

# A Regional GSI-Based Ensemble Kalman Filter Data Assimilation System for the Rapid Refresh Configuration: Testing at Reduced Resolution

KEFENG ZHU

*Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma*

YUJIE PAN

*Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma,  
and Nanjing University of Information Science and Technology, Nanjing, China*

MING XUE AND XUGUANG WANG

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

JEFFREY S. WHITAKER, STANLEY G. BENJAMIN, AND STEPHEN S. WEYGANDT

*NOAA/Earth System Research Laboratory, Boulder, Colorado*

MING HU

*NOAA/Earth System Research Laboratory, and Cooperative Institute for Research in Environmental Sciences,  
University of Colorado, Boulder, Colorado*

(Manuscript received 25 January 2013, in final form 2 May 2013)

## ABSTRACT

A regional ensemble Kalman filter (EnKF) system is established for potential Rapid Refresh (RAP) operational application. The system borrows data processing and observation operators from the gridpoint statistical interpolation (GSI), and precalculates observation priors using the GSI. The ensemble square root Kalman filter (EnSRF) algorithm is used, which updates both the state vector and observation priors. All conventional observations that are used in the operational RAP GSI are assimilated. To minimize computational costs, the EnKF is run at  $1/3$  of the operational RAP resolution or about 40-km grid spacing, and its performance is compared to the GSI using the same datasets and resolution. Short-range (up to 18 h, the RAP forecast length) forecasts are verified against soundings, surface observations, and precipitation data. Experiments are run with 3-hourly assimilation cycles over a 9-day convectively active retrospective period from spring 2010. The EnKF performance was improved by extensive tuning, including the use of height-dependent covariance localization scales and adaptive covariance inflation. When multiple physics parameterization schemes are employed by the EnKF, forecast errors are further reduced, especially for relative humidity and temperature at the upper levels and for surface variables. The best EnKF configuration produces lower forecast errors than the parallel GSI run. Gilbert skill scores of precipitation forecasts on the 13-km RAP grid initialized from the 3-hourly EnKF analyses are consistently better than those from GSI analyses.

## 1. Introduction

Variational methods, like three- or four-dimensional variational data assimilation schemes (3DVAR or

4DVAR; Parrish and Derber 1992; Courtier et al. 1998; Rabier et al. 2000), have been used successfully at operational numerical weather prediction (NWP) centers for over a decade. The typically used static, spatially homogeneous, and isotropic background error covariance (Purser et al. 2003) represents one of the main deficiencies in such schemes. The ensemble Kalman filter (EnKF; Evensen 1994) algorithm has gained much popularity, because of its ability to estimate flow-dependent error

---

*Corresponding author address:* Dr. Ming Xue, Center for Analysis and Prediction of Storms, University of Oklahoma, 120 David L. Boren Blvd., Norman, OK 73072.  
E-mail: mxue@ou.edu

covariance structures and its relative ease of implementation. Since its introduction into atmospheric data assimilation (DA) by Houtekamer and Mitchell (1998), numerous variations have been developed, including different formulations and algorithm refinements (e.g., Burgers et al. 1998; Evensen and van Leeuwen 2000; Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; Ott et al. 2004; Szunyogh et al. 2008) or combinations of the EnKF with variational methods to arrive at hybrid schemes (e.g., Hamill and Snyder 2000; Lorenc 2003; Buehner 2005; Wang et al. 2007b; Wang 2010). Some of these algorithms have been used in operational global forecast systems to provide ensemble-based background error covariance (Wang 2010; Raynaud et al. 2011; Bonavita et al. 2012), as well as initial conditions (Houtekamer and Mitchell 2005; Whitaker et al. 2008; Hamill et al. 2011).

Meng and Zhang (2011) reviewed the recent progress with EnKF applications in limited-area models (LAMs). For LAMs, the application of EnKF initially focused on the convective scale through observing system simulation experiments (OSSEs) under the perfect model assumption (Snyder and Zhang 2003; Zhang et al. 2004; Tong and Xue 2005). In those tests, EnKF using simulated Doppler radar radial wind alone was able to accurately estimate wind, temperature, and moisture fields for convective storms. When reflectivity data were also assimilated into a model including complex ice microphysics, the EnKF analyses and subsequent forecasts were further improved (Tong and Xue 2005). The application of EnKF to the assimilation of real radar data is also encouraging. For example, Dowell et al. (2004) used EnKF to assimilate radar data from a single Doppler radar and the analyzed storm location and strength were comparable to those in dual-Doppler wind analyses. By assimilating both surface station and radar data using EnKF and a full-physics model, Lei et al. (2009) were able to reproduce the main storm cell of a supercell storm in their deterministic forecast. Aksoy et al. (2010) launched 30-min ensemble forecasts after 60-min assimilation of radial velocity data (Aksoy et al. 2009) using an EnKF and a cloud model. They found that the forecast skill generally decayed on the time scale of tens of minutes for several types of storms. In Snook et al. (2012), ensemble forecasts launched using the full-physics Advanced Regional Prediction System (ARPS; Xue et al. 2003) from EnKF analyses that included data from four X-band Collaborative Adaptive Sensing of the Atmosphere (CASA) radars and five Weather Surveillance Radar-1988 Dopplers (WSR-88Ds), were able to predict high probability of surface vorticity centers within a mesoscale convective system that matched reported tornado locations more than 2 h after the end of radar data assimilation.

Mesoscale applications of LAM EnKFs have also progressed from OSSEs (Zhang et al. 2006) to observing system experiments (OSEs) using real data (Meng and Zhang 2008a). In Meng and Zhang (2008b), month-long tests with the assimilation of sounding data every 12 h showed that the EnKF outperformed a corresponding 3DVAR simulation. When multiple physics sets were used in the forecast ensemble, results for temperature and moisture fields were further improved. These multiple-physics results are consistent with those of Fujita et al. (2007), who also examined the benefits of using multiple physics sets.

The purpose of this paper is to document the development and testing of a regional EnKF system intended for future operational implementation for the Rapid Refresh (RAP) model. The RAP, which replaced the Rapid Update Cycle (RUC; Benjamin et al. 2004) in May 2012, uses the Advanced Research core of the Weather Research and Forecasting Model (WRF-ARW; Skamarock et al. 2005) and the gridpoint statistical interpolation (GSI) 3DVAR (Wu et al. 2002) for its DA. So far, real-time or quasi-operational mesoscale EnKF applications are still limited in the literature. Torn and Hakim (2008) describe results from 2-yr pseudo-operational regional EnKF runs. Compared with other operational center deterministic forecasts, their WRF-based EnKF system has slightly larger errors in wind and temperature but smaller errors in moisture. In their case, since half of their domain was centered over Washington State is over the ocean and their system did not assimilate any satellite data, their experimental results are encouraging. In another attempt, Bonavita et al. (2008) developed a regional local ensemble transform Kalman filter (LETKF; Hunt et al. 2007) system for the Italian National Meteorological Service. Different covariance inflation schemes were examined in their 15-day-long tests. Their results show that their LETKF performed better than the operational 3DVAR system. The EnKF system we present in this paper is developed for potential RAP operational application, which requires high-frequency updating. In addition, the system is designed to be capable of assimilating all available observations in the National Centers for Environmental Prediction (NCEP) operational data stream and takes advantage of all data-processing capabilities in the operational GSI system; the EnKF results are also directly compared with those of parallel GSI runs.

The rest of this paper is organized as follows. In section 2, a brief introduction of the RAP system is presented together with an overview of the EnKF system. The designs of the test experiments are given in section 3. In section 4, deterministic forecasts from the 3-hourly ensemble mean EnKF analyses are compared with

corresponding forecasts initialized from parallel GSI 3DVAR analyses assimilating the same datasets at the same 40-km resolution. Sensitivity to the localization scheme and the impacts of using multiple physics suites are also discussed in that section. Additionally, forecasts on the 13-km RAP grid, starting from interpolated 40-km EnKF and GSI analyses, are compared in section 5. A summary is given in section 6.

## 2. The prediction model, EnKF system, and verification tool

### a. *The Rapid Refresh system*

One way to improve the accuracy and utility of short-range weather forecasts is to more frequently assimilate the latest available observations and to make available rapidly updated forecasts. The NCEP operational hourly updated RUC system was designed using this approach (Benjamin et al. 2004). Since its first implementation in the 1990s with a 60-km grid spacing and a 3-hourly update cycle, it has been widely used for aviation and severe weather forecasting. Interested readers are referred to Benjamin et al. (2004) for an evaluation of the earlier version of RUC and to Benjamin et al. (2010), in which the impacts of various conventional observations on the accuracy of RUC forecasts are examined.

The RAP is a replacement for the RUC system, which is based on the nonhydrostatic WRF-ARW dynamic core and uses similar, but updated physical parameterizations compared to the RUC. The RAP was implemented operationally at NCEP on 1 May 2012. The GSI, a unified 3D variational DA system used for both global and regional applications (Kleist et al. 2009), is used within RAP for the hourly analysis update. The horizontal resolution of the RAP is  $\sim 13$  km and it has 50 vertical levels extending up to 10 hPa at the model top. The RAP system assimilates more observations than the RUC, including satellite radiance data, and the RAP has a larger domain, covering all of North America. The physics options used by the operational RAP include the Grell-G3 cumulus parameterization, Thompson microphysics, Rapid Radiative Transfer Model (RRTM) longwave radiation, Goddard shortwave radiation, Mellor–Yamada–Janjić (MYJ) turbulent mixing, and the RUC–Smirnova land surface model. Details on these schemes can be found in Benjamin et al. (2009).

The RAP (along with its predecessor, the RUC) employs a digital filter initialization (DFI) before launching the forecast. The DFI provides a simple, yet effective, way to filter out high-frequency noise during the initial period of model integration; the filtered model state tends to be more balanced (Lynch and Huang 1992; Chen and Huang 2006). There are three available DFI

options in the WRF model: digital filter launching (DFL; Lynch and Huang 1994), Diabatic DFI (DDFI; Huang and Lynch 1993), and twice DFI (TDFI; Lynch 1997). In the operational version of RAP, TDFI is used, which includes an adiabatic backward integration step and a full-physics forward integration step. An important issue with the TDFI option is the need to turn off diabatic processes in the backward integration; for mesoscale and convective applications, diabatic processes, including the precipitation processes, can be critically important, and neglecting them can lead to large errors.

In this study, we choose to use DFL, which does not include the adiabatic backward integration step. However, the original WRF DFL implementation had a problem; the original implementation filters the atmospheric variables but not the land surface model fields. When the forecast is relaunched from the filtered state at the center of the filter window, the unfiltered land surface fields from the beginning of the filter window are used. Such an inconsistency between soil and atmospheric variables occasionally caused model integration failures. In our current implementation, the land surface fields are also subject to the same filtering, and the modification improves the model integration stability.

### b. *The regional EnKF system based on GSI*

As mentioned earlier, our EnKF system is specifically designed for the RAP, using its operational data stream. To facilitate the running of a large number of experiments that seek to find the best settings of the EnKF DA system, all the results presented in this paper employ 3-hourly assimilation cycles. We note here that for the EnKF analysis, data input and output (I/O, including reading and writing of all the ensemble grids in and out of the GSI preprocessor and the EnKF code) represent a large portion of the overall computational cost. Although testing the EnKF system with hourly EnKF cycles is preferred given that the operational RAP system uses hourly cycles, doing this more than doubles the total computational cost and is deferred to future research. Figure 1 presents a flowchart of the RAP EnKF system described and tested in this paper. The EnKF system contains three major parts: 1) the 40-km-domain 3-hourly EnKF analysis–forecast cycles that include the GSI observation processing component for observation innovation calculations (not the actual 3DVAR analysis), 2) the EnKF analysis component, and 3) the ensemble forecast component. Longer, 18-h deterministic forecasts are launched from the ensemble mean analyses every 3 h on the 40-km grid, and from interpolated analyses at 0000 and 1200 UTC to produce forecasts on a 13-km grid for verification purposes. The analysis variables of the EnKF system are the Cartesian

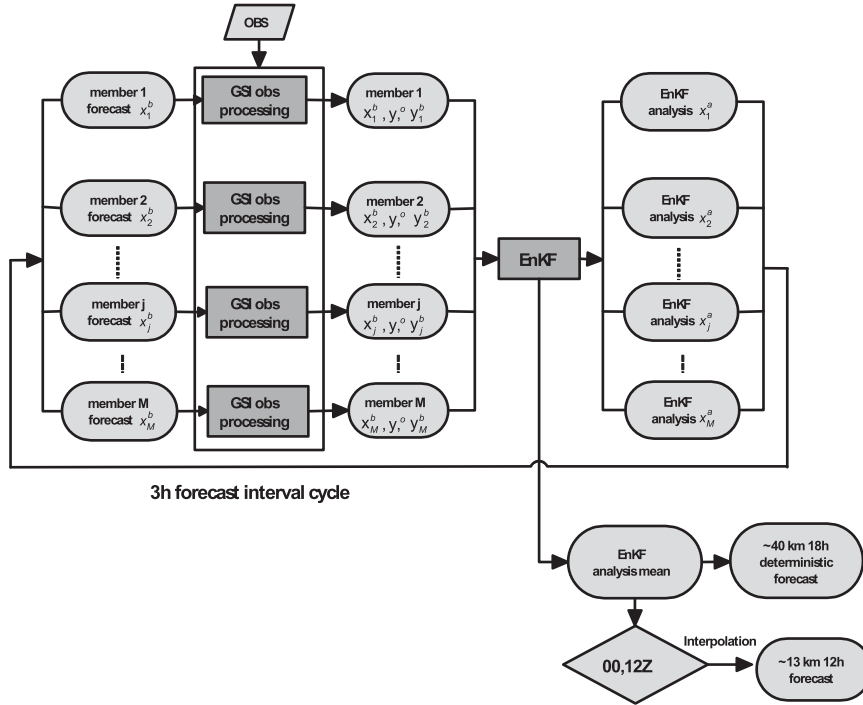


FIG. 1. Flowchart of the regional GSI-based EnKF system.

components of horizontal wind,  $U$  and  $V$  (in map projection space), water vapor mixing ratio, potential temperature, perturbation geopotential, and the perturbation dry air mass in the model column. Additional details on the EnKF algorithm and a description of the verification procedure are given below.

