A Hybrid ETKF-3DVAR Data Assimilation Scheme for the WRF Model. Part II: Real Observation Experiments

Xuguang Wang[1], Dale M. Barker[2], Chris Snyder[2], Thomas M. Hamill[3]

[1]*CIRES Climate Diagnostics Center, University of Colorado, and Physical Sciences Division, NOAA Earth System Research Laboratory, Boulder, CO*

[2]*Mesoscale and Microscale Meteorology Division, NCAR, Boulder, CO*

[3]*Physical Sciences Division, NOAA Earth System Research Laboratory, Boulder, CO*

Revised and resubmitted as an article to *Monthly Weather Review*

April 15 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

The hybrid ensemble transform Kalman filter (ETKF)-three dimensional variational (3DVAR) data assimilation system developed for the Weather Research and Forecasting (WRF) model was further tested with real observations, as a follow-up for the observation-system simulation experiment (OSSE) conducted in Part I.

A domain encompassing North America was considered. Because of limited computational resources and a large number of experiments conducted, the forecasts and analyses employed relatively coarse grid spacing (200 km) to emphasize synoptic scales. As a first effort to explore the new system with real observations, relatively sparse observation dataset consisting of radiosonde wind and temperature during 4 weeks of January 2003 were assimilated. The 12-hour forecasts produced by the hybrid analysis produced less root-mean-square error than the 3DVAR. The hybrid improved the forecast more in the western part of the domain than the eastern part. It also produced larger improvements in the upper troposphere. The overall magnitude of the ETKF ensemble spread agreed with the overall magnitude of the background forecast error. For individual variables and layers, the consistency between the spread and the error was less than the OSSE in Part I. Given the coarse resolution and relatively sparse observation network adopted in this study, caution is warranted when extrapolating these results to operational applications

A case study was also performed to further understand a large forecast improvement of the hybrid during the 4-week period. The flow-dependent adjustments produced by the hybrid extended a large distance into the eastern Pacific data-void region. The much improved analysis and forecast by the hybrid in the data void

subsequently improved forecasts downstream in the region of verification. Although no moisture observations were assimilated, the hybrid updated the moisture fields flow-dependently through cross-variable covariances defined by the ensemble, which improved forecasts of cyclone development.

1. Introduction

A hybrid ensemble transform Kalman filter (ETKF)-three dimensional variational (3DVAR) data assimilation system has been recently developed for the Weather Research and Forecasting (WRF) model (Wang et al. 2008, hereafter "Part I"). It is based on the existing WRF 3DVAR (Skamarock et al. 2005, Ch. 9; Barker et al. 2003, 2004). Unlike 3DVAR, which uses a static covariance model to estimate the background forecast errors, the hybrid system combines ensemble covariances with the 3DVAR static covariances to provide a flow-dependent estimate of the background error statistics. The ensemble-based estimates of covariances are incorporated into WRF 3DVAR using the extended control variable method (Lorenc 2003; Buehner 2005; Wang et al. 2007a). The ensemble perturbations are updated by the computationally efficient ETKF (Wang and Bishop 2003; Wang et al. 2004; Wang et al. 2007b).

As discussed in Part I, besides the benefit of including the flow-dependent ensemble covariance in the assimilation, our interest in hybrid schemes also stems from several other factors: (i) the hybrid scheme is simple to implement in the framework of existing operational variational schemes; (ii) it is potentially less expensive than the ensemble Kalman filter (EnKF, e.g., Hamill 2006), as the perturbations are updated by the less expensive ETKF and variational update of the ensemble mean may not scale linearly with the number of observations; and (iii) it may be more robust than the EnKF if one can only afford to run a small ensemble or if the model error is large (Wang et al. 2007b; 2008 in preparation), although in that case the benefit relative to standard 3DVAR can be expected to be small.

As an initial test of the newly developed system, an observation system simulation experiment (OSSE) was conducted in Part I. It was found that the analysis of the hybrid was more accurate than that of the 3DVAR. The OSSE assumed that the WRF model was perfect, the observation-error covariance was perfectly known, and the lateral boundary condition (LBC) ensemble perfectly represented the LBC uncertainty. When assimilating real observations, however, none of these assumptions will be valid. In this study (Part II), we will test the hybrid system with the assimilation of real observations and again compare its short-range forecast performance to that of the 3DVAR.

Studies exploring the use of hybrid ensemble-variational schemes for assimilating real observations are rather limited. Barker (1999) reported initial test results of combining only a single breeding ensemble member (Toth and Kalnay 1997) with the UK Meteorological Office's global 3DVAR system and suggested substantial forecast improvement would be possible if more ensemble members were used. Buehner (2005) constructed a hybrid system based upon the Canadian global 3DVAR system and tested it with real observations. Both ensembles from their global ensemble Kalman filter (EnKF; Houtekamer et al. 2005) and from "perturbed 3DVAR" were used. Their results showed small forecast improvements and suggested revisiting the problem with increasing ensemble size. On the other hand, numerous studies have tested the EnKF with real observations. Encouraging results have been obtained for both global (e.g, Whitaker et al. 2004; Whitaker et al. 2008; Szunyogh et al. 2007; Houtekamer et al. 2005) and limited-area numerical weather prediction (NWP) applications (e.g, Meng and Zhang 2007; Torn and Hakim 2008; Fujita et al. 2007; Dowell et al. 2004). These studies

suggest benefits of flow-dependent ensemble covariance even in the real-observation scenario.

Given the encouraging results of previous work, we further test the newly developed WRF hybrid ETKF-3DVAR data assimilation system with real observations in this study. It represents the first investigation of the hybrid ETKF-3DVAR method for a limited-area NWP model with real observations. As a first effort to understand the new system and given limited computational resources, the experiments were conducted with relatively coarse 200 km resolution and a subset of observations. Caution is warranted when extrapolating these results to operational applications, since many regional-scale applications employ much finer resolutions and denser observations,

In section 2, we describe how the real-data experiments are designed. Section 3 will first present results comparing WRF 3DVAR with WRF hybrid using general diagnostic tools, and then present a case study to further understand the differences between the two approaches. Section 4 concludes and provides further discussion.

2. Experiment Design

a. Model, observations, ensemble configurations, and verification techniques

In this study, we ran WRF with the same model configuration as the OSSE in Part I. The chosen domain covers North America and the surrounding oceans (see Fig. 1 in Part I). As in the OSSE, in order to conduct a large number of experiments to find optimal tunable parameters (section 3c in Part I; section 2c in Part II) using the limited computational resources available, we ran WRF with a 200-km grid spacing on a 45x45 horizontal grid with 27 vertical levels. The model top was at 50 hPa.

The experiments began at 0000 UTC 1 January 2003 and lasted for four weeks. Observations were assimilated every 12 hours beginning at 1200 UTC 1 January. The observations consisted of real radiosonde winds and temperatures, taken from the operational observation dataset from the National Center for Environmental Prediction (NCEP)[1]. Figure 1 of Part I shows a snapshot of the horizontal observation distribution. Observation errors were assumed to be uncorrelated and the observation error statistics were obtained from the NCEP operational observation dataset. No additional quality control was applied. Figure 2 of Part I shows the vertical profile of the observations errors. Note that for the purpose of facilitating comparison, the only difference between Part I and Part II in terms of the observations was that Part II adopted real observation values and Part I adopted simulated values.

For the hybrid experiment, we ran a 50-member ensemble, with perturbations updated using the ETKF (Wang et al. 2008). The initial ensemble at the very beginning of the data assimilation cycles (0000 UTC 1 January 2003) and the LBC ensembles during the cycles were generated by adding random 50 perturbations to the NCEP "Final" analyses (FNL; https://dss.ucar.edu/datazone/dsszone/ds083.2). These perturbations were drawn from a normal distribution having the same covariance as the WRF 3DVAR NCEP background-error covariance. Figure 3 of Part I shows typical values of the spread of such 50 random perturbations in the WRF domain including the lateral boundaries. This method of generating the LBC ensembles was proposed and tested by Torn et al. (2006) for their WRF EnKF system.

[1] See http://www.emc.ncep.noaa.gov/mmb/data_processing/prepbufr.doc for a detailed description of the observations.

For the 3DVAR experiment, the initial background forecast at 1200 UTC 1 January 2003 was taken from the 12-hour ensemble-mean forecast at that time so that both the hybrid and the 3DVAR experiments started with the same background forecast. The LBCs during the cycles for the 3DVAR were from the NCEP FNL analyses.

To evaluate the forecast errors of the hybrid and 3DVAR, we interpolated the 12-hour forecasts from either the 3DVAR or the hybrid analysis to the radiosonde observation locations, and compared them to the observations. As in the OSSE, the verification domain was within the inner quarter of the total domain (see Fig. 1 of Part I). The verification statistics were collected after first 5 days' data assimilation cycles.

b. Tuning the 3DVAR static background-error covariance

Since the default WRF 3DVAR NCEP covariance may not be the optimal model of static background-error covariances for the current data assimilation experiment, we re-calculated the static error covariances. We used the WRF ETKF background-forecast (12-hour) ensembles that were generated corresponding to the current real-observation experiment settings mentioned in the previous subsection.

Specifically, we ran the ETKF ensemble forecasts every 12 hours. The ensemble-mean background forecast was updated by the WRF 3DVAR, using the radiosonde observations and the default static error covariance. The ensemble perturbations were updated by the ETKF. We then added the updated perturbations to the updated ensemble mean to generate an ensemble of analyses and started 12-hour ensemble forecasts. The procedures were repeated for 4 weeks. We removed the first 5 days, collected the rest of the 12-hour ensemble forecasts and re-calculated the static background-error covariance

$\mathbf{B}$. In constructing $\mathbf{B}$, linear balance between mass and wind fields and horizontal homogenous error covariances were assumed. For details on calculating the static covariance for WRF 3DVAR, please refer to Skamarock et al. (2005, Chapter 9). Finally, we re-ran the 3DVAR experiment using the newly generated $\mathbf{B}$. As shown in Table 1, the 12-hour forecasts initialized by the 3DVAR analysis using this tuned static covariance were slightly more accurate than using the default background-error covariance. As in the OSSE, the results were also not sensitive to whether we used a 4-week period or a 2-week period of ETKF forecasts to calculate $\mathbf{B}$. In the following 3DVAR and hybrid data assimilation experiments, we will use this newly generated static background-error covariance.

c. Hybrid data assimilation experiments

The hybrid ETKF-3DVAR scheme for WRF was described in Part I. For details on the ETKF ensemble generation scheme and on how the ensemble was incorporated during the variational update through extending the control variables, please refer to section 2 of Part I and references therein. Here we only briefly describe the range of hybrid experiments associated with different parameters.

Like the OSSE in Part I, the hybrid experiments in this study were conducted with various combinations of two tunable parameters. One was the weighting factor $1/\beta_1$, which defined the weight placed on the static background-error covariances. To conserve the total background-error variance, the weight placed on the ensemble covariance was given by $(1-1/\beta_1)$ (Hamill and Snyder 2000; Etherton and Bishop 2004; Wang et al.

2007b). We used 5 different values for the weighting factor, $1/\beta_1 = 1.0, 0.8, 0.5, 0.2,$ and 0.0.

The other tunable parameter was the horizontal scale of the covariance localization applied on the ensemble covariance that was used to ameliorate the effect of sampling error on the analysis, i.e., the "covariance localization" (Hamill et al. 2001). As discussed in section 2a of Part I, to save computational costs, we only included the horizontal covariance localization in the current hybrid system. The detrimental effect of the sampling error in the vertical covariances estimated by the ensemble was ameliorated by the use of the static covariance in addition to the ensemble covariance in the cost function. In the current system, the horizontal covariance localization was modeled applying the recursive filter (Hayden and Purser 2005) on the extended control variables during the variational minimization. The correlation length scales of the recursive filter determine the degree to which the ensemble covariance is localized. For each of the four weighting factors $1/\beta_1 = 0.8, 0.5, 0.2,$ and 0.0, we used four ensemble covariance localization scales $S_e = 707, 1414, 2828,$ and $4242\ km$, where S_e is the e-folding scale of the asymptotic Gaussian response function of the filter. For further details on the meaning of $1/\beta_1$ and S_e, please refer to sections 2a and 3c of Part I.

