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

Xuguang Wang

CIRES Climate Diagnostics Center, University of Colorado, and Physical Sciences Division, NOAA Earth System Research Laboratory, Boulder, Colorado

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

Submitted as an article to Monthly Weather Review
Dec. 18 2008

Corresponding author address:

Dr. Xuguang Wang
NOAA Earth System Research Lab, Physical Sciences Division
325 Broadway, R/PSD1
Boulder, CO, 80305-3328
xuguang.wang@noaa.gov
Phone: (303) 497-4434

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 hybridized 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, 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 and in the second set of experiments, it 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 were 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 due to the extra localizations used in the EnSRF.

1. Introduction

The idea of incorporating ensemble covariances into three-dimensional variational (3DVAR) data assimilation systems, hereafter called hybrid data assimilation schemes, has been a subject of a number of recent papers (e.g., Hamill and Snyder 2000, Lorenc 2003, Etherton and Bishop 2004, Buehner 2005, Wang et al. 2007ab, Wang et al. 2008ab). Compared to 3DVAR schemes that utilize stationary background error covariances (e.g., Parrish and Derber 1992; Courtier et al. 1998; Gauthier et al. 1998; Cohn et al. 1998), the presumed benefits of the hybrid schemes are their ability to provide flow-dependent estimates of the background-error covariances so that the observations and the background are more appropriately weighted during the assimilation. Previous studies ranging from simple model tests (e.g., Hamill and Snyder 2000, Etherton and Bishop 2004, Wang et al 2007a) to real Numerical Weather Prediction (NWP) model tests (e.g., Buehner 2005, Wang et al. 2008ab) have demonstrated superior data assimilation performance of the hybrid method relative to 3DVAR.

Another framework of using ensembles in data assimilation is the ensemble Kalman filter (EnKF) based approach. Numerous studies have been conducted on this approach and encouraging results have also been found (Evensen 1994; Houtekamer and Mitchell 1998, 2001; Anderson 2001; Szunyogh et al. 2005; Torn et al. 2006; Houtekamer et al. 2005, 2008; Whitaker et al. 2008; Szunyogh et al. 2007; Miyoshi et al. 2007; Snyder and Zhang 2003; Dowell et al. 2004; Tong and Xue 2005; Meng and Zhang 2008; Liu et al. 2007; Dirren et al. 2007; Reichle et al. 2002; Keppenne and Rienecker 2002; Yang et al. 2008). For reviews please see Evensen 2003, Lorenc 2003, Hamill 2006, and Ehrendorfer 2007. A natural question one would ask is: how does the hybrid

compare to the EnKF? Previous studies on the hybrid (e.g., Lorenc 2003, Buehner 2005, Wang et al. 2007ab, Wang et al. 2008ab) discussed that unlike the EnKF that adopts a framework completely different from the long- existing operational variational scheme, one advantage of the hybrid scheme is that it can be implemented with minor changes to the existing variational codes. However, 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 with the ensemble generated by the ensemble transform Kalman filter (ETKF; Wang and Bishop 2003; Wang et al. 2004; Wang et al. 2007a) and the ensemble square-root filter (EnSRF; Whitaker and Hamill 2002), one of the well-tested EnKF-based approaches, using a primitive equation, two-layer model with a perfect-model assumption. They found that the hybrid 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 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.

The hybrid and the EnSRF schemes will be described in section 2. Section 3 provides a description on the experiment design. Section 4 describes the results and tests conducted to understand the difference of the two schemes. Section 5 concludes the paper.

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 $\overline{\mathbf{x}}^a$ using

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

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

$$\mathbf{HP}^b\mathbf{H}^T = (1-\alpha)\left(\rho_s^{p\times p} \circ \mathbf{HP}^e\mathbf{H}^T\right) + \alpha\left(f\mathbf{HBH}\right)^T, \quad (2)$$

$$\mathbf{P}^b\mathbf{H}^T = (1-\alpha)\left(\rho_s^{n\times p} \circ \mathbf{P}^e\mathbf{H}^T\right) + \alpha\left(f\mathbf{BH}\right)^T, \quad (3)$$

where $\mathbf{P}^e\mathbf{H}^T$ and $\mathbf{HP}^e\mathbf{H}^T$ are calculated from the K ETKF ensemble forecast perturbations ($\mathbf{x}_k^{e'}$, $k = 1,\cdots,K$). Note in the second terms of eq. (2) and (3), covariance localization was applied through Schur product $\circ$ between a correlation matrix and the raw ensemble covariance (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 $\mathbf{HBH}^T$ and $\mathbf{BH}^T$ 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 $trace\left(\mathbf{R}^{-1/2}\mathbf{HP}^t\mathbf{H}^T\mathbf{R}^{-1/2}\right)$). As in eq. (21) of Wang et al. (2007a), the rescaling factor f was determined dynamically. The user-tunable factor α, $0 \le \alpha \le 1$, determines the relative weights placed on the static and the ensemble covariances. As discussed later, an inflation factor was applied so that the ETKF ensemble forecast variance was equal to the total forecast error variance in the observation space also. Designed this way, the weighting factor α preserves the total variance (e.g., Etherton and Bishop 2004). Note that although we updated the mean using the classic optimum interpolation formula (Schlatter 1975), under the current experiment design, it will

provide the same solution as if we had updated the mean state in a 3D variational 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. The ETKF transforms the matrix of background ensemble perturbations $\mathbf{X}^b$, whose columns contain K background ensemble perturbations, $\mathbf{x}_k^{'b}$, $k=1,\cdots,K$, into a matrix of analysis perturbations $\mathbf{X}^a$, whose columns contain K analysis perturbations, $\mathbf{x}_k^{'a}$, $k=1,\cdots,K$. The transformation happens through the post-multiplication by the matrix $\mathbf{T}$, that is,

$$\mathbf{X}^a = \mathbf{X}^b\mathbf{T}. \tag{4}$$

The latest ETKF formulation (Wang et al. 2007a) is

$$\mathbf{X}^a = \Pi\mathbf{X}^b\mathbf{C}\left(\rho\mathbf{\Gamma}+\mathbf{I}\right)^{-1/2}\mathbf{C}^T, \tag{5}$$

