

## Tests of an Ensemble Kalman Filter for Mesoscale and Regional-Scale Data Assimilation. Part IV: Comparison with 3DVAR in a Month-Long Experiment

ZHIYONG MENG\* AND FUQING ZHANG<sup>+</sup>

*Department of Atmospheric Sciences, Texas A&M University, College Station, Texas*

(Manuscript received 30 May 2007, in final form 29 January 2008)

### ABSTRACT

In previous works in this series study, an ensemble Kalman filter (EnKF) was demonstrated to be promising for mesoscale and regional-scale data assimilation in increasingly realistic environments. Parts I and II examined the performance of the EnKF by assimilating simulated observations under both perfect- and imperfect-model assumptions. Part III explored the application of the EnKF to a real-data case study in comparison to a three-dimensional variational data assimilation (3DVAR) method in the Weather Research and Forecasting (WRF) model. The current study extends the single-case real-data experiments over a period of 1 month to examine the long-term performance and comparison of both methods at the regional scales. It is found that the EnKF systematically outperforms 3DVAR for the 1-month period of interest in which both methods assimilate the same standard rawinsonde observations every 12 h over the central United States. Consistent with results from the real-data case study of Part III, the EnKF can benefit from using a multischeme ensemble that partially accounts for model errors in physical parameterizations. The benefit of using a multischeme ensemble (over a single-scheme ensemble) is more pronounced in the thermodynamic variables (including temperature and moisture) than in the wind fields. On average, the EnKF analyses lead to more accurate forecasts than the 3DVAR analyses when they are used to initialize 60 consecutive, deterministic 60-h forecast experiments for the month. Results also show that deterministic forecasts of up to 60 h initiated from the EnKF analyses consistently outperform the WRF forecasts initiated from the National Centers for Environmental Prediction final analysis field of the Global Forecast System.

### 1. Introduction

The ensemble Kalman filter (EnKF; Evensen 1994), which estimates the background error covariance with a short-term ensemble forecast, is drawing increasing attention in the data assimilation community (Evensen 2003; Lorenc 2003; Hamill 2006; Zhang and Snyder 2007). Progress has lately been made toward implementing ensemble-based data assimilation into weather prediction models with real observations (Dowell et al. 2004; Whitaker et al. 2004, 2008; Barker 2005; Houteka-

mer et al. 2005; Hakim and Torn 2008; Fujita et al. 2007).

Application of the EnKF in real-world data assimilation with large-scale models has made significant progress. Houtekamer et al. (2005) implemented the EnKF into a large-scale model that includes a standard operational set of physical parameterizations by assimilating real observations from a fairly complete observational network. The 6-h data assimilation cycles were conducted for about half a month, and model error was treated by additive random perturbations sampled from the background error statistics (BES) of a simplified version of their three-dimensional variational data assimilation (3DVAR) method. The performance of the EnKF was found to be comparable to that of the 3DVAR system. Whitaker et al. (2004) and Compo et al. (2006) demonstrated that the EnKF is well suited to historical reanalysis problems and outperforms 3DVAR. Whitaker et al. (2008) further implemented the EnKF in a reduced-resolution version of the Global Forecast System (GFS) at the National Centers for En-

---

\* Current affiliation: Department of Atmospheric Sciences, School of Physics, Peking University, Beijing, China.

<sup>+</sup> Current affiliation: Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania.

---

*Corresponding author address:* Dr. Fuqing Zhang, Department of Meteorology, The Pennsylvania State University, 503 Walker Bldg., University Park, PA 16802.  
E-mail: fzhang@psu.edu

vironmental Prediction (NCEP) with all of the real-time observations used by the NCEP operational data assimilation system except for satellite radiance. Their 40-day, 6-hourly data assimilation experiments show that the EnKF works better than the operational 3DVAR system configured in the same way. The 48-h forecast from the ensemble data assimilation system was as accurate as the 24-h forecast from the 3DVAR system.

In previous parts of this study series, a mesoscale, ensemble-based data assimilation system has been shown to be promising in increasingly realistic environments. First, the EnKF was used to simulate observations both under a perfect-model assumption (Zhang et al. 2006, hereafter Part I) and in the presence of significant model error (Meng and Zhang 2007, hereafter Part II). The EnKF was then used in a mesoscale convective vortex (MCV) case study of real-world data assimilation (Meng and Zhang 2008, hereinafter Part III). As a natural extension of the case study involving real-world data assimilation with the EnKF implemented in the Weather Research and Forecasting (WRF) Model, a month-long rawinsonde data assimilation experiment is performed in this study to compare the behavior of the EnKF in a more general manner to WRF-3DVAR. Model error is treated using the “multischeme” ensemble (see Parts II and III) and the relaxation method for covariance inflation [Zhang et al. 2004, their Eq. (5)]. Different from the standard inflation method (Anderson 2001) in which all points in the prior field are inflated, this relaxation method only inflates the covariance at updated points via a weighted average between the prior perturbation and the posterior perturbation.

The rest of this work is presented as follows: The methodology is briefly introduced in section 2. The realism of the EnKF is examined in section 3. Section 4 compares the performance of the EnKF with 3DVAR and to the NCEP GFS final (FNL) analysis. The benefit of using a multischeme ensemble is further demonstrated in section 5 via comparison to a single-scheme experiment. Conclusions and discussions are presented in section 6.

## 2. Methodology

### a. The model

The same version of the Advanced Research WRF (ARW; Skamarock et al. 2005) with two domains and one-way nesting is used in this study as in Part III. The coarse domain covers the contiguous United States with  $45 \times 64$  grid points and a grid spacing of 90 km (Fig. 1), which is the same as that in Part III. The inner

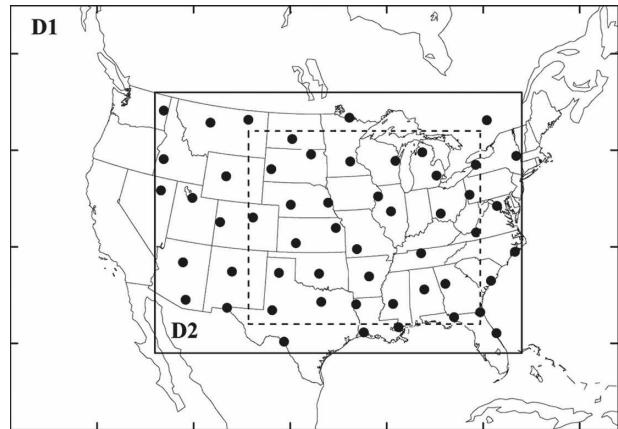


FIG. 1. Map of model domain. The solid circles denote the rawinsonde observations to be assimilated. The dashed box shows where the verification is performed.

domain is extended from covering only the central United States in Part III (the dashed box in Fig. 1) to encompass most of the continental United States with  $115 \times 82$  grid points and a grid spacing of 30 km (solid box in Fig. 1). Both model domains have 27 vertical layers, and the model top is set at 100 hPa. Except for the multischeme ensemble runs (refer to Table 2 of Part III), all forecasts use the Grell–Devenyi cumulus scheme (Grell and Devenyi 2002), WRF single-moment (WSM) six-class microphysics with graupel (Hong et al. 2004), and the Yonsei State University (YSU) scheme (Noh et al. 2003) for planetary boundary layer processes. The NCEP FNL analyses are used to create the initial and boundary conditions.

