

Tests of an Ensemble Kalman Filter for Mesoscale and Regional-Scale Data Assimilation. Part I: Perfect Model Experiments

FUQING ZHANG, ZHIYONG MENG, AND ALTUG AKSOY

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

(Manuscript received 4 August 2004, in final form 25 July 2005)

ABSTRACT

Through observing system simulation experiments, this two-part study exploits the potential of using the ensemble Kalman filter (EnKF) for mesoscale and regional-scale data assimilation. Part I focuses on the performance of the EnKF under the perfect model assumption in which the truth simulation is produced with the same model and same initial uncertainties as those of the ensemble, while Part II explores the impacts of model error and ensemble initiation on the filter performance. In this first part, the EnKF is implemented in a nonhydrostatic mesoscale model [the fifth-generation Pennsylvania State University–NCAR Mesoscale Model (MM5)] to assimilate simulated sounding and surface observations derived from simulations of the “surprise” snowstorm of January 2000. This is an explosive East Coast cyclogenesis event with strong error growth at all scales as a result of interactions between convective-, meso-, and subsynoptic-scale dynamics.

It is found that the EnKF is very effective in keeping the analysis close to the truth simulation under the perfect model assumption. The EnKF is most effective in reducing larger-scale errors but less effective in reducing errors at smaller, marginally resolvable scales. In the control experiment, in which the truth simulation was produced with the same model and same initial uncertainties as those of the ensemble, a 24-h continuous EnKF assimilation of sounding and surface observations of typical temporal and spatial resolutions is found to reduce the error by as much as 80% (compared to a 24-h forecast without data assimilation) for both observed and unobserved variables including zonal and meridional winds, temperature, and pressure. However, it is observed to be relatively less efficient in correcting errors in the vertical velocity and moisture fields, which have stronger smaller-scale components. The analysis domain-averaged root-mean-square error after 24-h assimilation is $\sim 1.0\text{--}1.5\text{ m s}^{-1}$ for winds and $\sim 1.0\text{ K}$ for temperature, which is comparable to or less than typical observational errors. Various sensitivity experiments demonstrated that the EnKF is quite successful in all realistic observational scenarios tested. However, as will be presented in Part II, the EnKF performance may be significantly degraded if an imperfect forecast model is used, as is likely the case when real observations are assimilated.

1. Introduction

The ensemble-based data assimilation method [ensemble Kalman filter (EnKF); Evensen 1994], which uses short-term ensemble forecasts to estimate the flow-dependent background error covariance, has recently been implemented in various atmospheric and oceanic models. These models vary from idealized examples based on simplified equation sets to those based on the complete, primitive equations with assimilation of real observations (Houtekamer and Mitchell 1998,

2001; Hamill and Snyder 2000; Keppenne 2000; Anderson 2001; Mitchell et al. 2002; Keppenne and Rienecker 2002; Whitaker and Hamill 2002; Zhang and Anderson 2003; Snyder and Zhang 2003; Houtekamer et al. 2005; Whitaker et al. 2004; Dowell et al. 2004; Zhang et al. 2004; Aksoy et al. 2005). These experimental studies demonstrated the feasibility and effectiveness of the EnKF for different scales and flows of interest and the advantages of using the EnKF over existing data assimilation schemes, which assume stationary, isotropic background error covariance. This present study seeks to exploit the potential of using the EnKF to assimilate simulated sounding and surface observations for mesoscale and regional-scale numerical weather prediction systems, which often include dynamics and interactions among convective, meso-, and subsynoptic scales.

Corresponding author address: Dr. Fuqing Zhang, Dept. of Atmospheric Sciences, Texas A&M University, 3150 TAMU, College Station, TX 77843-3150.
E-mail: fzhang@tamu.edu

Recently, short-term ensemble forecasts generated with different sets of initial perturbations were used to examine the dynamics and structure of mesoscale error covariance of the 24–25 January 2000 surprise snowstorm (Zhang 2005). In the ensemble forecast initiated with rescaled random perturbations, initial errors grow from smaller-scale, largely unbalanced and uncorrelated perturbations to larger-scale, quasi-balanced disturbances within 12–24 h. Comparable ensemble spread is found in ensemble forecasts initialized with balanced random perturbations or with gridpoint random perturbations. In all ensemble forecasts, the error growth is maximized in the vicinity of the strongest mean potential vorticity (PV) gradient and over the area of active moist convection, consistent with the lower predictability in these regions (Zhang et al. 2002, 2003). Consequently, the initially largely uncorrelated, mostly random errors evolve into strong coherent structures with spatial correlation not only within individual variables (autocovariance) but also between different forecast variables (cross covariance), especially over the region of strong cyclogenesis and along the upper-level front. The error covariance is highly anisotropic. Dramatic differences in magnitude, structure, and sign are found between covariances estimated from the same set of ensemble forecasts but verified at different times. The structure of the mesoscale error covariance is ultimately determined by the underlying governing dynamics and the associated error growth.

The spatial and cross covariance estimated from the short-term ensemble forecast has the potential to spread observational information nonuniformly to both observed and unobserved variables at different vertical layers (e.g., from the upper troposphere to the surface and vice versa) with a horizontal radius of influence potentially greater than 1000 km. The flow-dependent nature of the error growth dynamics and the covariance structure further demonstrates the necessity to use anisotropic and flow-dependent representations of background error covariance for mesoscale and regional-scale data assimilation.

The current study seeks to examine the significance and the effectiveness of the error covariance estimated from the short-term ensemble forecasts for mesoscale and regional-scale data assimilation for the same event as in Zhang (2005). Section 2 introduces the forecast model and the formulation and configuration of the EnKF. The truth simulation and the reference forecast ensemble are presented in section 3. Performance of the control EnKF experiment is examined in section 4. Forecast error growth from ensembles with and without the EnKF is discussed in section 5. The sensitivity experiments to EnKF configuration, data coverage, fre-

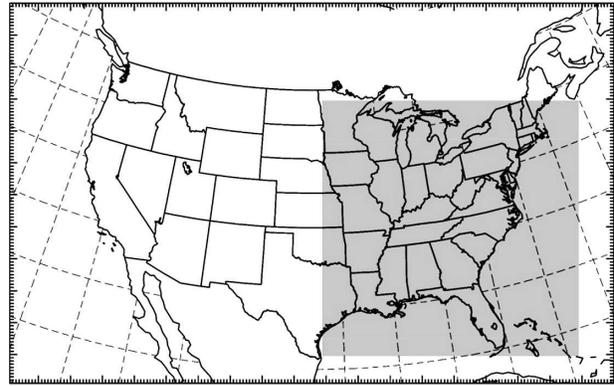


FIG. 1. Map of the model domain. Only data in shaded areas are assimilated and analyzed.

quency, and uncertainty of observations are presented in section 6. Summary and conclusions are presented in section 7. The impacts of model error and ensemble initiation on the filter performance will be explored in Meng and Zhang (2005; manuscript submitted to *Mon. Wea. Rev.*, hereafter Part II).