1) THE ENKF ALGORITHM AND COVARIANCE LOCALIZATION

The EnKF system uses the ensemble square root filter (EnSRF) algorithm described by Whitaker and Hamill (2002). The code base is the same as that used by the Global Forecast System (GFS) EnKF system (Whitaker et al. 2008). It was modified and linked to the regional GSI for RAP and the RAP forecast model, including codes for the reading and writing of RAP model grids, and for handling RAP specific datasets. The standard analysis equations for the ensemble mean and ensemble perturbations of the EnSRF follow Eqs. (4) and (5) of Whitaker and Hamill (2002) and have been used in many studies. The Kalman gain for the  $j$ th state variable  $x_j$  and  $m$ th observation  $y_m$  is given by

$$K_{m,j} = \rho_{mj} \frac{\text{Cov}(x_j^b, y_m^b)}{\text{Cov}(y_j^b, y_m^b) + R_m}, \quad (1)$$

where  $\text{Cov}(x_j^b, y_m^b)$  is the error covariance between background state variable  $x_j^b$  and observation prior  $y_m^b$ ,

and  $\text{Cov}(y_m^b, y_m^b)$  is the background error variance of the observation prior. The error variance of observation  $y_m$  is  $R_m$ . The corresponding Kalman gain for the ensemble perturbations is equal to  $\alpha K_{m,j}$ , where  $\alpha$  is the square root filter factor given by Eq. (13) of Whitaker and Hamill (2002). In Eq. (1),  $\rho_{m,j}$  is the covariance localization coefficient, defined in our case as functions of time and space according to

$$\rho_{m,j} = \rho_t \rho_h \rho_v = \text{taper} \left\{ \frac{|t_m - t_a|}{t_{\text{cut}}} \right\} \times \text{taper} \left\{ \frac{r_{m,j}}{r_{\text{cut}}} \right\} \times \text{taper} \left\{ \frac{|\ln(p_m) - \ln(p_j)|}{\ln(p_{\text{cut}})} \right\}, \quad (2)$$

where  $\rho_t$ ,  $\rho_h$ , and  $\rho_v$  are the localization coefficients in time, and the horizontal and vertical directions, respectively. Here, we introduce time localization to allow for the assimilation of observations not taken exactly as the analysis time, but within a time window  $[t_a - t_{\text{cut}}, t_a + t_{\text{cut}}]$  around the analysis time  $t_a$ . In addition,  $t_m$  is the time of the  $m$ th observation. Coefficient  $\rho_t$  decreases as the observation time  $t_m$  deviates from the analysis time and decreases to zero outside the time window. With this capability, the EnSRF is actually a four-dimensional algorithm (4DEnSRF), as is also discussed by Wang et al. (2013). We set  $t_{\text{cut}}$ ,  $r_{\text{cut}}$ , and  $\ln(p_{\text{cut}})$  to be the cutoff radii in time and in the

horizontal and vertical, respectively. The horizontal distance between the  $j$ th state variable and  $m$ th observation is  $r_{mj}$ , while  $p_j$  and  $p_m$  are the pressure values of the  $j$ th state variable and  $m$ th observation,

$$\text{taper}(r) = \begin{cases} \left( [(-0.5r + 0.5) \times 2r + 0.625] \times 2r - \frac{5}{3} \right) \times (2r)^2 + 1.0 & 0.0 \leq r \leq 0.5 \\ \left( \left( \left[ \left( \frac{r}{6} - 0.5 \right) \times 2r + 0.625 \right] \times 2r + \frac{5}{3} \right) \times 2r - 5 \right) \times 2r + 4 - \frac{1}{3r} & 0.5r \leq 1.0 \\ 0 & r > 1.0 \quad \text{or} \quad r < 0.0 \end{cases} \quad (3)$$

## 2) COVARIANCE INFLATION FOR THE ENSRF ALGORITHM

To account for the sampling error associated with the limited ensemble size, and model error that is inadequately represented by the forecast ensemble, as well as other factors that can cause ensemble underdispersion and filter divergence, we apply multiplicative covariance inflation (Anderson and Anderson 1999) in our EnSRF. The inflation factor consists of two parts: one that is horizontally homogeneous (on constant pressure levels) but height (pressure) dependent, and one that is adaptive and related to the amount of variance reduction by observations (Anderson 2009; Whitaker and Hamill 2010). In equation form, it is

$$\mathbf{x}'^a = (\beta_1 + \beta_2)\mathbf{x}^a, \quad (4)$$

where  $\mathbf{x}'^a$  on the right- and left-hand sides are the ensemble perturbations before and after inflation, respectively. The coefficients are

$$\beta_1 = b \times \text{taper} \left( \frac{|\ln p_a - \ln p_{\text{sfc}}|}{\ln p_{\text{covcut}}} \right) \quad \text{and} \quad (5)$$

$$\beta_2 = \sqrt{c \frac{\sigma_b^2 - \sigma_a^2}{\sigma_a^2} + 1}, \quad (6)$$

where  $b$  and  $c$  are constants whose values are given in Table 1 for various experiments. The cutoff distance in terms of log pressure is  $\ln p_{\text{covcut}}$ , and is set to 5 in our experiments. The pressure (in hPa) at the analysis grid point is  $p_a$  and at the corresponding surface level is  $p_{\text{sfc}}$ , while  $\sigma_b^2$  and  $\sigma_a^2$  are the prior and posterior (or background and analysis) variances, respectively, for the variable being analyzed. In Eq. (5) the taper function is also given by Eq. (3), as before. With this function, the static inflation factor  $\beta_1$  has a maximum value of  $b$  at the level of surface pressure and decreases with height with a scale length of  $\ln p_{\text{covcut}}$ . Additionally, in our tests,

respectively. Here, the taper function is based on Eq. (4.10) of Gaspari and Cohn (1999), which is commonly used to define spatial localization functions [see Eq. (3) below]:

when  $|\ln p_a - \ln p_{\text{sfc}}| / \ln p_{\text{covcut}} > 0.75$ , that is when  $\ln p_{\text{sfc}} > 0.75 + \ln p_a$ ,  $\beta_1$  is set to 0; not doing so has led to too large of a spread in temperature near the model top and occasional model integration instability. In Eq. (6), we call  $c$  the spread recovery factor. When  $c = 1$ ,  $\beta_2 = \sigma_b / \sigma_a$  (which is usually larger than 1), so the factor acts to fully recover the spread found in the background ensemble. When  $c = 0$ ,  $\beta_2 = 1$ , and if the static component is also zero ( $b = 0$ ), then there is no inflation according to Eq. (4).

### c. The MET verification tool

The performance of our EnKF system is evaluated by verifying short-range forecasts against observations. For the verification, we employ the Model Evaluation Tools (METs) developed by the Development Testbed Center (DTC; Brown et al. 2009), which contains comprehensive verification metrics for both deterministic and probabilistic forecasts. Even though only a few metrics are examined in this paper, the use of a well-tested, community-supported, verification package improves the reliability of the verification results.

For our experiments, the 40-km deterministic forecasts launched from the ensemble mean analyses (after applying DFL) are verified against upper-air and surface observations. Upper-air sounding observations are used to verify model relative humidity (RH), temperature  $T$ , and model wind components  $U$  and  $V$ , while surface observations are used to verify surface pressure, 2-m RH, 2-m temperature, and 10-m  $U$  and  $V$ . Root-mean-square error (RMSE) is used as the verification metric. For forecasts on the interpolated 13-km grid, we focus on the precipitation verification, using the NCEP stage IV precipitation data (Lin and Mitchell 2005). The Gilbert skill score (GSS) (Gandin and Murphy 1992), also known as the equitable threat score (ETS), is also used.

## 3. Design of data assimilation experiments

A 9-day retrospective period from 8 to 16 May 2010 was selected as our testing period. This period features

TABLE 1. List of data assimilation experiments. The single set of physics is the same as group p1 in Table 2. In the vertical localization column, the second factor is for surface pressure and the first factor is for other variables. The upward-pointing arrow ( $\nearrow$ ) indicates a value that is increasing with height, and the height-dependent localization radius function is  $r_1 - r_2 \times \text{taper}[\ln(p_{\text{ob}}) - \ln(1020)]/2.0$ , where  $r_1$  and  $r_2$  are given before and after the arrow ( $\nearrow$ ) in the table.

Expt	Horizontal localization radius $r_{\text{cut}}$ (km)	Vertical localization radius $\ln p_{\text{cut}}$	$b$ in Eq. (5)	$c$ in Eq. (6)	Physics schemes
EnKF_Ctr	1000	1.1 (RH, $T$ , $U$ , and $V$ ) 1.6 (Ps)	0.1	0.9	Single set
GSI_Ctr			N/A		
EnKF_S	800	1.0 (RH, $T$ , $U$ , and $V$ ) 1.2 (Ps)	0.1	0.9	Single set
EnKF_L	1200	1.2 (RH, $T$ , $U$ , and $V$ ) 2.0 (Ps)	0.1	0.9	Single set
EnKF_AI	1000	As EnKF_Ctr	0.0	0.98	Single set
EnKF_CtrHDL	700 $\nearrow$ 1050	1.1/4 $\nearrow$ 1.1/2 (RH and $T$ ) 1.1/2 $\nearrow$ 1.1 ( $U$ and $V$ ) 1.6 (Ps)	0.1	0.9	Single set
EnKF_Mult5	As EnKF_CtrHDL	As EnKF_CtrHDL	0.1	0.9	Multiple sets, 5 groups
EnKF_Mult10	As EnKF_CtrHDL	As EnKF_CtrHDL	0.1	0.9	Multiple sets, 10 groups

a wide variety of weather, including the 10 May Oklahoma tornado outbreak featuring strong mesoscale forcing, a mesoscale convective system (MCS) during the night of 11 May, a cold-season-type Front Range upslope low pressure precipitation event, and some southeast-propagating MCSs across Texas. A key characteristic for this period is the existence of propagating baroclinic disturbances and associated surface weather phenomena, as well as several episodes of precipitation events.

Our data assimilation experiments are run in continuous 3-hourly updated cycles throughout the period; they start at 0000 UTC 8 May 2010 and end at 2100 UTC 16 May 2010. Figure 2 shows the domain used by the 40-km grid-spacing DA and forecast experiments; the domain has  $207 \times 207$  horizontal grid points and covers the entire North American region. The 13-km grid-spacing domain used for the high-resolution forecasting is slightly smaller, as indicated by the smaller box in Fig. 2a, and has  $532 \times 532$  horizontal grid points. Both grids use 50 vertical levels. From each of the 3-hourly cycled EnKF ensemble mean analyses, an 18-h-long forecast is launched (after applying DFL). To compare with the GSI 3DVAR scheme currently employed by the operational RAP system, parallel experiments using the GSI 3DVAR system are also run. Both EnKF and GSI experiments use almost the same configurations as the current operational RAP except for the use of 40-km instead of 13-km horizontal grid spacing. For the EnKF system, 40 ensemble members are employed. The lateral boundary conditions are based on the NCEP GFS forecasts, available every 3 h and updated every 6 h. For the ensemble forecasts within the EnKF cycles, perturbations created using the WRF-3DVAR “random CV” option (Barker 2005) following Torn et al. (2006) and

Wang et al. (2008) are added to the GFS forecast boundary conditions and to the GFS analysis at 0000 UTC 8 May 2010 to create the initial ensemble. Such perturbations have spatial structures similar to the background error covariance derived from the National Meteorological Center (NMC, now known as NCEP) method (Parrish and Derber 1992) and have been used in other mesoscale EnKF studies as initial perturbations (e.g., Meng and Zhang 2008b). Our earlier tests with a single observation showed that flow-dependent covariances are well developed about 1 day into the EnKF assimilation cycles.