As in the OSSE (section 2b of Part I), an inflation factor Π was applied to the ETKF ensemble perturbations to ensure that on average the background-error variance estimated from the spread of ensembles about the ensemble mean was consistent with the background-error variance estimated from the differences between the ensemble mean and the observations (eq. (6) of Part I). In the current real-observation experiment, the inflation factor accounts not only for the ETKF's systematic underestimation of the error

variance owing to the limited ensemble size, but also for other mis-represented error sources. These include the errors from the model and the deficiencies in the LBC ensembles. One goal of this paper is to see if the hybrid scheme can improve upon the 3DVAR scheme with such a simple method to account for these error sources in the ensembles. Another factor in the ETKF, denoted as ρ, estimated the fraction of the forecast-error variance projected onto the ensemble subspace. Both factors were estimated adaptively as discussed in Part I.

3. Results

In this section, we first evaluate the performance of the hybrid analysis and the ETKF ensemble spread using data collected over the verification period. Then we present a case study to understand further the differences between the hybrid and the 3DVAR.

a. Verification of the forecasts

To evaluate the performance of the 3DVAR and hybrid analyses, we examined characteristics of the forecasts from the analyses generated by both schemes. The forecasts were evaluated against all the radiosonde wind and temperature observations in the verification domain.

1) 12-HOUR FORECAST ERROR OF THE HYBRID WITH RESPECT TO $1/\beta_1$ and S_e

Table 2 shows the root-mean-square (rms) difference between the 12-hour forecasts from the hybrid and the observations (called 12-hour forecast fit to observations, or simply 12-hour forecast error). The rms differences of the wind and temperature forecasts to observations are shown as a function of the weighting coefficient $1/\beta_1$ and the localization scale S_e. Note that throughout the paper, the fit of wind to the observations is defined as the square root of the averaged squared zonal and meridional wind fit to the observations, unlike some of the other studies (Houtekamer et al. 2005; Whitaker et al. 2008; Torn and Hakim 2008) where individual zonal and meridional components were examined separately. For most of the combinations of $1/\beta_1$ and S_e, except when $1/\beta_1 = 0$ for temperature, and $1/\beta_1 = 0$, $S_e = 4242\ km$ for wind, the hybrid forecasts were more accurate than the 3DVAR. The optimal parameters were $1/\beta_1 = 0.5$, $S_e = 2828\ km$ for the wind, and $1/\beta_1 = 0.8$, $S_e = 1414\ km, 2828\ km$ for the temperature.

As discussed in Part I's OSSE, the slight improvement of the $1/\beta_1 = 1.0$ experiment (full 3DVAR covariance updating the mean of the ETKF ensemble forecasts) over the 3DVAR experiment (a single-member forecast-assimilation cycle) was presumably because the background forecast from the ensemble mean was more accurate than the single control forecast. Further examining columns of Table 2, when $1/\beta_1$ was decreased from 1.0 to intermediate values, the forecast error also decreased. This indicated further improvement of the analysis when the ensemble covariance was incorporated. The improvement of the best-performing hybrid over the $1/\beta_1 = 1.0$ experiment was larger than the improvement of the $1/\beta_1 = 1.0$ experiment over the 3DVAR experiment, suggesting the improvement of the best-performing hybrid over 3DVAR was mainly due to incorporating the flow-dependent ETKF ensemble

covariances during the data assimilation. When $1/\beta_1$ was reduced to 0, the analyses became worse than when $1/\beta_1 = 1$, especially when the localization scales were large. This suggested when the background-error covariance was fully estimated by the ensemble, appropriate covariance localization needed to be applied. Otherwise, the detrimental effect of the sampling error will overwhelm the advantage of flow-dependent estimate of the error covariance by the ensemble.

Further examination of Table 2 indicates that, the forecast error was less sensitive to the ensemble covariance localization scales when the static covariance was included in addition to the ensemble covariance, as in Part I's OSSE. For example, the range of the temperature forecast error as the localization scale varied was 0.15 K when $1/\beta_1 = 0$ but only about 0.01~0.02 K when $1/\beta_1 = 0.2, 0.5, 0.8$.

2) FURTHER COMPARISON OF FORECAST AND ANLYSIS FIT TO OBSERVATIONS BETWEEN HYBRID AND 3DVAR

Next we further compare forecast errors and analysis fit to observations between the hybrid and 3DVAR. We chose the hybrid with $1/\beta_1 = 0.5$ and $S_e = 2828\ km$ and compared it with the 3DVAR. Hereafter we will term forecasts from hybrid analyses as "hybrid forecasts" and correspondingly "3DVAR forecasts" for forecasts from 3DVAR analyses. Figure 1 shows the time series of the rms 12-hour forecast fit to observations and the rms analysis fit to observations in the verification domain. Note that assimilated observations were used in the verification, so the analysis fit to observations is not a measure of analysis errors, rather it is measure of how much each scheme draws to the observation during the assimilation. The forecast fit to observations, however, is a

measure of forecast errors and thus was used to evaluate the performance of the hybrid and the 3DVAR. Figure 1 shows that the 3DVAR draws closer to the observation during the assimilation. However, for most of the time the hybrid forecasts were more accurate than the 3DVAR forecasts. At day 9, the hybrid was dramatically better, which will be explored in section 3c below.

Figure 2 shows the vertical profile of the biases of the 12-hour forecast and analysis fits to observations and the rms 12-hour forecast and analysis fits to the observations. The bias for wind was calculated as the square root of the sum of the squared zonal wind bias and the squared meridional wind bias. The value for each pressure level was calculated by averaging errors collected within a layer 50 hPa above and below that pressure level. For the wind forecast, the hybrid 12-hour forecast fit the observations better than 3DVAR for all levels. The largest improvement of the hybrid over 3DVAR was located ~ 200 hPa-300 hPa. For the temperature forecast, the hybrid 12-hour forecast fit the observations better than 3DVAR above 900 hPa. Below 900 hPa, the hybrid showed no improvement relative to the 3DVAR. Figure 2 also shows there was significant bias in the 12-hour temperature forecast at the lower troposphere below 900 hPa for both the hybrid and 3DVAR. This bias accounted for a significant fraction of the rms forecast fit to the observations (similar results were documented in Whitaker et al. 2008 comparing the NCEP 3DVAR with an ensemble filter using the Global Forecast System model). We speculate that this is because of the systematic errors in the boundary-layer, surface-layer, and land-surface parameterizations in WRF. Also consistent with Fig. 1, the 3DVAR draws more close to observations during the assimilation.

The OSSE results in Part I demonstrated that the improvement of the hybrid over 3DVAR was more pronounced in the western part than the eastern part of the continent. Table 3 summarizes the rms 12-hour forecast fit to the observations for the western and the eastern of the verification region for this real-observation experiment. The western and eastern regions were divided by the central longitude of the WRF domain. Consistent with the results of the OSSE, the hybrid forecast improved upon 3DVAR forecast in the western part of the verification region more than the eastern part. Vertical profiles of the rms 12-hour forecast fit to observations plotted for the western and eastern domains separately showed the same results (not shown). As discussed in Part I, one reason may be the hybrid improved the analysis more upstream over the data-sparse eastern Pacific. In section 3c, we consider a period of January 2003 when the hybrid had a dramatically smaller forecast error than the 3DVAR, which will demonstrate how the hybrid can correctly extrapolate the land-based observations into the data-sparse ocean region and thus improve the forecast downstream.

Figure 3 shows the rms analysis and forecast fits to the observations as a function of forecast lead times up to 48 hours averaged over the verification domain. The forecasts initialized from the hybrid analysis were more accurate. The hybrid improved wind forecast more than the temperature forecast. As discussed in section 4b of Part I, since the same LBCs were used to run the forecasts for both the hybrid and the 3DVAR, the influence of the LBCs becomes more dominant over the verification region with increasing forecast lead times. Therefore, the impact of the hybrid analysis should be better indicated by forecasts with relatively short lead times.

3) HYBRID SENSITIVITY TO LBC ENSEMBLE PERTURBATIONS

As stated in section 2a, the LBC ensemble perturbations for all previous experiments were generated by drawing random perturbations from the default 3DVAR static covariance. Figure 3 of Part I shows typical values of the spread of 50 such random perturbations at the lateral boundaries. We performed extra experiments to see if the 12-hour forecast error from the hybrid would vary significantly if the magnitude and the length scale of the perturbations were changed. In one experiment, we multiplied the perturbations generated by the default 3DVAR by 1.5. In the other experiment, we retuned the horizontal length scale of the static covariance so that the scale of the perturbation was comparable to synoptic scales, following the LBC ensembles used in the real-time WRF EnKF system (Torn and Hakim 2008) at the University of Washington (Ryan Torn 2007, personal communication). For both experiments, we used $1/\beta_1 = 0.5$ and $S_e = 2828\ km$.

Table 4 shows the rms 12-hour forecast fit to the observations over the verification domain for the two new experiments along with the hybrid whose LBC ensemble perturbations were generated by the default 3DVAR covariance. The performance of the hybrid was not sensitive to the chosen magnitude and length scale of the perturbations. This could be because, as suggested by the OSSE experiment by Torn et al. (2006), the verification region was in the inner quarter of the domain, which was less affected by the LBCs than the outer region. This also could be because all three choices of the LBC perturbations were crude ways to model the LBC uncertainty.

b. Verification of the ensemble spread

In this section, we measure the relationship of the ETKF 12-hour ensemble spread to the 12-hour background forecast error to evaluate how well the spread estimated the background forecast error.

As mentioned in section 2c and section 2b of Part I, there are two tunable coefficients in the ETKF, the inflation factor (Π) and the factor (ρ) that estimate the fraction of the background error variance projected onto the ensemble subspace. In the real-data experiment, these factors were intended to ameliorate the systematic underestimate on the error variance by the ETKF not only due to the limited ensemble size but also due to other mis-represented error sources such as the model error. Both factors were determined adaptively in the ETKF. Figure 4 shows the factors Π and ρ during the 4-week data assimilation period. The factors started to converge to constant values after 5 days (10 cycles). The average values after day 5 for Π and ρ were 8.6 and 11% respectively. As expected, the inflation factor Π was larger and the factor ρ was smaller than those in Part I's OSSE, due to other sources of system errors such as the model error in the real-data experiments. The inflation factors in Parts I and II were both smaller than the inflation factor used by the previous ETKF experiment in Wang and Bishop (2003) and were both larger than that used in Wang et al. (2007b). In both Parts I and II we used an improved formula of the ETKF as introduced in the Appendix A of Wang et al. (2007b). The improved formula requires smaller inflation factor than the old formula (Wang and Bishop 2003; Wang et al. 2004) as discussed in Wang et al. (2007b). In general, the magnitude of the inflation factor depends on the ensemble size, the model dynamics, the observation network and the errors in the model.

Next we define the metric to verify the ETKF ensemble spread. If the ensemble performs optimally, the innovation covariance should satisfy

$$\left\langle \left(\mathbf{y}^o - \mathbf{H}\overline{\mathbf{x}}^b\right)\left(\mathbf{y}^o - \mathbf{H}\overline{\mathbf{x}}^b\right)^T \right\rangle - \mathbf{R} = \mathbf{H}\mathbf{P}^e\mathbf{H}^T, \quad (1)$$

(e.g., Gelb 1974, eq. 9.1-15; Houtekamer et al. 2005). In eq. (1), $\mathbf{P}^e$ is the background-error covariance estimated from the ensemble, $\mathbf{H}$ is the observation operator, $\mathbf{R}$ is the observation error covariance matrix, $\mathbf{y}^o$ is a vector of observations, and $\overline{\mathbf{x}}^b$ is the ensemble-mean forecast. The symbol $\langle\cdot\rangle$ represents expectation. Therefore, examining the diagonal elements of the matrices on the left and right hand sides of (1) provides a measure on the skill of the ensemble spread. If the magnitude of the diagonal elements on the right and left sides of eq. (1) are similar, then ensemble spread is said to be consistent with the background error at the observation locations. In the following calculation, the expectation was estimated by an average over many observation times and locations.