where $\mathbf{C}$ contains the eigenvectors and $\mathbf{\Gamma}$ the eigenvalues of the $K\times K$ matrix $\left(\mathbf{X}^b\right)^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{X}^b/K$, and $\mathbf{I}$ is the identity matrix. For more details on the derivation of eq. (5), please refer to Wang et al. (2007a) and references therein. The scalar factor Π is the inflation factor and the scalar factor ρ is the fraction of the forecast-error variance projected onto the ensemble subspace. Both factors aim to ameliorate the systematic underestimate on the analysis-error variance by the ETKF due to the limited ensemble size. Wang et al. (2007a) provided details on how to estimate these two factors adaptively for each data assimilation cycle using the innovation statistics. The goal of Π is to ensure that the variance of the ensemble forecast initialized from the analysis perturbations $\mathbf{X}^a$ in eq. (5) is consistent with the true background-error variance in the observation space (under the norm of $trace\left(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{P}^t\mathbf{H}^T\mathbf{R}^{-1/2}\right)$).

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 $\mathbf{X}^b$ using the additive noise method. The background ensemble perturbation $\mathbf{X}^b$ ($\mathbf{x}_k^{'b}$, $k = 1, \cdots, K$.) is constructed as

$$\mathbf{x}_k^{'b} = \sqrt{(1-\alpha)}\mathbf{x}_k^{'e} + \sqrt{\alpha f}\mathbf{e}_k, \tag{6}$$

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 $\mathbf{P}^e\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}^e\mathbf{H}^T$ in eq. (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). Also note that in eq. (6), we adopted random noise consistent with the static covariance $\mathbf{B}$, which is similar to Houtekamer et al. (2005, 2008) 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. 6, Fig. 1b shows the rms background error and the ensemble spread as a function of latitude for an experiment with the hybrid scheme ($\alpha = 0.4$ and localization scale = 15000 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 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. While the constant inflation uniformly expanded the spread, the additive noise has larger magnitude in the mid-latitude (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:

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{K}_i \left(y_i - \mathbf{H}_i \overline{\mathbf{x}}^b \right), \tag{7}$$

$$\mathbf{x}_k'^a = \left(\mathbf{I} - \tilde{\mathbf{K}}_i \mathbf{H}_i \right) \mathbf{x}_k'^b \Pi . \tag{8}$$

Note that here $\mathbf{H}_i$ maps the state vector to the ith observation space. In eq. (7), $\mathbf{K}_i$ is the Kalman gain modified by the covariance localization,

$$\mathbf{K}_i = \left(\rho_s^{n\times 1} \circ \mathbf{P}^b \mathbf{H}_i^T \right) \left(\mathbf{H}_i \mathbf{P}^b \mathbf{H}_i^T + \mathbf{R}_{ii} \right)^{-1}. \quad (9)$$

As in the hybrid, horizontal localization utilizes the approximately Gaussian-shaped function of Gaspari and Cohn (1999). $\tilde{\mathbf{K}}_i$ in eq. (8) is called the "reduced" Kalman gain matrix (Whitaker and Hamill 2002). For serial assimilation,

$$\tilde{\mathbf{K}}_i = \left(1 + \sqrt{\frac{\mathbf{R}_{ii}}{\mathbf{H}_i \mathbf{P}^b \mathbf{H}_i^T + \mathbf{R}_{ii}}} \right)^{-1} \mathbf{K}_i . \quad (10)$$

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. (8) 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 $\mathbf{x}'^b_k$ in the EnSRF (eq. 8). 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. 6), and in this case $\mathbf{x}'^e_k$ is the kth the ensemble forecast perturbation generated from ensemble forecasts initialized by the analysis ensemble 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-model 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 the "true" atmospheric state is described by the forecast dynamics at T127 resolution. All data assimilation experiments were conducted at T31 resolution. In other words, we assume 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. Please see Hamill and Whitaker (2005) for detailed descriptions on the characteristics of the model and model errors due to unresolved scales. Also as discussed in Hamill and Whitaker (2005), this setup was designed to produce large model errors so as 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 component.

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 standard deviation of 8.75 $Jkg^{-1}K^{-1}$ for interface π and 0.875 $Jkg^{-1}K^{-1}$ for surface π, respectively, the same values used in Wang et al. 2007a.

Observation errors were constructed to be independent spatially and temporally, and observations were assimilated every 24h.

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 a150-day period, and the error statistics were evaluated over the last 100 days. The statistical significance of the following results test 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 experiment, the static background-error covariance $\mathbf{B}$ was formed and the random noise $\mathbf{e}_k$ was drawn from a large inventory of historical forecast errors over many separate times. We call this 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. The static background error covariance matrix $\mathbf{B}$ was then constructed by directly calculating the covariance of this large inventory of the forecast error samples. The 4th column of Table 1 shows the rms analysis error of the experiment where we ran a single-member forecast and analysis cycle using only the static covariance obtained from the last iteration. We denote this experiment "3DVAR" since it used a static covariance like 3DVAR. 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 section 4a and 4b.

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 since in reality the true state can never be known. The 3DVAR, hybrid and EnSRF data assimilation schemes may benefit differently from these assumptions. To answer this question, 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. Results of the second set of experiments using this new inventory of random noise and new static covariance are described in section 4c.

4. Results

a. 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 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 kinetic energy (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 EnSRF was more sensitive to the localization scales than the hybrid. 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 best-performing hybrid and EnSRF with 50-member ensembles are summarized in the lower panel 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 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 the hybrid is less sensitive to the ensemble size than the EnSRF, consistent with Wang et al. (2007a).

The hybrid running 200-member and 50-member ensembles both outperformed the "3DVAR" for all three norms considered. The 200-member EnSRF outperformed 3DVAR except for the π_s norm, which is different from Hamill and Whitaker (2005) where the EnSRF was better than the 3DVAR for all three norms. The differences of the current experiment design and that of the Hamill and Whitaker (2005) are that (1) Hamill and Whitaker (2005) did not assimilate π_s observations, and probably more importantly

(2) Hamill and Whitaker (2005)'s static covariance for 3DVAR was formed from 200 historical forecast errors with covariance localization whereas the static covariance here was formed iteratively from 6941 historical forecast errors with no localization. The 50-member EnSRF did not outperform the "3DVAR".

b. 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 the EnSRF. The superior performance of the hybrid over the EnSRF shown above must then arise from algorithmic differences between the two schemes. Although the assumption made in forming the random noise inventory and also the static covariance is not realistic since 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 4a.