The observation datasets used in this study are almost the same as those used by the operational RAP. They include the surface observations from land-reporting stations, mesonets, ships and buoys, etc., while the upper-air observations include land- or sea-launched radiosondes, data from commercial aircrafts, wind profilers, radar velocity–azimuth display (VAD) winds, and satellite retrieval winds. The GPS precipitable water (PW) and satellite radiance data are not used in the experiments reported here; initial testing indicated the need for improved bias correction, and results including radiance data will be reported upon in the future. Figure 2 shows an example of the horizontal distributions of several major observation types within the current testing domain. In general, observations cover both land and sea and are much denser over land. The radiosonde data are the most evenly distributed observation type over land and are also used for forecast verification. There are great concentrations of observations at the surface (see Fig. 2b) and at the jet level around 200 hPa. The latter is mainly due to commercial aircraft (see Fig. 2c) observations and satellite-retrieved winds (see Fig. 2d). Figure 3 shows the average number of

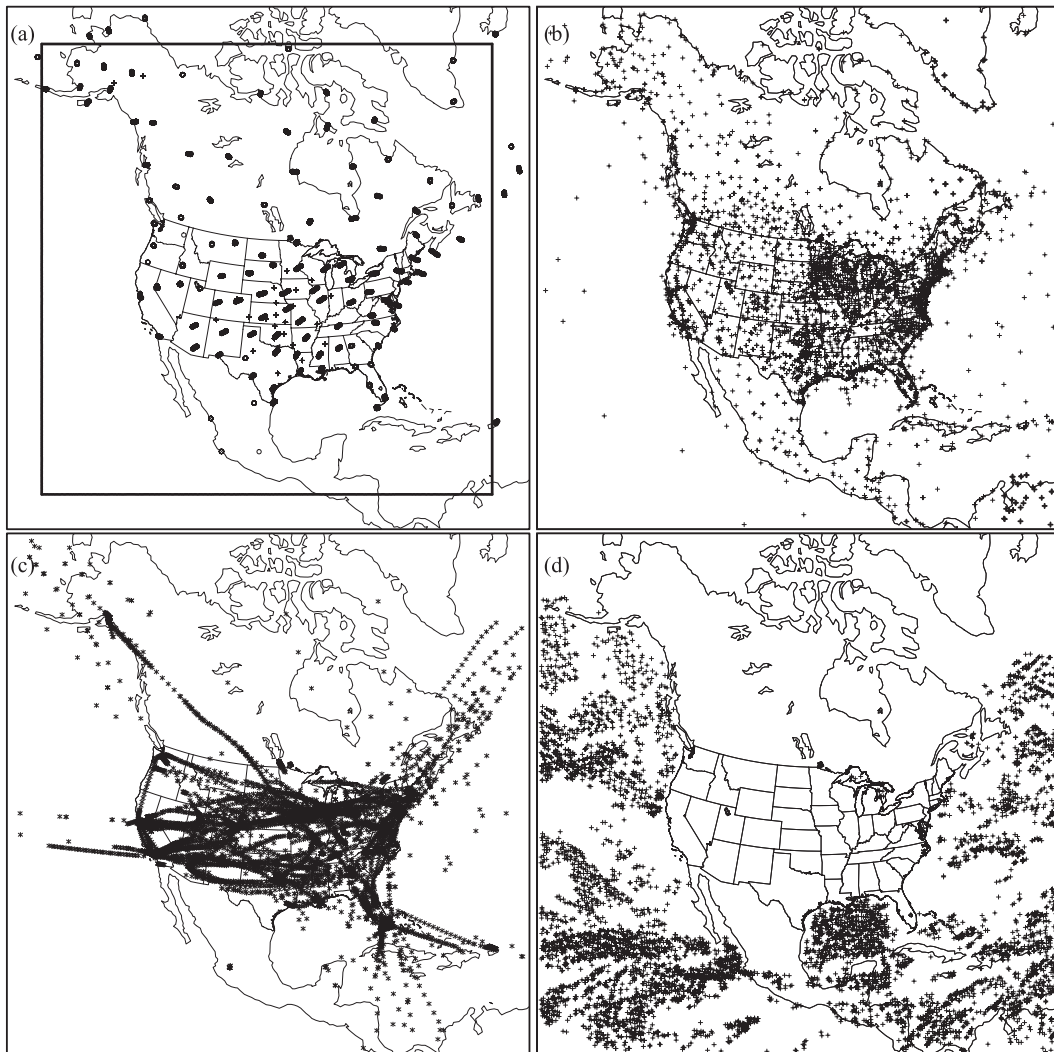


FIG. 2. Examples of the horizontal distributions of observations collected in real time 1.5 h before and 1 h after 0000 UTC 14 May 2010 as used by the real-time RAP system and in tests of this study: (a) sounding (circles, displacement of the circles is due to balloon drifting) and profiler (plus signs); (b) surface stations over land and from ships, buoys, etc.; (c) aircraft observations; and (d) satellite retrieved winds. The smaller box in (a) is the domain used by the 13-km forecast.

observations and their time distributions during the testing period. Not surprisingly, there are more observations at 0000 and 1200 UTC than at other times. Usually, the real-time RAP system collects data from 1.5 h before and 0.5 h after the time of analysis. However, for 0000 and 1200 UTC it waits half an hour longer for more data (such as sounding data) to arrive. In our tests, the datasets assimilated at 3-hourly intervals are the datasets collected and used by the operational hourly RAP system; as a result, observations that arrived in real time outside the 2-h (2.5 h for 0000 and 1200 UTC) windows are not used.

Table 1 lists all the experiments to be presented in this paper. These experiments can be categorized into three

groups: the first group has two control experiments using EnKF (EnKF\_Ctr) and GSI 3DVAR (GSI\_Ctr) as the analysis schemes, respectively. Single observation tests are first carried out to examine behaviors of the analysis schemes. The second group examines variations in the covariance localization scales and inflation factors in an attempt to find the best configurations. The control experiment EnKF\_Ctr uses a 1000-km cutoff radius in the horizontal and  $\ln p_{\text{cut}} = 1.1$  (which is about 7.6 km) in the vertical. Experiments EnKF\_L and EnKF\_S use a larger and a smaller horizontal and vertical localization radius, respectively. The inflation includes both static and adaptive parts in EnKF\_S and EnKF\_L. Experiment EnKF\_AI (AI for adaptive inflation) is the same as EnKF\_Ctr but

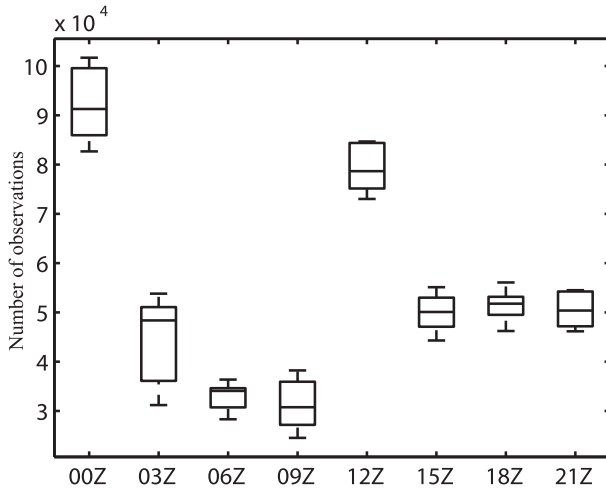


FIG. 3. Box plot of the total number of observations assimilated at different times of the day during the testing period. The central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers are extended to the most extreme numbers.

includes adaptive inflation only and uses a larger value for coefficient  $c$  (0.98) in the adaptive inflation factor. The next experiment, EnKF\_CtrHDL, uses height-dependent localization radii in both the horizontal and vertical, with the radii increasing with height. The third group examines the impact of using multiple suites of

physics parameterizations in the WRF forecast ensemble. Two experiments with different numbers of physics suites are tested. Table 2 gives the physics scheme combinations for the 10 suites used by EnKF\_mult10. In this experiment, the 40 ensemble members are evenly divided into 10 groups of 4, with each group using one suite of parameterizations. EnKF\_mult5 includes five groups of eight, employing the first five suites given in Table 2. Physics suite p1 in Table 2 is used by all of the single-physics suite experiments.

#### 4. Results of experiments

In this section, we will first examine the results of single observation tests, then compare the control EnKF results against the GSI results. The sensitivity to the localization and inflation schemes and the impacts of multiple physics suites are then examined. All verifications in this section are carried out at 40-km grid spacing.

##### a. Single-observation tests

EnKF differs from the 3DVAR approach in its use of flow-dependent background error covariance derived from the forecast ensemble. Single-observation tests provide a good view of the background error covariance structures since the spatial spread of observation innovation is primarily determined by the spatial correlation in

TABLE 2. Multiple-physics combinations (suites) used by experiment EnKF\_mult10. The single-physics experiments use suite p1 while EnKF\_mult5 uses suites p1–p5.

Physics suite	Longwave radiation	Shortwave radiation	Surface layer	PBL	Cumulus
p1	RRTM	Goddard shortwave	Monin–Obukhov (Janjić)	Mellor–Yamada–Janjić TKE	Grell 3D ensemble
p2	RRTM	Dudhia	Monin–Obukhov	Yonsei University	Kain–Fritsch (new Eta Model)
p3	RRTM	Goddard shortwave	MYNN surface layer	Mellor–Yamada–Nakanishi–Niino 2.5-level TKE	Grell–Devenyi ensemble
p4	GFDL (Eta Model) longwave	GFDL (Eta Model) shortwave	Monin–Obukhov (Janjić)	Mellor–Yamada–Janjić TKE	Grell 3D ensemble
p5	RRTM	Goddard shortwave	Monin–Obukhov (Janjić)	Mellor–Yamada–Janjić TKE	Grell–Devenyi ensemble
p6	RRTM	Goddard shortwave	Quasi-normal scale elimination (QNSE) surface layer	QNSE PBL	Kain–Fritsch (new Eta Model)
p7	RRTM	Goddard shortwave	Monin–Obukhov (Janjić)	Mellor–Yamada–Janjić TKE	Betts–Miller–Janjić
P8	GFDL (Eta Model) longwave	GFDL (Eta Model) shortwave	Monin–Obukhov (Janjić)	Mellor–Yamada–Janjić TKE	Betts–Miller–Janjić
P9	GFDL (Eta Model) longwave	GFDL (Eta Model) shortwave	QNSE surface layer	QNSE PBL	Kain–Fritsch (new Eta Model)
p10	RRTM	Goddard shortwave	Pleim–Xiu land surface model (ARW)	Asymmetric Convective Model version 2 (ACM2; Pleim) PBL (ARW)	Grell 3D ensemble



the covariance while correction to other variables is determined by the cross-variable covariance. A single temperature observation is assumed to be located at the Norman, Oklahoma, radiosonde site at the 600-hPa level, and is 1 K above the background value. Figure 4 shows the analysis increments of temperature and wind created by this single temperature observation. In this test, the backgrounds are the 3-h forecasts after 5 days of 3-hourly EnKF analysis cycles employing the full set of observations; the GSI analysis uses the mean of the 3-h ensemble forecasts as the background; therefore, the ensemble mean EnKF and GSI analyses used the same background.

As expected, EnKF gives the temperature increment (shaded) with flow-dependent structures, which are more stretched along the direction of the isotherms (Fig. 4a) while the GSI temperature increment has a circular shape, reflecting its static, isotropic spatial covariance structures (Fig. 4b). The maximum temperature increment is right at the location of the observation in GSI but is slightly off centered in the EnKF analysis. The increment maxima are also found at the observation level (cf. to the second row). The flow-dependent structure is more obvious below the observation level at 650 hPa. In Fig. 4c, the temperature increment is seen to roughly follow a temperature trough east of the observation location and the maximum temperature increment is located northeast of the observation, apparently reflecting spatially inhomogeneous flow-dependent covariance. In comparison, the analysis increment of GSI is obviously unaffected by the first-guess temperature pattern; the increment contours are circular, and the magnitude is smaller away from the observation level. The differences between the EnKF and GSI analyses in the vertical cross section are also obvious. Figures 4e and 4f show that, for EnKF, the major axis of the temperature increment is inclined to the west with height, which is consistent with the temperature contour inclination at the observation point. For GSI, the temperature increment in the vertical cross section is not exactly elliptic but follows the terrain-following coordinate surfaces somewhat and decreases away from the observation location. This is due to the use of a recursive filter in the GSI to model the background error covariance (Wu et al. 2002), where the increment tends to be spread along coordinate surfaces.

The wind increments created by the single temperature observation are also worth examining. In GSI, there are five control variables: streamfunction  $\Psi$ ; velocity potential  $\chi$ ; temperature  $T$ ; relative humidity RH; and surface pressure  $p_{\text{sf}}$ . Except for the streamfunction and relative humidity, all other variables are divided into the balanced and unbalanced parts. When analyzing temperature observations, wind increments are induced from

the balance relation between the balanced temperature and streamfunction according to

$$T_b = G\Psi, \quad (7)$$

where  $G$  contains the regression coefficients between the streamfunction and temperature. In GSI,  $G$  is nearly zero at the observation level, positive below and negative above. Correspondingly, the wind increment at the observation level is nearly zero (Fig. 4b), cyclonic below (Fig. 4d) and anticyclone above (Fig. 4f). This regression more or less reflects the hydrostatic and geostrophic relations; where there is a positive observation increment in temperature, geopotential height is increased above the observation, inducing an anticyclonic geostrophic circulation (and cyclonic below). Compared to the GSI analysis, the pattern of the EnKF wind increment is more complicated but is mainly oriented parallel to the temperature gradient near the observation point, consistent with the findings of Buehner (2005). The increments do also show an anticyclonic pattern above the observation and a cyclonic pattern below (Fig. 4e), and the increments are an order of magnitude stronger than those of GSI (note the 0.1 and 0.01  $\text{m s}^{-1}$  contour intervals used by the two). Overall, we see physically consistent and strongly flow-dependent increments produced by the EnKF analyses while the GSI analysis increments are more isotropic and mainly reflect the large-scale geostrophic balance between the temperature and wind fields.