Figure 5 shows the vertical profile of the 12-hour ensemble spread (square root of the diagonal elements of the right side of eq. (1) averaged for each level) versus the square root of the 12-hour innovation variance minus the observation error variance (square root of the diagonal elements of the left side of eq. (1) averaged for each level) for the wind and temperature. The hybrid shown in Fig. 5 corresponded to the weighting factor $1/\beta_1 = 0.5$ and the localization scale $S_e = 2828\ km$. The value at each pressure level was calculated by averaging the data collected within a layer that was 50 hPa above and below that level. As stated above, the statistics were collected over all observation sites within the verification domain over the verification period. Measured under the norm that defined the inflation factor (Wang and Bishop 2003; eq. (10) in Part I), the

overall spread matched the overall first-guess error by construction. Checking wind and temperature individually in Fig. 5, the overall spread of the wind was over-dispersive and the temperature spread was under-dispersive. Further examining individual levels, we found that, for both the wind and the temperature, the ensemble spread was under-dispersive in the lower and upper levels and over-dispersive in between.

Compared to the OSSE results of Part I, the ensemble spread in the real-observation experiment was less representative of the background forecast error. For example, as shown on Fig. 5, the ensemble spread of the temperature near the surface was much smaller than the background forecast error. This was consistent with the systematic model bias presumably due to the deficiency in parameterizing the boundary layer, surface layer and the land surface processes. Since all 50-member ensembles were run with the same set of physics schemes, such error was not properly represented in the ensemble. In addition, the initial state uncertainty for the land surface was not perturbed, which could also lead to spread deficiency at low levels (Sutton et al. 2006). The lack of ensemble spread due to mis-representation of these additional error sources can be spatially inhomogeneous, an effect that was not captured by an inflation that was calculated based on innovation variances summed over all variables and the entire domain.

It is possible that the spread-skill consistency may be improved if we explore other methods to account for these error sources. An inflation factor that is not only adaptive in time but also adaptive in space (Anderson 2007) can be used to account for the inhomogeneity of the spread deficiency. Alternatively, one can employ different physical parameterizations for different members (e.g., Fujita et al. 2007, Houtekamer et

al. 1996) or using additive noise method (e.g., Whitaker et al. 2008; Hamill 2006; Houtekamer et al. 2005; Mitchell et al. 2002). To perturb the land surface state, one can use land surface state analyses from different sources or include the land surface state in the ensemble update.

c. Case study

Previous studies have shown that flow-dependent estimates of background-error covariances are particularly helpful in the data-sparse regions and in the analysis of unobserved variables (Hamill and Snyder 2000; Snyder and Zhang 2003; Whitaker et al. 2008; Part I). Part I concluded that ensemble-based estimate of background errors provided varying, flow-dependent adjustments to the observations. Sometimes the ensemble covariance produced large increments in data void regions because of the diagnosed strong correlation of the background errors in these regions with those in the data-dense region. In comparison, the 3DVAR used a fixed, isotropic background error covariance model and the update for the data-void region was constrained by the fixed length scale in the static covariance. The improved analysis from the ensemble-based methods in data sparse regions subsequently improved the forecast downstream.

Another potential benefit of the ensemble-based covariance estimates is a more explicit coupling of the moisture to other state variables. In the WRF 3DVAR system, the humidity is weakly coupled with temperature field since the static covariance model does not explicitly include cross-covariances between humidity and other state variables. In contrast, the flow-dependent dynamical relationship between the moisture field and other fields is easily represented by the cross-variable covariances of the ensemble. We

hypothesize that the hybrid system may be able to benefit from all these advantages of the ensemble covariance. In this section, we discuss such potential advantages of the hybrid over the 3DVAR through a case study.

We chose to understand why the hybrid improved the 12-hour forecast valid at 1200 UTC 09 January 2003 (day 9), a date when the hybrid suddenly began to produce forecasts that were dramatically lower in error than the 3DVAR (Fig. 1). In the following diagnostics, we focused on determining the factors that contributed to the improvement of the wind forecast by the hybrid.

We first plot the difference of the 12-hour wind forecasts between the 3DVAR and the hybrid valid at 1200 UTC 09 January 2003. Figure 6 (solid and dotted lines) shows the vertical profile of the 12-hour wind forecast fit to observations averaged over the verification domain. The improvement of the hybrid over the 3DVAR peaked at 300 hPa.

A spatial map of the difference of the 12-hour wind forecast between the hybrid and the 3DVAR at 300 hPa is shown in Fig. 7. The largest difference was in the eastern Pacific and southwest of California. Relative to the hybrid, the 3DVAR wind forecast showed an anti-cyclonic anomaly in the eastern Pacific. (Hereafter, we call the difference of the 3DVAR and hybrid forecasts or analyses, defined as 3DVAR minus hybrid, the anomaly). This wind forecast anomaly penetrated into the verification domain, extending from southern California to the northern Mexico. At the radiosonde sites, 3DVAR showed the largest westerly anomaly at Vandenberg and largest northerly anomaly at El Paso at 300 hPa.

We further verified the forecasts over this region by comparing the 12-hour wind forecasts at 1200 UTC 9 January 2003 to the soundings at Vandenburg and El Paso. Figure 8 shows that while the 3DVAR wind forecast deviated from the sounding by as much as 37 ms^{-1}, the hybrid fit the sounding closely. Thus, the 3DVAR anti-cyclonic anomaly shown in Fig. 7 was largely associated with errors in the forecast from the 3DVAR analysis.

To understand these dramatic differences, we looked back a few data assimilation and forecast cycles prior to find the original source of the 3DVAR anomaly in the 300 hPa wind. Figure 9 shows the evolution of the 3DVAR anomaly in 300 hPa height over the 24-hour period between 1200 UTC 08 and 1200 UTC 09 January. Note that the positive 300-hPa geopotential anomaly in Fig. 9e corresponds to the anti-cyclonic wind anomaly in Fig. 7. Collectively, Fig. 9 indicates that the ridge anomaly in 3DVAR started to appear at the analysis time at 1200 UTC 08 January 2003, when it was centered at about 127 W, 22N (Fig. 9b). This anomaly grew during the subsequent 12-hour forecast (Fig. 9c). It became stronger after the analysis at 0000 UTC 9 January 2003 (Fig. 9d), and was further intensified during the next 12-hour forecast, leading to the ridge anomaly (Fig. 9e), and the wind forecast anomaly (Fig. 7) in 3DVAR at 1200 UTC 09 January 2003.

To understand the original difference in the 300 hPa height analyses shown in Fig. 9b, we examined the 300 hPa height increments at 1200 UTC 08 January 2003 for both the hybrid and the 3DVAR. It was found that the hybrid, through utilizing ensemble-based covariances, updated the geopotential height over the eastern Pacific more extensively than the 3DVAR (not shown).

Next we conducted a series of diagnostics to understand the growth of the geopotential anomaly (Fig. 9c) and its impact on the 12-hour forecast valid at 1200 UTC 9 January 2003. Examination of the forecast from 1200 UTC 8 January 2003 every 3 hours until 0000 UTC 9 January 2003 confirmed that the region of the geopotential anomaly at about 127W, 22N in Fig. 9b grew quickly and formed the ridge anomaly in Fig. 9c. A further examination of the accumulated precipitation during the 12-hour forecast period ending at 0000 UTC 9 January 2003, showed that the 3DVAR was raining heavily during this period in this region (Fig. 10a) while the hybrid was not (Fig. 10b).

The additional precipitation and diabatic heating in the forecast from the 3DVAR analysis might be expected to produce ridging, consistent with the positive 3DVAR geopotential anomaly. To confirm this hypothesis, we removed the rain in the 3DVAR by replacing the water vapor mixing ratio in the 3DVAR analysis at 1200 UTC 8 January 2003 with that from the hybrid analysis. A revised 3DVAR 12-hour forecast was then generated. The resulting accumulated precipitation ending at 0000 UTC 9 January 2003 is shown in Fig. 11a and the corresponding 12-hour 300 hPa height anomaly forecast valid at 0000 UTC 9 January 2003 is shown in Fig. 11b. The precipitation in the revised forecast, especially in the region where the ridge anomaly grew rapidly, was significantly reduced, as was the corresponding geopotential anomaly.

To see if these apparent improvements in the moisture analysis would also benefit subsequent 3DVAR analyses and forecasts, we then continued the 3DVAR data assimilation cycle through 1200 UTC 9 January 2003. Fig. 6 (dash-dotted line) shows the fit of the corresponding wind forecast to the observations in the verification domain.

The advantage of the hybrid over 3DVAR at 300 hPa was reduced by 73%, through the improvements in the earlier moisture analysis in 3DVAR.

The above diagnostics suggest that the difference in the water-vapor mixing ratio analyses at 1200 UTC 8 January 2003 over the eastern Pacific was a main factor that led to the ridge anomaly in the 3DVAR forecast on 1200 UTC 9 January 2003. To understand how the hybrid adjusted the moisture field over the data-void eastern Pacific using the observations over the west coast, Fig. 12a shows the 700 hPa water vapor mixing ratio increment by the hybrid after assimilating all observations at 1200 UTC 8 January 2003. In the background forecast, the moisture gradient extending from the center of the domain to the southwest of the domain was associated with a warm front. A positive increment was associated with the moist air penetrating from the southeast corner, while negative increments appeared along the warm front and in the warm moist air mass. In other words, the hybrid extrapolated information from the wind and temperature observations over land in order to dry the troposphere along the warm front. The drying along the warm front reached the data-void eastern Pacific where the 3DVAR forecast produced spurious rain. Further diagnostics (not shown) also revealed that the hybrid dried all of the lowest 12 model levels in the same region, with increments up to 1 g/kg.

As an additional demonstration of the flow-dependent update by the hybrid, Fig. 12b shows the low-level moisture increment for another case at 1200 UTC 6 January 2003. The increment of the mixing ratio again aligns with the moisture gradient. In this case, there was negative increment at the moisture bulge at the southwest corner and positive increment downstream. Similar plots for both cases (not shown) for the 3DVAR

analysis showed negligible moisture increment and increments for the wind and temperature were also localized as was defined by the static correlation length scale.

4. Conclusions and Discussions

As a follow-up to the OSSEs of Part I, this paper has presented further tests of the hybrid ETKF-3DVAR data assimilation system developed for WRF by assimilating real observations. The experiments were conducted in a region surrounding North America. A coarse resolution of 200km horizontal grid spacing was used. Radiosonde wind and temperature observations for a 4-week period starting 1200 UTC 1 January 2003 were assimilated.

Our results showed that the hybrid analyses produced more accurate 12-hour forecasts than the 3DVAR. The improvements from the hybrid were larger over western North America than eastern North America. Vertical profiles of the 12-hour forecast error showed that, for the wind, the hybrid produced the largest improvement by 9%~11% relative to the 3DVAR at 200-300 hPa. For the temperature, the hybrid improved upon the 3DVAR by 3% on average at all layers except in the lower troposphere, where both the hybrid and the 3DVAR showed large systematic errors. The performance of the hybrid was generally not sensitive to the magnitude and the scale of LBC ensemble perturbations generated by the static covariance, perhaps because of the large domain used.

Using the inflation and subspace projection factors of Wang et al. (2007b), the overall ensemble spread was tuned to agree in magnitude with the overall 12-hour forecast error. For individual variables and layers, the consistency between spread and

error was less than that in the OSSE. As discussed in section 3b, this suggested more sophisticated methods other than a global inflation are needed to account for the system errors.

A case study was performed to understand a particular situation when the hybrid system outperformed the 3DVAR. The hybrid was able to make significant adjustments to the background over the data-void eastern Pacific using the observations over land according to the background flow. Although only wind and temperature observations were assimilated, the hybrid successfully updated the moisture field flow-dependently through cross-variable covariances defined by the ensemble. The changes in the moisture analysis improved the subsequent wind and temperature analyses and forecasts downstream in the verification region.

