETKF transforms the matrix of background ensemble perturbations $\mathbf{X}^b$ into a matrix of analysis perturbations $\mathbf{X}^a$ by using a transformation matrix. Assuming that the covariance of the raw forecast ensemble perturbations was equal to the true forecast-error covariance, then the transformation matrix is derived so that the outer product of the transformed perturbations was equal to the true analysis error covariance. The same ETKF formula described in Wang et al. (2007a) was adopted here.

Unlike the perfect-model experiment (Wang et al. 2007a) where background ensemble perturbation $\mathbf{X}^b$ is formed only from the ETKF ensemble forecasts, in this imperfect-model experiment, following Hamill and Whitaker (2005), we account for the model error in $\mathbf{X}^b$ by using the additive noise method. The background ensemble perturbation $\mathbf{X}^b$ ($\mathbf{x}_k'^b$, $k = 1, \ldots, K$) is constructed as

$$\mathbf{x}_k'^b = \sqrt{(1-\alpha)}\mathbf{x}_k'^e + \sqrt{\alpha f}\mathbf{e}_k, \qquad (4)$$

where $\mathbf{x}_k'^e$ is the ensemble-forecast perturbation generated from analysis ensemble updated by the ETKF method, and $\mathbf{e}_k$ is a random sample drawn from the large inventory of the historical forecast errors that form the static covariance $\mathbf{B}$, which will be described in section 3b. Note that $\mathbf{x}_k'^e$ is used to compute $\mathbf{P}^e\mathbf{H}^{\mathrm{T}}$ and $\mathbf{H}\mathbf{P}^e\mathbf{H}^{\mathrm{T}}$ in Eqs. (2) and (3). Therefore, the relative weight of the ETKF perturbation and the random perturbation in the background ensemble perturbation is consistent with the weight of the ETKF ensemble covariance and the static covariance in the background-error covariance used to update the mean [Eqs. (2) and (3)]. Designed this way, the estimated background-error covariances used for the perturbation update and the mean update are more consistent. Also note that, in Eq. (4), we adopted random noise consistent with the static covariance $\mathbf{B}$. These random perturbations can also be considered drawn from the eigenvector space of $\mathbf{B}$. This method is similar

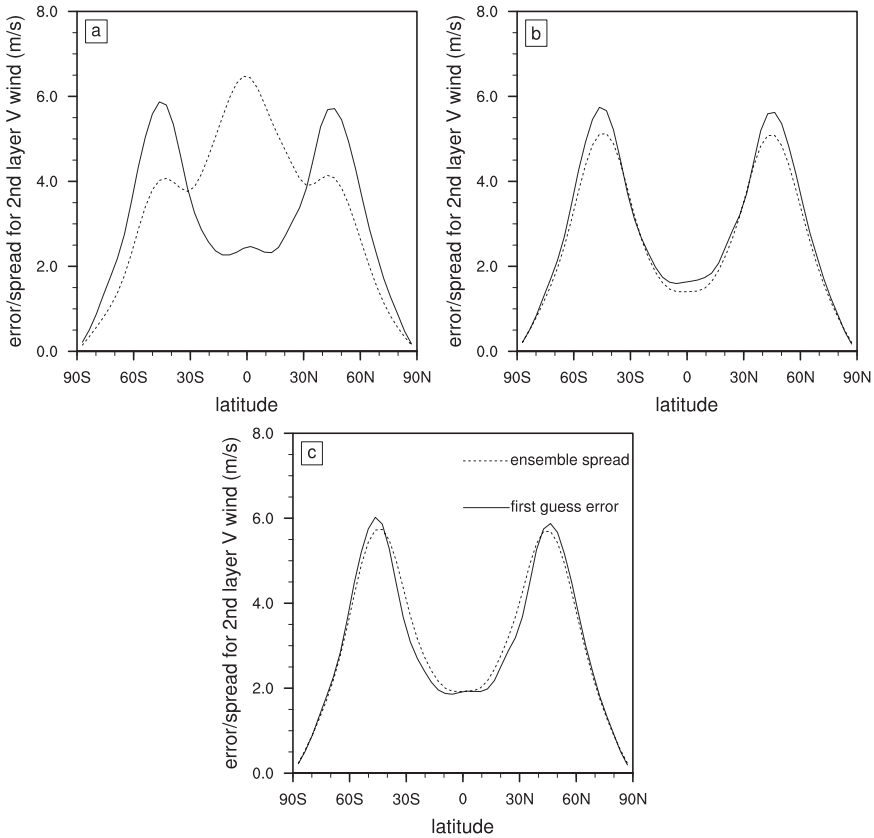

FIG. 1. Zonally averaged RMS first-guess second-layer meridional wind error (solid) and background ensemble spread (dotted) as a function of latitude for (a) the hybrid, with localization scale of 15 000 km and no additive error; (b) the hybrid, with localization scale of 15 000 km and additive error to parameterize model error in the background ensemble; and (c) EnSRF, with localization scale of 5000 km and additive error to parameterize model error in the background ensemble. The weighting coefficient of (a)–(c) is 0.4.

to Houtekamer et al. (2005, 2009), where random noise drawn from the 3DVAR static covariance was used to parameterize model errors. In practice, such random perturbations are easy to obtain.

To test the background ensemble perturbation in Eq. (4), Fig. 1b shows the root-mean-square (rms) background error and the ensemble spread as a function of latitude for an experiment with the hybrid scheme ($\alpha = 0.4$ and the localization scale = 15 000 km). As a comparison, another experiment where no additive noise was applied and thus a global constant inflation was used to parameterize model error is shown in Fig. 1a. Consistent with the findings in Hamill and Whitaker (2005), with a globally constant inflation, the background ensemble spread was abnormally large in the tropics, whereas by using the additive noise method the background ensemble spread better matched the latitudinal variation of the background errors. As explained by Hamill and Whitaker (2005), the actual growth of model error depends on the dynamics and grows more rapidly in the midlatitudes. Although the constant inflation uniformly expanded the spread, the additive noise has larger magnitude in the midlatitude (not shown).

b. The EnSRF analysis scheme

As opposed to the hybrid, which assimilates observations simultaneously, the EnSRF serially assimilates observations. The ensemble perturbations updated by the previous observations are used to model the background-error covariance for assimilating the next observation (for details, see Whitaker and Hamill 2002). Similarly, the updated mean from the assimilation of the previous observation is used as the prior state for the assimilation of the next observation. The EnSRF update equations for assimilating the ith single observation y_i are as follows:

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{K}_i(y_i - \mathbf{H}_i\overline{\mathbf{x}}^b) \quad \text{and} \tag{5}$$

$$\mathbf{x}_k'^a = (\mathbf{I} - \tilde{\mathbf{K}}_i\mathbf{H}_i)\mathbf{x}_k'^b\Pi. \tag{6}$$

Note that in the previous equations, $\mathbf{H}_i$ maps the state vector to the ith observation space. In Eq. (5), $\mathbf{K}_i$ is the Kalman gain modified by the covariance localization

$$\mathbf{K}_i = (\boldsymbol{\rho}_s^{n\times 1} \circ \mathbf{P}^b\mathbf{H}_i^{\mathrm{T}})(\mathbf{H}_i\mathbf{P}^b\mathbf{H}_i^{\mathrm{T}} + \mathbf{R}_{ii})^{-1}. \tag{7}$$

As in the hybrid, horizontal localization utilizes the approximately Gaussian-shaped function of Gaspari and Cohn (1999). In Eq. (6), $\tilde{\mathbf{K}}_i$ is called the reduced Kalman gain matrix (Whitaker and Hamill 2002). For serial assimilation,

$$\tilde{\mathbf{K}}_i = \left(1 + \sqrt{\frac{\mathbf{R}_{ii}}{\mathbf{H}_i\mathbf{P}^b\mathbf{H}_i^{\mathrm{T}} + \mathbf{R}_{ii}}}\right)^{-1}\mathbf{K}_i. \tag{8}$$

As in the ETKF, an adaptive inflation Π is used to ensure that the variance of the ensemble forecast initialized from the analysis perturbations $\mathbf{x}'^a$ in Eq. (6) is consistent with the true background-forecast-error variance in the observation space.

We also use additive noise to account for model error in the background ensemble perturbations $\mathbf{x}_k'^b$ in the EnSRF [Eq. (6)]. For the purpose of a parallel comparison with the hybrid, the background ensemble perturbation in the EnSRF is constructed the same as that in the hybrid [Eq. (4)]; in this case, $\mathbf{x}_k'^e$ is the kth ensemble-forecast perturbation generated from ensemble forecasts initialized by the analysis ensemble and updated by the EnSRF method. Figure 1c also illustrates that, with the additive noise method, the background ensemble spread for the EnSRF also can represent the latitudinal variation of the background forecast errors.

3. Experiment design

a. *Model, model error, observations, ensemble configuration, and verification methods*

In this study, we ran a dry, global, two-layer primitive equation model (Zou et al. 1993). It was previously used in Hamill et al. (2001), Hamill and Whitaker (2005), and Wang et al. (2007a) for ensemble data assimilation experiments in both perfect- and imperfect-model contexts. The model is spectral, and the model state vector includes coefficients of vorticity and divergence at two levels and coefficients of two layer thicknesses $\Delta\pi_1$ and $\Delta\pi_2$, where π is the Exner function. There is a simple, zonal wavenumber 2 terrain. The model is forced by Newtonian relaxation to a prescribed interface Exner function. A fourth-order Runge–Kutta scheme is used for numerical integration, and ∇^8 hyperdiffusion is used. The parameters chosen are the same as in Hamill and Whitaker (2005).

We assume that the "true" atmospheric state is described by the forecast dynamics at T127 resolution. All data assimilation experiments were conducted at T31 resolution (the number of complex spectral coefficients predicted by the model is 3168, and the dimension of the model in the Gaussian grid is 27 648). In other words, we assume that our data assimilation and forecast system is only able to resolve scales T31 and larger. The short-term model error in T31 resolution is thus due to the lack of representation of the interaction with the unresolved scales (for detailed descriptions on the characteristics of the model and model errors due to unresolved scales, see Hamill and Whitaker 2005). Also as discussed in Hamill and Whitaker (2005), this setup was designed to produce large model errors in order to provide a stringent test on the ability of the two schemes to deal with model errors. Model errors here are dominated by random rather than systematic components.

Observations of interface π and surface π were taken at a set of nearly equally spaced locations on a spherical geodesic grid (Fig. 2 of Wang et al. 2007a). The 362 observations of each consisted of the T127 true state plus errors drawn from a distribution with zero mean and a standard deviation of 8.75 J kg^{-1} K^{-1} for interface π and 0.875 J kg^{-1} K^{-1} for surface π, the same values used in Wang et al. (2007a). Observation errors were constructed to be independent spatially and temporally. As in Hamill and Whitaker (2005), given that the error doubling time of the model at T31 is 3.78 days, observations were assimilated every 24 h.

Following Hamill and Whitaker (2005), we first ran both systems with 200 ensemble members. Then, to study the sensitivity of each scheme to ensemble sizes, we ran 50 members. The ensemble was initialized with random draws from the model climatology. The data assimilation was conducted for a 150-day period, and the error statistics were evaluated over the last 100 days. The statistical significance of the following results was evaluated with a paired sample t test with the temporal correlation of the data taken into account (Wilks 2006, p. 455).

b. *Formation of static background-error covariance and inventory of random noise*

In the first set of experiments, the static background-error covariance $\mathbf{B}$ and the random noise $\mathbf{e}_k$ were constructed from a large inventory of historical forecast errors over many separate times. We call this inventory the "forecast minus truth" inventory. Following Wang et al. (2007a), an iterative procedure was taken to construct such an inventory to form the static covariance $\mathbf{B}$ that produced the smallest analysis errors. In the final iteration, 6541 samples of 24-h forecast errors were collected.

TABLE 1. The rms errors in KE norm (m s^{-1}), upper-layer Exner function thickness norm $\Delta\pi_2$ (J kg^{-1} K^{-1}), and surface Exner function norm π_s (J kg^{-1} K^{-1}) for the analyses of the hybrid, EnSRF, and static for the (top) 200- and (bottom) 50-member ensembles. Only the best-performing hybrid and EnSRF are shown. The fourth column shows the absolute and relative improvement of the hybrid over the EnSRF. The last column is the confidence level at which the rms errors of the hybrid and EnSRF are different. The optimal weighting coefficients and localization length scales are shown in parentheses in the second and third columns (e.g., 0.4 means a weight of 0.4 is placed on the static covariance and 15 k means 15 000 km).