### b. Control EnKF and GSI experiments

The RAP system had been run experimentally in real time for several years at the National Oceanic and Atmospheric Administration/Earth System Research Laboratory (NOAA/ESRL) before being officially implemented at NCEP in May 2012. In our experiments, we borrow from the standard configurations of the 13-km RAP and made some adjustments for our 40-km grid-spacing tests. One of the changes, as mentioned in section 2, is to use the DFL with a single forward step employing full physics. In the tests below, a single 40-min filter window centered at 20 min of forecast time is used. Compared to TDFI, the modified DFL improves the forecast accuracy of surface variables for both GSI and EnKF (not shown).

Using the modified DFL as an initialization scheme, we compare forecasts initialized from the GSI analyses and EnKF ensemble mean analyses. Figures 5 and 6 show the average RMSEs of 3- and 18-h forecasts, respectively, verified against sounding data. It can be seen that the RMSEs of EnKF\_Ctr are overall lower than those of GSI\_Ctr except for the temperature at the upper levels, where the EnKF error can be  $\sim 0.1$  K larger.

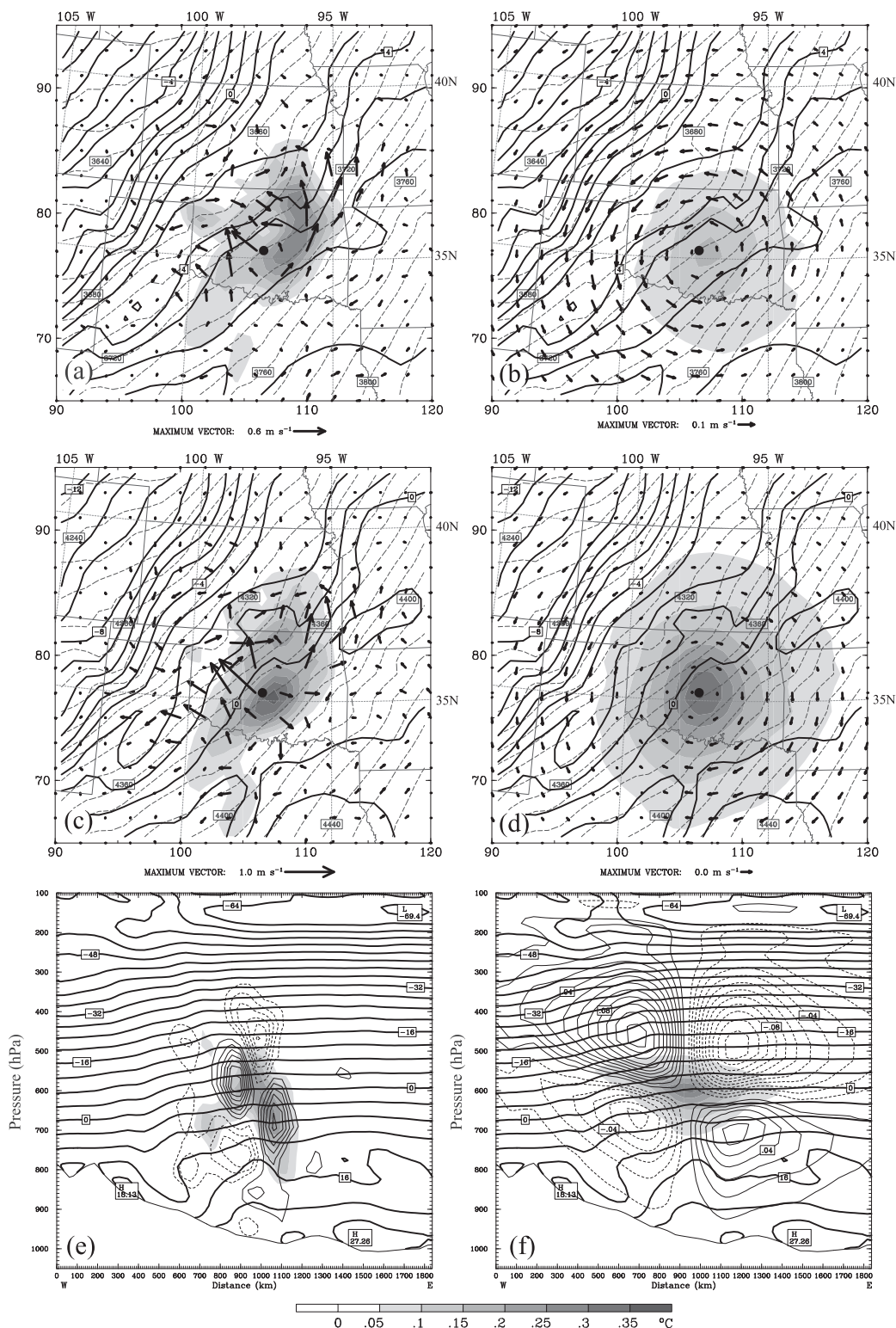


FIG. 4. Analysis increments of single observation tests using (left) EnKF and (right) GSI 3DVAR. A hypothetical temperature observation of 1 K above the background is located at 600 hPa at Norman, as shown by the black dot in (a),(b). (a),(b) The solid and dashed contours are the first-guess temperature (K) and geopotential height (m) at the 600-hPa level, respectively. Shading indicates the temperature analysis increment together with vector wind analysis increments. (c),(d) As in (a),(b), but at the 650-hPa level. (e),(f) The increment fields in an east–west vertical cross section through the observation point. Shading indicates the temperature increment. The thin contours (solid for positive and dashed for negative) are for the north–south wind increment while thick solid contours are for the first-guess temperature.

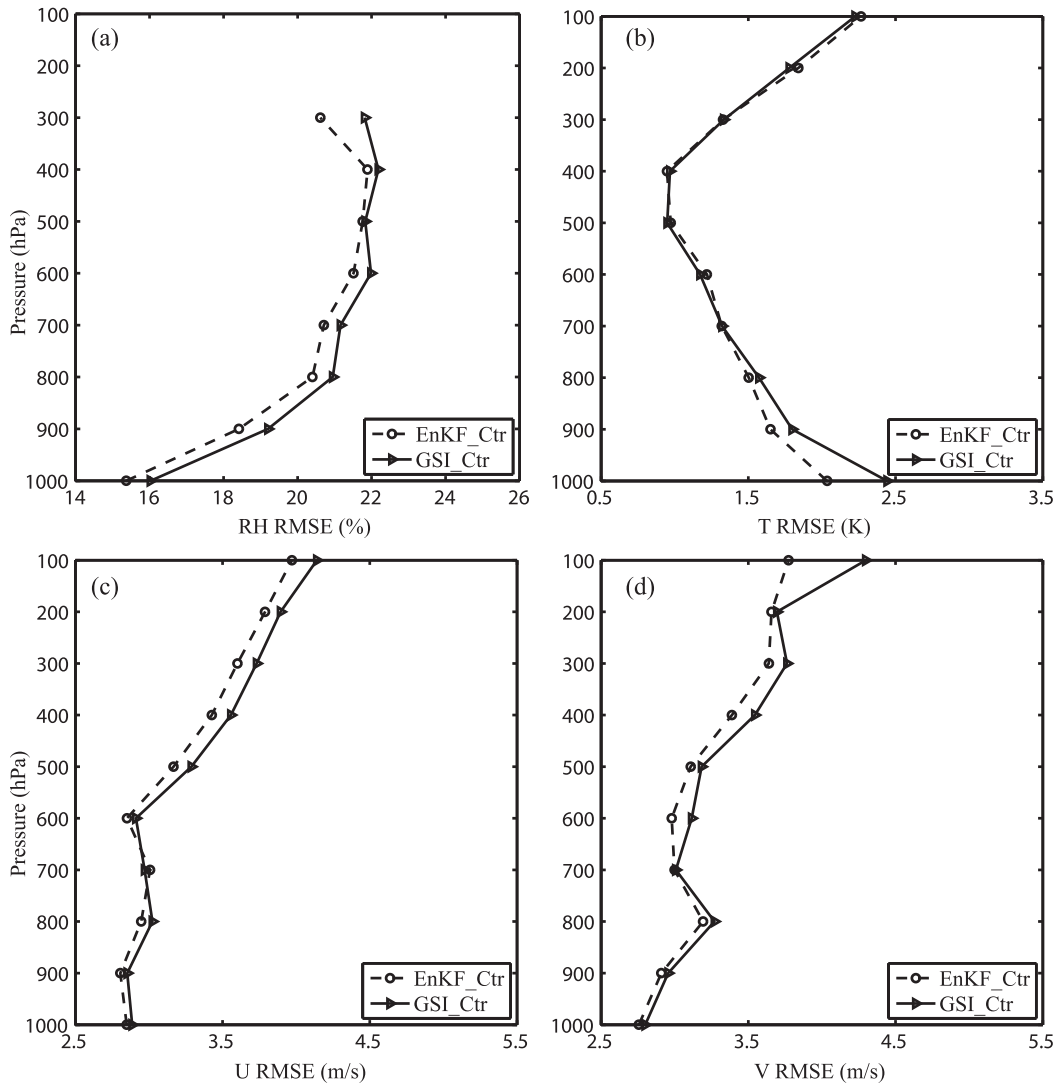


FIG. 5. Vertical profiles of 3-h forecast RMSEs from EnKF\_Ctr and GSI\_Ctr, verified against upper-air sounding data for (a) RH, (b)  $T$ , and the (c)  $U$  and (d)  $V$  wind components.

The improvement by the EnKF over GSI seems to be the largest for the RH field. Except for slightly higher RMSEs at the upper levels in later hours, EnKF\_Ctr maintains its advantage over GSI\_Ctr throughout the 18 h of forecast. For temperature, EnKF\_Ctr is clearly better at the lower levels, suggesting a better analysis of the boundary layer structures. Table 3 lists the relative percentage improvement (RPI) compared to GSI\_Ctr, defined as the RMSE of GSI\_Ctr minus the RMSE of an EnKF run normalized by the RMSE of GSI\_Ctr. A positive RPI indicates the improvement of EnKF over GSI. It can be seen that the temperature RMSEs are improved by 16.7% at 1000 hPa, the largest among the variables. For the wind components, the improvement is not as large as for temperature at the lower levels but is

generally positive, except for a slight degradation for  $V$  at the surface. For nearly all forecast hours and height levels, EnKF\_Ctr outperforms GSI\_Ctr (Table 3, Figs. 5 and 6).

The surface variables are verified against surface observations and shown in Fig. 7 for forecast hours 3–18. For 2-m  $T$  and 10-m  $U$ , EnKF\_Ctr is better than GSI\_Ctr at almost all forecast hours (see Fig. 7). However, for surface pressure, 2-m RH and 10-m  $V$ , EnKF\_Ctr is slightly worse than GSI\_Ctr. This may be due to the constant, rather large, 1000-km horizontal localization scale used in EnKF\_Ctr. In GSI, the horizontal covariance decorrelation scale is generally smaller near the surface but larger at the upper levels. The sensitivity to localization scale will be examined next.