1) *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 the one hand, the ETKF ensemble perturbations have superior balance since the update of the perturbations does not involve covariance localization. On the other hand, due to 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 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 while still using the EnSRF to obtain the ensemble mean analyses. The results are summarized in Table 2. 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 Table 2, in the only case where 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.

2) *Effect of sampling error in the estimation of model-error covariance when updating the mean*

We then used a single-observation test to understand the difference of the two methods in updating the ensemble mean. Figure 3 shows the increment by assimilating a single second layer thickness ($\Delta\pi_2$) observation located at (47N, 108W) that was $3\,J\,kg^{-1}\,K^{-1}$ smaller than the background forecast. Fig. 3a shows the increment of the hybrid with a weighting coefficient of 0.4 and a localization scale of 15000 km. Note such 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-hour ensemble forecast at the 123th 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 in order to understand the difference in the update of the ensemble mean. Since 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 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. 4). Comparing Fig. 3a and Fig. 3b shows that the length scale of the EnSRF increment was shorter than that of the hybrid. Since the hybrid analysis was more accurate, Fig. 3a and 3b thus suggest observational influence that appeared to be physically important in the hybrid was missing by the EnSRF. When the localization scale was increased to 15000 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. 4) the difference of increments between the hybrid and the EnSRF with 15000 km localization scale. Figure 4 shows that the difference was in relatively small spatial scale and the magnitude was about one tenth of the increment. The EnSRF appeared to have noisier increment at longer distance from the observation. Since 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 Fig. 4 were thus due to the treatment of model error when updating the mean.

From eqs. (1)-(3) and (6), (7), and (9), 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 errors, we plot (Fig. 5) the spatial correlation of the static covariance (that was built from a large sample of perturbations as described in section 3b) and the correlation of limited (200) samples drawn from the static covariance. Fig. 5 shows that applying a localization scale

of 15000 km, the sampling error is hardly corrected, which presumably explains the difference in Fig. 4. Figure 5 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.

3) *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 covariance localization applied, the serial update and the simultaneous update are equivalent. However, a serial update with the same Gaspari-Cohen (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 using the serial EnSRF with localization. In the second experiment, at each assimilation time we first computed the covariance using the same 200 samples and then apply the same localization on this sample covariance. We then simultaneously assimilated all observations. Localization scales of 5000 km and 15000 km were tried. The rms analysis errors of the experiments are shown in Table 3. We found that for both localizations 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

15000 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 km and 15000 km and the conclusion was the same.

4) *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 diagnostics 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 re-ran forecasts from the ensemble-mean analysis up to 24-h lead, producing output every hour. We then calculated the hourly π_s tendency. Figure 6 shows the globally averaged absolute hourly π_s tendency for all analysis times and all hourly tendency snapshots during the 24-h forecast period. 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. 6 as a comparison.

As discussed in Lorenc (2003), covariance localization can damage the wind-mass balance. Compared 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 error. These extra covariance localizations thus can make the EnSRF analyses less

balanced than the hybrid analyses. Experiments in section 4b1) suggested 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.

c. 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 "3DVAR", the hybrid and the EnSRF may profit to a different extent from such assumptions.

To test this hypothesis, we now consider assimilation results 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 re-ran the "3DVAR", 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, 3DVAR, hybrid and the EnSRF all performed worse (compared with results in Table 1).

The relative improvements of the hybrid and EnSRF over the 3DVAR were both larger. This result together with that in section 3a demonstrated the reliability of the hybrid since it outperformed 3DVAR no matter how the static covariance was formed. The hybrid still performed statistically significantly better than the EnSRF. The absolute and relative improvements of the hybrid over the EnSRF were smaller though.

Results from sections 4a and 4b suggested 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 error in model-error parameterization became less of a factor.

5. Conclusions and Discussions

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 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 static background-error covariance. In the background ensembles of the hybrid and the EnSRF, the model error was parameterized using additive noise method. To test the sensitivity of the

performances of the data assimilation schemes to the sources of additive noise, in the first set of experiments, the additive noise was drawn from a large inventory of historical forecast errors and in the second set of experiments, it 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 due to 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. Since 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. Also model errors in full numerical weather prediction models can be caused by many other factors other than truncation error. In the future when both the EnSRF and the hybrid are maturing with the

real NWP models, a direct comparison of the EnSRF and hybrid with an operational NWP setting should be conducted.

The encouraging results of the hybrid as compared with the 3DVAR 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 Buenher (2005), Wang et al. (2007ab; 2008ab), and Zhang et al. (2008) the idea of combining ensemble covariance with static covariance can be extended to the four-dimensional variational (4DVAR) framework. The incorporation of the ensemble covariance may improve the initial background error covariance and thus improve the 4DVAR analysis.

Acknowledgements.

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.

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., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, J. Hamrud, A. Hollingsworth, F. Rabier, and M. Fisher, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quart. J. Roy. Meteor. Soc.*, **124**, 1783-1807.

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.

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.

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

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res*., **99(C5)**, 10143-10162.

—, 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **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.

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

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

—, 2006: Ensemble based atmospheric data assimilation. Chapter 6 of *Predictability of Weather and Climate*, R. Hagedorn and T. N. Palmer, eds. Cambridge Press, 702 pp.

— 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.

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, 2008: Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev*., in press, http://ams.allenpress.com/archive/1520-0493/preprint/2008/pdf/10.1175_2008MWR2737.1.pdf.

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, H., J. Anderson, Y-H. Kuo, C. Snyder and A. Caya, 2007: Evaluation of a nonlocal quasi-phase observation operator in assimilation of CHAMP radio occultation refractivity with WRF. *Mon. Wea. Rev*., in press. Available from hliu@ucar.edu

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: Test 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). *Sci. Online Lettter Atmos.*, **3,** 89-92

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 vs. ensemble Kalman filtering for land data assimilation. *J. Hydrometeor*., **3**, 728-740.

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, J. A. Yorke, 2006: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. Submitted to *Tellus*. Available at http://www.weatherchaos.umd.edu/papers/Szunyogh_et_all_2007.pdf

Tong, M. and X. Ming, 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.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp., second edition.