200 member	Hybrid	EnSRF	EnSRF – hybrid	Static	Confidence level
KE	3.938 (0.4, 15 k)	4.237 (0.4, 5 k)	0.299 (7%)	4.509	>99%
$\Delta\pi_2$	5.997 (0.4, 15 k)	6.317 (0.4, 5 k)	0.32 (5%)	6.631	>99%
π_s	0.341 (0.6, 15 k)	0.404 (0.6, 15 k)	0.063 (16%)	0.379	>99%
50 member	**Hybrid**	**EnSRF**	**EnSRF – hybrid**	**Static**	**Confidence level**
KE	4.153 (0.6, 15 k)	4.671 (0.4, 5 k)	0.518 (11%)	4.509	>99%
$\Delta\pi_2$	6.192 (0.6, 15 k)	6.826 (0.4, 5 k)	0.634 (9%)	6.631	>99%
π_s	0.355 (0.8, 15 k)	0.515 (0.4, 5 k)	0.16 (31%)	0.379	>99%

The static background-error covariance matrix **B** was then constructed by directly calculating the covariance of this large inventory of the forecast-error samples. The fifth column of Table 1 shows the rms analysis error of the experiment where we ran a single-member forecast and analysis cycle by using only the static covariance obtained from the last iteration. We denote this experiment as static because it used a static covariance such as 3DVAR and OI. Note that the static covariance and the random noise were generated from the same inventory, which will provide a clean comparison and thus reveal the fundamental differences of the hybrid and the EnSRF schemes. These results are presented in sections 4 and 5.

As discussed in Wang et al. (2007a), the static covariance produced by the above method is likely to be much better than the static covariances formulated for operational 3DVAR. The random noise inventory where parameterized model error is drawn is also not obtainable because in reality the true state can never be known. The static, hybrid, and EnSRF experiments may benefit differently from these assumptions. To test this hypothesis, we form another random noise inventory where we use the analysis to estimate the truth. In other words, instead of collecting the forecast errors (forecast minus truth), we collect the differences between the forecast and the corresponding analysis and calculate the static covariance from the new inventory. We call this the "forecast minus analysis" inventory, which can be more realistically obtained. This is similar to the National Meteorological Center (NMC) method by Parrish and Derber (1992), where the perturbations used to form the static covariance were obtained from the collections of the difference of the 48- and 24-h forecasts. Results of the second set of experiments using this new inventory of random noise and new static covariance are described in section 6.

4. Analysis errors with the static covariance and random noise formed from forecast-minus-truth inventory

We first examine the analysis errors of the different DA schemes by using the static covariance and random noise formed from the forecast-minus-truth inventory described in section 3b. Figure 2 shows the root-mean-square analysis errors in the KE upper-layer Exner function thickness ($\Delta\pi_2$) and surface Exner function (π_s) norms for the hybrid and the EnSRF schemes as functions of the localization scale and the weighting factor for 200-member ensembles. The weighting coefficients tried are 0.2, 0.4, 0.6, and 0.8, and the localization length scales tried are 3000, 5000, 15 000, and 25 000 km. The EnSRF was more sensitive to the localization scales than the hybrid, which means a careful tuning of the covariance localization length is needed for the EnSRF in order to find the optimal performance. The best-performing hybrid and EnSRF from Fig. 2 are summarized in Table 1. It is shown that the best-performing hybrid was statistically significantly better than the best-performing EnSRF. The hybrid improved upon the EnSRF by 7%, 5%, and 16% for the kinetic energy, upper-layer Exner function thickness, and surface Exner function norms, respectively.

To measure the sensitivity of the rms analysis errors of the hybrid and EnSRF with respect to the ensemble size, we also ran both schemes with 50-member ensembles. The results of the best-performing hybrid and EnSRF with 50-member ensembles are summarized in the bottom of Table 1. The best-performing hybrid was still statistically significantly better than the best-performing EnSRF. The relative improvement of the hybrid over EnSRF running 50-member ensembles was 11%, 9%, and 31% for the three norms, which is larger than running 200-member ensembles. The rms analysis

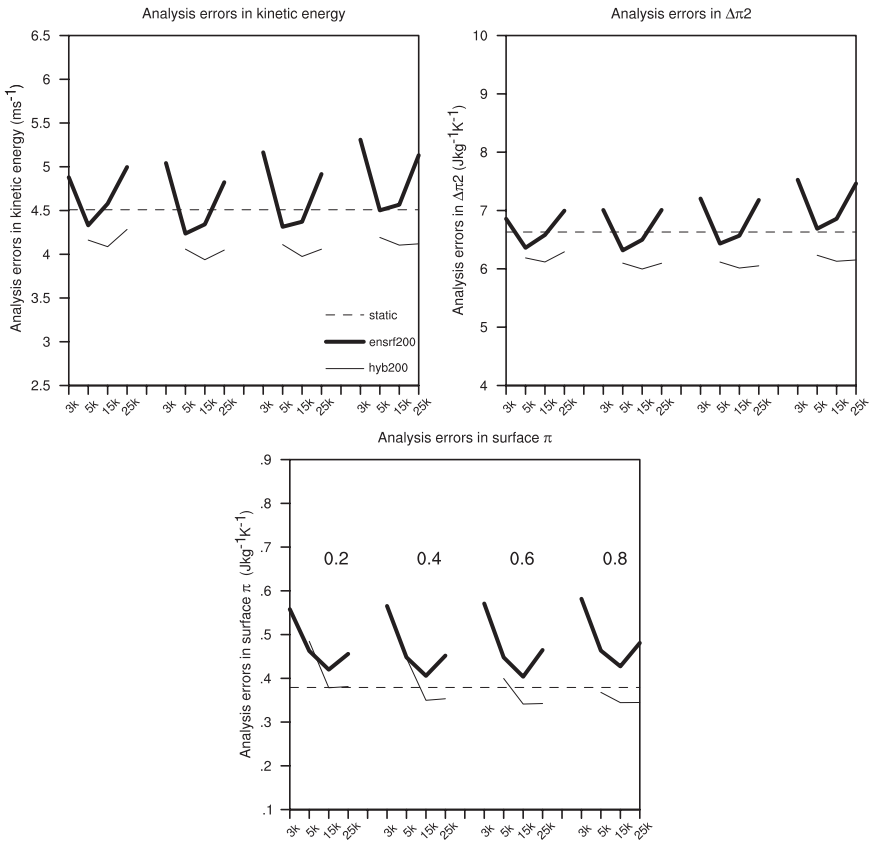

FIG. 2. RMS analysis error for the KE second-layer thickness $\Delta\pi_2$ and surface Exner function π_s norms as a function of localization scales and weighting coefficients for the hybrid (solid thin), EnSRF (solid thick), and static (dashed). The weighing coefficients tried are 0.2, 0.4, 0.6, and 0.8. For the hybrid, localization scales of 5000, 15 000, and 25 000 km were tried for each of weighting coefficients. For the EnSRF, localization scales of 3000, 5000, 15 000, and 25 000 km were tried for each of weighting coefficients.

error of the best-performing 50-member hybrid was comparable or even smaller than that of the best-performing 200-member EnSRF. These results indicate that the hybrid is less sensitive to the ensemble size than the EnSRF, which is consistent with Wang et al. (2007a).