### b. The EnKF and 3DVAR

The EnKF and 3DVAR methods used in this study are the same as those in Part III (refer to sections 2b and 2c of Part III for details). The EnKF uses an ensemble size of 40 and the covariance relaxation method [Zhang et al. 2004, their Eq. (5)] to inflate the analysis error covariance with a relaxation coefficient of 0.7 (which means 70% of the perturbation of the analysis ensemble comes directly from the prior), a value that gives the best performance in the real-data case study of Part III. WRF-3DVAR (Barker et al. 2004) uses the newly generated  $CV = 5$  (CV5) background error statistics and the National Meteorological Center (NMC) method (Parrish and Derber 1992; Xiao and Sun 2007). This method uses statistics describing differences between the sixty 24- and 12-h WRF forecasts valid at the same time, performed on the fine domain, and initialized twice daily (0000 and 1200 UTC) for the entire month of May 2003. The control variables are streamfunction, velocity potential, unbalanced pressure, and

relative humidity (Xiao and Sun 2007). The CV5 option is used for the control 3DVAR experiment because the BES are based on the WRF forecast difference for the preceding month, which is demonstrated to perform slightly better than using the default BES option (CV3) both for the case study (Part III) and for the month-long experiment (not shown). Both data assimilation methods are performed only in domain 2.

#### 1) DATA TO BE ASSIMILATED

For this month-long application, only standard rawinsonde observations available every 12 h are assimilated. Part III showed that only marginal benefit might be gained through assimilation of additional data from surface and profiler observations. As in Part III, the Gaspari and Cohn (1999) fifth-order correlation function with a radius of influence of 30 grid points (i.e., 900 km) in the horizontal direction and 15 sigma levels in the vertical direction is used for covariance localization (based on the MCV case study examined in Part III). Rawinsonde observation errors are height-dependent as defaulted in WRF-3DVAR (Table 1 of Part III).

The rawinsonde observations of June 2003 go through a quality control procedure similar to that described in Part III. In detail, the quality control contains two steps—basic quality control through the 3DVAR preprocessor and so-called “errormax” checking through the prerun of 3DVAR—to guarantee that 3DVAR and the EnKF use exactly the same observations. First, 12-hourly rawinsonde observations are processed by the observation preprocessor of WRF-3DVAR for basic quality control such as range, domain, persistency, extreme-value, and buddy (horizontal consistency) checks. Then the FNL analysis fields at 12-h intervals are interpolated to the WRF model grid and used as the first guess in a prerun of WRF-3DVAR to assimilate the observations that have gone through the basic quality control procedure. In the prerun, an errormax quality control is performed by dropping the observations whose absolute differences from the first guess are more than 5 times larger than the corresponding observation errors. The observations that pass the second quality control procedure will then be assimilated in the following 3DVAR and EnKF experiments. In the output file containing the ingested observations, the original wind speed, wind direction, and relative humidity are transformed into wind components ( $u$  and  $v$ ) and the mixing ratio of water vapor ( $q$ ). Consequently, the assimilated variables from rawinsondes become  $u$ ,  $v$ ,  $q$ , and temperature ( $T$ ). The assumed observation error of rawinsondes is the default in the WRF-3DVAR system given in Table 1 of Part III. Observations from 53 rawinsondes within the inner

domain are assimilated every 12 h with locations denoted as solid circles in Fig. 1.

#### 2) ENSEMBLE INITIAL AND BOUNDARY CONDITIONS

The initial reference analyses of both the coarse and inner domains are interpolated from the FNL analyses. The initial ensemble perturbations are generated with balanced perturbations randomly drawn from the WRF-3DVAR background uncertainty as described in section 2b of Part III (refer also to Barker et al. 2004; Barker 2005), which is also used in Houtekamer et al. (2005). For the subsequent analysis and forecast cycles, the initial and boundary conditions of the coarse domain are regenerated every 12 h using FNL analyses that have been perturbed with the 3DVAR background uncertainties, whereas the initial conditions of the inner domain are recycled from the posterior of the EnKF. Data assimilation is only performed on the 30-km inner domain. The boundary conditions are provided by the coarse domain ensemble through one-way nesting, as if they were provided by a global ensemble forecast system with the same forecast model as the inner domain. Moreover, centering the coarse domain ensemble with the FNL analyses every 12 h can prevent the model from drifting away from the FNL analyses.

The relative importance of initial and boundary perturbations on the 12-h forecast error covariance is assessed in a series of 12-h forecast experiments wherein either the initial or boundary conditions are perturbed. It is found that the forecast ensemble spread due to boundary and initial perturbations on the 12-h background error covariance is approximately similar over the verification domain (Fig. 1).

#### c. Experiment design

The control EnKF experiment for this month-long test uses a multischeme ensemble (hereafter also referred to as EnKF\_m; Table 1) with an ensemble size of 40 because the case study in Part III found that a multischeme ensemble performs better than a single-scheme ensemble. The multischeme ensemble uses a combination of three different microphysics parameterizations, three different cumulus parameterizations, and three different boundary layer parameterization schemes, the detailed configuration of which is listed in Table 2 of Part III. EnKF\_m is initiated at 0000 UTC 1 June 2003, the first observations are assimilated at 1200 UTC 1 June, and the analysis and forecast cycles continue for the entire month of June until 1200 UTC 30 June 2003.

In comparison with EnKF\_m, standard WRF-3DVAR (hereinafter also referred to as 3DVAR\_WRF—note

TABLE 1. List of primary assimilation and forecast experiments.

Experiment	Description
EnKF_m	The WRF ensemble for the EnKF uses different combinations of physical parameterization schemes for different members (refer to Table 2 in Part III for detailed configuration).
EnKF_s	The WRF ensemble for the EnKF uses the same combination of physical parameterization schemes for all members.
3DVAR_WRF	The control WRF-3DVAR experiment with the CV5 BES option. The WRF forecast for the prior uses the same physical parameterizations as in EnKF_s.
FNL_GFS	The 12-hourly WRF forecasts with the same configuration as in 3DVAR_WRF initializing from GFS FNL analyses interpolated to WRF grid.
EnKF_mDF	The 12-hourly single deterministic WRF forecasts with the same configuration as in 3DVAR_WRF initializing from posterior of EnKF_m
EnKF_sDF	The 12-hourly single deterministic WRF forecasts with the same configuration as in 3DVAR_WRF initializing from posterior of EnKF_s.

that 3DVAR\_WRF denotes this particular experiment, whereas WRF-3DVAR denotes the 3DVAR system in WRF; Table 1) data assimilation is performed using one model instead of an ensemble. 3DVAR\_WRF is started from the same time as EnKF\_m, and the first data are also assimilated after 12 h of integration. For 3DVAR\_WRF, the initial reference analysis of both domains and the subsequent initial and boundary conditions of the coarse domain regenerated every 12 h are all interpolated from the GFS FNL analysis.