2. Forecast model and EnKF

The study uses the nonhydrostatic fifth-generation Pennsylvania State University–National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5) (Dudhia 1993). The model domain has 190×120 horizontal grid points with 30-km grid spacing and covers the continental United States (Fig. 1). There are 27 layers in the terrain-following vertical coordinate with model top at 100 hPa and vertical spacing smallest within the boundary layer. The model has a total of 10 prognostic variables including three Cartesian velocity components (u , v , w), pressure perturbation (p'), temperature (T), and mixing ratios for water vapor (q), cloud water (q_c), rainwater (q_r), cloud ice (q_i), and graupel (q_g). Details and references on the model configuration can be found in Zhang et al. (2002, hereafter ZSR02). The state dimension of the forecast model is $\sim 10^7$. Observations are taken only from the shaded area in Fig. 1 and only state vectors in this inner box are updated and analyzed.

The EnKF was first proposed for geophysical applications by Evensen (1994). The implementation of the EnKF used in the current study follows closely that of Snyder and Zhang (2003). As in standard Kalman filter,

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^f), \quad (1)$$

where \mathbf{x}^f represents the prior estimate or first guess, \mathbf{x}^a is the posterior estimate or analysis, \mathbf{y} is the observation vector, \mathbf{H} is the observation operator that returns ob-

served variables given the state, and \mathbf{K} is the so-called Kalman gain matrix defined as

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1}, \quad (2)$$

where \mathbf{P}^f and \mathbf{R} represent the background and observational error covariance, respectively. In the EnKF, the flow-dependent \mathbf{P}^f is estimated through an ensemble of short-range forecasts. Observations are taken sequentially with uncorrelated observations errors. Further background on the EnKF can be found in Snyder and Zhang (2003) and references therein.

One modification to Snyder and Zhang (2003) is to adopt the covariance relaxation method proposed in Zhang et al. (2004) in which a modified analysis deviation ($\mathbf{x}_{\text{new}}^a$)' is computed by “relaxing” or weighting (\mathbf{x}^a)' and (\mathbf{x}^f)':

$$(\mathbf{x}_{\text{new}}^a)' = (1 - \alpha)(\mathbf{x}^a)' + \alpha(\mathbf{x}^f)', \quad (3)$$

where deviations from the mean are denoted by primes, and $\alpha = 0.5$ is used in this study. The modified analysis deviations are then used as initial conditions for the ensemble forecasts to the next assimilation time. Since the analysis (posterior) deviation (\mathbf{x}^a)' is smaller than the forecast (prior) deviation (\mathbf{x}^f)', reflecting the reduction of uncertainty after assimilating observations, the use of (3) will overestimate (inflate) the uncertainty in the analysis, as an alternative to the covariance inflation used by Anderson (2001). It is worth noting that, both the covariance relaxation using Eq. (3) and the covariance inflation of Anderson (2001) are ad hoc ways of dealing with the tendency of the spread of a small ensemble to underestimate the true error of the ensemble mean. Another alternative is to use an EnKF configuration with a pair of ensembles (Houtekamer and Mitchell 1998; Houtekamer et al. 2005), which does not require any “adjustable” inflation or relaxation parameters.

In addition, a covariance localization method using the Gaspari and Cohn (1999) compactly supported fifth-order correlation function is performed in the full three-dimensional physical space. The covariance is set to be zero if the total gridpoint distance [$r = (r_x + r_y + r_\sigma)^{0.5}$, where r_x , r_y , and r_σ are distance in number of grid points in the x , y , and σ (or z) directions, respectively] is greater than 30, equivalent to a horizontal distance of 900 km.

3. The truth simulation and the reference forecast ensemble

The truth simulation and the reference forecast ensemble are produced by randomly perturbing the ref-

erence analysis at 0000 UTC 24 January 2000. The perturbations used are directly derived from the background error covariance of the MM5 three-dimensional variational data assimilation (3DVAR) system (Barker et al. 2004). The reference analysis is generated using the National Centers for Environmental Prediction (NCEP)–NCAR reanalysis. The MM5 3DVAR analysis (and thus the initial perturbations) is performed on a transformed streamfunction field (Barker et al. 2003). Forty random perturbations of the streamfunction, which are consistent with the background error covariance used by the MM5 3DVAR system, are selected and then transformed to derive the horizontal wind (u and v), temperature (T), and pressure perturbations (p') (Barker et al. 2003, 58–59). The derived initial wind, temperature, and pressure perturbations are thus geostrophically balanced. The use of the 3DVAR background error covariance to generate the initial ensemble for the EnKF can also be found in Houtekamer et al. (2005). The domain-averaged standard deviation (STD) of such perturbations is approximately 1 m s^{-1} for u and v , 0.5 K for T , 0.4 hPa for p' and 0.2 g kg^{-1} for q . Other prognostic variables (vertical wind w , mixing ratios of cloud water q_c , rainwater q_r , snow q_s , and graupel q_g) are not perturbed in the MM5 3DVAR system used here. These perturbations are then added to the reference analysis at 0000 UTC 24 January 2000 to generate a 40-member reference forecast ensemble that is integrated for 36 h with boundary conditions provided by the NCEP–NCAR reanalysis updated every 12 h.

The truth simulation is generated in the same manner (i.e., the same model and the same initial uncertainties) as one of the members of the reference forecast ensemble but with a different realization of random perturbations. The truth simulation is used to generate observations and is also used as the reference to evaluate the performance of the EnKF. Only state vectors and observations in the shaded region of Fig. 1, an area of $2400 \text{ km} \times 2400 \text{ km}$, are analyzed and assimilated. Selection of the shaded areas (instead of the total model domain) as the analysis domain is to minimize the impact of using the same boundary conditions for the integration of both the truth and the reference forecast ensemble. Over the 36-h integration, state variables inside the shaded domain have little influence from the model lateral boundary conditions.

Figure 2 shows the mean sea level pressure (MSLP) and model-derived reflectivity at the 12-, 24-, and 36-h forecast times from the truth simulation (upper panels) and the reference forecast ensemble mean (lower panels). Corresponding geopotential heights, PV, and vector winds at 300 hPa are displayed in Fig. 3. The truth