The hybrid running 200-member and 50-member ensembles both outperformed the static for all three norms considered. The 200-member EnSRF outperformed the static, except for the π_s norm, which is different from Hamill and Whitaker (2005), where the EnSRF was better than the static for all three norms. The differences between the current experiment design and that of Hamill and Whitaker (2005) are as follows: 1) Hamill and Whitaker (2005) did not assimilate π_s observations and (probably more importantly) 2) Hamill and Whitaker's (2005) static covariance was formed from 200 historical forecast errors with covariance localization, whereas the static covariance here was formed iteratively from 6541 historical forecast errors with no localization. The 50-member EnSRF did not outperform the static.

5. Why is the hybrid better than the EnSRF?

The source of parameterized model errors is the same in both the ensemble-mean and ensemble-perturbation updates for the hybrid and EnSRF. The superior performance of the hybrid over the EnSRF shown earlier must then arise from algorithmic differences between the two schemes. Although the assumption made in forming the random noise inventory and the static covariance is not realistic because we assumed we knew the truth, this set of experiment still offers opportunities for understanding the underlying differences between the two schemes. In this section, we describe experiments designed to elucidate which of the differences between the hybrid and EnSRF algorithms contributed to the better analysis in the hybrid than the EnSRF as shown in section 4.

TABLE 2. The rms errors in KE norm (m s^{-1}), upper-layer Exner function thickness norm $\Delta\pi_2$ (J kg^{-1} K^{-1}), and surface Exner function norm π_s (J kg^{-1} K^{-1}) for the analyses updated by the EnSRF scheme, the hybrid scheme, and the mixed EnSRF–ETKF scheme (ensemble updated by the ETKF and ensemble mean updated by the EnSRF). A localization scale of 5000 km and a weighting coefficient of 0.4 were used. Both 50- and 200-member ensemble runs were tried.

200 member	EnSRF	Mixed EnSRF – ETKF	Hybrid
KE	4.237	4.312	4.059
$\Delta\pi_2$	6.317	6.339	6.098
π_s	0.448	0.460	0.446
50 member	**EnSRF**	**Mixed EnSRF – ETKF**	**Hybrid**
KE	4.671	4.709	4.268
$\Delta\pi_2$	6.826	6.799 (<80% confidence)	6.286
π_s	0.515	0.563	0.518

a. *Effect of differences in the ensemble-perturbation update*

One difference between the hybrid and the EnSRF is in the update of the ensemble perturbations. On one hand, the ETKF ensemble perturbations in the hybrid have superior balance because the update of the perturbations does not involve covariance localization. On the other hand, because of the global nature of the ETKF, the ETKF perturbations will do a poorer job of resolving the spatial inhomogeniety of the error covariance (Wang and Bishop 2003). To understand if the differences in the ensemble-perturbation update were an important factor in making the hybrid analysis more accurate than the EnSRF, we replace the ensemble-perturbation update in the EnSRF by the ETKF method but still use the EnSRF to obtain the ensemble-mean analyses. The results are summarized in Table 2. Comparing the second and the third columns, it was found that with the ETKF updating the ensemble perturbations, the analysis was no better than using the EnSRF to update the ensemble perturbations. Note that in the only case in Table 2 in which the EnSRF analysis ($\Delta\pi_2$ norm for 50-member ensemble) was inferior, the difference between the EnSRF and the hybrid was not statistically significant.

b. *Effect of sampling error in the estimation of model-error covariance when updating the mean*

The last column of Table 2 shows that when we use the same perturbation update, the ETKF, but use the EnSRF and the hybrid to update the mean, the analysis produced by using the hybrid method to update the mean is more accurate. This suggests that the difference of the two methods in the update of the ensemble mean contributed to the better performance of the hybrid. We then used a single-observation test to understand the difference of the two methods in updating the ensemble mean. Figures 3a–c show the increment by assimilating a single second-layer thickness $\Delta\pi_2$ observation located at 46.64°N, 108°W that was 3 J kg^{-1} K^{-1} smaller than the background forecast. Figure 3a shows the increment of the hybrid with a weighting coefficient of 0.4 and a localization scale of 15 000 km. Note that such a combination of the weighting coefficient and the localization scale produced the best hybrid analyses. The flow-dependent ensemble $\mathbf{x}_k'^e$ was from the 24-h ensemble forecast at the 123rd cycle of the best-performing 200-member hybrid experiment. The exact same flow-dependent ensemble was used in the single-observation experiment for the EnSRF to understand the difference in the update of the ensemble mean. Because the best-performing 200-member EnSRF for the $\Delta\pi_2$ norm was using the 5000-km localization and a weighting coefficient of 0.4, in the following we first compared the hybrid increment with the EnSRF increment by using these parameters (Fig. 3b). Then, we further compared the increment of the hybrid with the increment of the EnSRF, adopting the same localization scale and weighting coefficient as the hybrid (Fig. 3c). Comparing Figs. 3a,b shows that the length scale of the EnSRF increment was shorter than that of the hybrid. Because the hybrid analysis was more accurate, Figs. 3a,b thus suggest that the observational influence that appeared to be physically important in the hybrid was missed by the EnSRF. When the localization scale was increased to 15 000 km—the same used in the flow-dependent ensemble part in the hybrid—the EnSRF increment appeared more similar to that of the hybrid. Further examination, however, reveals (Fig. 3d) the difference of increments between the hybrid and the EnSRF with a 15 000-km localization scale. Figure 3d shows that the difference was in relatively small spatial scale and also that the magnitude was about one-tenth of the increment. The EnSRF appeared to have noisier increments at longer distances from the observation. Figures 3e–h repeat the same analysis but with the observation located at 41.51°N, 80.27°W, and the results are similar to Figs. 3a–d. Because the inputs for the flow-dependent part of the ensemble and the localization scales applied were the same for the hybrid and the EnSRF, the difference shown in Figs. 3d,h were thus due to the treatment of model error when updating the mean.