EnKF\_m is also compared to the GFS FNL analyses—which are created with a data assimilation method similar to 3DVAR by assimilating the whole set of operational observations including satellite data (Parrish and Derber 1992)—verified at the same times and interpolated to the same WRF model grid (Fig. 1). It is further compared with subsequent 12-h deterministic WRF forecasts from the FNL analyses using the same model configuration as that in 3DVAR\_WRF. This experiment will be referred to as FNL\_GFS hereinafter (Table 1).

#### d. Verification metrics

Verification metrics used in this study are similar to those in Houtekamer et al. (2005) and Whitaker et al. (2008). All experiments are verified against rawinsondes that have passed the quality control procedure at nine standard pressure levels: 925, 850, 700, 500, 400, 300, 250, 200, and 150 hPa. To simplify the description, the RMS fit of the verified field to rawinsonde observations is referred to as RMSE, and the RMS ensemble standard deviation is referred to as RMS ensemble spread. The mean denotes an average over all verifying observations for the entire month of June 2003. In addition to the RMSE of individual variables, the root-mean-difference total energy (RM-DTE), which combines errors of  $u$ ,  $v$  and  $T$ , is also used for verification. The DTE is defined as

$$DTE = 0.5(u'u' + v'v' + kT'T'), \quad (1)$$

where the prime denotes the difference between the observations and the verified fields and  $k = C_p/T_r$  ( $C_p = 1004.7 \text{ J kg}^{-1} \text{ K}^{-1}$  and the reference temperature  $T_r = 290 \text{ K}$ ). Root-mean-difference kinetic energy (RM-DKE) is also used to denote wind error through combining errors from  $u$  and  $v$  by only using the first two items on the right-hand side of Eq. (1).

This study uses several metrics in addition to RMSE and RM-DTE. For example, bias is defined here as the mean difference between the observation and the prior interpolated to the observation location. Relative error reduction is used to describe the relative performance of data assimilation for different variables and different experiments. The relative improvement of  $A$  over  $B$  is defined as

$$(\text{RMSE}^B - \text{RMSE}^A)/\text{RMSE}^B \times 100\%. \quad (2)$$

To be consistent with Part III, all experiments are verified in the Bow Echo and Mesoscale Convective Vortex Experiment (BAMEX) field experiment domain (dashed box in Fig. 1; the same area as the inner domain in Part III). Verification in a smaller domain will reduce the influence of lateral boundary conditions, although qualitatively similar results are also obtained if verified over the entire nested domain 2.

### 3. Ensemble realism examination

In general terms, if an ensemble is sampling the distribution from which the observation is sampled, the ensemble can be regarded as being realistic. In this section, the realism of the ensemble is examined using rank histograms (Anderson 1996; Hamill and Colucci 1997). A rank histogram describes the extent to which an ensemble encompasses the verifying data by ranking the verifying data in the sorted ensemble. It describes the relative frequency at which a verifying observation

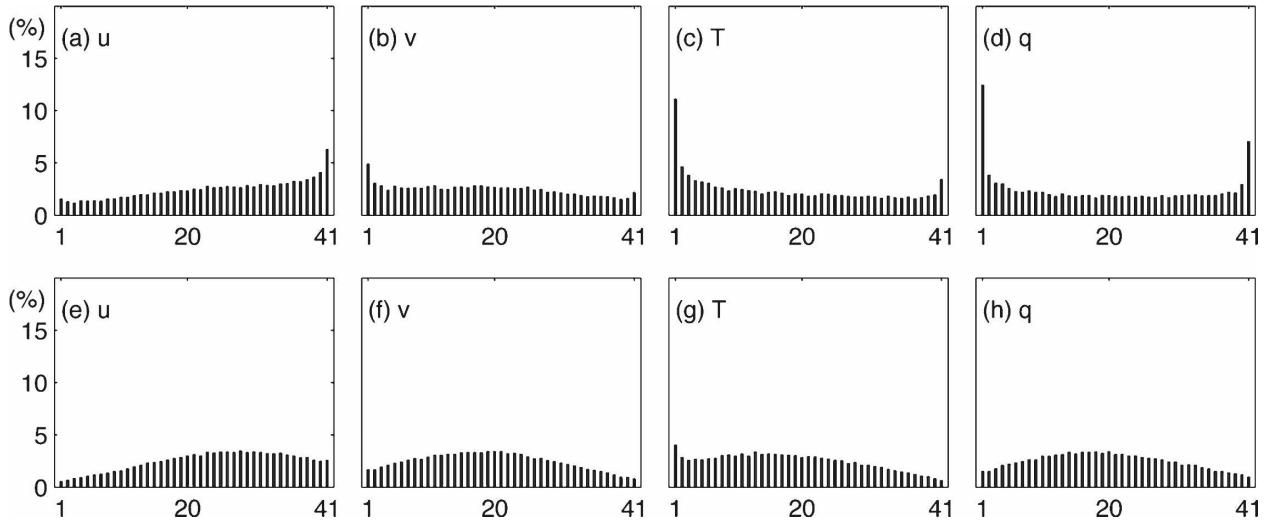


FIG. 2. Rank histogram for (a)  $u$ , (b)  $v$ , (c)  $T$ , and (d)  $q$  of the prior ensemble in EnKF\_m. The y coordinate denotes the relative frequency of the verifying observation. The x coordinate denotes the bins formed by the ensemble. (e)–(h) As in (a)–(d), but the observation error is accounted for.

falls into categories formed by a sorted ensemble. The reliability of an ensemble can be diagnosed by the shape of its rank histogram. A flat shape implies that the observation can be taken as a random member of the ensemble, and consequently the ensemble is reliable. A  $U$  shape suggests the ensemble spread is insufficient, while a reversed (upside down)  $U$  shape indicates that the ensemble spread is too large.

According to Hou et al. (2001), a histogram can be regarded as fairly flat if the adjusted missing rate is lower than 10%. To calculate the adjusted missing rate, one must first calculate the missing rate, which is the sum of the relative frequencies of the two extreme (the first and last) categories. Zhu et al. (1996) defines the adjusted missing rate as the difference between the expected missing rate [ $2/(N + 1)$ , where  $N$  is the ensemble size] and the missing rate. This gives a generalized metric for the realism of ensembles with different sizes. The rank histogram can also be used to examine the bias of an ensemble. If an ensemble has a positive (negative) bias, the relative frequency of the verifying observations will be shifted toward higher (lower) categories.

The histogram created with all of the observations within the verifying domain during the month without considering observation error (Figs. 2a–d) shows that the prior ensemble of EnKF\_m is generally reliable. The rank histogram of  $q$  is based on observations lower than 300 hPa because the mixing ratio of water vapor above 300 hPa is very small, and the ensemble spread is too small for the rank histogram to be meaningful. For an ensemble with 40 members, the expected missing rate is 5%, and thus the adjusted missing rates are less

than 10%. Wind components  $u$  and  $v$  have reasonable ensemble spread, whereas  $T$  and  $q$  are slightly insufficient in their ensemble spread (Figs. 2a–d). When observation error is accounted for, all frequencies are less than 5% (Figs. 2e–h). This result shows that the ensemble is reasonable, though the ensemble spread is slightly overestimated because of slightly overestimated observation error or prior ensemble spread.