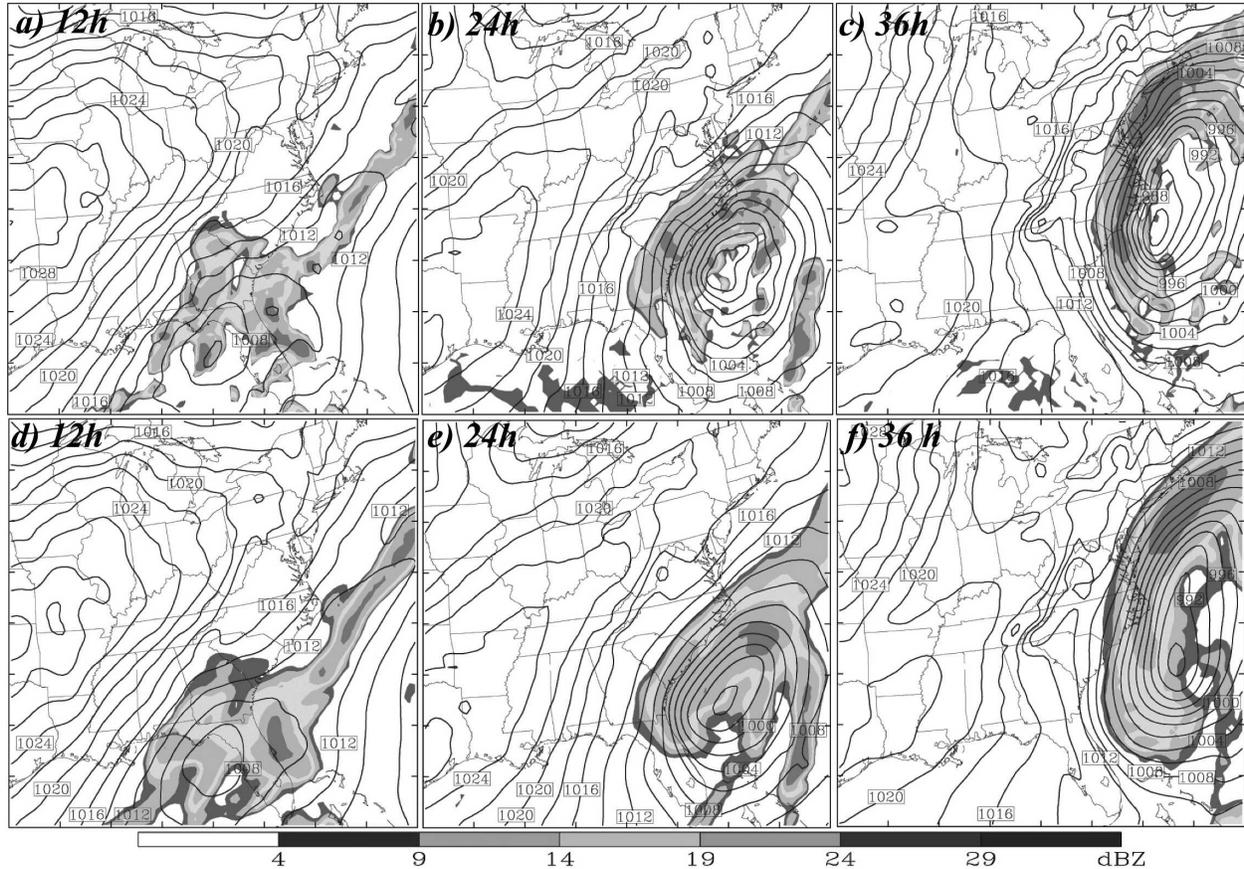


FIG. 2. The mean sea level pressure (MSLP, every 2 hPa) and model-derived reflectivity (shaded) at the 12-, 24-, and 36-h forecast times from the (a)–(c) truth simulation and (d)–(f) reference forecast ensemble mean.

simulation is chosen from 50 different random realizations to compare most favorably to observations of this event in terms of the location and strength of the surface cyclone (Figs. 3a, b of ZSR02) and 300-hPa short-wave trough (Fig. 2 of ZSR02) and the onshore precipitation band (Figs. 3c, d of ZSR02).

After 12 h of simulation, the reference forecast ensemble mean, which is used as the first guess in the following EnKF experiments, has noticeable difference from the truth simulation in all fields. In addition to a ~ 1.5 hPa weaker surface incipient cyclone (differences of wind vectors and MSLP are shown in Fig. 4a), the incipient inland precipitation from the Gulf Coast across Georgia to South Carolina in the reference forecast ensemble mean is much weaker (Fig. 2a versus Fig. 2d). Moreover, the reference forecast ensemble mean of the 300-hPa short-wave PV trough is slightly but systematically shifted to the east (Figs. 3a,d and 5a).

At 24 h, the maximum differences of MSLP and winds associated with the surface cyclone between the ensemble mean and the truth simulation are as large as 5 hPa and 12.5 m s^{-1} , respectively (Figs. 2b,e and 4b).

Moreover, the reference forecast ensemble mean (Fig. 2e) also misses the strong inland precipitation across the Carolinas seen in the reference run (Fig. 2b) and radar observations (Fig. 3a of ZSR02). Associated with a systematic eastward shift of the upper-level PV trough (fronts) in the ensemble mean forecast, the maximum PV and wind differences at 300-hPa reached an amplitude of 2.5 PVU and 22.5 m s^{-1} , respectively (Figs. 3b,e and 5b). Growth of maximum difference along fronts is consistent with the error evolution in the quasigeostrophic model examined by Snyder et al. (2003). After 36 h of simulation, due to the strong diabatic destruction of the upper-level PV as the cyclone reaches its peak intensity, the maximum PV difference at 300 hPa (Fig. 5c) is slightly smaller than that at 24 h (Fig. 5b). Nevertheless, the maximum MSLP difference between the truth simulation and ensemble mean is as high as 8.5 hPa in addition to the even stronger dislocation of the surface cyclone and precipitation band (Figs. 2c,f and 4c).

The evolution of the forecast error growth revealed from the difference between the truth simulation and