Wang, X. 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, C. H. Bishop: 2007a: A comparison of hybrid ensemble transform Kalman filter-OI 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/3D-Var 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*., in press, available from http://www.cdc.noaa.gov/people/xuguang.wang/wrfhybda_part1.pdf

—, D. M Barker, C. Snyder, and T. M. Hamill, 2008b: A hybrid ETKF – 3DVAR data assimilation scheme for the WRF Model. Part II: real observation experiments. *Mon. Wea. Rev*., in press, available from http://www.cdc.noaa.gov/people/xuguang.wang/wrfhybda_part2.pdf

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

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.***,** Accepted.

Table captions

TABLE 1. Rms errors in kinetic energy norm (KE, units $m\ s^{-1}$), upper layer Exner function thickness norm ($\Delta\pi_2$, units $J\ kg^{-1}\ K^{-1}$) and surface Exner function norm (π_s, units $J\ kg^{-1}\ K^{-1}$) for the analyses of the hybrid, EnSRF and 3DVAR. Upper panel is for 200-member ensembles and lower panel is for 50-member ensembles. Only the best performing hybrid and EnSRF are shown. The 4th 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.

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

TABLE 3. RMS analysis errors for the experiments where 200 random samples were used to build the background error covariance. In "serial" experiment, the observations are assimilated serially with a fixed localization applied for each observation. In "simultaneous" experiment, observations are assimilated simultaneously with the localization applied for the covariance before assimilation. Both 5000 km and 15000 km localizations were tried.

TABLE 4. Same as Table 1 except using the newly constructed random noise inventory and static covariance

Figure captions

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) Hybrid with weighting coefficient of 0.4, localization scale of 15000 km and no additive error; (b) as (a) but with additive error to parameterize model error in the background ensemble; (c) EnSRF with weighting coefficient of 0.4, localization scale of 5000 km, and with additive error to parameterize model error in the background ensemble.

Fig. 2 RMS analysis error for the kinetic energy (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 3DVAR (dashed). The weighing coefficients tried are 0.2, 0.4, 0.6 and 0.8. For the hybrid, for each of weighting coefficients, localization scales of 5000 km, 15000 km and 25000 km were tried. For the EnSRF, for each of weighting coefficients, localization scales of 3000 km, 5000 km, 15000 km and 25000 km were tried.

Fig. 3 A snapshot (at the 123rd analysis cycle) of ensemble mean upper-layer thickness ($\Delta\pi_2$) increment for a single $-3Jkg^{-1}K^{-1}$ $\Delta\pi_2$ observation increment located at the black dot. The black lines are the contours of the background $\Delta\pi_2$. The color in (a) is the increments ($Jkg^{-1}K^{-1}$) for the hybrid with weighting coefficient of 0.4 and localization scale of 15000 km; and the color in (b) is that of the EnSRF with weighting coefficient of 0.4 and localization scale of 5000 km.

Fig. 4 Same as Fig. 3, except that the colors scale now emphasizes small values. (a) is the same as (a) in Fig. 3, except using the new color scale. (b) is for the EnSRF with weighting coefficient of 0.4 and localization scale of 15000 km; (c) is the difference in the increments between (b) and (a).

Fig. 5 Spatially lagged correlation along $45^{o}N$ 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 (dash-dotted) and 15000 km localization (dotted). (a) and (b) are for the first and second sets of 200-member random samples.

Fig. 6 Mean absolute surface Exner function (π_s) tendency ($Jkg^{-1}K^{-1}h^{-1}$) averaged globally, over the subsequent twenty-three 1-h forecast period and over all time. White bar is for the hybrid with weighting coefficient of 0.4 and localization scale of 15000 km. Black bars are for the EnSRF with weighting coefficient of 0.4 and localization scales of 5000 km and 15000 km. The grey bar is for the truth run.

TABLE 1. Rms errors in kinetic energy norm (KE, units $m\ s^{-1}$), upper layer Exner function thickness norm ($\Delta\pi_2$, units $J\ kg^{-1}\ K^{-1}$) and surface Exner function norm (π_s, units $J\ kg^{-1}\ K^{-1}$) for the analyses of the hybrid, EnSRF and 3DVAR. Upper panel is for 200-member ensembles and lower panel is for 50-member ensembles. Only the best performing hybrid and EnSRF are shown. The 4th 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.

200-member	Hybrid	EnSRF	EnSRF-Hybrid	3DVAR	Confidence level
KE	3.938	4.237	0.299 (7%)	4.509	> 99%
$\Delta\pi_2$	5.997	6.317	0.32 (5%)	6.631	> 99%
π_s	0.341	0.404	0.063 (16%)	0.379	> 99%

50-member	Hybrid	EnSRF	EnSRF-Hybrid	3DVAR	Confidence level
KE	4.153	4.671	0.518 (11%)	4.509	> 99%
$\Delta\pi_2$	6.192	6.826	0.634 (9%)	6.631	> 99%
π_s	0.355	0.515	0.16 (31%)	0.379	> 99%

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

200 member	EnSRF	Mixed EnSRF-ETKF
KE	4.237	4.312
$\Delta\pi_2$	6.317	6.339
π_s	0.448	0.460

50 member	EnSRF	Mixed EnSRF-ETKF
KE	4.671	4.709
$\Delta\pi_2$	6.826	6.799 (<80% confidence)
π_s	0.515	0.563

TABLE 3. RMS analysis errors for the experiments where 200 random samples were used to build the background error covariance. In "serial" experiment, the observations are assimilated serially with a fixed localization applied for each observation. In "simultaneous" experiment, observations are assimilated simultaneously with the localization applied for the covariance before assimilation. Both 5000 km and 15000 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

15000 km localization	Serial	Simultaneous
KE	5.28	5.31
$\Delta\pi_2$	7.70	7.69(<90% confidence)
π_s	0.48	0.50

TABLE 4. Same as Table 1 except using the newly constructed random noise inventory and static covariance.