Assuming the observation error is independent of forecast error and the model is perfect, the optimality of the ensemble requires the innovation covariance equal to the sum of the background error covariance and observation error covariance (Dee 1995; Houtekamer et al. 2005; Whitaker et al. 2008):

$$\langle (\mathbf{y}^o - \mathbf{H}\bar{\mathbf{x}}^b)(\mathbf{y}^o - \mathbf{H}\bar{\mathbf{x}}^b)^T \rangle = \mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}, \quad (3)$$

where  $\mathbf{y}^o$  denotes observation,  $\mathbf{H}$  denotes observation operator,  $\bar{\mathbf{x}}^b$  denotes prior forecast,  $\mathbf{T}$  denotes the transpose of a matrix,  $\mathbf{P}^b$  denotes background error covariance, and  $\mathbf{R}$  denotes observation error covariance. This metric of ensemble realism has also been used in the presence of model error (Houtekamer et al. 2005; Whitaker et al. 2008; Torn and Hakim 2008). As in the above-mentioned studies, here only the diagonals of both sides of Eq. (3) are considered. For this criterion, if the innovation variance (or prior RMSE) is similar to the sum of background and observation error variance (or predicted RMSE; see Houtekamer et al. 2005), the prior ensemble can be regarded as generally consistent with the verifying observation.

In this experiment, the month-averaged predicted RMSE (black dotted–dashed lines in Fig. 3) is generally

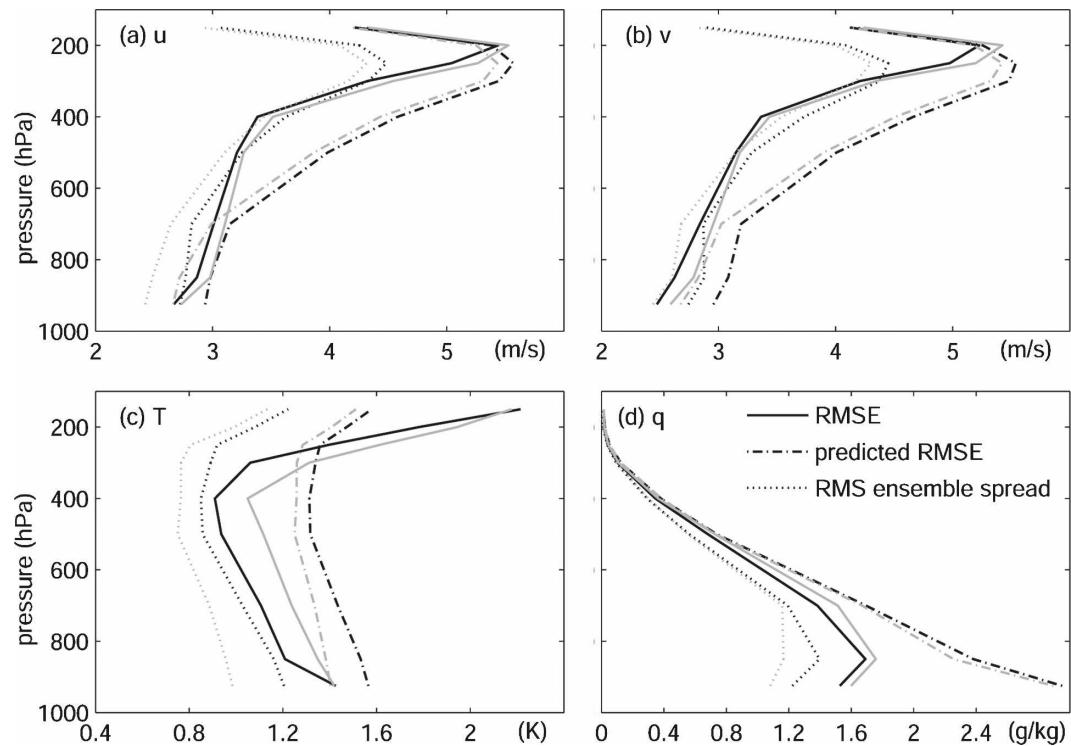


FIG. 3. Vertical distribution of the month-averaged RMSE (solid), predicted RMSE (dotted–dashed), and RMS ensemble spread (dotted) of the prior for (a)  $u$ , (b)  $v$ , (c)  $T$ , and (d)  $q$  in EnKF\_m (black) and EnKF\_s (gray).

larger than the prior RMSE (black solid lines in Fig. 3), especially below 300 hPa. The maximum difference between the prior and predicted RMSE is  $1 \text{ m s}^{-1}$  for  $u$  and  $v$  and  $0.4 \text{ K}$  for  $T$ . Smaller predicted RMSE is observed above 300 hPa in  $T$  with a maximum magnitude of about  $0.8 \text{ K}$ . This result is comparable to other large-scale studies (Houtekamer et al. 2005; Whitaker et al. 2008). Houtekamer et al. (2005) demonstrates a closer match between both sides of Eq. (3), especially in the lower troposphere, probably because they used a simplified version of 3DVAR covariance to represent the model error. In Whitaker et al. (2008), larger deficiencies in the ensemble spread are observed at upper and lower levels. The deficiency at upper levels in our study is similar in magnitude to that observed in Whitaker et al. (2008). The relatively larger predicted RMSE in this study is consistent with the slightly reversed  $U$  shape of the histogram shown in Figs. 2e–h and could be caused by slight overestimation of both the prior ensemble spread and the observation error, the latter being the default in the formally released WRF-3DVAR system (Xiao and Sun 2007).

Two additional EnKF experiments are performed (for the first 5 days only) to further examine the discrepancy between the predicted and actual prior RMSE. One decreases the prior ensemble spread with-

out covariance relaxation; the other halves the assumed observation error defaulted in WRF-3DVAR. The predicted RMSE in both experiments has a closer match to the prior RMSE, but the prior RMSE in both experiments is larger than that of EnKF\_m over the same period (not shown). This suggests that a better match between predicted and actual prior RMSE may not necessarily lead to better filter performance in the real-data case.

#### 4. Comparison between the EnKF and 3DVAR

##### a. Comparison of standard prior and posterior

Figures 4–7 show the comparison of the performance of EnKF with WRF-3DVAR over this 1-month period. EnKF\_m in general has consistently lower RMSE in both priors (Fig. 4) and posteriors (Fig. 5) for both winds and thermodynamic variables (except for comparable RMSE for the posterior of  $q$ ). EnKF\_m also generally has smaller posterior and prior RMSEs than those of both the interpolated FNL analysis and a subsequent 12-h WRF forecast from the interpolated analysis (FNL\_GFS; Figs. 4 and 5). Relative to 3DVAR\_WRF, FNL\_GFS has smaller prior RMSEs but larger posterior RMSEs for  $T$  and  $q$  and comparable posterior RMSEs for wind components. Further-