From Eqs. (1)–(5) and (7), when updating the mean, the hybrid used a static covariance model to represent the model error, whereas the EnSRF used a limited sample drawn from the static covariance and then applied a covariance localization to that sample covariance. To reveal the differences in the two treatments of model

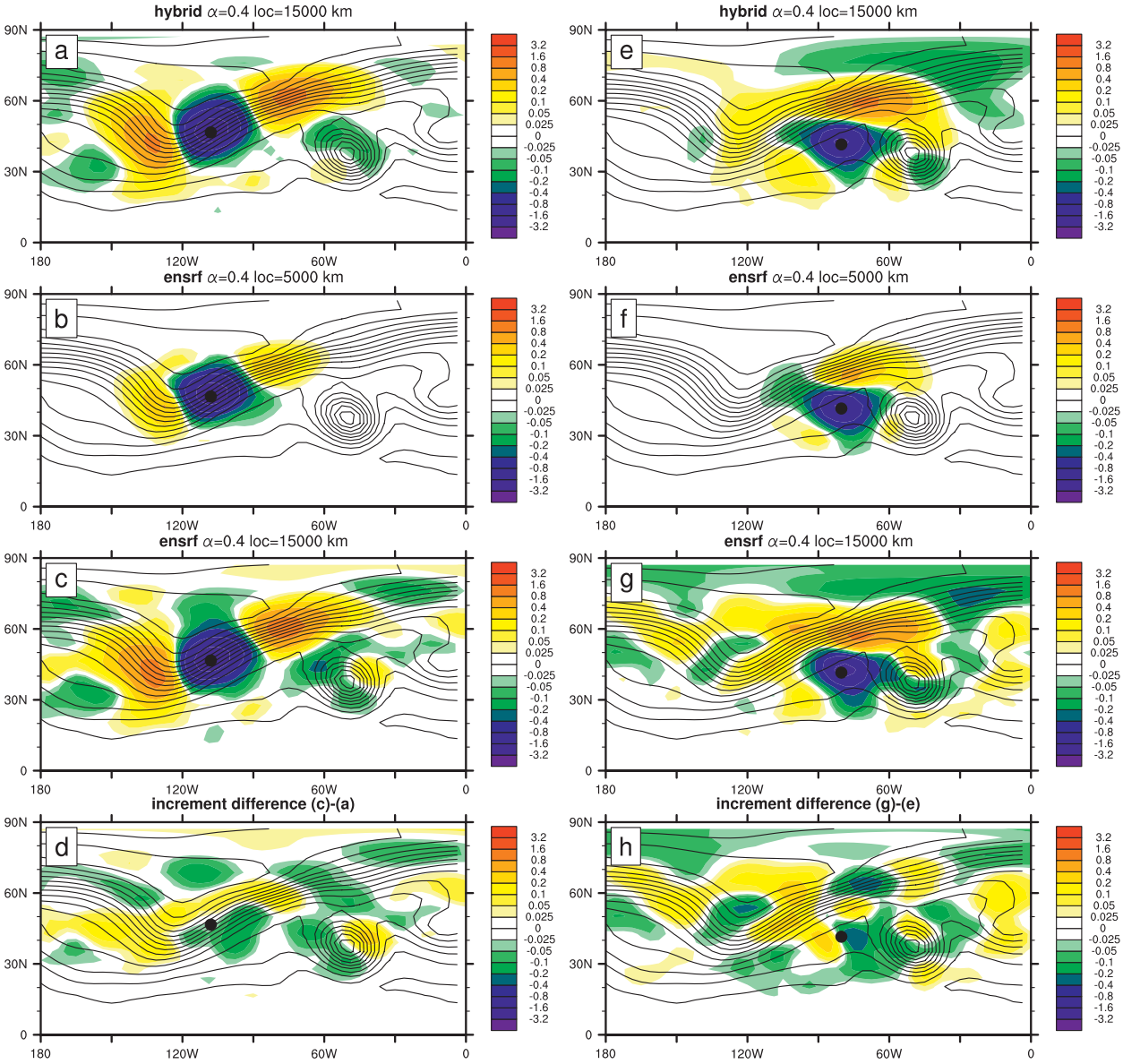

FIG. 3. A snapshot (at the 123rd analysis cycle) of ensemble-mean upper-layer thickness $\Delta\pi_2$ increment for a single -3 J kg^{-1} K^{-1} $\Delta\pi_2$ observation increment located at the black dot at (a)–(d) 46.64°N, 108°W and (e)–(h) 41.51°N, 80.27°W. The black lines are the contours of the background $\Delta\pi_2$. The color shades in (a) and (e) are the increments (J kg^{-1} K^{-1}) for the hybrid with weighting coefficient of 0.4 and localization scale of 15 000 km; the color shades in (b) and (f) are those of the EnSRF with weighting coefficient of 0.4 and localization scale of 5000 km; the color shades in (c) and (g) are those of the EnSRF with weighting coefficient of 0.4 and localization scale of 15 000 km; and the color shades in (d) and (h) are the difference in the increments between (c) and (a) and (g) and (e), respectively. Note that contour interval of the color shades emphasizes small values.

errors, we plot (Fig. 4) the spatial correlation of the static covariance (built from a large sample of perturbations, as described in section 3b) and the correlation of limited (200) samples drawn from the static covariance. Figure 4 shows that applying a localization scale of 15 000 km hardly corrects the sampling error, which presumably explains the difference between Figs. 3d,h. Figure 4 also shows that, although applying a localization of 5000 km can reduce the sampling error at the long distance, it degrades the correlation at shorter distances.

c. *Effect of serial and simultaneous updates*

Another difference in the update of the mean is that the EnSRF used serial assimilation of observations, whereas the hybrid assimilates observations simultaneously. If the observation error is uncorrelated and no