In this pilot study, we employed a coarse resolution model and a relatively sparse observation network. Future work should explore the effectiveness of the hybrid in a configuration closer to real-world regional scale implementation, using a higher-resolution model and dense observations. In general, the benefits of the hybrid over 3DVAR may diminish when model error is large and when the observation network becomes more uniform and denser (e.g., Hamill and Snyder 2000; Whitaker et al. 2004, 2008).

As discussed in Part I, our choice of a coarse resolution in this study emphasized the synoptic scales. The real-world regional scale applications use a much higher resolution. The mesoscale features that appear at higher resolution are known to be sensitive to the details of imperfect model parameterizations, and thus the hybrid, whose covariance estimates partially depend on the accuracy of the forecast model, may not

perform significantly better than 3DVAR. On the other hand, increasing the model resolution might be expected to reduce model errors at synoptic scales and thus improve the quality of the ensemble-estimated covariances and thus the analyses from the hybrid. Also note that the mesoscale is more poorly observed than larger scales, and does not exhibit as strongly the balances assumed by the 3DVAR covariance model, in which case the hybrid's ensemble covariances may provide a further advantage.

Given the limitation of our experiment design, our result in this study thus is not a direct analog to the real-world applications. Parameters such as weighting coefficients and localization length scales that are optimal for this study may not be optimal if the system were run at a higher resolution and with more observations and thus likely need to be retuned.

Note also that results of Buehner (2005) for a similar hybrid assimilation system applied to global analyses were less encouraging in terms of the relative performance of 3DVAR and hybrid. Since there were many differences between those experiments and ours in terms of numerical models, ensemble generation methods, ensemble size, observation networks and verification regions, it is unclear which of these explain the relative performance difference.

Further development and tests of the hybrid system are warranted given the encouraging results in this study and its appealing characteristics, such as its straightforward implementation within existing variational schemes and the efficient update of the ensemble perturbations by the ETKF. Besides, with further development of the ensemble, the benefit of the hybrid may also improve. For example, using multiple physics or stochastic physics to account for model errors may provide flow-dependent

representation of the model error, which may improve the skill of the ensemble to estimate the background forecast error. The performance of the ensemble may also improve if we use short-range ensemble forecasts from a global ensemble data assimilation system to generate the LBC ensembles, instead of using the random perturbations drawn from a static covariance.

As an initial effort to test and understand the newly developed hybrid system, we compare the hybrid with the 3DVAR. We recommend direct and thorough comparisons with other data assimilation techniques such as the EnKF and the 4DVAR so as to understand the relative advantages and disadvantages of different techniques in realistic NWP settings.

Acknowledgements

The participation of the first author was funded by National Science Foundation grants ATM-0205612, ATM-0205655 and a NOAA THORPEX grant. Discussions with Jeff Whitaker improved the manuscript. Yongsheng Chen kindly provided scripts to implement the perturbed lateral boundary conditions for WRF. We appreciate the constructive comments from both reviewers.

References

Anderson, J. 2007: Spatially-varying adaptive covariance inflation for an ensemble filter. Submitted to *Tellus*. Available from jla@ucar.edu.

Barker, D. M., 1999: Var scientific development paper 25: the use of synoptic-dependent error structure in 3DVAR. UK MET office technical reports, available from U.K. Met Office, FitzRoy Road, Exeter, Devon, EX1 3PB, United Kingdom.

—, W. Huang, Y.-R. Guo, and A. Bourgeois, 2003: A three-dimensional variational (3DVAR) data assimilation system for use with MM5. NCAR Tech. Note. NCAR/TN-453 + STR, 68pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO, 80307.]

—, —, —, —, and Q. N. Xiao, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897-914.

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.

Dowell, D. C., F. Zhang, L. J. Wicker, C. Snyder and N. Andrew 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.

Fujita, T., D. J. Stensrud and D. C. Dowell, 2007: Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Mon. Wea. Rev.*, **135**, 1846-1868.

Gelb, A., 1974: Applied Optimal Estimation. MIT Press, 374pp.

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.

Hayden, C. M., and R. J. Purser, 1995: Recursive filter objective analysis of meteorological fields: Applications to NESDIS operational processing. *J. Appl. Meteor.*, **34**, 3-15.

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

—, H. L. Mitchell, 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.

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, 2007: 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., accepted. Available from http://www.met.tamu.edu/people/faculty/fzhang/papers/MengZhang2007MWR2.pdf

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance, and model error representation in an ensemble Kalman filter. *Mon. Wea. Rev*., **130**, 2791-2808.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRF Version 2. NCAR Tech. Note 468+STR, National Center for Atmospheric Research, Boulder, CO, 88 pp.

Sutton, C., T. M. Hamill and T. T. Warner, 2006: Will perturbing soil moisture improve warm-season ensemble forecasts? A proof of concept. *Mon. Wea. Rev*., **134**, 3174-3189.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, 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 from http://www.weatherchaos.umd.edu/papers/Szunyogh_et_all_2007.pdf.

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

—, and —, 2008: Performance characteristics of a pseudo-operational ensemble Kalman filter. *Mon. Wea. Rev*., accepted, available from http://www.atmos.washington.edu/~hakim/papers/torn_hakim_realtime_enkf.pdf

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.

—, D. Barker, C. Snyder, T. M. Hamill, 2008a: A hybrid ETKF – 3DVAR data assimilation scheme for the WRF Model. Part I: observing system simulation experiment. Submitted to *Mon. Wea. Rev*., available from http://www.cdc.noaa.gov/people/xuguang.wang/wrfhybda_part1.pdf

—, C. Snyder, and T.M. Hamill, 2007a: On the theoretical equivalence of differently proposed ensemble/3D-Var hybrid analysis schemes. *Mon. Wea. Rev.*, **135**, 222-227.

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

Whitaker S. J., 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.

Table Caption:

TABLE 1. Root-mean-square (rms) fit of the 12-hour wind and temperature forecasts to the radiosonde observations for the 3DVAR experiments with default and the tuned static background error covariances.

TABLE 2. Root-mean-square (rms) fit of the 12-hour wind and temperature forecasts to the radiosonde observations for the hybrid with various combinations of the weighing coefficients $1/\beta_1$ and the covariance localization scales S_e. Please see text for the definition of $1/\beta_1$ and S_e. Numbers in the parentheses indicate the percentage improvement relative to the 3DVAR with tuned static covariance. The smallest rms fits are highlighted. For $1/\beta_1 = 1.0$, experiments do not depend on S_e.

TABLE 3. Rms fit of the 12-hour forecasts to the radiosonde observations for the 3DVAR and the hybrid ($1/\beta_1 = 0.5$, $S_e = 2828\ km$) over the western and eastern verification regions. The 3rd column is the absolute and percentage (in parentheses) improvement of the hybrid relative to the 3DVAR.

TABLE 4. Sensitivity of the rms 12-hour hybrid forecast fit to observations to the magnitude and the scale of the LBC ensemble perturbations generated by the static covariance. The hybrid corresponds to $1/\beta_1 = 0.5$ and $S_e = 2828\ km$. (1) First row shows the rms forecast fit to observations for the LBC ensemble perturbations generated by the default 3DVAR. (2) Second row is for increasing the magnitude of the perturbations

generated in (1) by a factor of 1.5. (3) Third row is for retuning the length scale of the perturbation to be comparable to synoptic scales.

Figure Captions:

Fig. 1. Time series of the rms fit of the 12-hour forecasts to the radiosonde observations (thin) and the rms fit of the analysis to the radiosonde observations (thick) over the verification domain for 3DVAR (dotted) and the hybrid with $1/\beta_1 = 0.5$ and $S_e = 2828\ km$ (solid). (a) wind. (b) temperature.

Fig. 2. Vertical profiles of the rms fits of the 12-hour forecasts (red thin) and analysis (blue thin) to the radiosonde observations and biases of the 12-hour forecasts (thick red) and analysis (thick blue) fits to the radiosonde observations for 3DVAR (dotted) and hybrid with $1/\beta_1 = 0.5$ and $S_e = 2828 km$ (solid). (a) wind. (b) temperature.

Fig. 3. The rms fit of the forecast to observations as a function of forecast lead times up to 48 hours for the hybrid (solid) and the 3DVAR (dotted) for (a) wind and (b) temperature.

Fig. 4. The inflation factor (Π) and the factor (ρ) of percentage projection of the first guess error variance onto the ensemble subspace for the ETKF in the hybrid experiment with $1/\beta_1 = 0.5$ and $S_e = 2828\ km$.

Fig. 5 Vertical profiles of the ETKF ensemble spread (dotted) versus the square root of the innovation variance minus the observation error variance (solid) for (a) wind and (b) temperature.

Fig. 6 Vertical profiles of the rms fit of the 12-hour wind forecast valid at 1200 UTC 9 January 2003 to the radionsonde observations over the verification domain for hybrid with $1/\beta_1 = 0.5$ and $S_e = 2828km$ (solid), 3DVAR (dotted) and 3DVAR whose water vapor mixing ratio analysis was replaced by that of the hybrid at 1200 UTC 8 January 2003 (dash-dotted).

Fig. 7 The difference of the 12-hour wind forecast between the 3DVAR and the hybrid with $1/\beta_1 = 0.5$ and $S_e = 2828km$ valid at 1200 UTC 9 January 2003. Solid lines are the 12-hour 300 hPa height forecast of the hybrid (m). Vectors are the wind forecast difference (3DVAR minus hybrid). The shades are the magnitude of the wind forecast difference.

Fig. 8 Fit of the 12-hour wind forecast to the soundings valid at 1200 UTC 9 January 2003 for the 3DVAR (dotted) and hybrid with $1/\beta_1 = 0.5$ and $S_e = 2828\ km$ (solid). (a) zonal wind fit to the sounding at Vandenburg, CA. (b) meridional wind fit to the sounding at El Paso, NM.

Fig. 9 The difference of the analysis and the first guess for the 300 hPa geopotential height between the 3DVAR and the hybrid from 1200 UTC 8 January 2003 to 1200 UTC 9 January 2003 every 12 hours. (a) difference of the background at 1200 UTC 8 January 2003. (b) difference of the analysis at 1200 UTC 8 January 2003. (c) difference of the background at 0000 UTC 9 January 2003. (d) difference of the analysis at 0000 UTC 9

January 2003. (e) difference of the 12-hour forecast at 1200 UTC 9 January 2003. Black contours are the 300 hPa geopotential height (m) for the hybrid. Color shades are the difference (3DVAR minus hybrid) of the 300 hPa geopotential height (m).

Fig. 10 12-hour forecasts of the accumulated precipitation valid at 0000 UTC 9 January 2003 for (a) 3DVAR and (b) hybrid.

Fig. 11 (a) same as Fig. 10a, and (b) same as Fig. 9c, except that plots represent the forecast simulations where the 3DVAR water vapor mixing ratio analysis at 0000 UTC 9 January 2003 was replaced with that of the hybrid.

Fig. 12 700 hPa water vapor mixing ratio increment by the hybrid at (a) 1200 UTC 8 January 2003 and (b) 1200 UTC 6 January 2003. Black lines are the first guess of 700 hPa water vapor mixing ratio. Color shades are the increment of the mixing ratio by assimilating radiosonde wind and temperature observations.

TABLE 1. Root-mean-square (rms) fit of the 12-hour wind and temperature forecasts to the radiosonde observations for the 3DVAR experiments with default and the tuned static background error covariances.

	Wind ($\sqrt{U^2+V^2}$, ms^{-1})	T (k)
Default 3DVAR	6.40	1.89
Tuned 3DVAR	6.39	1.87

TABLE 2. Root-mean-square (rms) fit of the 12-hour wind and temperature forecasts to the radiosonde observations for the hybrid with various combinations of the weighing coefficients $1/\beta_1$ and the covariance localization scales S_e. Please see text for the definition of $1/\beta_1$ and S_e. Numbers in the parentheses indicate the percentage improvement relative to the 3DVAR with tuned static covariance. The smallest rms fits are highlighted. For $1/\beta_1 = 1.0$, experiments do not depend on S_e.