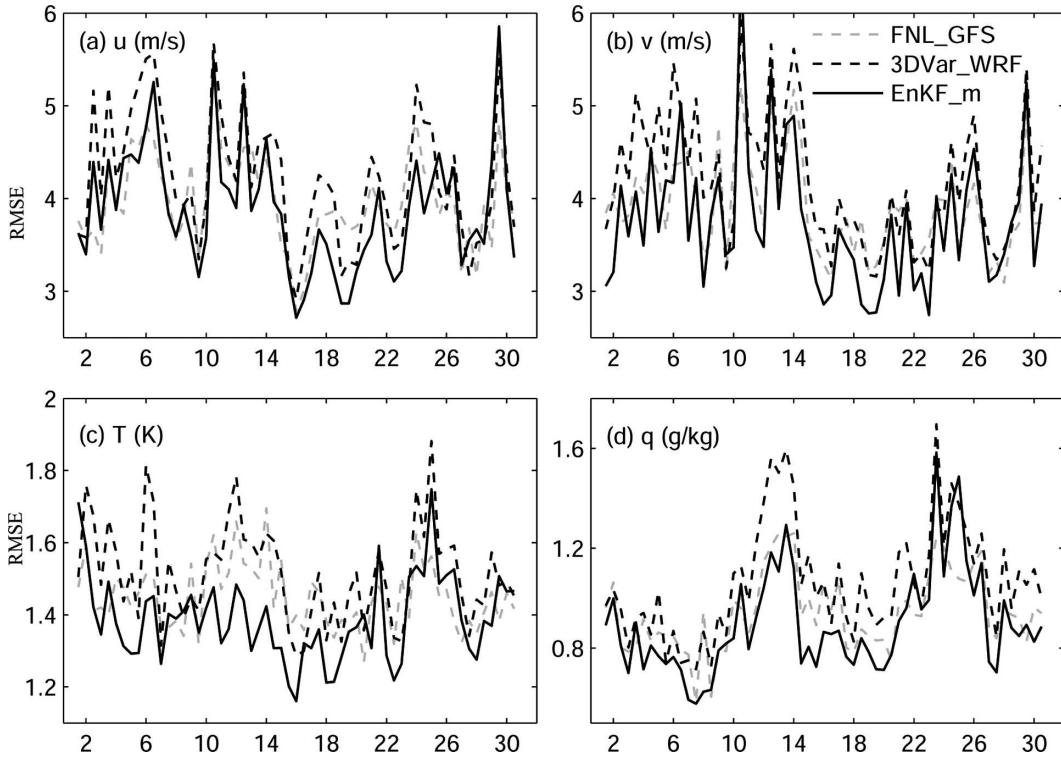


FIG. 4. Time evolution of domain-averaged RMSE of the 12-hourly prior of (a)  $u$ , (b)  $v$ , (c)  $T$ , and (d)  $q$  for EnKF\_m (black solid), 3DVAR\_WRF (black dashed), and FNL\_GFS (gray dashed).

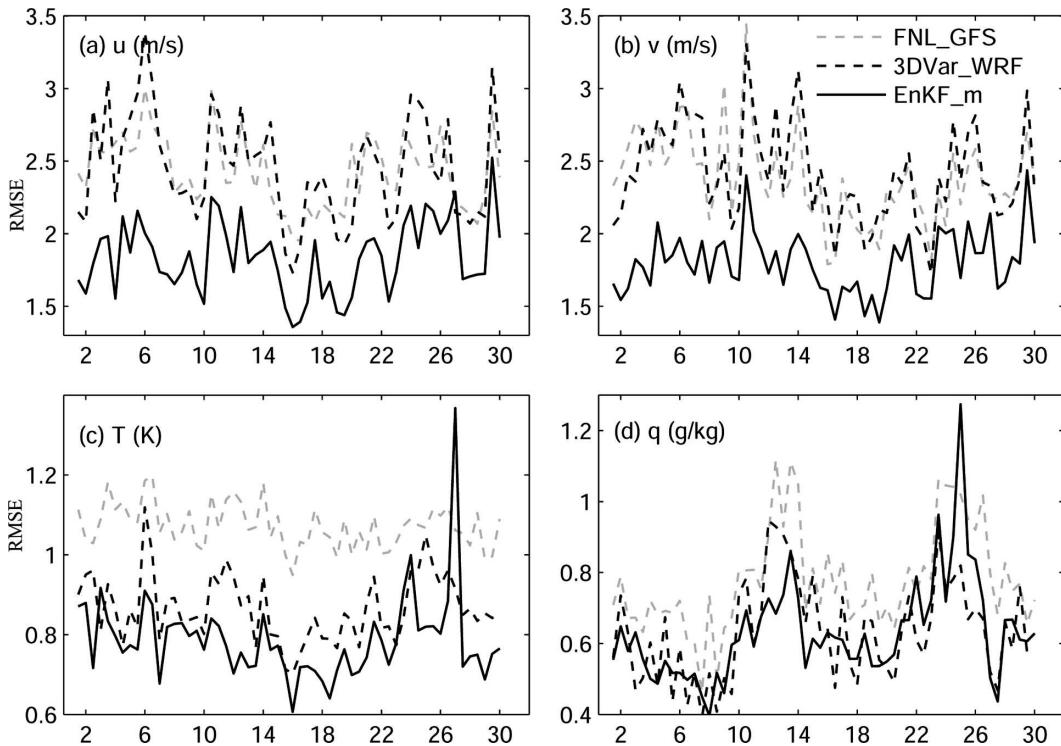


FIG. 5. As in Fig. 4, but for 12-hourly posterior.

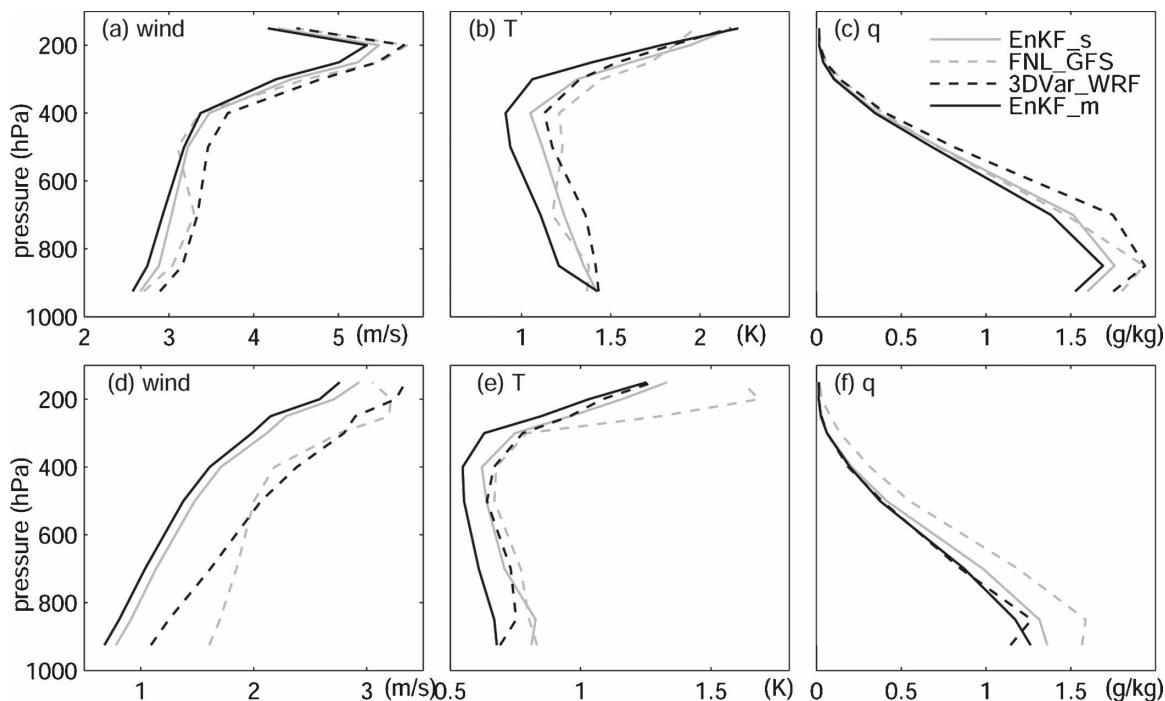


FIG. 6. Vertical distribution of the month-averaged (a)–(c) prior and (d)–(f) posterior RMSE of horizontal wind,  $T$ , and  $q$  for EnKF\_m (black solid), 3DVAR\_WRF (black dashed), FNL\_GFS (gray dashed), and EnKF\_s (gray solid).