200-member	Hybrid	EnSRF	EnSRF-Hybrid	3DVAR	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%

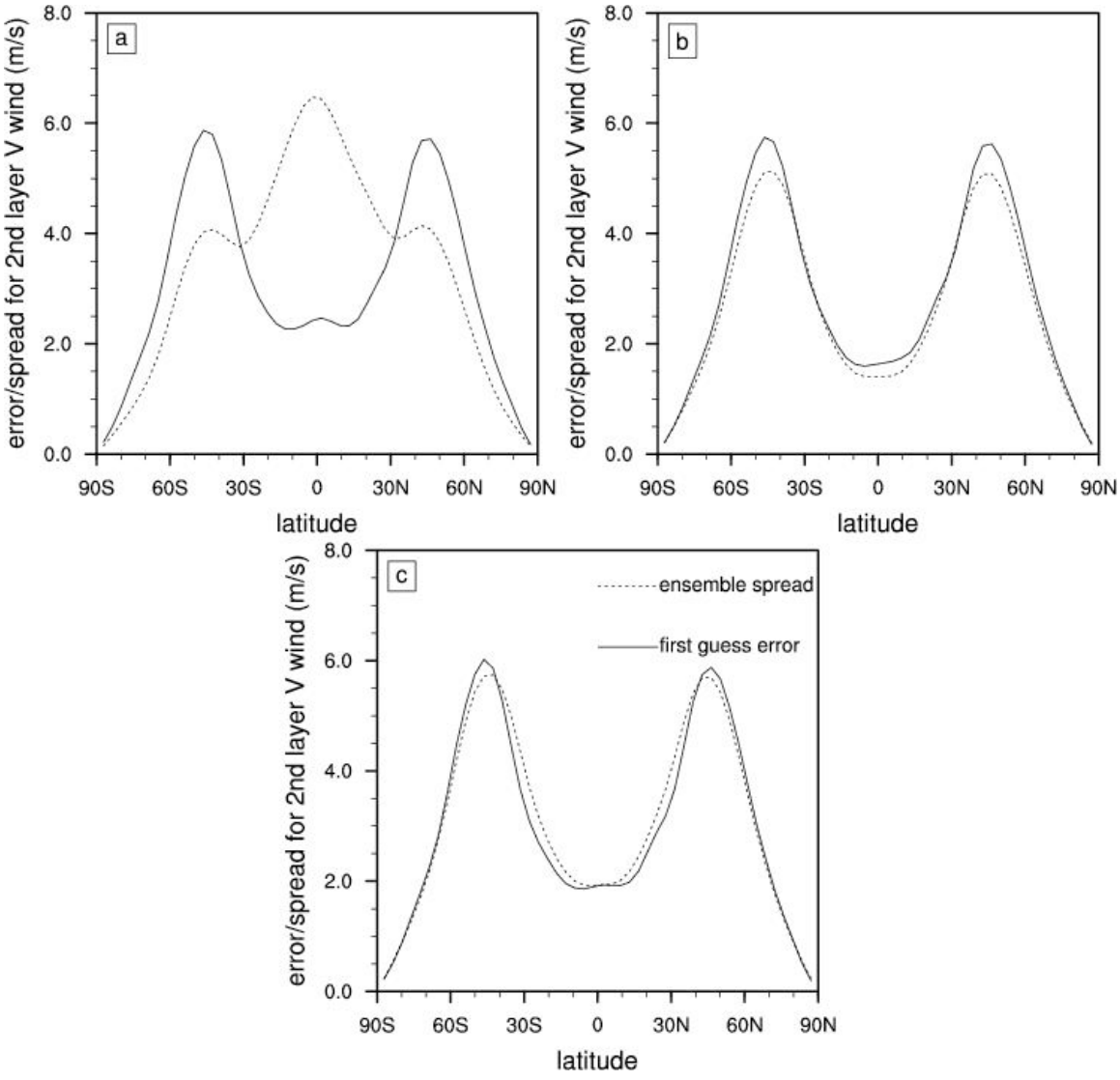

Fig. 1 Zonally averaged RMS first guess second layer meridional wind error (solid) and background ensemble spread (dotted) as a function of latitude, for (a) Hybrid with weighting coefficient of 0.4, localization scale of 15000 km and no additive error; (b) as (a) but with additive error to parameterize model error in the background ensemble; (c) EnSRF with weighting coefficient of 0.4, localization scale of 5000 km, and with additive error to parameterize model error in the background ensemble.