Wind (ms^{-1})	$S_e = 4242\ km$	2828 *km*	1414 *km*	707 *km*
$1/\beta_1 = 1.0$	6.255 (2.2)	—	—	—
0.8	5.991 (6.3)	6.001 (6.1)	6.046 (5.4)	6.186 (3.2)
0.5	5.997 (6.2)	**5.960 (6.8)**	5.998 (6.2)	6.146 (3.9)
0.2	5.964 (6.7)	6.010 (6.0)	6.045 (5.4)	6.160 (3.6)
0.0	6.457 (-1.0)	6.327 (1.0)	6.241 (2.4)	6.201 (3.0)

T (K)	$S_e = 4242\ km$	2828 *km*	1414 *km*	707 *km*
$1/\beta_1 = 1.0$	1.858 (0.6)	—	—	—
0.8	1.818 (2.7)	**1.813 (3.0)**	**1.813 (3.0)**	1.827 (2.2)
0.5	1.821 (2.6)	1.818 (2.7)	1.816 (2.8)	1.823 (2.5)
0.2	1.851 (1.0)	1.845 (1.3)	1.829 (2.1)	1.841 (1.5)
0.0	2.034 (-8.8)	1.979 (-5.9)	1.921 (-2.8)	1.886 (-0.9)

TABLE 3. Rms fit of the 12-hour forecasts to the radiosonde observations for the 3DVAR and the hybrid ($1/\beta_1 = 0.5$, $S_e = 2828\ km$) over the western and eastern verification regions. The 3rd column is the absolute and percentage (in parentheses) improvement of the hybrid relative to the 3DVAR.

Wind (ms^{-1})	3DVAR	HYBRID	IMPROVEMENT
WEST	7.143	6.567	0.576 (8.1)
EAST	5.713	5.423	0.290 (5.1)

T (k)	3DVAR	HYBRID	IMPROVEMENT
WEST	2.008	1.929	0.079 (3.9)
EAST	1.763	1.735	0.028 (1.6)

TABLE 4. Sensitivity of the rms 12-hour hybrid forecast fit to observations to the magnitude and the scale of the LBC ensemble perturbations generated by the static covariance. The hybrid corresponds to $1/\beta_1 = 0.5$ and $S_e = 2828\ km$. (1) First row shows the rms forecast fit to observations for the LBC ensemble perturbations generated by the default 3DVAR. (2) Second row is for increasing the magnitude of the perturbations generated in (1) by a factor of 1.5. (3) Third row is for retuning the length scale of the perturbation to be comparable to synoptic scales.

	Wind (ms^{-1})	T (K)
Default 3DVAR	5.960	1.818
Magnitude multiplied by 1.5	6.005	1.819
Retuned length scale	5.956	1.816

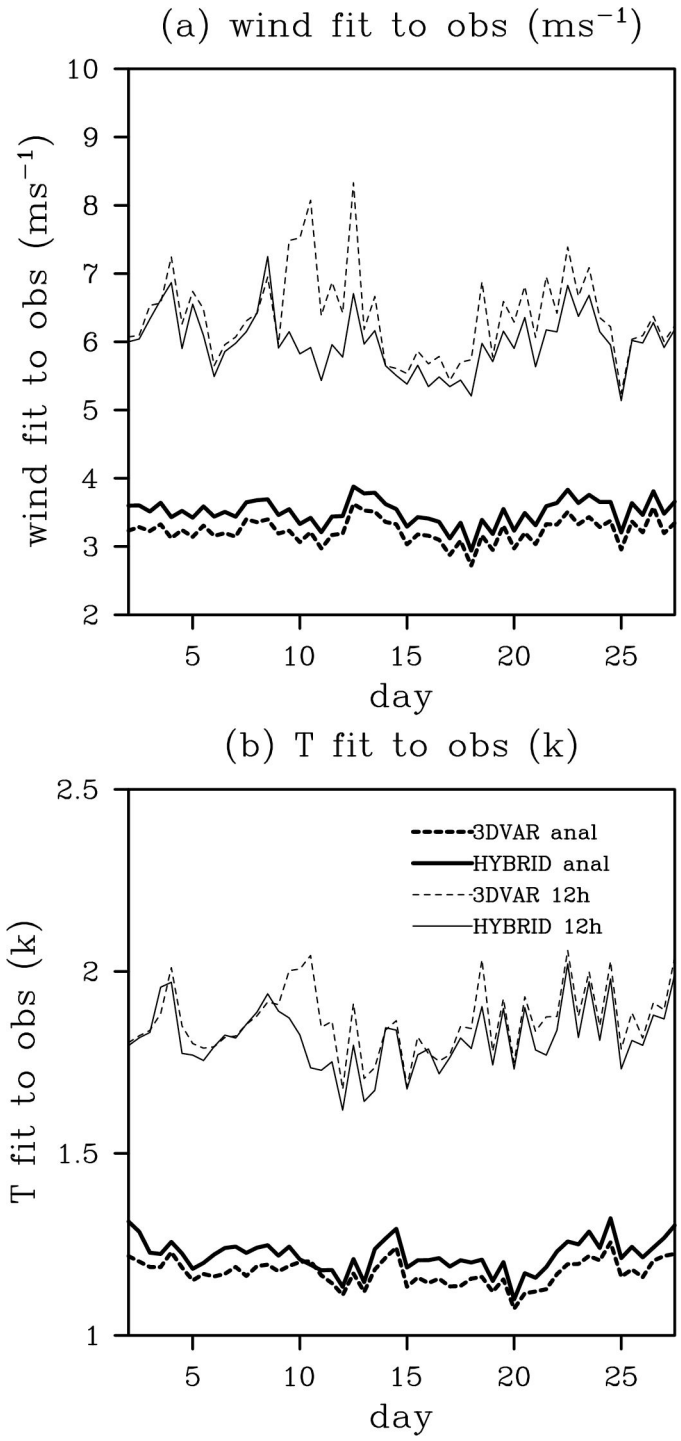

Fig. 1. Time series of the rms fit of the 12-hour forecasts to the radiosonde observations (thin) and the rms fit of the analysis to the radiosonde observations (thick) over the verification domain for 3DVAR (dotted) and the hybrid with $1/\beta_1 = 0.5$ and $S_e = 2828\ km$ (solid). (a) wind. (b) temperature.

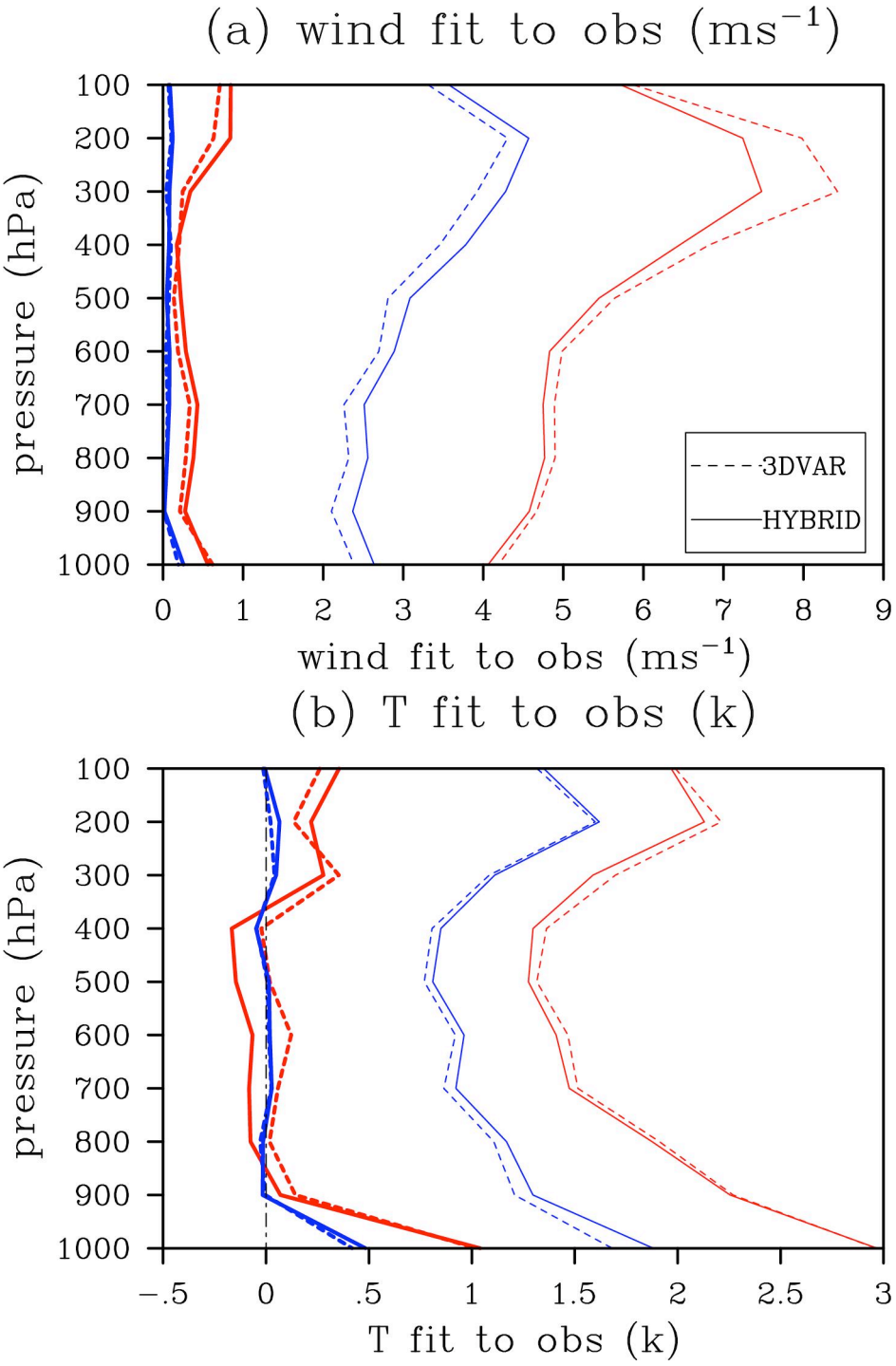

Fig. 2. Vertical profiles of the rms fits of the 12-hour forecasts (red thin) and analysis (blue thin) to the radiosonde observations and biases of the 12-hour forecasts (thick red) and analysis (thick blue) fits to the radiosonde observations for 3DVAR (dotted) and hybrid with $1/\beta_1 = 0.5$ and $S_e = 2828 km$ (solid). (a) wind. (b) temperature.

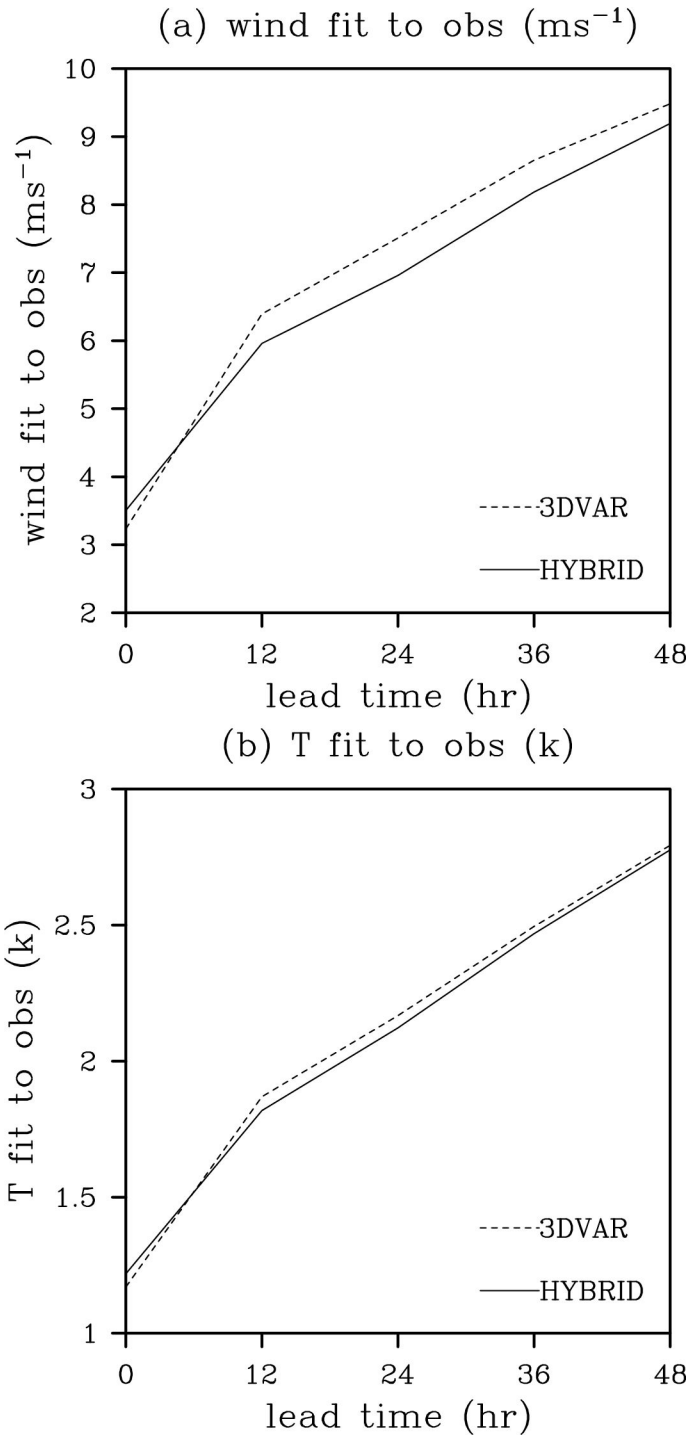

Fig. 3. The rms fit of the forecast to observations as a function of forecast lead times up to 48 hours for the hybrid (solid) and the 3DVAR (dotted) for (a) wind and (b) temperature.