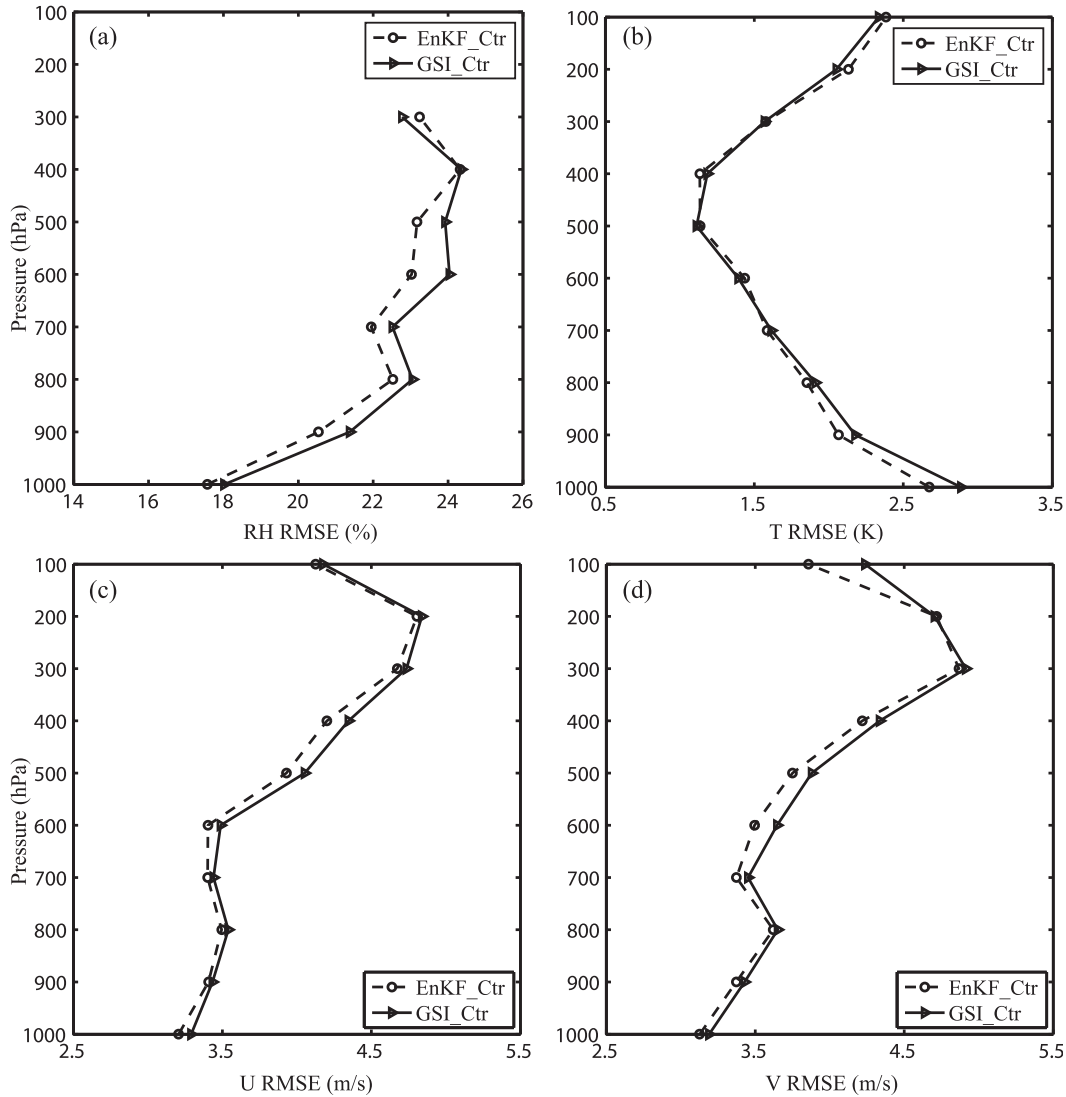


FIG. 6. As in Fig. 5, but for 18-h forecast RMSEs.

*c. Sensitivity to covariance inflation and localization*

The covariance inflation and localization factors are two of the most sensitive parameters in the EnKF. Unfortunately, their optimal values are difficult to determine theoretically. Their optimal values are related to the size of the ensemble and the number and type of observations assimilated, as well as the error characteristics of the forecast model. Other factors including the initial and lateral boundary conditions and their perturbations can also affect the optimal settings. The next set of experiments (EnKF\_L, EnKF\_S, EnKF\_AI, EnKF\_CtrHDL) shows a few of many tests that we have conducted in an attempt to determine the “optimal” settings.

Figure 7 shows the surface forecast RMSEs of these and the control experiments. Note that given the

essentially infinite degrees of freedom with the choice of inflation and localization factors and their functional forms, the “optimality” is only relative, and is only so within the parameter space examined.

As described earlier, our covariance inflation includes two parts. In EnKF\_Ctr, a 10% maximum static inflation [ $b = 0.1$  in Eq. (5)] plus an adaptive inflation with a spread recovery factor of 0.9 [ $c = 0.9$  in Eq. (6)] are used. Other values of  $b$  and  $c$  had been tried, with values of  $b$  up to 0.2 and  $c$  as small as 0.8. The values used by EnKF\_Ctr were found to give the best results. In EnKF\_AI, we turned off the static inflation. To compensate, we used a larger spread recovery factor of 0.98. It can be seen from Fig. 7 that EnKF\_AI performs the worst among all experiments in terms of the surface variable forecast errors. As we can see in Fig. 8, the

TABLE 3. RPI compared to experiment GSI\_Ctr for 3-h forecasts averaged over all forecast cycles throughout the experiment period. Columns A, B, and C are for experiments EnKF\_Ctr, EnKF\_CtrHDL, and EnKF\_Mult5, respectively. The “surface” row has the RPIs of 2-m RH, 2-m temperature, and 10-m  $U$  and  $V$  components.

Level (hPa)	RH (%)			$T$ (%)			$U$ (%)			$V$ (%)		
	A	B	C	A	B	C	A	B	C	A	B	C
100				-1.8	-2.4	2.3	4.1	6.1	3.2	12.2	14.0	11.6
200				-3.4	-2.3	-1.4	2.7	3.2	4.0	1.0	2.1	3.3
300	5.5	5.7	7.2	0.3	-0.3	-1.2	3.5	2.0	2.5	3.3	2.4	3.3
400	1.3	2.4	4.5	2.1	3.6	2.6	3.7	2.7	5.0	4.4	4.6	4.8
500	0.3	1.1	2.1	-2.5	0.9	1.2	3.7	4.7	6.4	2.4	2.0	2.8
600	2.1	4.0	4.5	-4.0	-4.3	-2.0	2.1	4.0	5.3	4.4	6.0	6.9
700	2.1	3.9	6.2	0.2	0.2	2.3	-1.3	3.0	3.9	0.3	5.7	5.2
800	2.6	4.2	5.4	4.4	3.9	5.8	2.5	4.8	6.5	2.4	5.4	5.4
900	4.1	7.2	9.4	8.2	9.5	10.4	1.7	4.0	5.1	1.6	3.7	5.0
1000	4.1	6.2	8.9	16.7	18.0	20.7	1.3	3.5	4.5	1.5	1.4	4.2
Surface	-1.4	1.0	3.6	6.7	7.6	8.1	0.4	2.0	3.3	-0.3	1.3	3.5

adaptive inflation alone cannot produce sufficient spread; the spreads for most variables are about half of those of other experiments. In addition to these tests, we also ran experiments employing other inflation schemes, such as that based on innovation statistics (Wang and Bishop 2003; Wang et al. 2007a), and were not able to obtain better results. Therefore, we consider the inflation settings used by EnKF\_Ctr optimal and will use them in all remaining experiments.

We saw earlier that EnKF\_Ctr outperforms GSI\_Ctr in verifications against soundings for all variables, except for some variables at the surface. We speculated that the horizontal covariance localization scale may be too large near the surface. To investigate this issue, EnKF\_S is run using a smaller horizontal localization radius of 800 (instead of 1000) km and  $\ln p_{\text{cut}} = 1$  (rather than 1.1) (Table 1). Simulation EnKF\_L uses 1200 km and  $\ln p_{\text{cut}} = 1.2$  ( $\sim 8.3$  km), respectively.

Figure 7 shows that EnKF\_S performs better than EnKF\_Ctr for all surface variables. Errors using the larger localization scales in EnKF\_L are the largest. Full domain verification against sounding data was also performed (not shown). For RH, the smaller localization scale does help. For temperature, EnKF\_Ctr and EnKF\_S have similar levels of performance, and both are better than EnKF\_L.

Although EnKF\_S has a somewhat better performance than EnKF\_Ctr, further reduction of the localization scale is not recommended. Figure 8 shows that the forecast spread increases as the localization scale decreases. The spread of EnKF\_S for wind components is approaching the forecast RMSEs near 300 hPa (cf. Fig. 6), which can be too large after accounting for the expected error in the observations used for the verification (here, the RMSEs are actually the observation innovations). The spread in some tests using further reduced localization scales was even larger; too large spread created by localized inflation can create

discontinuities in the model fields, causing model integration instability. In addition, for  $U$  and  $V$  verifications against soundings, EnKF\_S shows better performance than EnKF\_Ctr at the lower levels but not at the upper levels. For example, the 3-h forecast RMSEs of  $U$  over 500 hPa are worse in EnKF\_S than in EnKF\_Ctr (Fig. 9a). The flows at the upper levels are usually smoother than at the lower levels, corresponding to typically larger horizontal correlation scales. In fact, the background error matrix computed using the NMC method in the GSI system has larger correlation scales at the upper levels than at the lower levels. These results suggest that perhaps localization scales that increase with height can give better results. This is tested in EnKF\_CtrHDL.

In EnKF\_CtrHDL, localization scale parameters  $r_{\text{cut}}$  and  $\ln(p_{\text{cut}})$  in Eq. (2) are made height dependent;  $r_{\text{cut}}$  at the model top is set to 1.5 times the value at the surface (i.e., increasing from the 700 km at the surface to 1050 km at the model top), according to the same taper function used in Eq. (2) (Table 1); in the vertical  $\ln(p_{\text{cut}})$  is designed similarly but with additional observation type dependency. For RH and temperature, it is set to a quarter of 1.1 and half of 1.1 for the surface and model top, respectively. For  $U$  and  $V$ , the vertical correlation length is twice that for RH and temperature. We set  $\ln(p_{\text{cut}})$  to 1.6 for surface pressure observations. The choices of these settings were guided by the correlations scales in the NMC-method-derived error statistics used by GSI and were further tuned based on dozens of sensitivity tests. We have tried to use the same height-dependent vertical localization for all variables; the results were not as good as in EnKF\_CtrHDL.

Figure 7 shows the surface forecast RMSEs of EnKF\_CtrHDL. Compared to EnKF\_S, which is the best for surface variables among earlier experiments, EnKF\_CtrHDL has slightly larger RMSEs for surface pressure and temperature and similar RMSEs for surface  $V$ , but

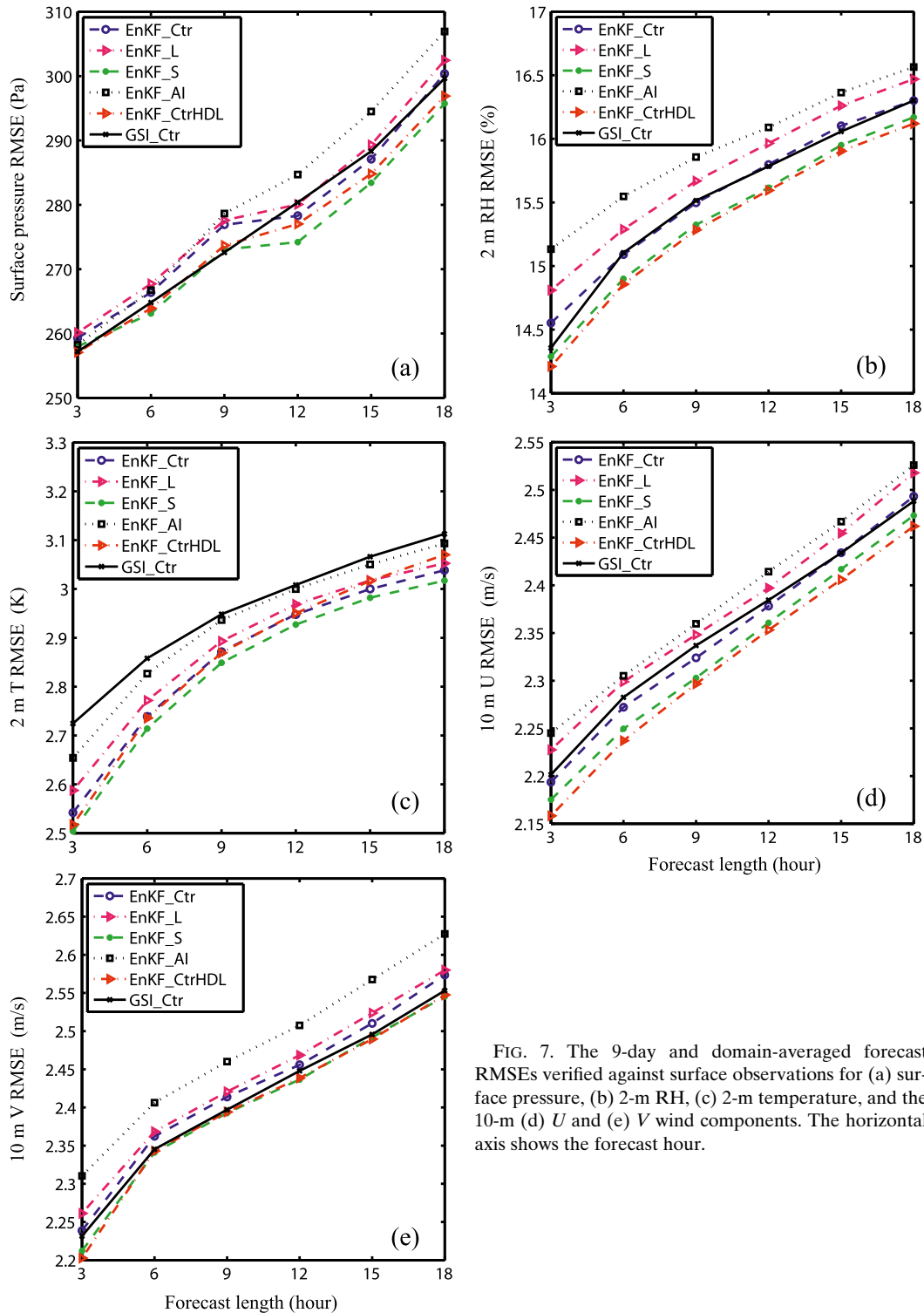


FIG. 7. The 9-day and domain-averaged forecast RMSEs verified against surface observations for (a) surface pressure, (b) 2-m RH, (c) 2-m temperature, and the 10-m (d)  $U$  and (e)  $V$  wind components. The horizontal axis shows the forecast hour.

smaller RMSEs for surface RH and  $U$ . The RPIs from EnKF\_CtrHDL using height-dependent localization are given in Table 3. The RPIs in the “surface” row for EnKF\_CtrHDL (column B) are all positive. In general,

EnKF\_CtrHDL is better than EnKF\_Ctr for surface variables.