more, Fig. 6 shows that EnKF\_m also consistently outperforms 3DVAR\_WRF and FNL\_GFS in the month-averaged vertical distributions of prior and posterior RMSE. The overall domain- and monthly-averaged prior RM-DTE of EnKF\_m, 3DVAR\_WRF, and FNL\_GFS is 4.26, 4.7, and 4.61  $\text{m s}^{-1}$  respectively (Fig. 7). The confidence level of these RMSE differences between experiments always exceeds 99%.

In terms of prior RM-DTE, EnKF\_m outperforms 3DVAR\_WRF by 9%. The degree of improvement of EnKF\_m over 3DVAR\_WRF is different for different variables. The domain-averaged prior RMSE of EnKF\_m for horizontal wind,  $T$ , and  $q$  is respectively 9%, 8%, and 16% smaller than that of 3DVAR\_WRF (Figs. 6a–c). Figures 6d–f show consistently better posterior of EnKF\_m relative to that of 3DVAR\_WRF and FNL\_GFS. In terms of posterior RMSE, EnKF\_m outperforms 3DVAR by about 25% in wind (Fig. 6d) and 9% in  $T$ , while the RMSE of  $q$  is similar in the experiments. 3DVAR\_WRF demonstrates smaller posterior error than FNL\_GFS in  $T$  and  $q$ , and they are comparable in winds.

Forecast bias, which is primarily caused by model error from deficient physics parameterizations (Whitaker et al. 2008) and can affect the performance of a data assimilation method, is significantly smaller than the corresponding RMSE for all experiments including

EnKF\_m, 3DVAR\_WRF, and FNL\_GFS, with a confidence level exceeding 99% (Fig. 8). Among the variables,  $u$  has consistently positive bias (Fig. 8a),  $v$  has negative bias (Fig. 8b), and  $T$  has negative (positive) bias above (below) 800 hPa, with the maximum bias

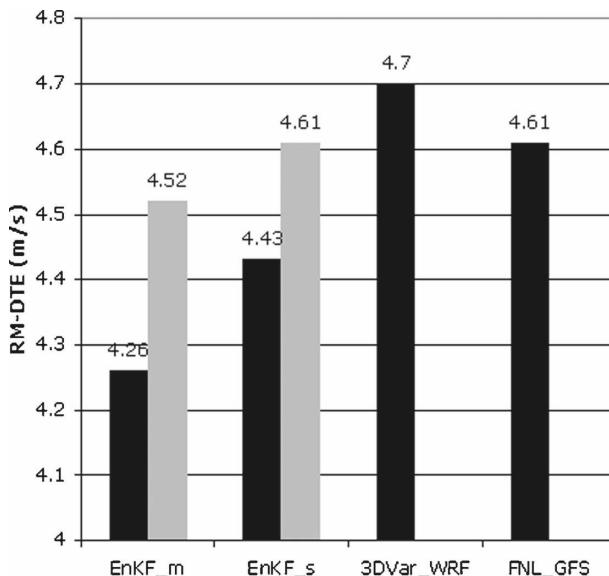


FIG. 7. Absolute 12-h forecast error of different experiments in terms of RM-DTE. The gray bars are the respective RM-DTE of EnKF\_mDF and EnKF\_sDF.

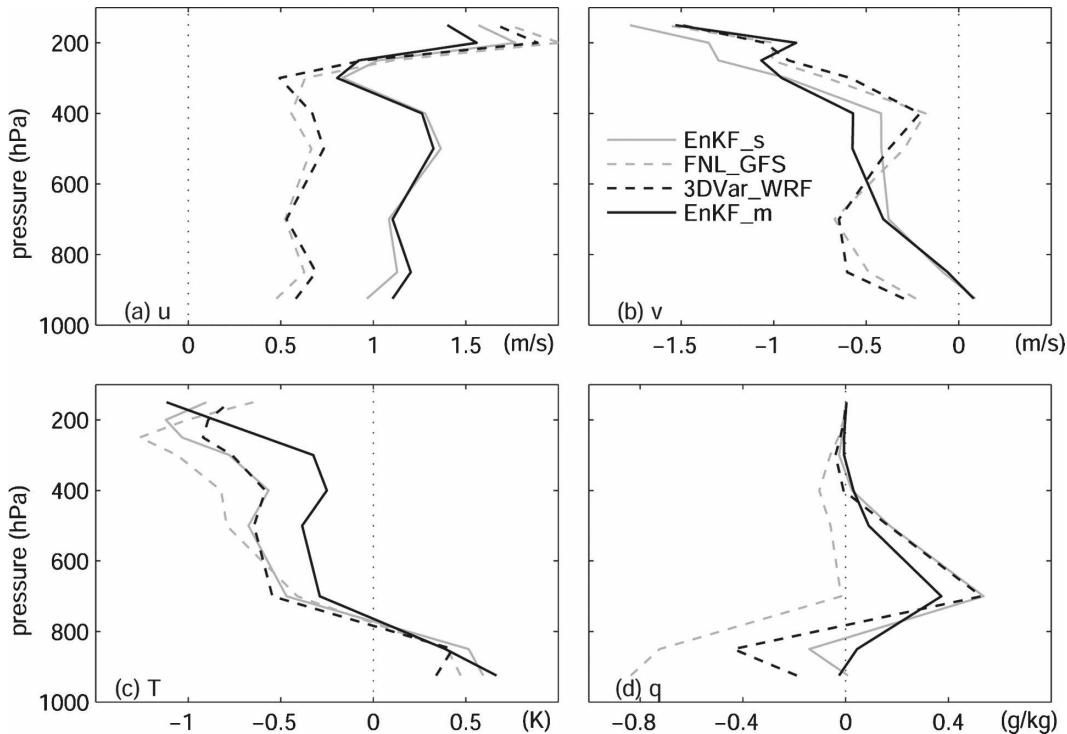


FIG. 8. Vertical distribution of the month-averaged bias of prior of (a)  $u$ , (b)  $v$ , (c)  $T$ , and (d)  $q$  for EnKF\_m (black solid), 3DVAR\_WRF (black dashed), FNL\_GFS (gray dashed), and EnKF\_s (gray solid). The zero line is plotted as a dotted line for reference.

around 200 hPa (Fig. 8c). For  $q$ , 3DVAR\_WRF shows mixed results, whereas EnKF\_m (FNL\_GFS) shows generally positive (negative) bias. EnKF\_m has generally smaller bias than 3DVAR\_WRF and FNL\_GFS except for  $u$ . In comparison with 3DVAR\_WRF, FNL\_GFS has a similar bias for wind components but a larger bias for  $T$  and  $q$ .