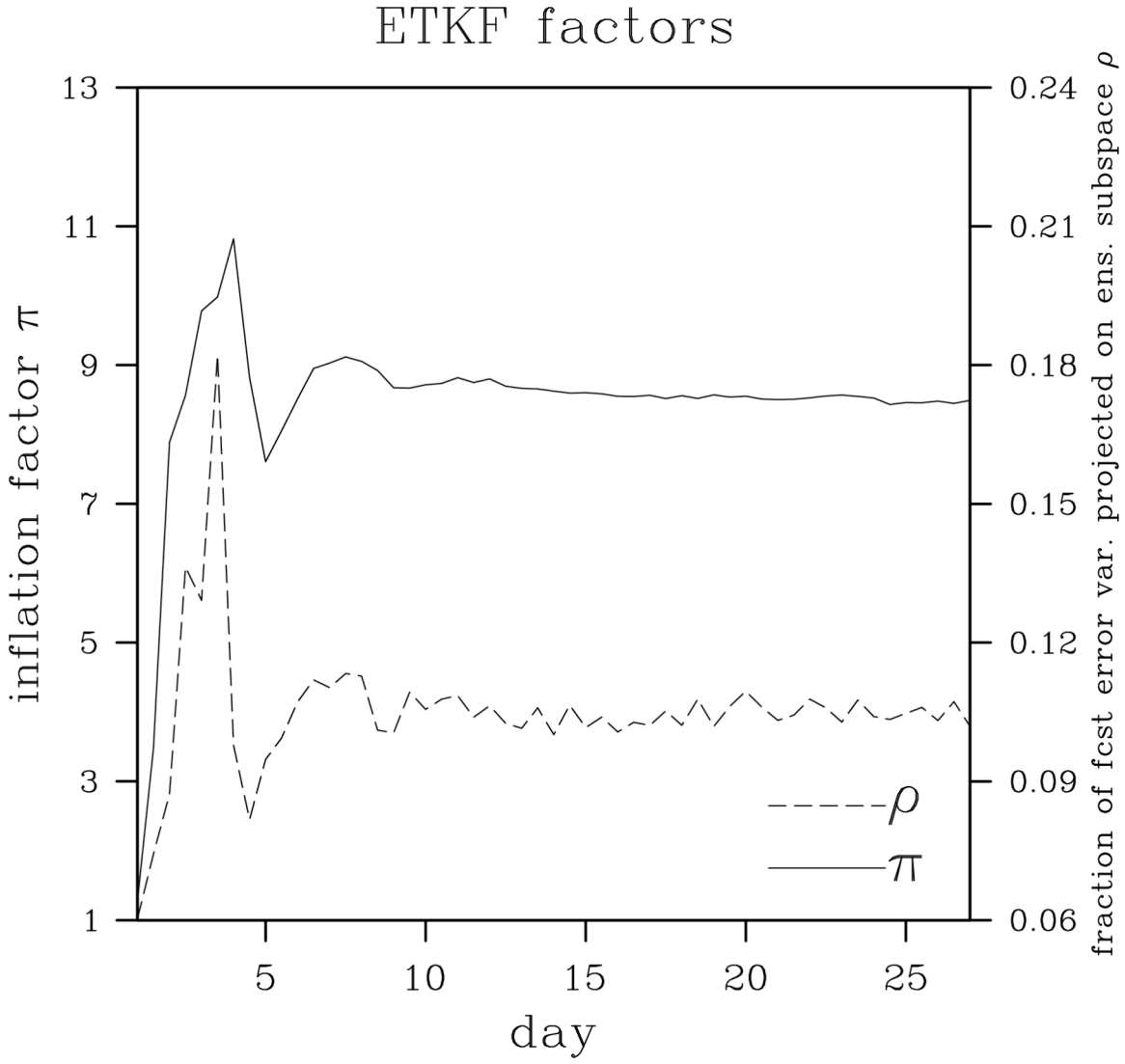

Fig. 4. The inflation factor (Π) and the factor (ρ) of percentage projection of the first guess error variance onto the ensemble subspace for the ETKF in the hybrid experiment with $1/\beta_1 = 0.5$ and $S_e = 2828\ km$.

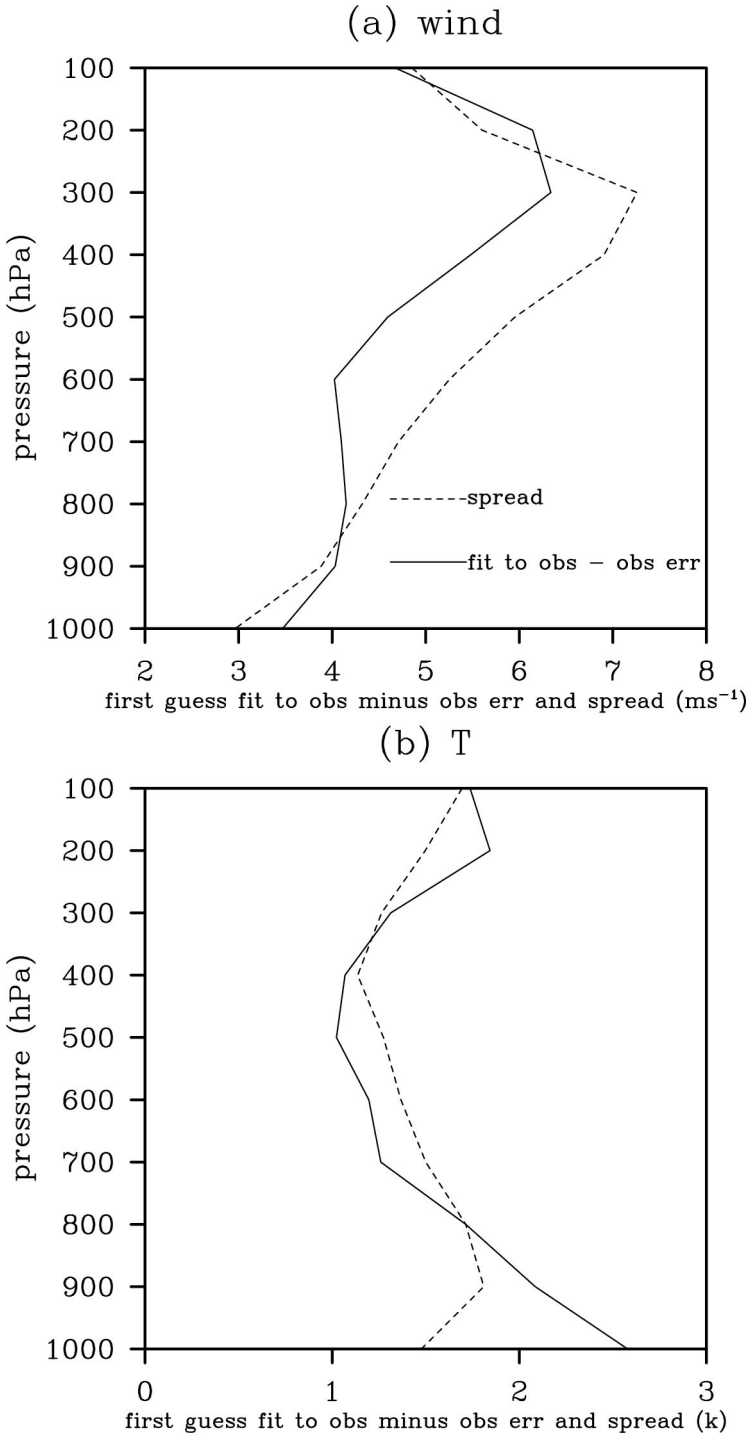

Fig. 5. Vertical profiles of the ETKF ensemble spread (dotted) versus the square root of the innovation variance minus the observation error variance (solid) for (a) wind and (b) temperature.

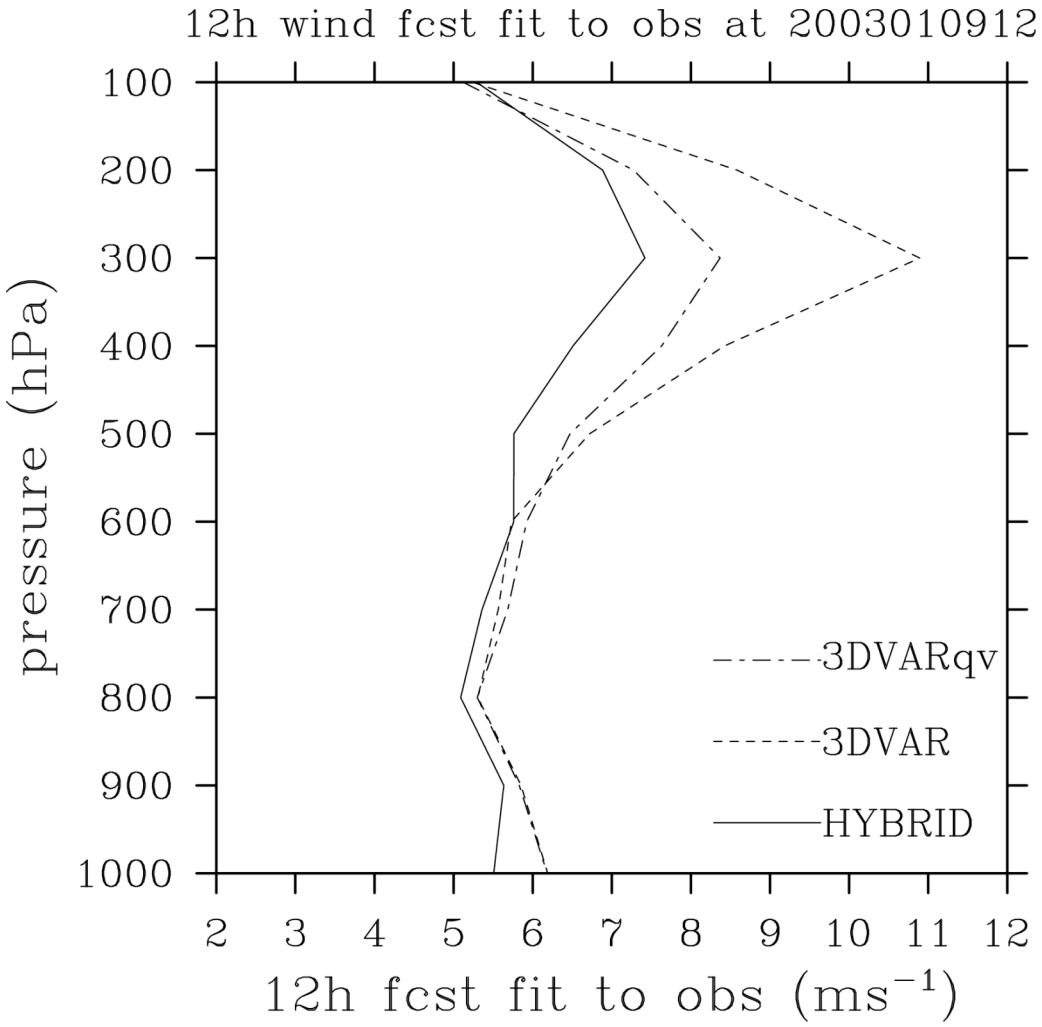

Fig. 6. Vertical profiles of the rms fit of the 12-hour wind forecast valid at 1200 UTC 9 January 2003 to the radionsonde observations over the verification domain for hybrid with $1/\beta_1 = 0.5$ and $S_e = 2828\ km$ (solid), 3DVAR (dotted) and 3DVAR whose water vapor mixing ratio analysis was replaced by that of the hybrid at 1200 UTC 8 January 2003 (dash-dotted).