For 3-h forecast RMSEs of  $U$  and  $V$  against soundings, EnKF\_CtrHDL outperforms EnKF\_S at all levels

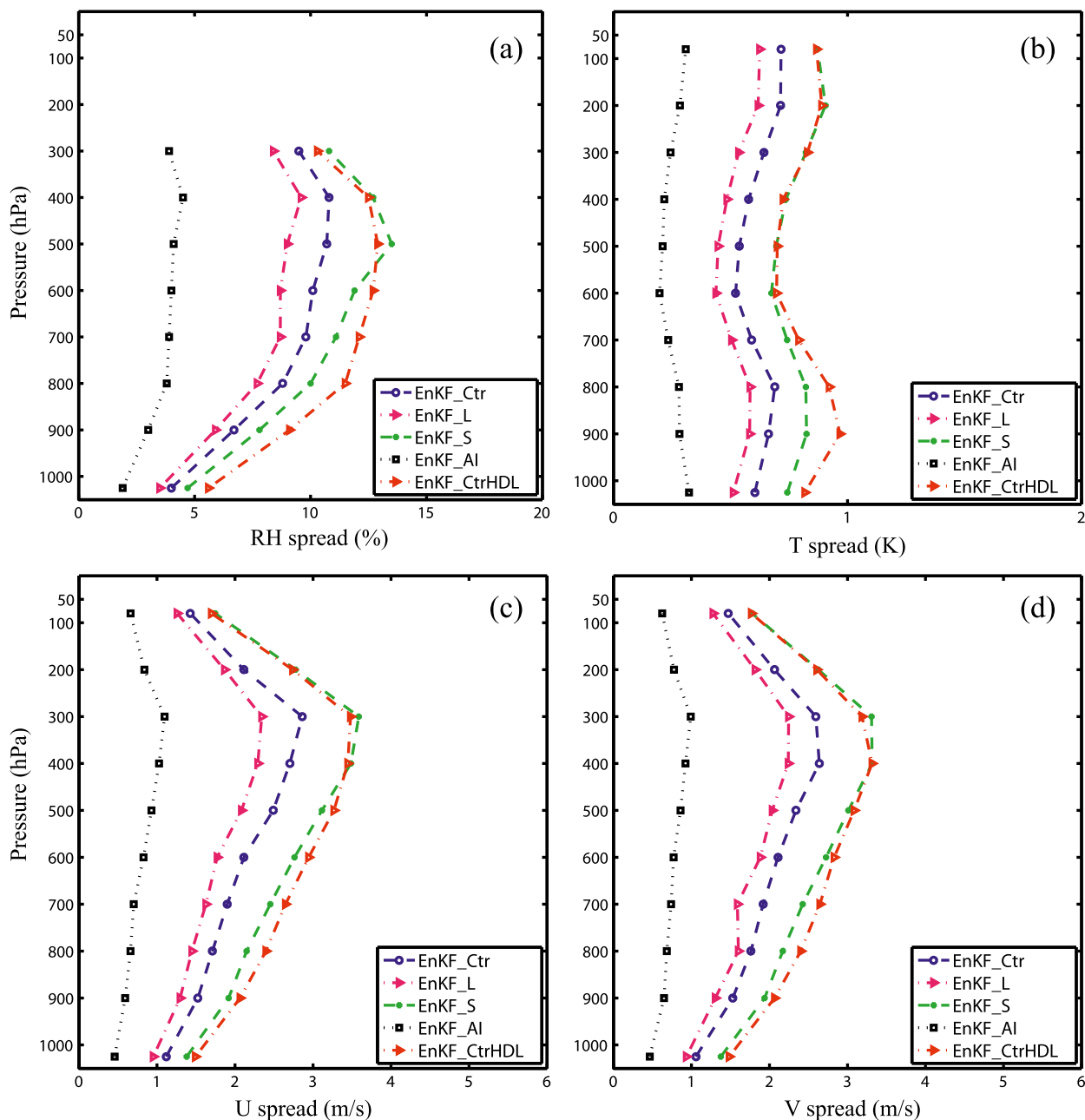


FIG. 8. Vertical profiles of 3-h forecast spread calculated in the sounding observation space (i.e., in terms of the observed variables) for (a) RH, (b) temperature, and the (c)  $U$  and (d)  $V$  wind components.

(Figs. 9a and 9b). Moreover, EnKF\_CtrHDL maintains its advantage over EnKF\_S throughout 18h of forecast (Figs. 9c and 9d). As discussed earlier, for the 3-h forecast RMSEs of the wind components, EnKF\_S is better than EnKF\_Ctr at the lower levels but slightly worse at the upper levels. The 3-h forecast RPIs for EnKF\_CtrHDL are mostly higher than those for EnKF\_Ctr (Table 3). For the wind components, there are larger improvements in EnKF\_CtrHDL than in EnKF\_Ctr for  $U$  and  $V$  at all

levels except for 300 and 400 hPa for  $U$  and 500 and 1000 hPa for  $V$ . For RH, the improvement in EnKF\_CtrHDL is always larger (column B versus column A under RH in Table 3), while for  $T$  there is larger improvement for about half of the levels. Overall, the height-dependent localization scheme, EnKF\_CtrHDL, improves the forecasts over the experiments using constant localizations. The height-dependent scheme is therefore used in two additional experiments with multiple physics suites.

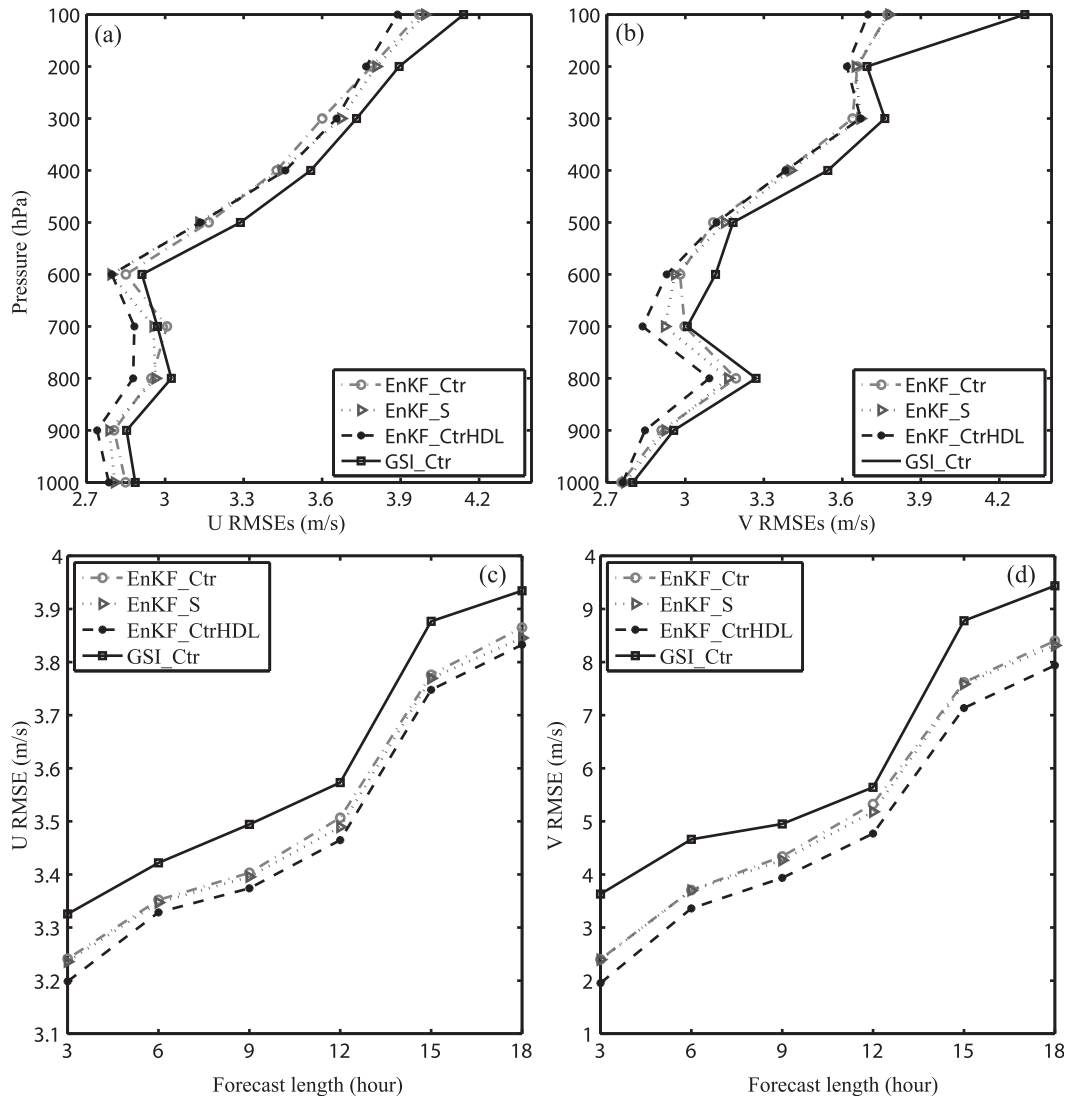


FIG. 9. Vertical profiles of 3-h forecast RMSEs verified against sounding data for the (a)  $U$  and (b)  $V$  wind components. (c),(d) As in (a),(b), but for domain-average RMSEs at different forecast hours.

*d. Impact of using multiple physics suites in forecast ensemble*

In the earlier tests, the effect of model errors on ensemble spread was considered indirectly through covariance inflation. Another approach to including model uncertainty, as investigated by, for example, Fujita et al. (2007), Meng and Zhang (2008b), and Snook et al. (2012), is to use different physics suites in the ensemble forecast model.

In this subsection, the impacts of using multiple physics schemes are examined. Compared to EnKF\_CtrHDL, which uses a single set of physics scheme, EnKF\_mult5 and EnKF\_mult10 have smaller surface RMSEs (Fig. 10). For example, for surface pressure

(Fig. 10a), EnKF\_CtrHDL has similar errors as GSI\_Ctr up to 9 h of forecast, but EnKF\_mult5 and EnKF\_mult10 clearly outperform GSI\_Ctr at all forecast hours. The improvement with the use of multiple physics is largest with surface temperature (Fig. 10b; see also Table 3) and the least with surface wind (Fig. 10c; see also Table 3). The use of 10 physics suites in EnKF\_mult10 produces generally smaller errors than using 5 suites, especially for the  $V$  wind component (Fig. 10e) but the differences here are generally smaller than the differences between the single- and multiple-physics approaches.

Figures 11a and 11b show the vertical profiles of 3-h-average RMSEs against soundings. Compared to the single-physics suite, the multiphysics runs do improve the 3-h RH and  $T$  forecasts at most levels. This can also



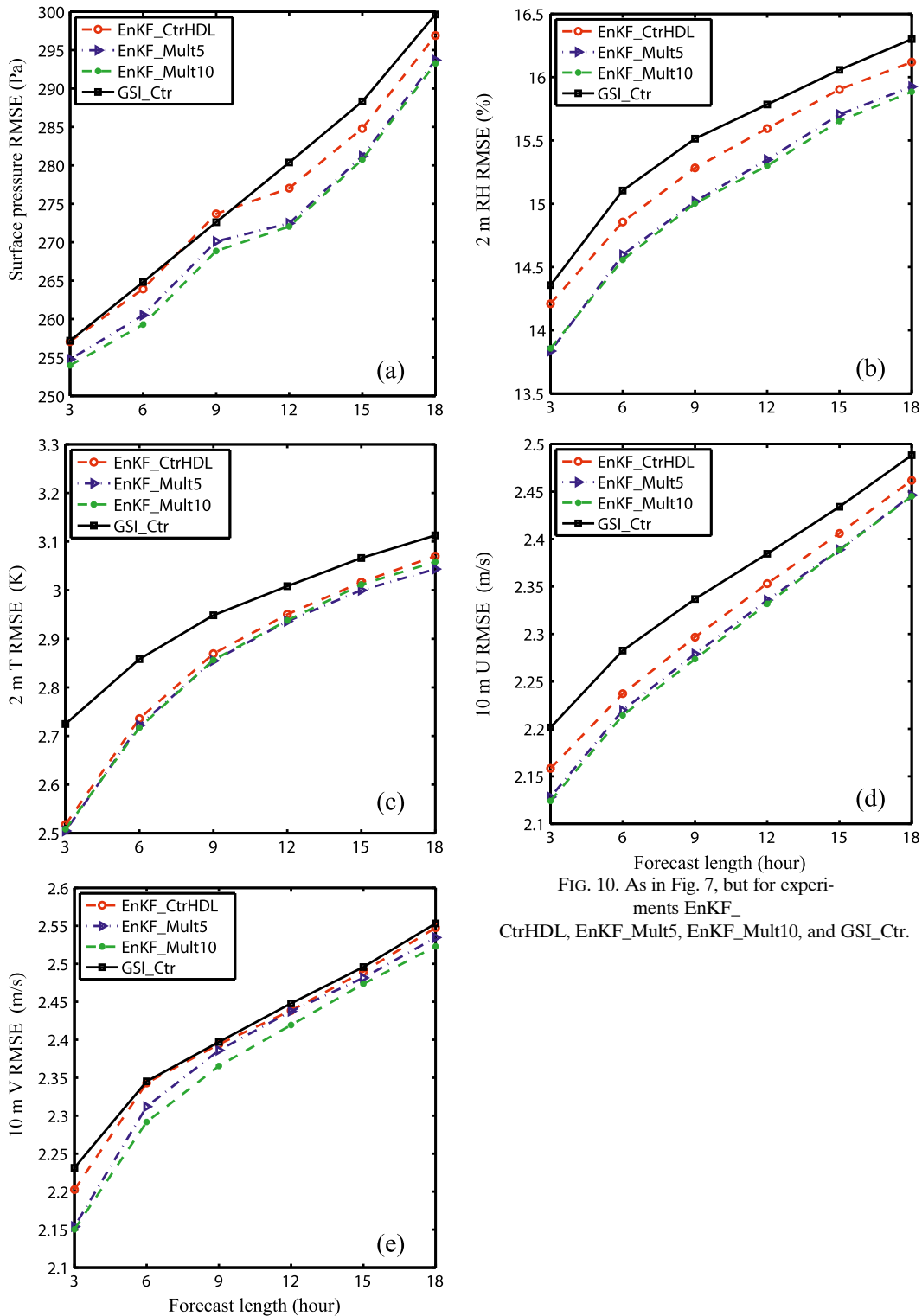


FIG. 10. As in Fig. 7, but for experiments EnKF\_CtrHDL, EnKF\_Mult5, EnKF\_Mult10, and GSI\_Ctr.