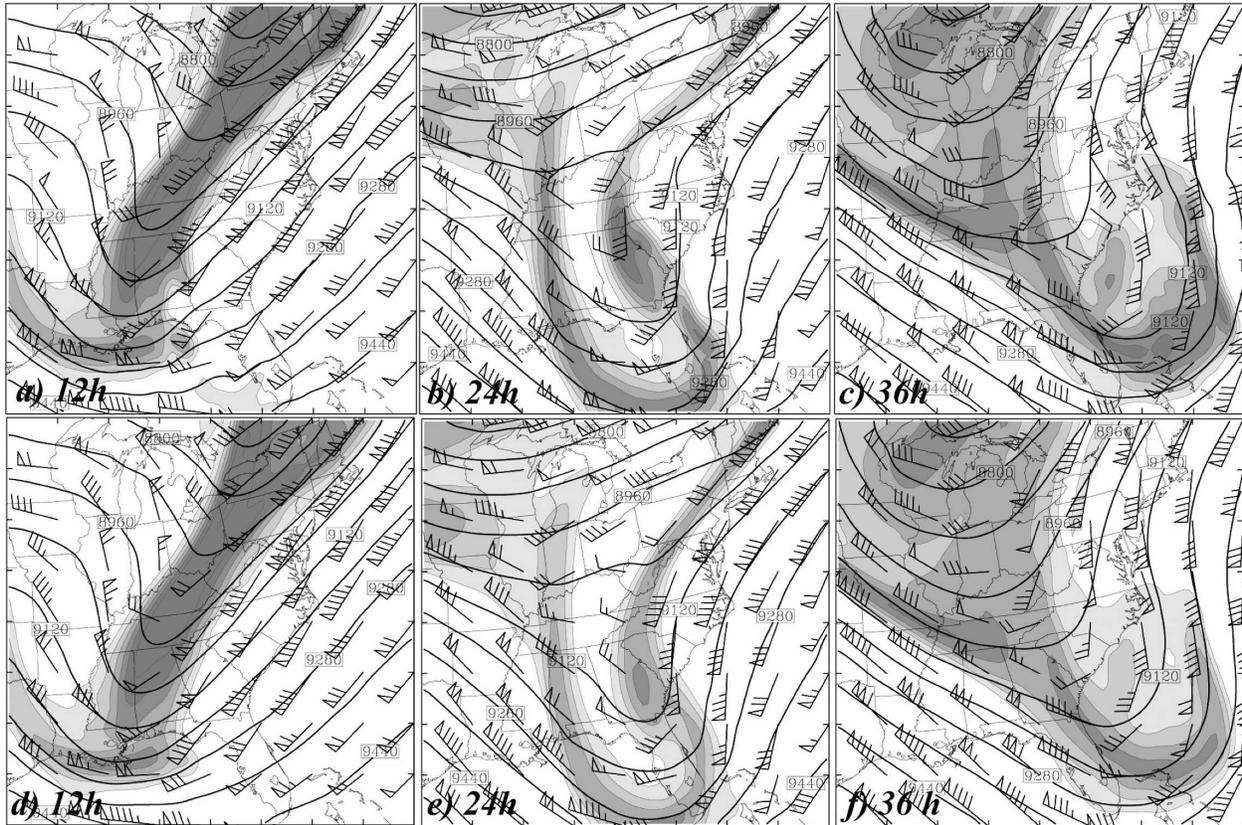


FIG. 3. As in Fig. 2 but for the geopotential heights (every 80 m), potential vorticity [shaded every 1 PVU (potential vorticity unit), where 1 PVU = $1.0 \times 10^{-6} \text{ m}^2 \text{ s}^{-1} \text{ K kg}^{-1}$], and vector winds (full barb 5 m s^{-1}) at 300 hPa.

the reference forecast ensemble mean can be best summarized in terms of difference total energy (DTE):

$$\text{DTE} = 0.5(u'u' + v'v' + kT'T'), \quad (4)$$

where primes denote difference between any two simulations and $k = C_p/T_r$ ($C_p = 1004.7 \text{ J kg}^{-1} \text{ K}^{-1}$ and the reference temperature $T_r = 270 \text{ K}$). The horizontal distributions of the (vertically averaged) root-mean (RM) of DTE (RM_DTE) at 12, 24, and 36 h are displayed in Figs. 6a–c. The initial RM_DTE from the random initialization of the ensemble forecast using the MM5 3DVAR method is $\sim 1.2 \text{ m s}^{-1}$ and is nearly constant across the domain (not shown). By 24 and 36 h, it has become greater than 4 m s^{-1} all across the Atlantic Coast with maxima of $\sim 16 \text{ m s}^{-1}$. Consistent with Figs. 2–5 and Zhang (2005), the maximum error growth occurs near the surface cyclone, the upper-level short-wave trough, and associated fronts and moist processes (Figs. 2 and 3).

Throughout the study, the reference forecast ensemble is used as a benchmark for the performance of EnKF and the evolution of the analysis error. It is also

regarded as the worst-case scenario in which no observations are assimilated.

4. The control EnKF experiment

In the control EnKF experiment with a 40-member ensemble (CNTL), simulated sounding and surface wind and temperature observations are taken from the truth simulation. Typical of the standard sounding and surface observational network over the continental United States, the sounding observations are spaced 300 km apart horizontally and at every sigma level; the surface observations are spaced every 60 km apart and are available at the lowest model level. We assume that the observations have independent, Gaussian random errors of zero mean and variance of 2.0 m s^{-1} for u and v , and 1.0 K for T . Sounding and surface observations are assimilated every 12 and 3 h, respectively. The forecast model is assumed to be perfect, namely, the same numerical model produces the forecasts and the truth simulation from which observations are taken. We begin assimilating observations at 12 h using the 12-h

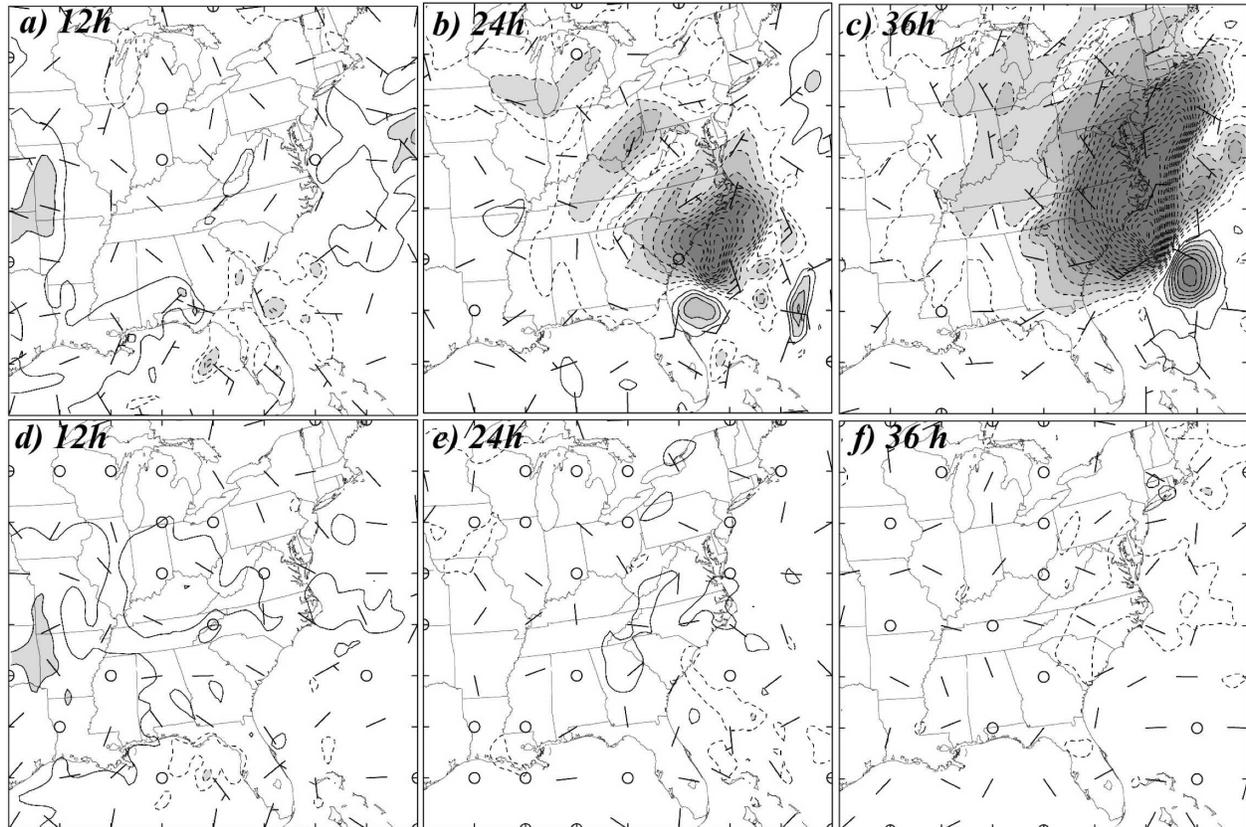


FIG. 4. Differences of wind vectors (full bars 5 m s^{-1}) and MSLP (every 0.5 hPa) between the truth simulation and the reference forecast ensemble mean at (a) 12, (b) 24, and (c) 36 h, and between the truth simulation and the EnKF mean analyses at (d) 12, (e) 24, and (f) 36 h.

short-term reference forecast ensemble as the first guess and to estimate the background error covariance.