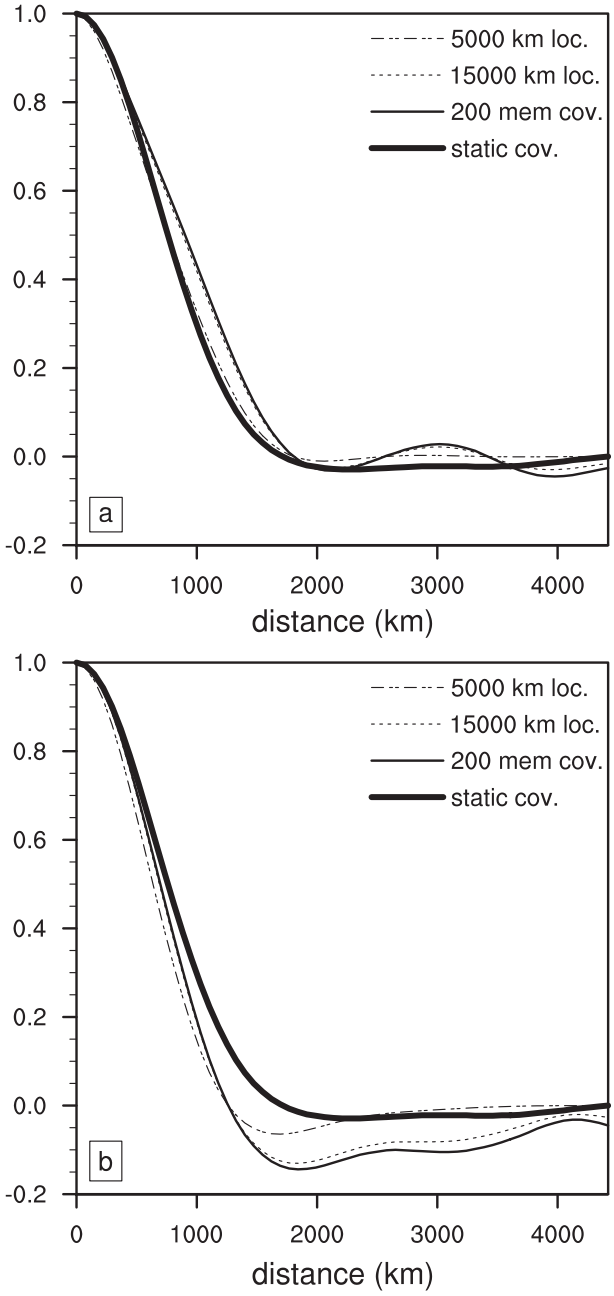

FIG. 4. Spatially lagged correlation along 45°N latitude of the second-layer thickness $\Delta\pi_2$ as a function of zonal distance for the static correlation (thick solid), correlation from two sets of 200-member random samples (thin solid), 200-member sample correlation with 5000-km localization (dashed–dotted), and 15 000-km localization (dotted) for the (a) first and (b) second sets of 200-member random samples.

covariance localization is applied, then the serial update and the simultaneous update are equivalent. However, a serial update with the same Gaspari and Cohn (1999) localization function applied when assimilating each observation is not equivalent to the simultaneous update when the same localization function is applied for the covariance once (Ehrendorfer 2007). To test if simultaneous update contributed to the better performance of the hybrid, we ran a few more experiments. In one experiment, at each assimilation time, we randomly picked 200 samples from the random perturbation inventory and conducted the assimilation by using the serial EnSRF with localization. In the second experiment, at each assimilation time, we first computed the covariance by using the same 200 samples and then applied the same localization on this sample covariance. We then simultaneously assimilated all observations. Localization scales of 5000 and 15 000 km were tried. The rms analysis errors of the experiments are shown in Table 3. We found that, for both localization scales, the simultaneous update performed no better than the serial update. Note that, for the only case where the simultaneous update appears to be a little better ($\Delta\pi_2$ norm and 15 000-km localization), the difference is not statistically significant. Table 3 also shows that with less-severe localization, the difference between the simultaneous and serial updates becomes smaller. We also tried several localization scales between 5000 and 15 000 km, and the conclusion was the same.

d. Initial condition balance

Spurious imbalances between the mass and momentum fields in the analysis increments can produce gravity wave noise and thus reduce the accuracy of the forecast and analysis. The mean absolute tendency of surface pressure is a useful diagnostic of the amount of imbalance for an analysis produced by a data assimilation scheme. For the two-layer model, the surface Exner function π_s is the quantity analogous to the surface pressure. To examine π_s tendency, we reran forecasts from the ensemble-mean analysis up to 24-h lead, producing output every hour. We then calculated the hourly π_s tendency. Figure 5 shows the globally averaged absolute hourly π_s tendency for all analysis times and all hourly tendency snapshots during the 24-h forecast period for the hybrid and EnSRF with localization scale of 15 000 km and weighting coefficient of 0.6. These parameters produced the smallest analysis errors for π_s. The EnSRF has larger π_s tendency value than the hybrid, which suggests the EnSRF ensemble-mean analyses were less balanced. The result for the truth run is also shown in Fig. 5 as a comparison.

As discussed in Lorenc (2003), covariance localization can damage the wind–mass balance. Relative to the hybrid, the EnSRF has two extra covariance localizations: one resides in ensemble-perturbation update, and the other resides in the localization of the covariance of the random noise that is used to parameterize the model

TABLE 3. The rms analysis errors for the experiments where 200 random samples were used to build the background-error covariance. In the serial experiment, the observations are assimilated serially with a fixed localization applied for each observation. In the simultaneous experiment, observations are assimilated simultaneously with the localization applied for the covariance before assimilation. Both the 5000- and 15 000-km localizations were tried.

5000-km localization	Serial	Simultaneous
KE	5.04	5.25
$\Delta\pi_2$	7.29	7.46
π_s	0.50	0.59
15 000-km localization	**Serial**	**Simultaneous**
KE	5.28	5.31
$\Delta\pi_2$	7.70	7.69 (<90% confidence)
π_s	0.48	0.50

error. These extra covariance localizations thus can make the EnSRF analyses less balanced than the hybrid analyses. Experiments in section 5a suggested that the superior balance of the ETKF perturbations may be compensated by its lack of local resolution of the error covariances. However, the extra covariance localization applied to the model-error covariance in the EnSRF, but not in the hybrid, can lead to larger analysis errors in the EnSRF.

6. Sensitivity to the type of samples used to form the static covariance and random sample inventory

In the previous experiments, the static covariance and the random noise that was used to parameterize model error were both constructed based on a large inventory of historical forecast error, where we assumed we knew the true atmospheric state, the forecast-minus-truth inventory. In other words, we assumed that the climatological distribution of the true forecast error is known. Of course the truth is unknown. The static, hybrid, and EnSRF may profit to a different extent from such assumptions.