#### b. Comparison of deterministic forecasts starting from posteriors

In this section, 60-h forecasts starting from 12-hourly analysis fields of all three experiments are performed. Instead of ensemble forecasts, 60-h single, deterministic forecasts are carried out starting from the 12-hourly posterior (mean) analyses of EnKF\_m for domain 2. Domain 1 uses the same regenerated initial and boundary conditions interpolated from FNL analysis as used in 3DVAR\_WRF and FNL\_GFS. This experiment is referred to as EnKF\_mDF (Table 1).

It is found that the 12-h-forecast RM-DTE in EnKF\_mDF is larger than in priors of EnKF\_m (which is the mean of the ensemble forecast), but the RM-DTE in EnKF\_mDF is still noticeably smaller than that in 3DVAR\_WRF and FNL\_GFS (Fig. 7). The RMSE of wind,  $T$ , and  $q$  at different output times (Fig. 9) shows

that EnKF\_mDF has generally smaller errors than both 3DVAR\_WRF and FNL\_GFS, especially in  $T$  and  $q$ . This result further demonstrates that improvement of the EnKF over 3DVAR may have come from a better prior estimate using the ensemble mean and/or the flow-dependent background error covariance. Nevertheless, a large portion of the improvement appears to have come from using an ensemble mean, some of which may be due to the ensemble smoothing effect discussed in Part III (their Fig. 9).

Figure 9 also shows that FNL\_GFS has a consistently smaller error in  $q$  than 3DVAR\_WRF. For horizontal wind and  $T$ , FNL\_GFS has slightly smaller errors at early stages, but it has larger errors than 3DVAR\_WRF after about 24 h of integration.

#### 5. Comparison between multi- and single-scheme ensembles in the EnKF

To further investigate the impact of using multiple physical schemes in an ensemble to account for model uncertainty from model physical parameterizations, experiment EnKF\_s (Table 1) uses the same set of (single-scheme) physical parameterizations as in 3DVAR\_WRF to provide a comparison with the use of the multischeme ensemble in EnKF\_m.

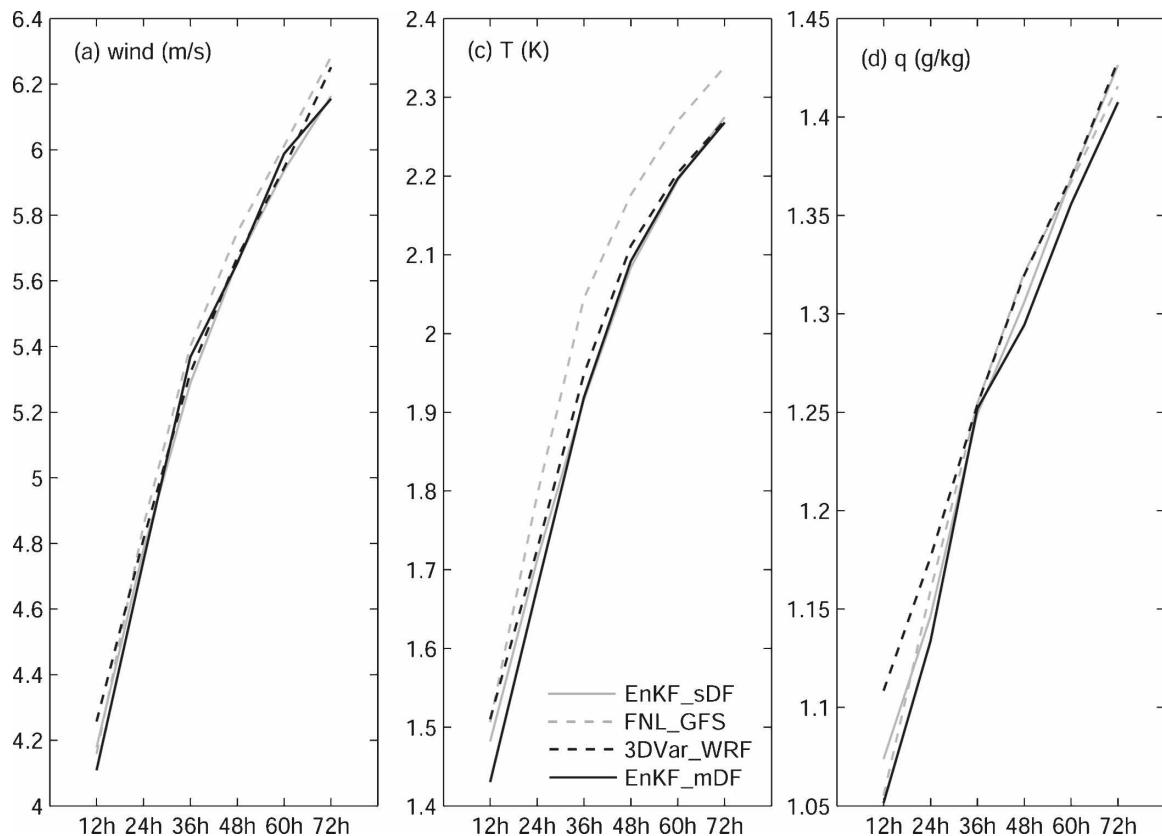


FIG. 9. Month- and domain-averaged forecast RMSE of (a) wind, (b)  $T$ , and (c)  $q$  for EnKF\_sDF (gray solid), FNL\_GFS (gray dashed), 3DVAR\_WRF (black dashed), and EnKF\_mDF (black solid).

EnKF\_s has a larger RMSE (gray versus black lines in Fig. 3), smaller RMS ensemble spread (dotted lines in Fig. 3), and consequently smaller predicted RMSE (dotted-dashed lines in Fig. 3) than EnKF\_m in each layer for  $u$ ,  $v$ ,  $T$ , and  $q$ . The prior RMSE of EnKF\_m is 3%, 3%, 8%, and 6% smaller than that of EnKF\_s for  $u$ ,  $v$ ,  $T$ , and  $q$ , respectively (Figs. 6a–c), and the improvement of EnKF\_m over EnKF\_s in  $T$  and  $q$  could be partially explained by its smaller bias (Fig. 8). This result indicates that a better match between the innovation variance and predicted RMSE (defined in section 3), as shown in EnKF\_s, does not necessarily lead to a smaller prior RMSE under the real-data applications when there is apparent model error. The relatively larger predicted RMSE in EnKF\_m, which is due to its larger RMS ensemble spread, could help to decrease the prior RMSE. Finally, EnKF\_s increases the prior RM-DTE by  $0.2 \text{ m s}^{-1}$  (Fig. 7) relative to EnKF\_m.