Differences in MSLP and surface winds between the ensemble mean analysis after the EnKF assimilation (EnKF analysis) and the truth simulation at 12, 24, and 36 h are displayed in Figs. 4d–f. At 12 h, after the first cycle of assimilating both the surface and sounding observations, there is only marginal, overall reduction of MSLP error (compared to the reference forecast ensemble) but errors in surface winds are significantly reduced (Fig. 4d versus Fig. 4a). At 300 hPa, not only are the errors in the winds reduced by $\sim 30\%$, but errors in PV (as signature of balanced dynamics) are also significantly reduced (Fig. 5d versus Fig. 5a). The overall improvement after the EnKF assimilation across the domain is clearly seen in the horizontal distribution of the (column averaged) RM_DTE in Fig. 6d. Compared to the RM_DTE of the mean forecast error of the reference ensemble at this time (Fig. 6a), we can see that the improvement is more pronounced in the vicinity of the upper-level short-wave trough than near the surface low, consistent with Figs. 4d and 5d.

At 24 h, after assimilating five sets of surface observations (every 3 h) and two sets of sounding observations (every 12 h), the EnKF analyses of the surface winds and MSLP and the 300-hPa winds and PV (not shown) approach those in the truth simulation (Figs. 2b and 3b). More specifically, the maximum analysis errors in surface winds and MSLP are $\sim 2.5 \text{ m s}^{-1}$ and 1 hPa , respectively (Fig. 4e), which represent 60%–80% reduction of the ensemble mean forecast error without the EnKF (Fig. 4b). A similar or even larger degree of improvement can also be seen in the analysis error distribution at 300 hPa (Fig. 5e versus Fig. 5b). The two local maxima of RM_DTE associated respectively with the upper-level front and the surface low in the forecast (Fig. 6b) are no longer noticeable in the DTE of the EnKF analysis (Fig. 6e).

Error reduction in both observed and unobserved (or derived) variables continues through 36 h, with more surface and sounding observations assimilated (Figs. 4f, 5f, and 6f). Most strikingly, compared to the 8.5-hPa MSLP forecast error without EnKF (Fig. 4c), the EnKF analysis of MSLP has become nearly indistinguishable

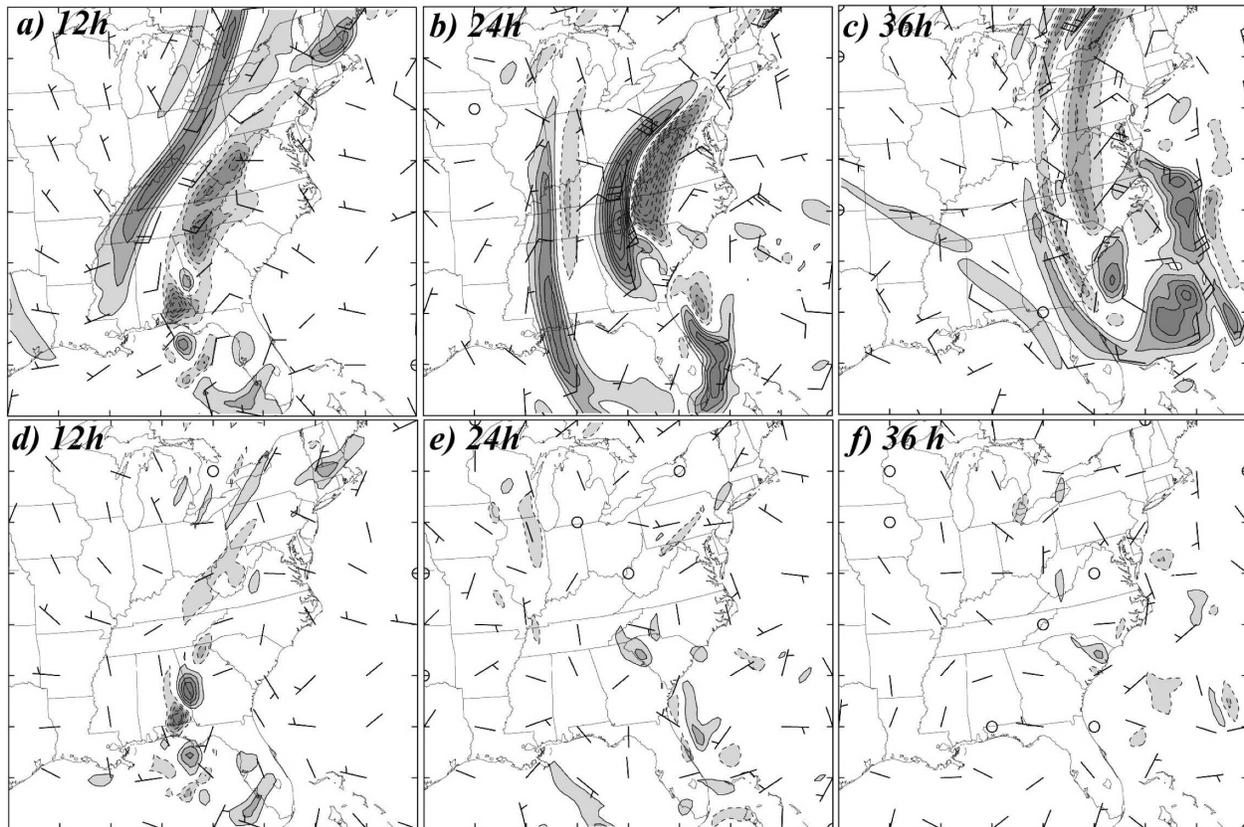


FIG. 5. As in Fig. 4 but for the difference of 300-hPa potential vorticity (every 0.5 PVU) and winds (full barb 5 m s^{-1}).

from that of the truth simulation. There are only a few small areas with the MSLP error greater than 1 hPa (Fig. 4f).

The vertical distribution of the mean analysis and forecast errors in terms of (horizontally averaged) RM_DTE, p' , w , and q at different times is shown in Fig. 7. For the RM_DTE (Fig. 7a), the forecast error of the reference ensemble gradually grows into a distinct double-peak structure over the 36-h forecast, becoming maximum in the upper and lower troposphere, respectively. The primary peak in the upper troposphere is consistent with the forecast-error statistics of operational ensemble prediction systems (e.g., Molteni et al. 1996) as well as in simplified dry systems (Hamill et al. 2002, 2003). The secondary peak in the lower troposphere is likely due to the lower-level fronts associated with strong moist processes. On the other hand, the analysis error exhibits nearly the same amplitude vertically throughout the troposphere, implying that the largest improvement occurs where the reference forecast ensemble has the largest forecast errors.