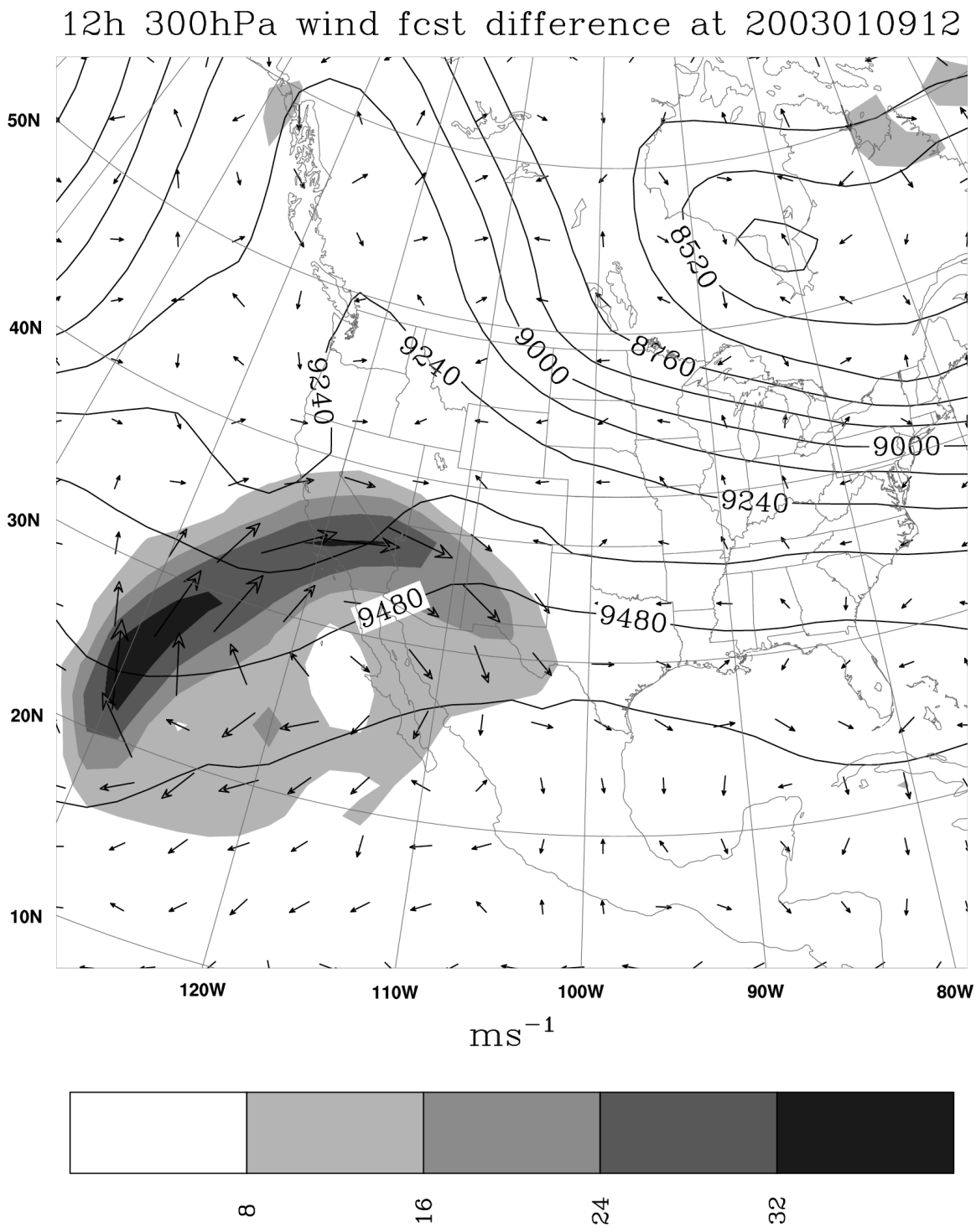

Fig. 7. The difference of the 12-hour wind forecast between the 3DVAR and the hybrid with $1/\beta_1 = 0.5$ and $S_e = 2828km$ valid at 1200 UTC 9 January 2003. Solid lines are the 12-hour 300 hPa height forecast of the hybrid (m). Vectors are the wind forecast difference (3DVAR minus hybrid). The shades are the magnitude of the wind forecast difference.

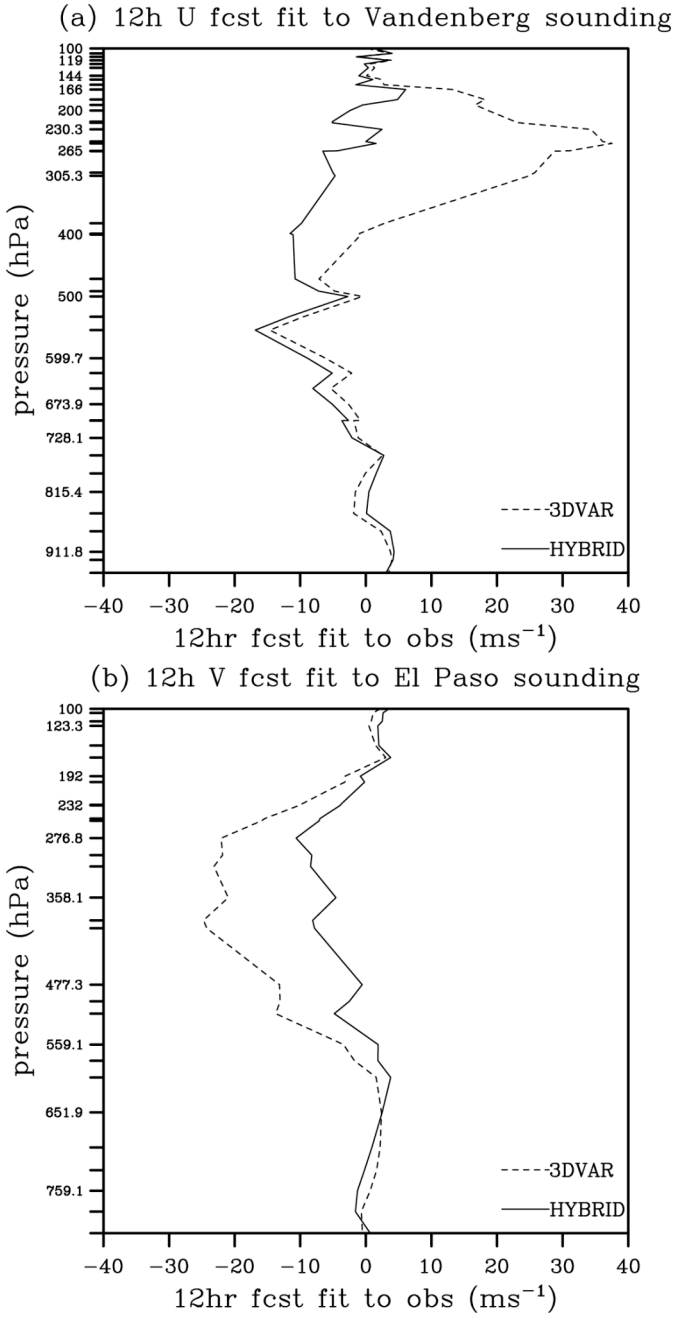

Fig. 8. Fit of the 12-hour wind forecast to the soundings valid at 1200 UTC 9 January 2003 for the 3DVAR (dotted) and hybrid with $1/\beta_1 = 0.5$ and $S_e = 2828\ km$ (solid). (a) zonal wind fit to the sounding at Vandenburg, CA. (b) meridional wind fit to the sounding at El Paso, NM.

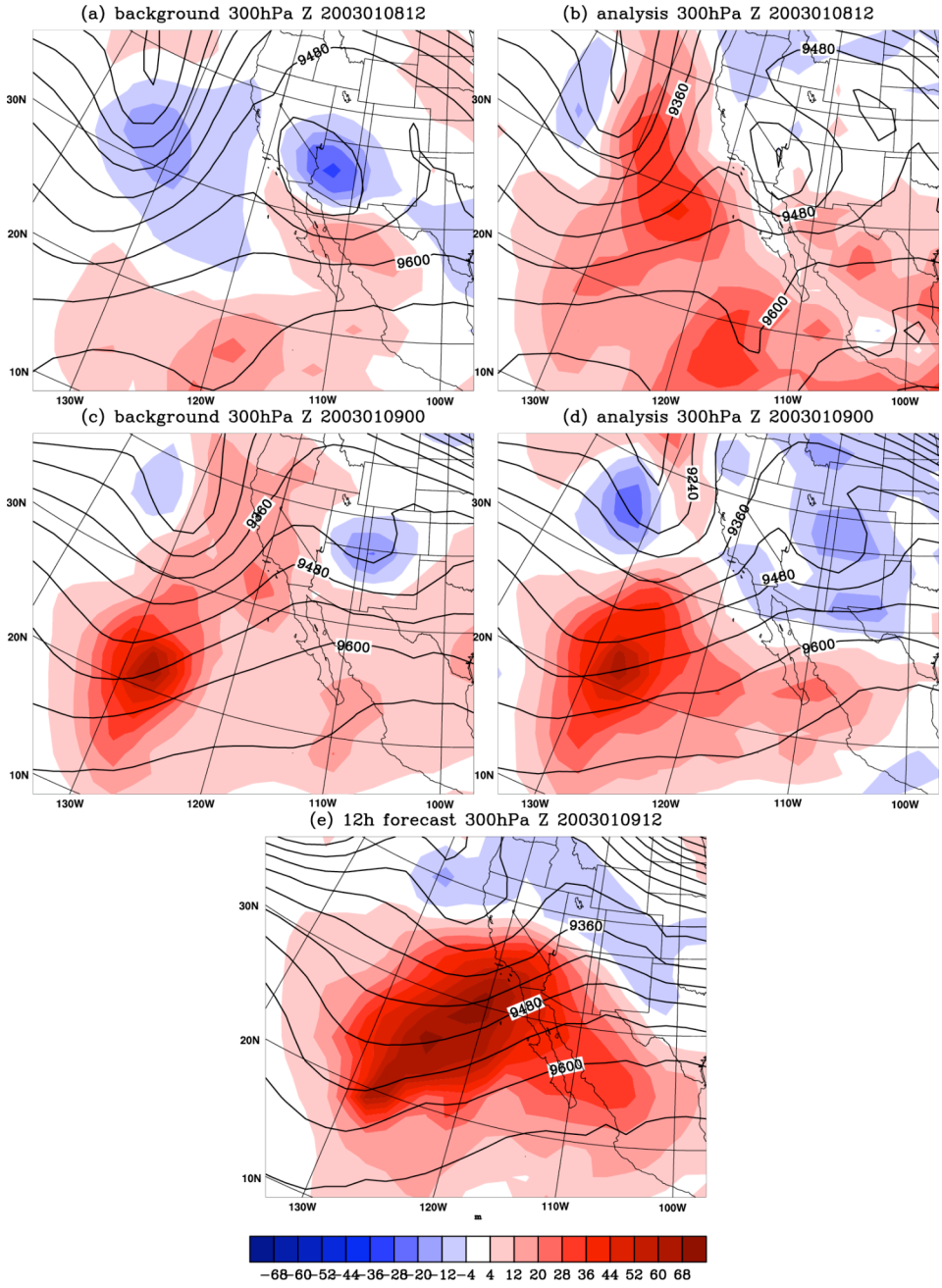

Fig. 9. The difference of the analysis and the first guess for the 300 hPa geopotential height between the 3DVAR and the hybrid from 1200 UTC 8 January to 1200 UTC 9 January 2003 every 12 hours. (a) difference of the background at 1200 UTC 8 January 2003. (b) difference of the analysis at 1200 UTC 8 January 2003. (c) difference of the background at 0000 UTC 9 January 2003. (d) difference of the analysis at 0000 UTC 9 January 2003. (e) difference of the 12-hour forecast at 1200 UTC 9 January 2003. Black contours are the 300 hPa geopotential height (m) for the hybrid. Color shades are the difference (3DVAR minus hybrid) of the 300 hPa geopotential height (m).