To test this hypothesis, we now consider assimilation results from using the forecast-minus-analysis inventory discussed in section 3b. The model error was then parameterized by drawing random noise from this new inventory and the static covariance was also recalculated from this new inventory. We then reran the static, the hybrid and the EnSRF experiments. The best-performing results for the hybrid and the EnSRF with 200-member ensembles are summarized in Table 4. As expected, the static, hybrid, and EnSRF all performed worse (relative to results in Table 1). The relative improvements of the hybrid and EnSRF over the static were both larger, indicating that the static is more prone to the quality of the random noise. This result along with

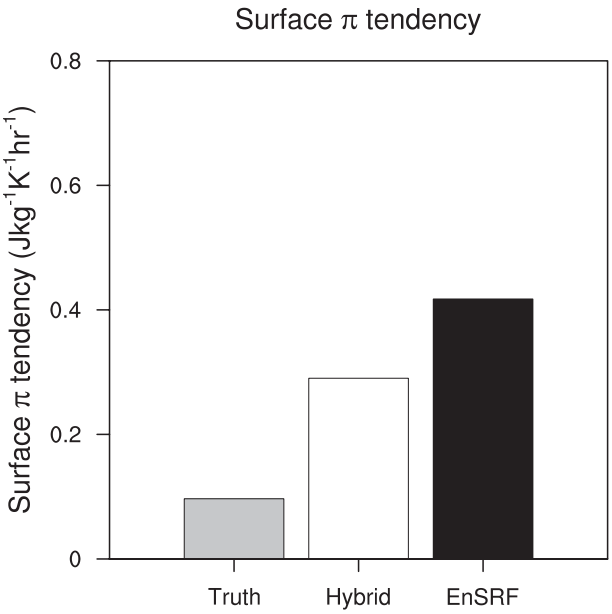

FIG. 5. Mean absolute surface Exner function π_s tendency (J kg^{-1} K^{-1} h^{-1}) averaged globally over the subsequent 23 1-h forecast periods and over all time for the hybrid and EnSRF with weighting coefficient of 0.6 and localization scale of 15 000 km. The gray bar is for the truth run.

the result in section 3a demonstrate the reliability of the hybrid because it outperformed static, no matter how the static covariance was formed. The hybrid still performed statistically significantly better than the EnSRF. However, the absolute and relative improvements of the hybrid over the EnSRF were smaller.

Results from sections 4 and 5 suggest that the inferior performance of the EnSRF relative to the hybrid was due to the sampling error in the model-error parameterization of the EnSRF when updating the mean. However, when the random noise was drawn from the more realistic forecast-minus-analysis inventory, sampling errors in the model-error parameterization became less of a factor.

7. Conclusions and discussion

As a follow-up to the perfect-model study of Wang et al. (2007a), we compared the skill of the hybrid and the EnSRF analysis schemes by using an observation-system simulation experiment in the presence of truncation model error. A two-layer global primitive equation model was used. The true state was a T127 nature run. The data assimilation was performed at T31 resolution. A simplified observation network was assumed, and imperfect observations were created by adding random noise to the nature run. In the hybrid scheme, the ensemble perturbations are updated by the ETKF and the ensemble mean is updated with a hybridized ensemble and

TABLE 4. As in Table 1, but using the newly constructed random noise inventory and static covariance.

200 member	Hybrid	EnSRF	EnSRF – hybrid	Static	Confidence level
KE	4.140	4.378	0.238 (5%)	5.129	>99%
$\Delta\pi_2$	6.257	6.459	0.202 (3%)	7.318	>99%
π_s	0.379	0.412	0.033 (7%)	0.457	>99%

static background-error covariance. In the background ensembles of the hybrid and the EnSRF, the model error was parameterized by using the additive noise method. To test the sensitivity of the performances of the data assimilation schemes to the sources of additive noise, the additive noise in the first set of experiments was drawn from a large inventory of historical forecast errors and the additive noise in the second set of experiments was drawn from a more realistic inventory of differences between forecasts and analyses. The static covariance was formed from these inventories accordingly.

The results demonstrated that the hybrid analysis was statistically significantly more accurate than the EnSRF analysis. The EnSRF was more sensitive to the ensemble size than the hybrid. Series of tests revealed that the less accurate analyses from the EnSRF were probably due to the sampling error in model-error parameterization during the mean update as well as the less-balanced initial conditions resulting from the extra covariance localization used in the EnSRF. However, the relative improvement of the hybrid over the EnSRF was smaller when the parameterized model error and the static covariance were generated from a more realistic inventory of differences between forecasts and analyses rather than from an inventory of historical forecast errors. Because, by design, the magnitude of the model error in this experiment is large (Hamill and Whitaker 2005), this result suggests that the advantage of the hybrid over the EnSRF may become smaller if we lack of an accurate specification of model error.

The simulated observational network is much simpler and more uniform than the real observing network. The number of observations relative to the number of degrees of the model is also very likely to be different from the real world. In general, the use of the flow-dependent ensemble covariance in data assimilation benefits data-sparse regions more than data-rich regions, and it benefits directly observed state variables more than indirectly observed variables. When the observational network is less uniform, long-distance ensemble correlation will be used to update data-void regions. Because the hybrid is less prone to sampling errors, as shown in this paper, the hybrid may show larger advantages over the EnSRF. On the other hand, because we used ETKF to generate ensembles for the hybrid and because of its global update, the ETKF may not resolve the nonuniform observational network as well as the uniform network, especially when the ensemble size is small.

Model errors in full numerical weather prediction models can be caused by many other factors. They are likely to be a combination of errors in physical parameterizations, misspecification of parameters, model truncations, and so on. Realistic model errors are also likely to have a systematic bias component and a stochastic component. A robust ensemble-based data assimilation system should account for both components. In this study, we only considered model errors due to model truncation, and the model error was accounted for by adopting the commonly used additive noise method. Additive noise can be drawn from the 3DVAR covariance or from the forecast-minus-analysis inventory described in section 6. Although these methods are easy to implement in operational data assimilation, they only provide a crude way of parameterizing model errors. For example, model errors can be flow-dependent, and the deviation of the forecast from the analysis approximates the model error only when the analysis is very close to the truth. Future work should conduct the comparison experiments by introducing different types of model errors through, for example, varying the forcing term of the model. Other methods, such as stochastic perturbations to physical tendencies and stochastic kinetic energy backscatter, that account for flow-dependent model errors (e.g., Buizza et al. 1999; Teixeira and Reynolds 2008; Palmer et al. 2005; Houtekamer et al. 2009) may also be tried. The advantage of the flow-dependent representation of the forecast error by the EnSRF may outweigh its deficiency resulting from sampling errors if we implement methods to accurately represent flow-dependent model errors.