For the pressure perturbation p' (Fig. 7b), the largest forecast error occurs near the surface. Consistently, through continuous analysis and forecast cycles, the

most error reduction occurs in the lower troposphere. For the vertical velocity field (Fig. 7c), the reference forecast ensemble mean error peaks at 400–500-hPa layer. Unlike the RM_DTE or p' , the forecast error in w follows closely the strength of w in the truth simulation (as an index of the intensity of the background cyclogenesis): the strongest forecast error occurs at ~ 24 h when there is strongest vertical motion in the truth simulation (not shown); there is an apparent decay of forecast error at 36 h when the surface cyclone has matured and begins to decay. Compared to the reference forecast ensemble mean, the overall error reduction for w at 36 h is 30%–40%. Error reduction comes not only from direct EnKF analyses at any given time but also from a better first guess due to the improvement in unobserved variables.

The ensemble forecast error for the moisture field q peaks at 800–900 hPa in association with the abundance of lower-level background moisture as well as moist convection (Fig. 7d). The peak error is approximately constant in the EnKF analysis with an overall error reduction of $\sim 50\%$ compared to the reference forecast ensemble.

The performance of the EnKF in this control experi-

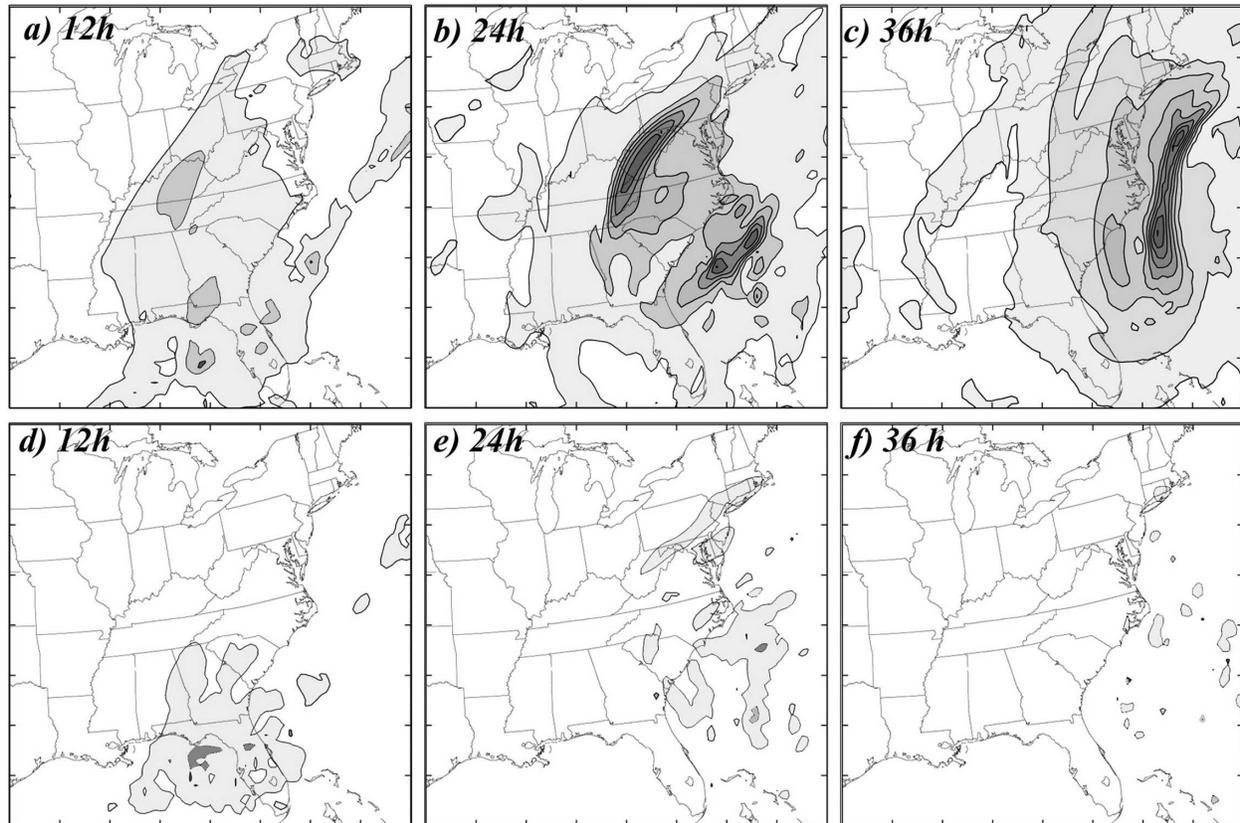


FIG. 6. As in Fig. 4 but for the root-mean of column-averaged DTE (RM_DTE; every 2 m s^{-1}).

ment is best summarized in Fig. 8, which shows the evolution of the domain-averaged root-mean-square (rms) errors in the EnKF analyses of the six prognostic variables (u , v , T , p' , w , q), the corresponding STD of the analysis ensemble, and the rms errors of the reference forecast ensemble. Compared to the reference forecast ensemble, over the 24-h assimilation period, the overall error reduction for the observed variables u , v , and T is $\sim 60\%$ – 80% . The overall analysis quality of all variables stays fairly constant throughout the EnKF, indicating that at later times, the error growth during the short-term (3 h) ensemble forecast will be approximately equal to the reduction of analysis error through assimilation of new observations. The final domain-averaged rms error after 24-h assimilation is ~ 1.0 – 1.5 m s^{-1} for winds and $\sim 1.0 \text{ K}$ for temperature, which is less than or at most comparable to typical observational errors. The unobserved variable p' has the biggest overall improvement with the 36-h analysis error being only one-sixth of the forecast error. Nearly 50% overall error reduction is observed in the moisture field. Again, there is relatively small (30%–40%) overall improvement in the vertical velocity field.

The difference in the degree of error reduction

among different variables is also examined through the comparison of the power spectra of analysis and forecast errors of the reference forecast ensemble and CNTL at different times (Fig. 9). The vertical velocity and moisture fields, for which the EnKF assimilation is the least effective, have the most error energy in smaller scales. The pressure field, which has the most power energy at larger scales, in general enjoys the biggest error reduction. As a result of stronger error reduction at larger scales, power spectra in all the variables (except for w) become increasingly flattened at smaller and smaller wavenumbers (“whitening”; Hamill et al. 2002; Daley and Menard 1993) through the EnKF assimilation (Fig. 9). In essence, the EnKF is very efficient in reducing errors at larger scales but less effective in reducing errors at smaller, marginally resolvable scales. The EnKF analyses of other water substances associated with clouds, which have the strongest smaller-scale variations, are found to be problematic (not shown), suggesting the accurate estimation of clouds with the current EnKF is not yet possible, at least for the current filter configuration and model resolution with parameterized moist convection.