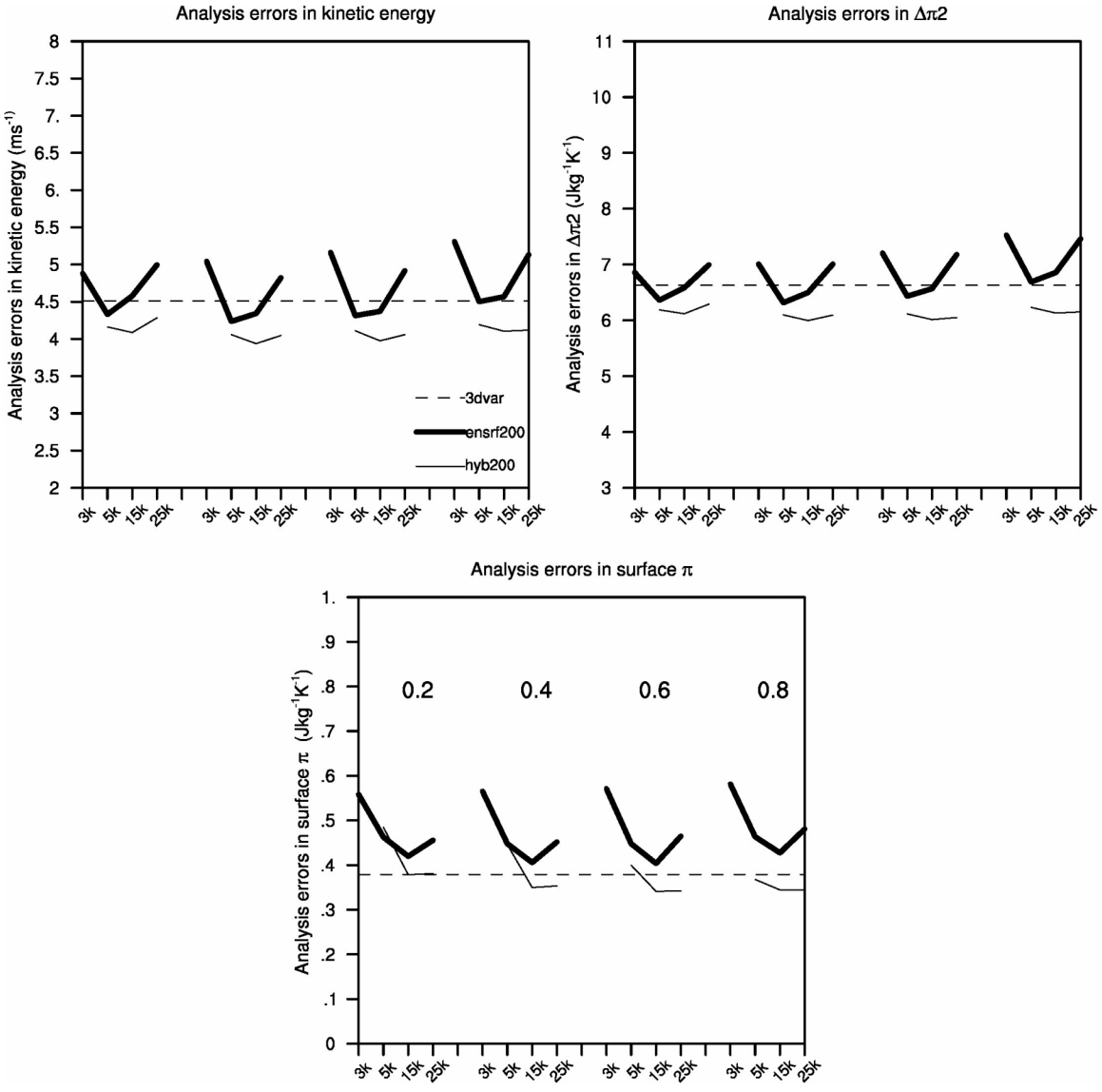

Fig. 2 RMS analysis error for the kinetic energy (KE), second layer thickness ($\Delta\pi_2$) and surface Exner function (π_s) norms as a function of localization scales and weighting coefficients for the Hybrid (solid thin), EnSRF (solid thick) and 3DVAR (dashed). The weighing coefficients tried are 0.2, 0.4, 0.6 and 0.8. For the hybrid, for each of weighting coefficients, localization scales of 5000 km, 15000 km and 25000 km were tried. For the EnSRF, for each of weighting coefficients, localization scales of 3000 km, 5000 km, 15000 km and 25000 km were tried.

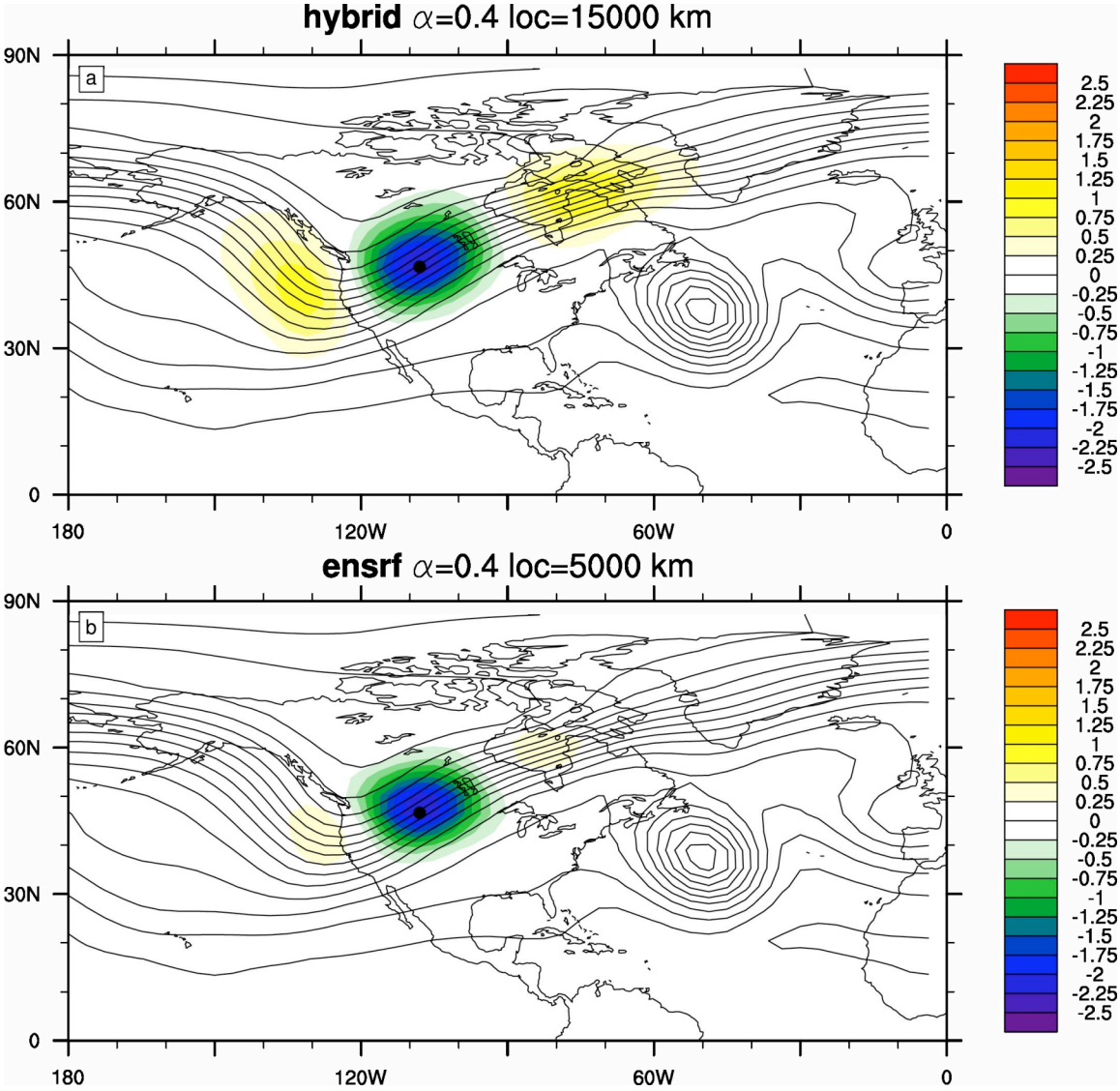

Fig. 3 A snapshot (at the 123rd analysis cycle) of ensemble mean upper-layer thickness ($\Delta\pi_2$) increment for a single $-3Jkg^{-1}K^{-1}$ $\Delta\pi_2$ observation increment located at the black dot. The black lines are the contours of the background $\Delta\pi_2$. The color in (a) is the increments ($Jkg^{-1}K^{-1}$) for the hybrid with weighting coefficient of 0.4 and localization scale of 15000 km; and the color in (b) is that of the EnSRF with weighting coefficient of 0.4 and localization scale of 5000 km.