be seen in Table 3. For RH, column C (EnKF\_Mult5) has consistently larger RPIs than column B (EnKF\_CtrHDL). For temperature, the RPI of EnKF\_CtrHDL is 18% at 1000 hPa while that of EnKF\_Mult5 is 20.7%,

the largest value among all tests. Though  $T$  at some levels still has negative RPI, most levels are improved over EnKF\_CtrHDL. The multiple-physics suites give overall better performance than does the single-physics

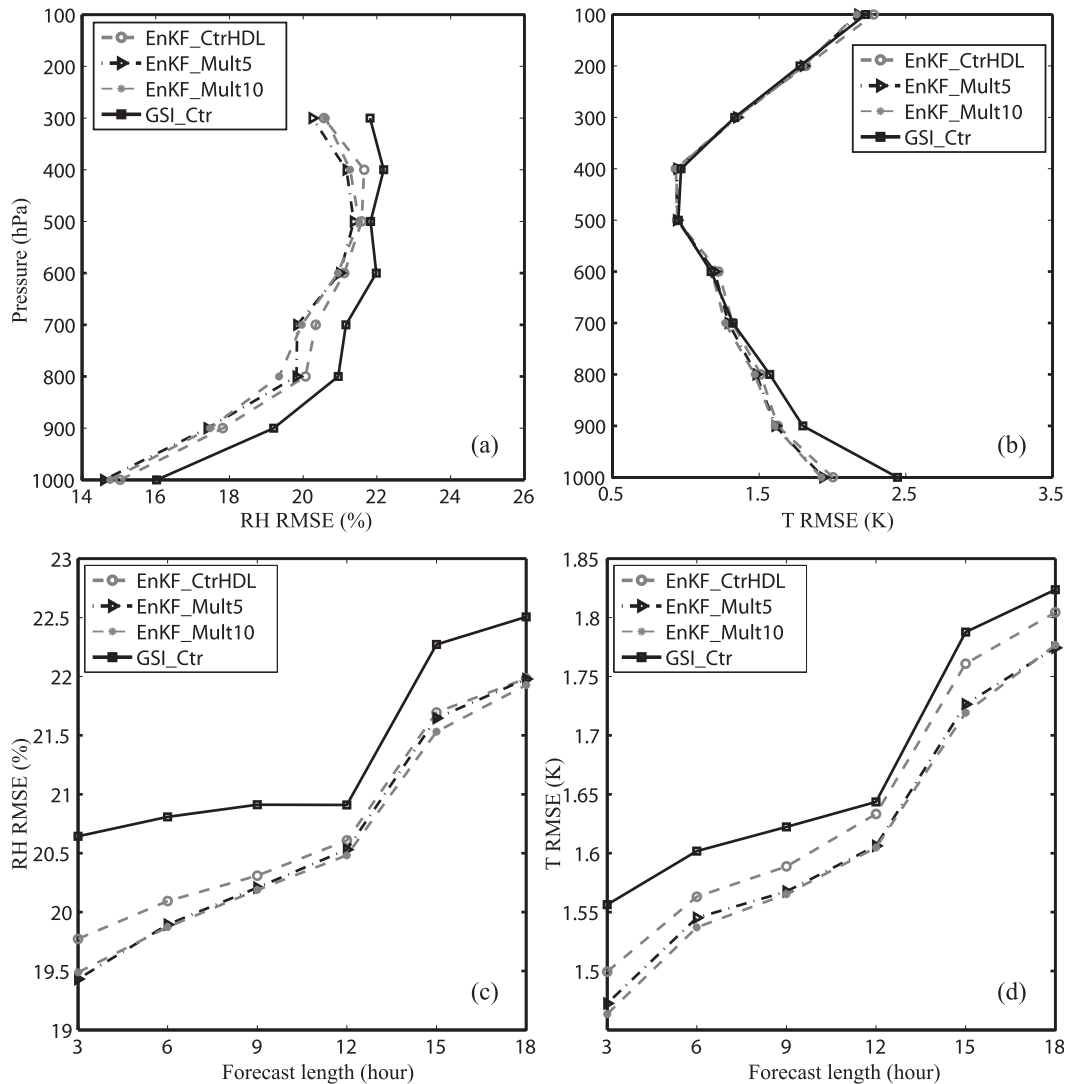


FIG. 11. As in Fig. 9, but for RH and temperature.

case throughout the 18 h of forecast for both RH and  $T$  (Figs. 11c and 11d), although the difference becomes smaller in later forecast hours for RH. For 3-h forecasts of the wind components, the multiple-physics suites improve the lower levels but not by as much as the upper levels (Table 3). In summary, the use of multiple-physics suites in the ensemble forecasts in the EnKF further improves the model forecasts over GSI compared to the use of a single-physics suite, making the EnKF-based forecasts better for almost all levels and at all forecast hours compared to the GSI-based forecasts.

### 5. Precipitation forecasts on a 13-km grid

In this section, precipitation forecasts on the 13-km grid initialized from the 40-km EnKF\_Mult5 and GSI\_Ctr

analyses are compared. For simplicity, in the following description, we refer to these experiments as EnKF and GSI, respectively. Due to the limitation of computer resources, we launched the forecasts only twice a day at 0000 and 1200 UTC. The precipitation forecasts are then verified against the NCEP stage IV precipitation data. GSSs calculated within the continental U.S. (CONUS) domain are presented.

Figure 12 shows the 1-h accumulated precipitation forecasts of two select cases during the testing period. The first case corresponds to the 10 May Oklahoma tornado outbreak. At a 13-km grid spacing, the convective storms are partially parameterized (using cumulus parameterization) and partially resolved by the grid. For both EnKF and GSI, the predicted precipitation is weaker than observed but EnKF is better than GSI at

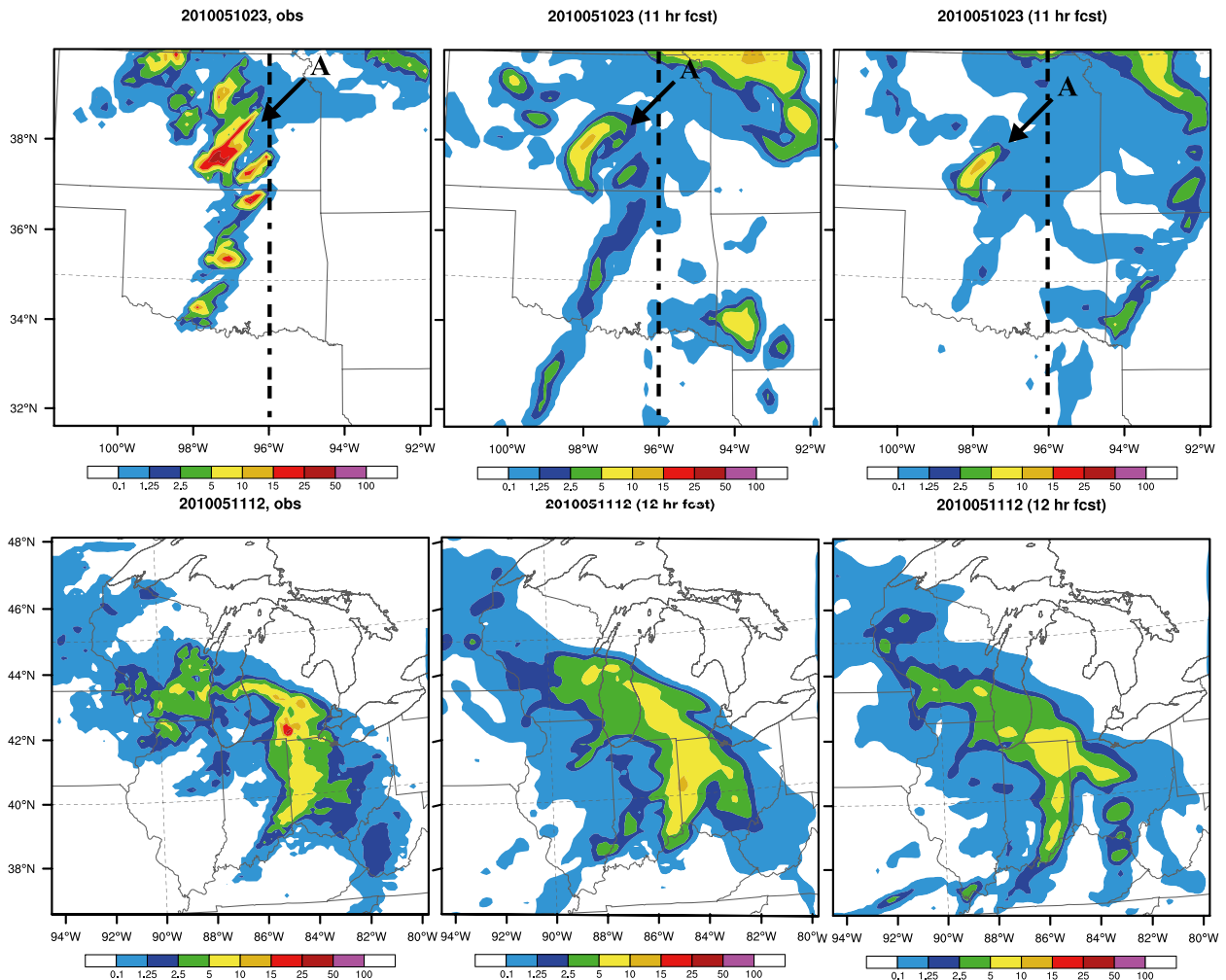


FIG. 12. Observed and forecast 1-h accumulated precipitation for select forecasts starting from two different times. (left) The NCEP stage IV precipitation; forecasts from (middle) EnKF\_Mult5 and (right) GSI\_Ctr analyses, with the two sets of forecasts starting at (top) 1200 UTC 10 May and (bottom) 0000 UTC 11 May 2010, at forecast lengths of 11 and 12 h, respectively.

predicting the locations and intensity of precipitation in general. For the 10 May case and at the 11th forecast hour shown (close to the times of tornado outbreaks associated with a series of supercell storms through central Oklahoma), EnKF predicts a stronger line of precipitation than GSI, at roughly the location of the convective line within Oklahoma (first row in Fig. 12). It also produces a larger convective storm labeled A in south-southwest Kansas that verifies better than GSI.

The second case is a fast-moving mesoscale convective system (MCS), which first formed in the central United States and propagated to the northeast. In this case, both EnKF and GSI predicted precipitation structures similar to the observations, but overall the pattern of stronger precipitation (above  $5 \text{ mm h}^{-1}$ ) is better in EnKF at 12 h (second row of Fig. 12). These examples illustrate the kinds of differences found between forecasts

with EnKF and GSI DA. The rather coarse 40-km resolution used by the data assimilation might have limited the impact of flow-dependent error covariance in EnKF, but the impact is clearly positive.

Figure 13 shows the GSSs averaged over all cases for thresholds of 0.1, 1.25, and 2.5 mm. Since EnKF and GSI share the same initial conditions at time 0000 UTC 8 May 2010, the statistics start at 0900 UTC. It can be seen that EnKF outperforms GSI on average for all forecast hours and thresholds shown. Therefore, the EnKF analyses performed on the 40-km grid, using multiple physics in the ensemble, improve precipitation forecasts produced on a 13-km grid.