In an examination of spectral characteristics of Kal-

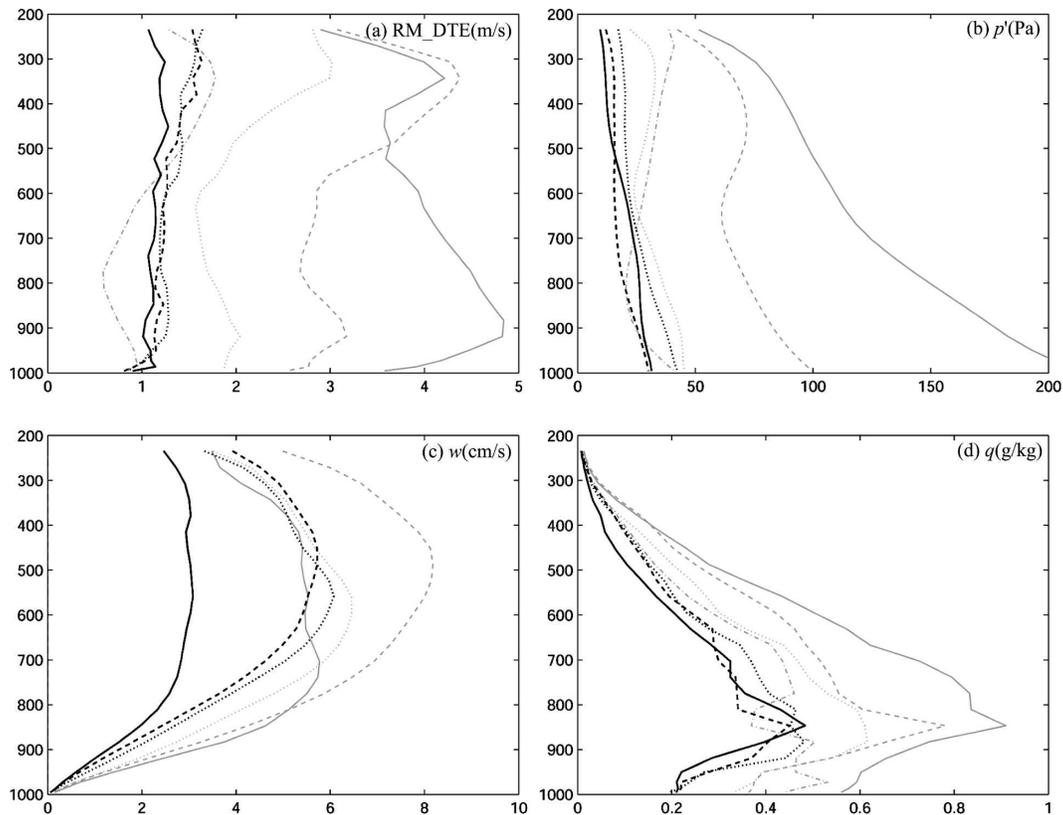


FIG. 7. Vertical distribution of the mean analysis errors of the control EnKF experiment (dark) and the mean forecast errors of the reference forecast ensemble (gray) for (a) (horizontally averaged) RM_DTE, (b) p' , (c) w , and (d) q valid at 12 (dotted), 24 (dashed), and 36 h (solid). Errors in the initial ensemble are denoted with dotted-dashed gray curves.

man filter systems, Daley and Menard (1993, their Fig. 2) showed that the Kalman filter has a much larger impact on the large scales than the smaller scales. Because the uncorrelated observational error is projected equally to all scales when the model-error spectrum is red, the observations are considered to be more accurate with respect to the background for the large scales than they are for the small scales. As discussed in Daley (1991, his Fig. 5.9), this is strictly applicable to univariate analysis in which the larger scales have the most error energy. The scale-dependent filter performance in the multivariate analysis is much more complex (Daley 1991, his Fig. 5.10), probably even more so for those unobserved variables using the flow-dependent background error covariance at the mesoscales for the current study. Besides the possible mechanisms discussed by Daley (1991, his book section 5.4–5.5), the scale- and variable-dependent filter performance may also be due to faster error saturation (thus shorter predictability) resulting in poorer estimate of the prior guess and background error covariance at the smaller, marginally resolvable scales. It could also arise from observations

that are too sparse to provide sufficient information for analysis at smaller scales while larger scales are influenced (corrected) by observations of similar (comparable) horizontal resolutions.

5. Forecast experiments with the EnKF analysis

To evaluate the performance of short-range ensemble forecasts with improved analyses and to examine the forecast error growth dynamics after the EnKF assimilation, two 40-member ensemble forecast experiments (“EF12H” and “EF24H”) are performed with the analyses from CNTL at 12 and 24 h as initial conditions. The horizontal distribution of the (vertically averaged) RM_DTE from the 12- and 24-h integration of EF12H and the 12-h integration of EF24H are shown in Fig. 10. For EF12H, which starts from the EnKF analysis cycle that assimilated only the observations at 12 h, there are noticeably smaller ensemble mean errors in RM_DTE at both 24 and 36 h compared to the reference forecast ensemble (Figs. 10a,b versus Figs. 6b,c). Even smaller RM_DTE error is found in the 12-h

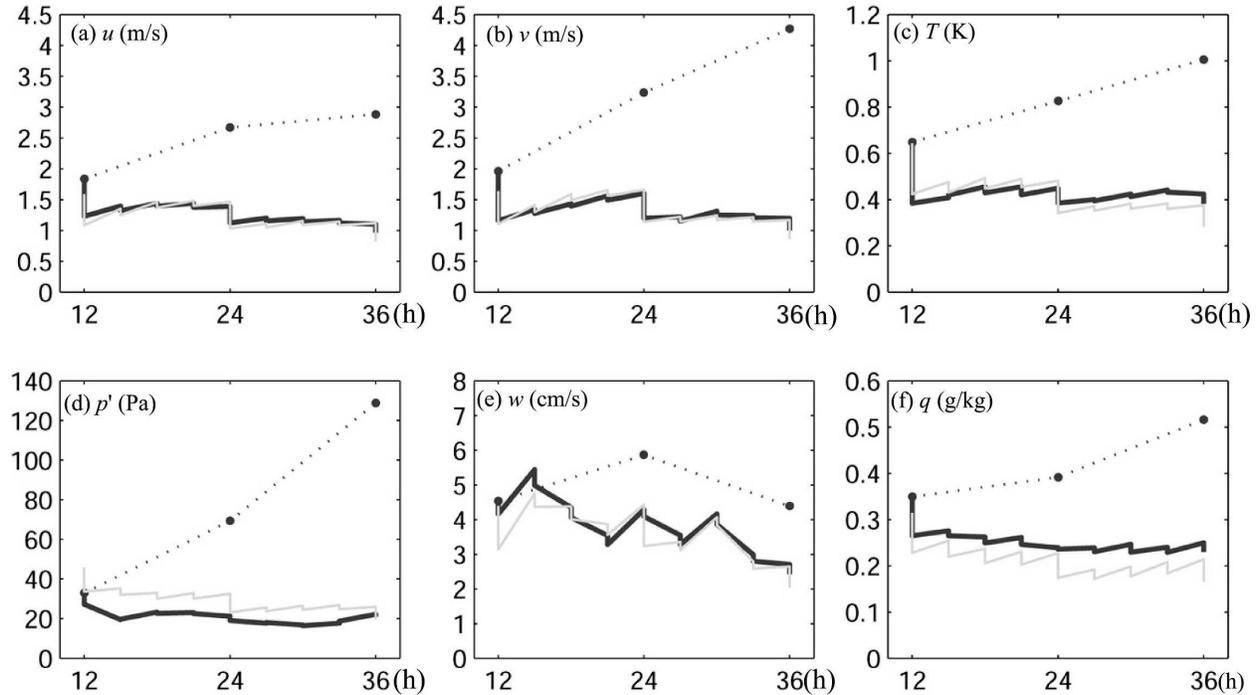


FIG. 8. Evolution of the domain-averaged root-mean-square errors of the EnKF analysis (solid black) with respect to the truth simulation for six prognostic variables (u , v , T , p' , w , q), the corresponding standard deviation (solid gray) of analysis ensemble, and the root-mean-square errors of the reference forecast ensemble (dotted black, computed every 12 h only).

ensemble forecast by EF24H, which starts with the EnKF analysis at 24 h (after a 12-h assimilation period; Fig. 10c). Compared to a maximum RM_DTE error of $\sim 16 \text{ m s}^{-1}$ just off the Atlantic coast in the reference forecast ensemble (Fig. 6c), the maximum RM_DTE error for the 12-h forecast of EF24H is merely $\sim 6 \text{ m s}^{-1}$ (Fig. 10c).