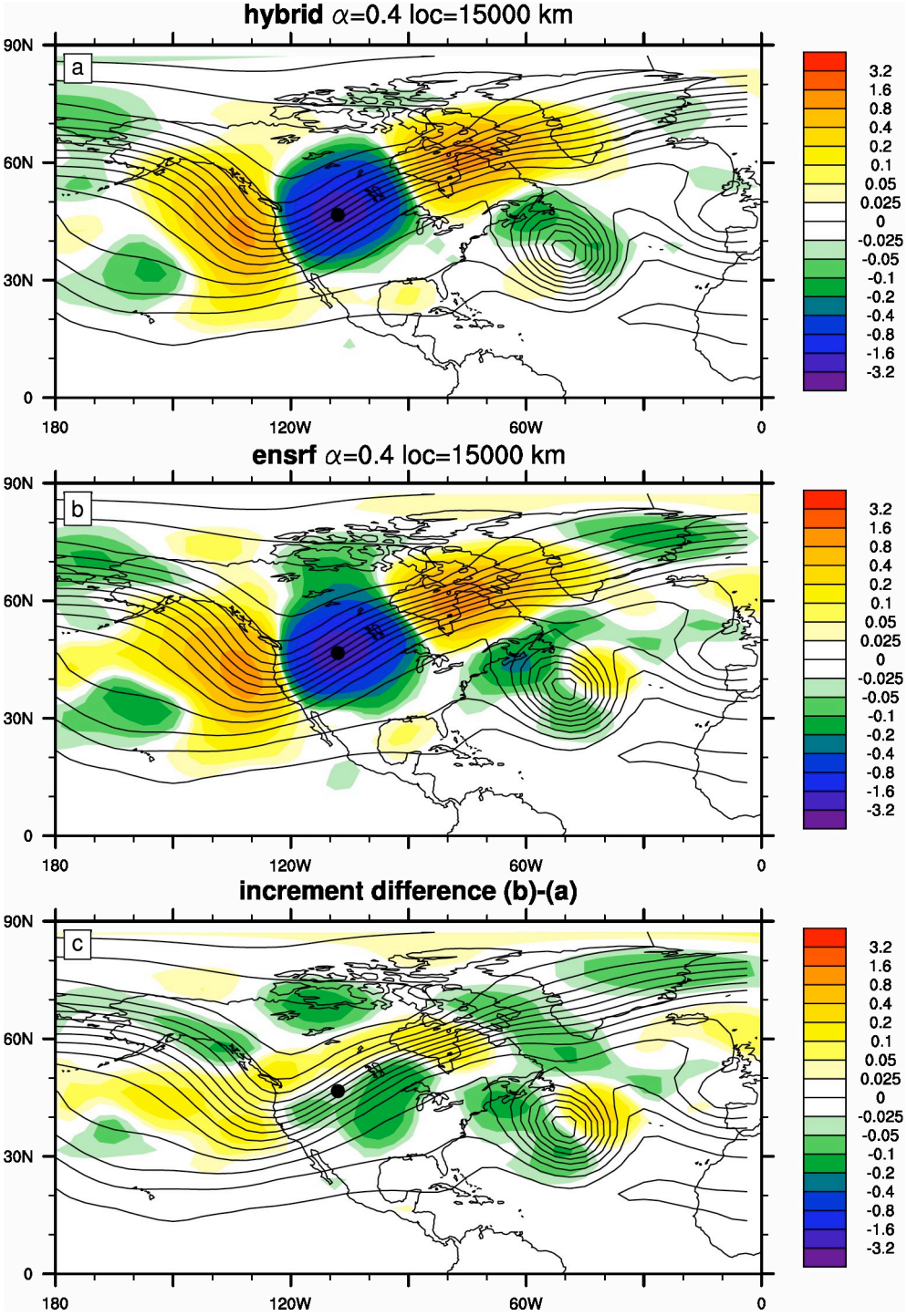

Fig. 4 Same as Fig. 3, except that the colors scale now emphasizes small values. (a) is the same as (a) in Fig. 3, except using the new color scale. (b) is for the EnSRF with weighting coefficient of 0.4 and localization scale of 15000 km; (c) is the difference in the increments between (b) and (a).