An additional experiment with the relaxation coefficient increased from 0.7 to 0.75 in EnKF\_s increases the ensemble spread of EnKF\_s to be similar to EnKF\_m, but it also slightly increases the prior RMSE. This ex-

periment suggests that the performance of the single-scheme EnKF in terms of RMSE is close to being optimum when the relaxation coefficient is around 0.7 for the current study, and the better performance of EnKF\_m over EnKF\_s does not merely come from a large ensemble spread.

Better performance of EnKF\_m due to using a multischeme ensemble relative to the single-scheme ensemble can also be seen in terms of the posterior RMSE (Figs. 6d–f). Similar to what is seen in prior RMSE, there is also a larger improvement in  $T$  (12%) and  $q$  (10%) than in wind (6%) in terms of the posteriors of EnKF\_m relative to those of EnKF\_s. Nevertheless, Fig. 6 also shows that even though EnKF\_s has larger error than that of EnKF\_m, its errors are still generally smaller than those of 3DVAR\_WRF and FNL\_GFS.

Single, deterministic 60-h forecasts (referred to as EnKF\_sDF in Table 1), which are initialized from the 12-hourly posteriors of EnKF\_s (gray solid lines in Fig. 9) instead of EnKF\_m, are also performed. The monthly-averaged RM-DTE of 12-h forecasts in EnKF\_sDF is  $4.61 \text{ m s}^{-1}$ , which is larger than that of EnKF\_mDF, smaller than 3DVAR\_WRF, and compa-

rable to FNL\_GFS. At later forecast stages, EnKF\_sDF error is continuously comparable to or larger than that of EnKF\_mDF and generally smaller than that in 3DVAR\_WRF and FNL\_GFS (Fig. 9).

## 6. Summary and discussion

The month-long performance of a WRF-based EnKF is examined and compared to WRF-3DVAR by conducting 12-hourly real rawinsonde data assimilation in June 2003. The ensemble uses multiple physical parameterization schemes to account for model parameterization error. WRF-3DVAR uses newly generated background error statistics via the NMC method in which a month of 12-h and 24-h forecast differences is utilized to estimate climatological background error statistics. To obtain stable performance of both the EnKF and 3DVAR, a new initial ensemble is created for the coarse domain at 12-h intervals based on the GFS FNL data for the EnKF, whereas the inner domain is continuously updated by the observational data through forecast and data assimilation cycling.

The realism of the ensemble is examined by using a rank histogram and by assessing the agreement between the innovation covariance and the summation of ensemble spread and observation error. The results show that the prior ensemble of the EnKF is generally reasonable. In this case, the sum of variances of ensemble and observation error does not have to be comparable to the innovation covariance to achieve a smaller forecast error. Results also show that slightly overestimated ensemble spread could help to improve the performance of the EnKF for this month-long real-data assimilation test.

The EnKF outperforms 3DVAR for all variables in the troposphere for the particular time period of this study. The prior (posterior) RM-DTE by the EnKF is 9% (25%) smaller than that of 3DVAR. Larger improvements are observed in  $q$  than in  $u$ ,  $v$ , and  $T$ . It is worth noting that because both the WRF model and WRF-3DVAR are changing rapidly, the conclusions drawn by using current systems may not be unequivocally generalized.

The EnKF benefits significantly from the use of a multischeme ensemble to account for model uncertainties in subgrid physical parameterization schemes with confidence level higher than 99%. The benefit of using a multischeme ensemble (over a single-scheme ensemble) is more pronounced in the thermodynamic variables (including temperature and moisture) than in the wind fields. The better performance in  $T$  and  $q$  could be partially explained by their smaller bias resulting from the use of multischemes.

On average, the EnKF analyses lead to more accurate forecasts than 3DVAR analyses when they are used to initialize 60 consecutive, deterministic 60-h forecast experiments for a month. Comparing the ensemble and deterministic forecasts initiated from EnKF analyses to those initiated with 3DVAR shows that a large portion of the improvement appears to come from using an ensemble mean and may be due in part to the ensemble smoothing effect. Deterministic forecasts starting from the EnKF analysis also consistently outperform forecasts initiated from the FNL analyses (in which all operational observations, including satellite data, have been ingested).

*Acknowledgments.* The authors are grateful to Chris Davis, Chris Snyder, Jeff Whitaker, Zoltan Toth, Steve Lord, Tom Hamill, and Jason Sippel for helpful discussions and comments on this subject. We thank Qingnong Xiao, Yongrun Guo, Dale Baker, and Ryan Torn for help on WRF-3DVAR and the EnKF. This research is sponsored by the NSF Grants 0205599 and 0620624 and by the Office of Navy Research under Grant N000140410471.

## REFERENCES

- Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1530.
- , 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Barker, D. M., 2005: Southern high-latitude ensemble data assimilation in the Antarctic mesoscale prediction system. *Mon. Wea. Rev.*, **133**, 3431–3449.
- , W. Huang, Y.-R. Guo, A. J. Bourgeois, and Q. N. Xiao, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897–914.
- Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh, 2006: Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. Amer. Meteor. Soc.*, **87**, 175–190.
- Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1144.
- 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.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162.
- , 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367.
- 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.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation

- functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.
- Grell, G. A., and D. Devenyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, **29**, 1693, doi:10.1029/2002GL015311.
- Hakim, G. J., and R. D. Torn, 2008: Ensemble synoptic analysis. *Synoptic–Dynamic Meteorology and Weather Analysis and Forecasting: A Tribute to Fred Sanders*, L. F. Bosart and H. B. Bluestein, Eds., Amer. Meteor. Soc., in press.
- Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation. *Predictability of Weather and Climate*, T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 124–156.
- , and S. J. Colucci, 1997: Verification of Eta–RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327.
- Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103–120.
- Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX'98 ensemble forecasts. *Mon. Wea. Rev.*, **128**, 73–91.
- Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Speak, and B. Hansen, 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: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. *Mon. Wea. Rev.*, **135**, 1403–1423.
- , and —, 2008: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study. *Mon. Wea. Rev.*, **136**, 522–540.
- Noh, Y., W.-G. Cheon, S.-Y. Hong, and S. Raasch, 2003: Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data. *Bound.-Layer Meteor.*, **107**, 401–427.
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- 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 NCAR/TN-468+STR, 88 pp.
- Torn, R. D., and G. J. Hakim, 2008: Performance characteristics of a pseudo-operational ensemble Kalman filter. *Mon. Wea. Rev.*, **136**, 3947–3963.
- Whitaker, J. S., G. P. Compo, X. Wei, 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.
- Xiao, Q., and J. Sun, 2007: Multiple-radar data assimilation and short-range quantitative precipitation forecasting of a squall line observed during IHOP\_2002. *Mon. Wea. Rev.*, **135**, 3381–3404.
- Zhang, F., and C. Snyder, 2007: Ensemble-based data assimilation. *Bull. Amer. Meteor. Soc.*, **88**, 565–568.
- , —, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253.
- , Z. Meng, and A. Aksoy, 2006: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: Perfect model experiments. *Mon. Wea. Rev.*, **134**, 722–736.
- Zhu, Y., G. Iyengar, Z. Toth, M. S. Tracton, and T. Marchok, 1996: Objective evaluation of the NCEP global ensemble forecasting system. Preprints, *15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., J79–J82.