The encouraging results of the hybrid, as compared to the experiment with static background-error covariance and the EnSRF in this study, and the fact that the hybrid is straightforward to implement in an operational variational system strongly suggest that the hybrid should be considered as a candidate for operational data assimilation. The relative merit of the hybrid is also a function of the quality of the 3DVAR scheme. Advanced 3DVAR schemes feature error correlation length scales tuned by carefully designed ensemble experiments and sophisticated balance constraints. Hence, designers of ensemble data assimilation schemes who have easy access to advanced 3DVAR schemes may find the hybrid more appealing.

As discussed in Buehner (2005), Wang et al. (2007a,b, 2008a,b), Liu et al. (2008a), and Zhang et al. (2009), the idea of combining ensemble covariance with static covariance can be extended to the 4DVAR framework.

The incorporation of the ensemble covariance may improve the initial background-error covariance and thus improve the 4DVAR analysis.

Acknowledgments. The first author was supported by NSF Grant ATM-0205612, a NOAA THORPEX Grant, and the innovative research grant from CIRES of University of Colorado. Craig Bishop acknowledges support from ONR Projects 0602435N, BE435-003, and N0001407WX30012.

REFERENCES

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

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.,* **131,** 1013–1043.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.,* **125,** 2887–2908.

Cohn, S. E., D. M. da Silva, J. Guo, M. Sienkiewiez, and D. Lamich, 1998: Assessing the effects of data selection with the DAO physical space statistical analysis system. *Mon. Wea. Rev.,* **126,** 2913–2926.

Courtier, P., and Coauthors, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quart. J. Roy. Meteor. Soc.,* **124,** 1783–1807.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.

Dee, D. P., and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.,* **124,** 269–295.

Dirren, S., R. D. Torn, and G. J. Hakim, 2007: A data assimilation case study using a limited-area ensemble Kalman filter. *Mon. Wea. Rev.,* **135,** 1455–1473.

Dowell, D. C., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.,* **132,** 1982–2005.

Ehrendorfer, M., 2007: A review of issues in ensemble-based Kalman filtering. *Meteor. Z.,* **16,** 795–818.

Etherton, B. J., and C. H. Bishop, 2004: Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error. *Mon. Wea. Rev.,* **132,** 1065–1080.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

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

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

Gauthier, P. C., L. Cherette, L. Fillion, P. Koclas, and S. Laroche, 1998: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: The global analysis. *Atmos.–Ocean,* **37,** 103–156.

Hamill, T. M., 2006: Ensemble based atmospheric data assimilation. *Predictability of Weather and Climate,* T. Palmer and R. Hagedorn, Eds., Cambridge Press, 124–156.

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

——, and J. S. Whitaker, 2005: Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Mon. Wea. Rev.,* **133,** 3132–3147.

——, ——, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.,* **129,** 2776–2790.

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

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

——, ——, G. Pellerin, M. Buehner, and M. Charron, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.,* **133,** 604–620.

——, ——, and X. Deng, 2009: Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev.,* **137,** 2126–2143.

Keppenne, C. L., and M. M. Rienecker, 2002: Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.,* **130,** 2951–2965.

Liu, C., Q. Xiao, and B. Wang, 2008a: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.,* **136,** 3363–3373.

Liu, H., J. Anderson, Y.-H. Kuo, C. Snyder, and A. Caya, 2008b: Evaluation of a nonlocal quasi-phase observation operator in assimilation of CHAMP radio occultation refractivity with WRF. *Mon. Wea. Rev.,* **136,** 242–256.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP – A comparison with 4D-VAR. *Quart. J. Roy. Meteor. Soc.,* **129,** 3183–3203.

Meng, Z., and F. Zhang, 2008: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study. *Mon. Wea. Rev.,* **136,** 522–540.

Miyoshi, T., S. Yamane, and T. Enomoto, 2007: Localizing the error covariance by physical distances within a local ensemble transform Kalman filter (LETKF). *SOLA,* **3,** 89–92.

Palmer, T. N., G. J. Shutts, R. Hagedorn, F. J. Doblas-Reyes, T. Jung, and M. Leutbecher, 2005: Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.,* **33,** 163–193.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.,* **120,** 1747–1763.

Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser, 2002: Extended versus ensemble Kalman filtering for land data assimilation. *J. Hydrometeor.,* **3,** 728–740.

Schlatter, T. W., 1975: Some experiments with a multivariate statistical objective analysis scheme. *Mon. Wea. Rev.,* **103,** 246–257.

Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.,* **131,** 1663–1677.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, and J. A. York, 2005: Assessing a local ensemble Kalman filter: Perfect model experiments with the NCEP global model. *Tellus,* **57A,** 528–545.

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

Teixeira, J., and C. A. Reynolds, 2008: Stochastic nature of physical parameterizations in ensemble prediction: A stochastic convection approach. *Mon. Wea. Rev.,* **136,** 483–496.

Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.,* **133,** 1789–1807.

Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited area ensemble Kalman filters. *Mon. Wea. Rev.,* **134,** 2490–2502.

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 S. J. Julier, 2004: Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.,* **132,** 1590–1605.

——, T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2007a: A comparison of hybrid ensemble transform Kalman filter–optimum interpolation and ensemble square root filter analysis schemes. *Mon. Wea. Rev.,* **135,** 1055–1076.

——, C. Snyder, and T. M. Hamill, 2007b: On the theoretical equivalence of differently proposed ensemble–3DVAR hybrid analysis schemes. *Mon. Wea. Rev.,* **135,** 222–227.

——, D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part I: Observation system simulation experiment. *Mon. Wea. Rev.,* **136,** 5116–5131.

——, ——, ——, and ——, 2008b: A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiments. *Mon. Wea. Rev.,* **136,** 5132–5147.

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

——, G. P. Compo, and T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.,* **132,** 1190–1200.

——, T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.,* **136,** 463–482.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 467 pp.

Yang, S.-C., E. Kalnay, B. Hunt, and N. E. Bowler, 2008: Weight interpolation of efficient data assimilation with the Local Ensemble Transform Kalman Filter. *Quart. J. Roy. Meteor. Soc.,* **135,** 251–262.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.,* **132,** 1238–1253.

——, M. Zhang, and J. A. Hansen, 2009: Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Adv. Atmos. Sci.,* **26,** 1–8.

Zou, X., A. Barcilon, I. M. Navon, J. Whitaker, and D. G. Cacuci, 1993: An adjoint sensitivity study of blocking in a two-layer isentropic model. *Mon. Wea. Rev.,* **121,** 2833–2857.

Zupanski, D., 1997: A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.,* **125,** 2274–2292.

Zupanski, M., 2005: Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Wea. Rev.,* **133,** 1710–1726.