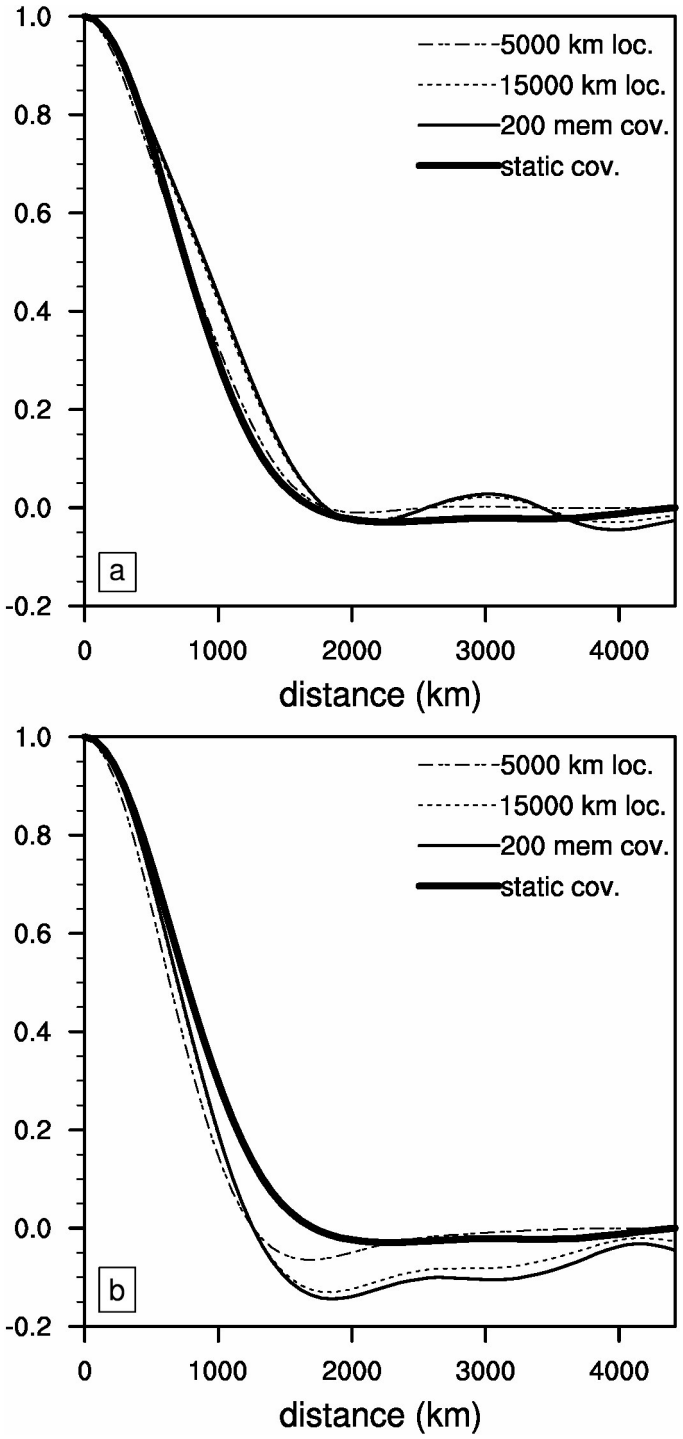

Fig. 5 Spatially lagged correlation along $45^{o}N$ latitude of the second layer thickness ($\Delta\pi_2$) as a function of zonal distance for the static correlation (thick solid), correlation from two sets of 200-member random samples (thin solid), 200-member sample correlation with 5000 km localization (dash-dotted) and 15000 km localization (dotted). (a) and (b) are for the first and second sets of 200-member random samples.

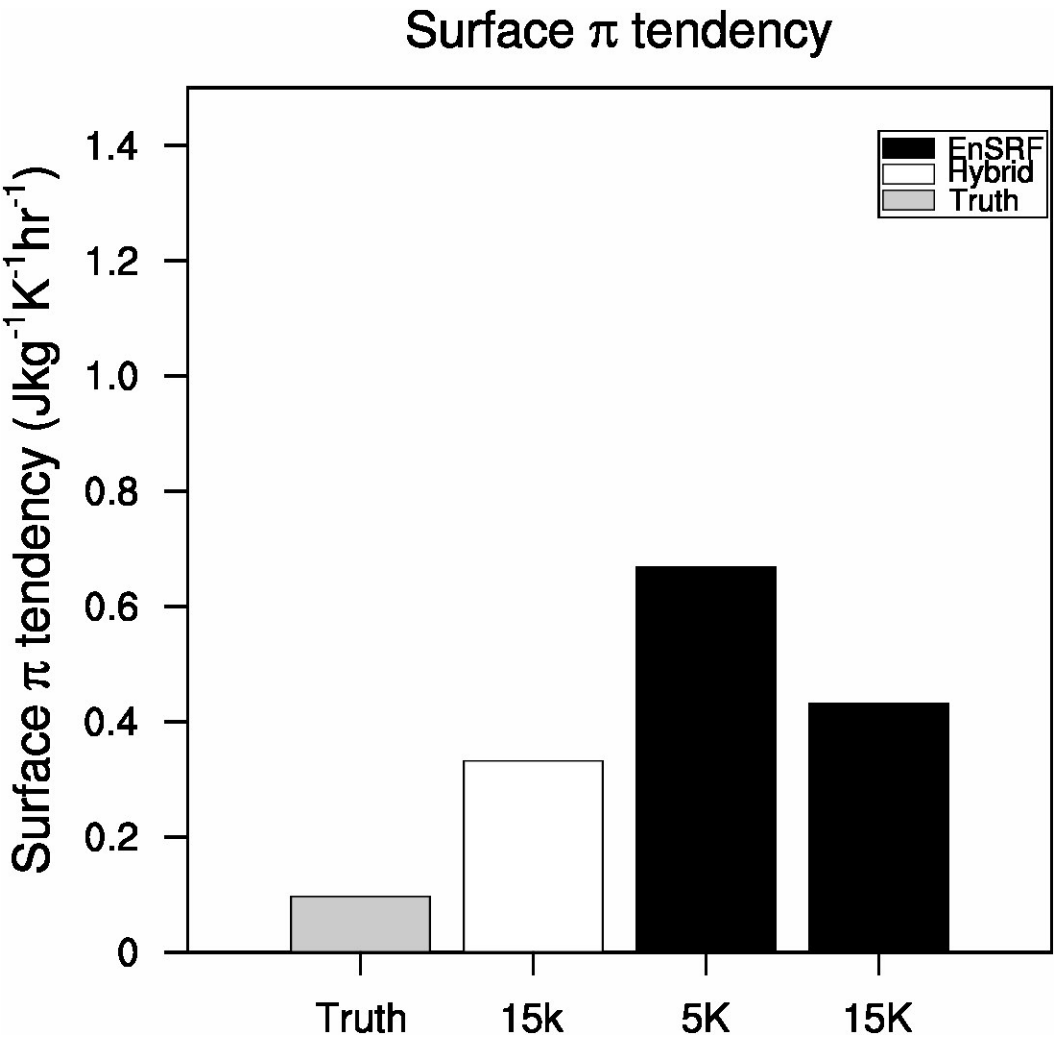

Fig. 6 Mean absolute surface Exner function (π_s) tendency ($Jkg^{-1}K^{-1}h^{-1}$) averaged globally, over the subsequent twenty-three 1-h forecast period and over all time. White bar is for the hybrid with weighting coefficient of 0.4 and localization scale of 15000 km. Black bars are for the EnSRF with weighting coefficient of 0.4 and localization scales of 5000 km and 15000 km. The grey bar is for the truth run.