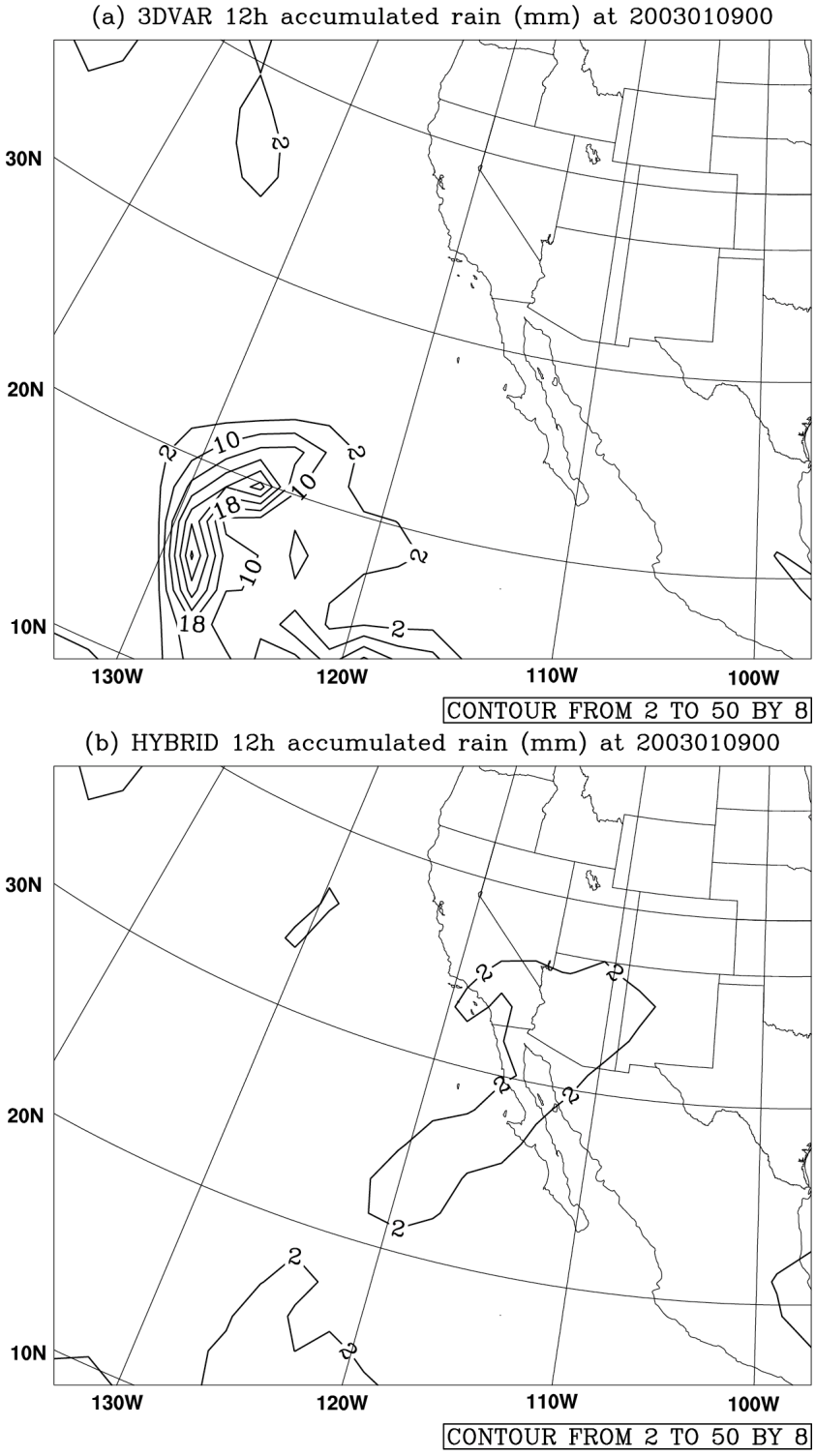

Fig. 10 12-hour forecasts of the accumulated precipitation valid at 0000 UTC 9 January 2003 for (a) 3DVAR and (b) hybrid.

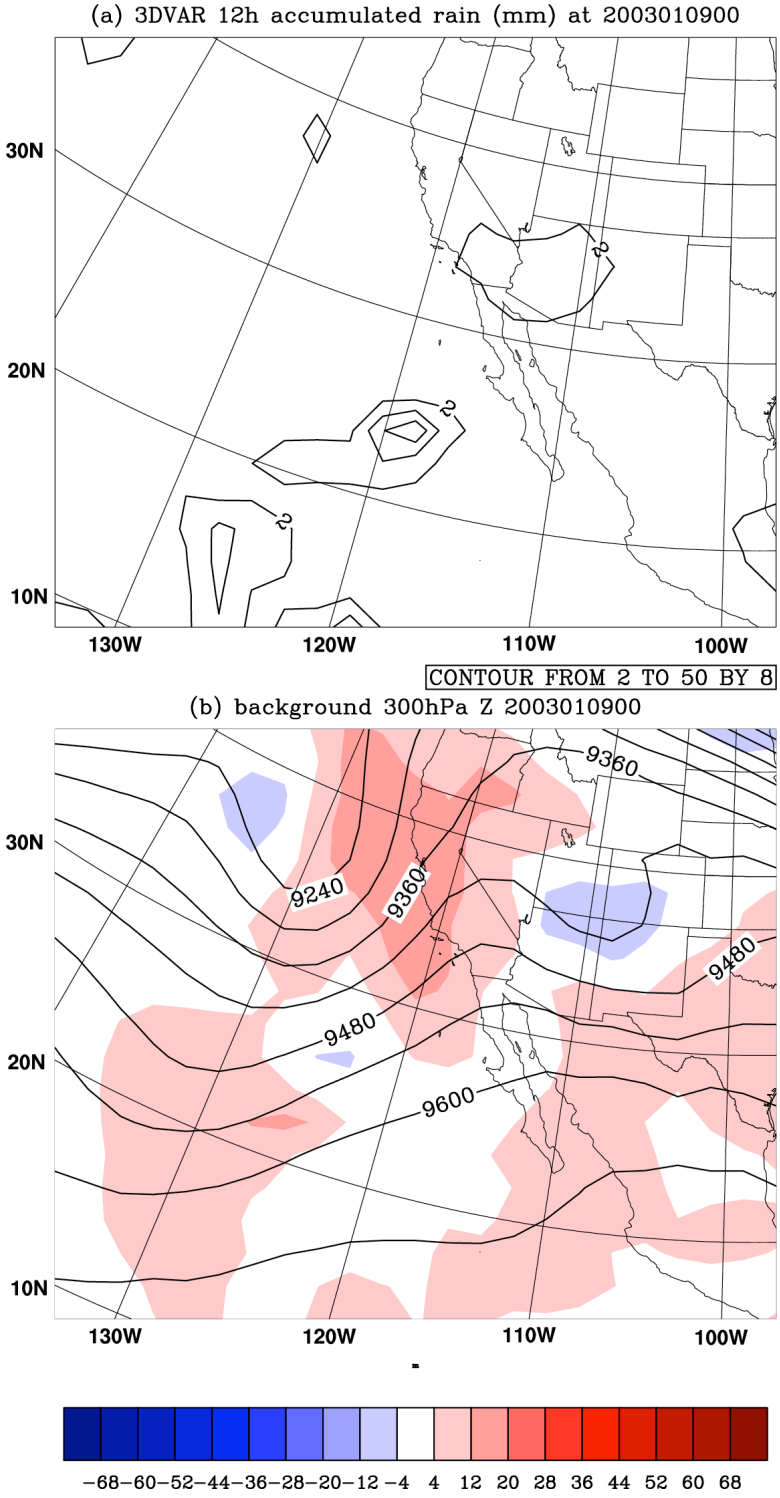

Fig. 11. (a) same as Fig. 10a, and (b) same as Fig. 9c, except that plots represent the forecast simulations where the 3DVAR water vapor mixing ratio analysis at 0000 UTC 9 January 2003 were replaced with that of the hybrid.

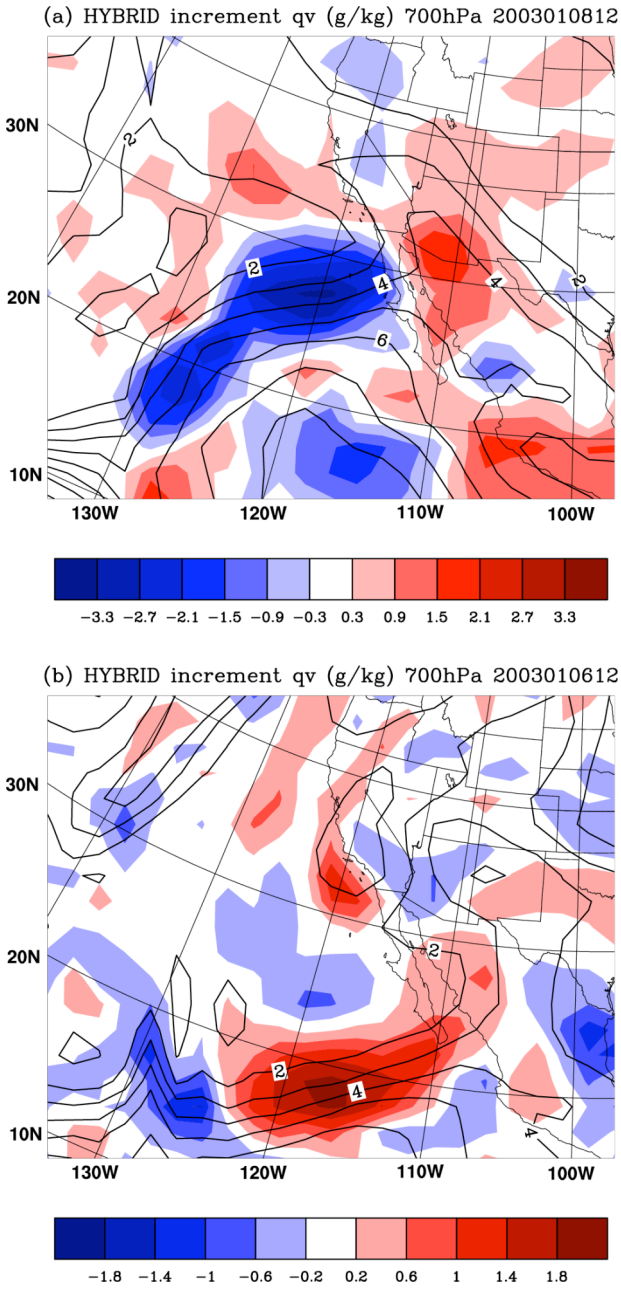

Fig. 12. 700 hPa water vapor mixing ratio increment by the hybrid at (a) 1200 UTC 8 January 2003 and (b) 1200 UTC 6 January 2003. Black lines are the first guess of 700 hPa water vapor mixing ratio. Color shades are the increment of the mixing ratio by assimilating radiosonde wind and temperature observations.