## 6. Summary and discussion

This paper describes a newly developed regional EnKF DA system based on the operational GSI 3DVAR

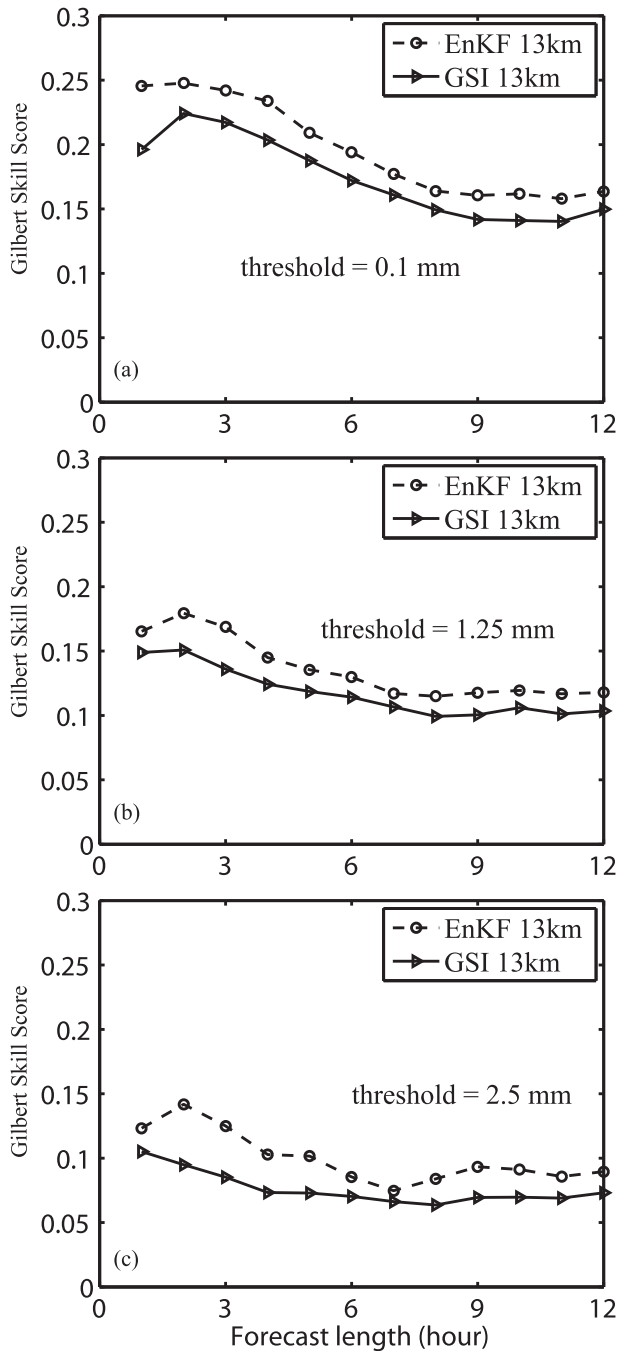


FIG. 13. Average hourly precipitation GSSs of all 13-km forecasts at different forecast lengths for thresholds of (a) 0.1, (b) 1.25, and (c) 2.5 mm h<sup>-1</sup>.

method, as well as its configuration and tuning for potential Rapid Refresh operational application. A 9-day-period containing many active convection events is selected to examine the performance of such a system through comparisons with parallel experiments using the GSI. The EnKF and GSI experiments use the same

observational datasets as the operational RAP system except that GPS PW, satellite radiance, and radar data are not included in the results presented here. To keep the computational cost manageable for potential operational implementation, the experiments were performed on a grid with a spacing of ~40 km and with assimilation intervals of 3 h. The performance of the DA systems is evaluated based on forecast RMSEs calculated against surface-state observations and upper-air sounding data, for forecasts ranging from 3 to 18 h.

The results indicate that the control EnKF experiment EnKF\_Ctr using fixed covariance localization radii of 1000 km in the horizontal and 1.1 in terms of logarithmic pressure (in hPa) (about 7.6 km) in the vertical, and also using a single suite of RAP physics parameterizations in the ensemble, is generally better than the GSI control experiment GSI\_Ctr in terms of the domain-averaged forecast errors throughout the 18 h of forecast but slightly worse in the forecast of some surface variables. Sensitivity experiments show that when the localization radii are reduced to 800 km and 1.0 in the horizontal and vertical, respectively, in EnKF\_S, errors at the lower levels are reduced but errors in the upper-level wind components become slightly larger. A height-dependent localization scheme in which the localization radii decrease with height was designed, and it led to smaller errors at both lower and upper levels in experiment EnKF\_CtrHDL. To better consider the model uncertainty, we further tested the use of multiple-physics suites in EnKF\_mult5 and EnKF\_mult10. The use of 5 or 10 physics suites among the 40 ensemble members in the forecast step further improves the EnKF performance, leading to smaller forecast errors for all variables at all model levels and forecast hours compared to GSI, and the use of 10 physics suites slightly outperforms 5 physics suites. The percentage improvements over GSI in terms of the 3-h forecast RMSEs indicate that the improvement is generally increased by the use of height-dependent localization and multiple physics suites (Table 3). The largest increment is 20.7% for temperature at 1000 hPa when multiple physics is used.

Forecasts were launched on a 13-km grid initialized from interpolated EnKF ensemble mean analyses at 0000 and 1200 UTC of each day. Hourly accumulated precipitation is better predicted in the EnKF experiments (results using five physics suites were shown) than the GSI experiment, in terms of both the predicted precipitation structures and Gilbert skill scores of hourly accumulated precipitation.

Despite the encouraging results, the EnKF system still should have room for further improvement. Future experiments will include additional GPS PW and satellite radiance data. For satellite radiance data, bias correction

appears to be an important issue and the positive impact from including the data over the data-dense continental United States is not automatic, as suggested by our earlier preliminary tests. For the eventual operationally implementation, it will be desirable to perform the EnKF DA at a higher resolution, either completely on the 13-km RAP grid (at a higher computational cost), or using dual resolutions where the ensemble forecasts remain at the lower resolution while the ensemble mean analysis update and a control forecast are performed at the native 13-km resolution. We expect even greater impacts from the flow-dependent error covariances in the EnKF when more transient convective features are captured by the higher-resolution grid. Efforts are also under way to implement and test a hybrid DA system using the EnKF to provide ensemble perturbations to a variational analysis system.

*Acknowledgments.* This work was primarily supported by the FAA Aviation Weather Research Program through Grant DOC-NOAA NA080AR4320904, and by MOST of China Grant 2013CB430100. Additional support was provided by NSF Grants AGS-0802888, OCI-0905040, AGS-0941491, AGS-1046171, and AGS-1046081, and by the NOAA Warn-on-Forecast program. The fourth author is also supported by NOAA THOPREX NA08OAR4320904 and NASA NIP NNX10AQ78G.

#### REFERENCES

- Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805–1824.
- , —, and —, 2010: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part II: Short-range ensemble Forecasts. *Mon. Wea. Rev.*, **138**, 1273–1292.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- , 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83.
- , and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.
- Barker, D. M., 2005: Southern high-latitude ensemble data assimilation in the Antarctic Mesoscale Prediction System. *Mon. Wea. Rev.*, **133**, 3431–3449.
- Benjamin, S. G., and Coauthors, 2004: An hourly assimilation-forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495–518.
- , and Coauthors, 2009: Technical review of rapid refresh/RUC project. NOAA/ESRL/GSD Internal Review. [Available online at [http://ruc.noaa.gov/pdf/RR-RUC-TR\\_11\\_3\\_2009.pdf](http://ruc.noaa.gov/pdf/RR-RUC-TR_11_3_2009.pdf).]
- , B. D. Jamison, W. R. Moninger, S. R. Sahn, B. E. Schwartz, and T. W. Schlatter, 2010: Relative short-range forecast impact from aircraft, profiler, radiosonde, VAD, GPS-PW, METAR, and Mesonet observations via the RUC hourly assimilation cycle. *Mon. Wea. Rev.*, **138**, 1319–1343.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- Bonavita, M., L. Torrisi, and F. Marcucci, 2008: The ensemble Kalman filter in an operational regional NWP system: Preliminary results with real observations. *Quart. J. Roy. Meteor. Soc.*, **134**, 1733–1744.
- , L. Isaksen, and E. Hólm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **138**, 1540–1559.
- Brown, B. G., J. H. Gotway, R. Bullock, E. Gilleland, and D. Ahijevych, 2009: The Model Evaluation Tools (MET): Community tools for forecast evaluation. Preprints, *25th Conf. on Int. Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Phoenix, AZ, Amer. Meteor. Soc., 9A.6. [Available online at <https://ams.confex.com/ams/pdfpapers/151349.pdf>.]
- 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.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Chen, M., and X. Y. Huang, 2006: Digital filter initialization for MM5. *Mon. Wea. Rev.*, **134**, 1222–1236.
- Courtier, P., and Coauthors, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quart. J. Roy. Meteor. Soc.*, **124**, 1783–1808.
- Dowell, D., 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.
- , and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, **128**, 1852–1867.
- 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.
- Gandin, L. S., and A. H. Murphy, 1992: Equitable skill scores for categorical forecasts. *Mon. Wea. Rev.*, **120**, 361–370.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.
- Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.
- , J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.*, **139**, 668–688.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- , and —, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269–3289.
- Huang, X.-Y., and P. Lynch, 1993: Diabatic digital-filtering initialization: Application to the HIRLAM model. *Mon. Wea. Rev.*, **121**, 589–603.

- Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126.
- Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691–1705.
- Lei, T., M. Xue, and T. Yu, 2009: Multi-scale analysis and prediction of the 8 May 2003 Oklahoma City tornadic supercell storm assimilating radar and surface network data using EnKF. Preprints, *13th Conf. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*, Phoenix, AZ, Amer. Meteor. Soc., 6.4. [Available online at <https://ams.confex.com/ams/pdfpapers/150404.pdf>.]
- Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. Preprints, *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2. [Available online at <https://ams.confex.com/ams/pdfpapers/83847.pdf>.]
- Lorenc, A., 2003: Modelling of error covariances by 4DVAR data assimilation. *Quart. J. Roy. Meteor. Soc.*, **129**, 3167–3182.
- Lynch, P., 1997: The Dolph–Chebyshev window: A simple optimal filter. *Mon. Wea. Rev.*, **125**, 655–660.
- , and X.-Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.*, **120**, 1019–1034.
- , and X. Y. Huang, 1994: Diabatic initialization using recursive filters. *Tellus*, **46A**, 583–597.
- Meng, Z. Y., and F. Q. Zhang, 2008a: 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.
- , and —, 2008b: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with 3DVAR in a month-long experiment. *Mon. Wea. Rev.*, **136**, 3671–3682.
- , and —, 2011: Limited-area ensemble-based data assimilation. *Mon. Wea. Rev.*, **139**, 2025–2045.
- Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428.
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003: Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances. *Mon. Wea. Rev.*, **131**, 1536–1548.
- Rabier, F., H. Jarvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143–1170.
- Raynaud, L., L. Berre, and G. Desroziers, 2011: An extended specification of flow-dependent background error variances in the Météo-France global 4D-Var system. *Quart. J. Roy. Meteor. Soc.*, **137**, 607–619.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. D. Powers, 2005: A description of the Advanced Research WRF version 2. NCAR Tech. Note NCAR/TN-468+STR, 88 pp. [Available online at [http://www.mmm.ucar.edu/wrf/users/docs/arw\\_v2.pdf](http://www.mmm.ucar.edu/wrf/users/docs/arw_v2.pdf).]
- Snook, N., M. Xue, and Y. Jung, 2012: Ensemble probabilistic forecasts of a tornadic mesoscale convective system from ensemble Kalman filter analyses using WSR-88D and CASA radar data. *Mon. Wea. Rev.*, **140**, 2126–2146.
- Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677.
- Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus*, **60A**, 113–130.
- Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.*, **133**, 1789–1807.
- Torn, R. D., and G. J. Hakim, 2008: Performance characteristics of a pseudo-operational ensemble Kalman filter. *Mon. Wea. Rev.*, **136**, 3947–3963.
- , —, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490–2502.
- Wang, S., M. Xue, and J. Min, 2013: A four-dimensional asynchronous ensemble square-root filter (4DENSURF) and tests with simulated radar data. *Quart. J. Roy. Meteor. Soc.*, **139A**, 805–819.
- Wang, X. G., 2010: Incorporating ensemble covariance in the Gridpoint Statistical Interpolation variational minimization: A mathematical framework. *Mon. Wea. Rev.*, **138**, 2990–2995.
- , and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158.
- , T. A. Hamill, J. S. Whitaker, and C. H. Bishop, 2007a: A comparison of hybrid ensemble transform Kalman filter–optimum interpolation and ensemble square root filter analysis schemes. *Mon. Wea. Rev.*, **135**, 1055–1076.
- , C. Snyder, and T. M. Hamill, 2007b: On the theoretical equivalence of differently proposed ensemble/VAR hybrid analysis schemes. *Mon. Wea. Rev.*, **135**, 222–227.
- , D. M. Barker, C. Snyder, and T. M. Hamill, 2008: A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiment. *Mon. Wea. Rev.*, **136**, 5132–5147.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- , and —, 2010: A simple state-dependent multiplicative inflation algorithm. *Fourth Ensemble Kalman Filter Workshop*, Rensselaerville, NY, The Pennsylvania State University. [Available online [http://hftp.psu.edu/EDA2010/Whitaker\\_inflation.pdf](http://hftp.psu.edu/EDA2010/Whitaker_inflation.pdf).]
- , —, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482.
- Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916.
- Xue, M., D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Phys.*, **82**, 139–170.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observations on the 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.