Evolution of the forecast errors of the six prognostic variables from these two forecast experiments (EF12H and EF24H) and the reference forecast ensemble as well as the analysis errors from CNTL is plotted in Fig. 11. Again, compared to the reference forecast ensemble, the positive effect of improved initial conditions using the EnKF analysis can be seen in both forecasts verified at 36 h in all prognostic variables shown. It is also seen that, with a longer assimilation period and thus more data being assimilated, the mean forecast error verified at 36 h of EF24H is considerably smaller than that of EF12H and the reference forecast ensemble.

6. Sensitivity experiments

a. Ensemble size, variance relaxation, and localization

Difference in error spectral distribution and error growth dynamics among the different state variables

will potentially result in inconsistencies between the analysis/forecast error and ensemble spread between different variables if the same localization or error inflation/relaxation is used for all state variables. For the control EnKF experiment (CNTL), the domain-averaged standard deviations (ensemble spread) of the ensemble forecast and EnKF analysis (Fig. 8) stay very close to the rms errors of the ensemble forecasts and EnKF analyses for all variables. There is no obvious filter divergence, which would be indicated by the growth of the ratio of ensemble mean error to ensemble spread. The (domain averaged) RM_DTE also agrees reasonably well with the STD of the analyses when a 20-member ensemble (“CNTL20”) is used to estimate the background error covariance (Fig. 12a). The difference of the analysis accuracy (in terms of RM_DTE) between CNTL and CNTL20 is rather insignificant ($\sim 0.1\text{--}0.2 \text{ m s}^{-1}$) throughout the assimilation. On the other hand, though much less accurate than CNTL and CNTL20, a 10-member ensemble EnKF experiment still performed reasonably well albeit with a significantly larger ratio of rms error to STD (not shown).

We observed that even though the pressure perturbation has the maximum overall improvement, the analysis error from CNTL20 after the EnKF assimilation at 12 h is greater than the forecast error at this time (not shown). The degradation occurred only in the

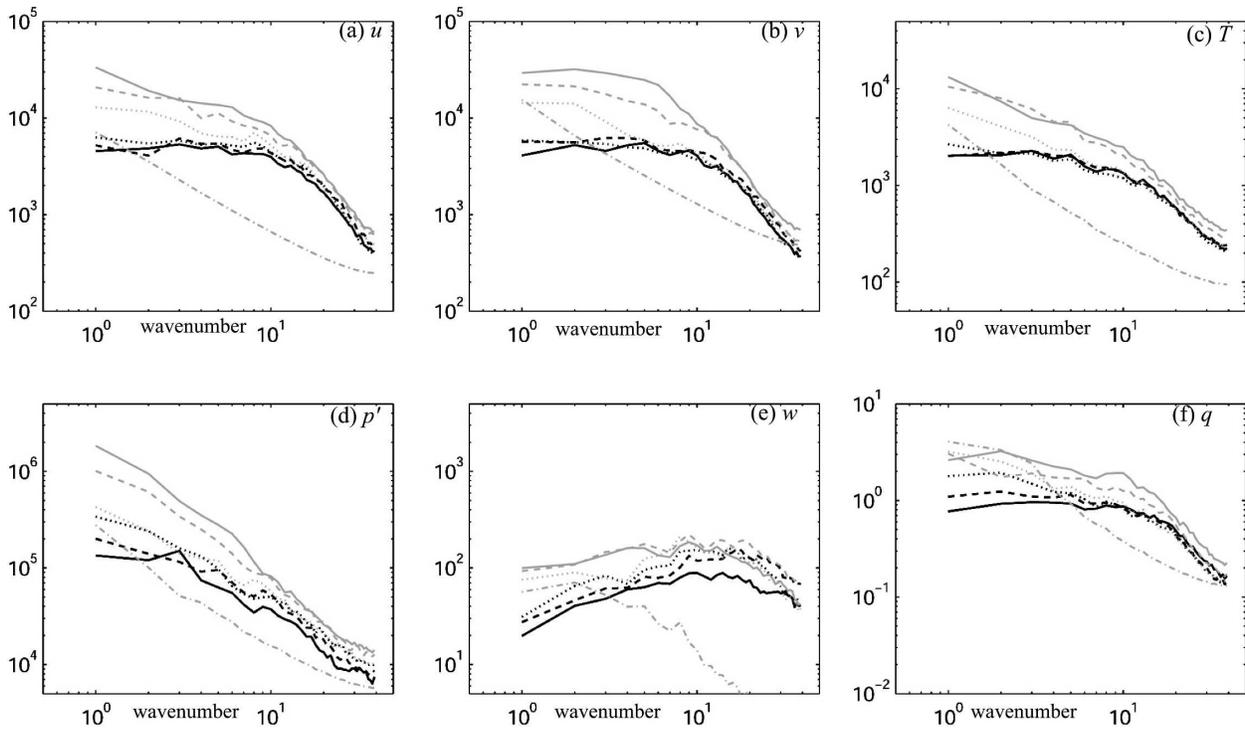


FIG. 9. Power spectrum analysis of the EnKF analysis errors (dark) and the reference forecast ensemble errors (gray) at 12 (dotted), 24 (dashed), and 36 h (solid) for six prognostic variables (u , v , T , p' , w , q). The minimum (maximum) wavenumber of 1 (40) corresponds to a horizontal wavelength of 2400 (60) km. Error spectra in the initial ensemble are denoted with dotted-dashed gray curves.

lower troposphere for the first assimilation cycle. To test whether the degradation is systematic, we examined three additional experiments: the control experiment with 40 members, an experiment similar to CNTL20 but with different random realizations, and an experiment similar to CNTL20 but with a different truth (discussed section 6c). In all three experiments, we observed that the degradation of the pressure analysis did not occur, suggesting that the covariance between the observed variables and the pressure field

may be unrepresentative of the true forecast error in pressure perturbations at this time when the ensemble has only 20 members.

The impact of the variance relaxation can be clearly seen in a 40-member EnKF experiment similar to CNTL but without the application of variance relaxation (“NOMIX”), which has larger overall RM_DTE and poorer agreement between RM_DTE and STD (Fig. 12b). The deficiency in the ensemble spread becomes even more severe when a 20-member ensemble

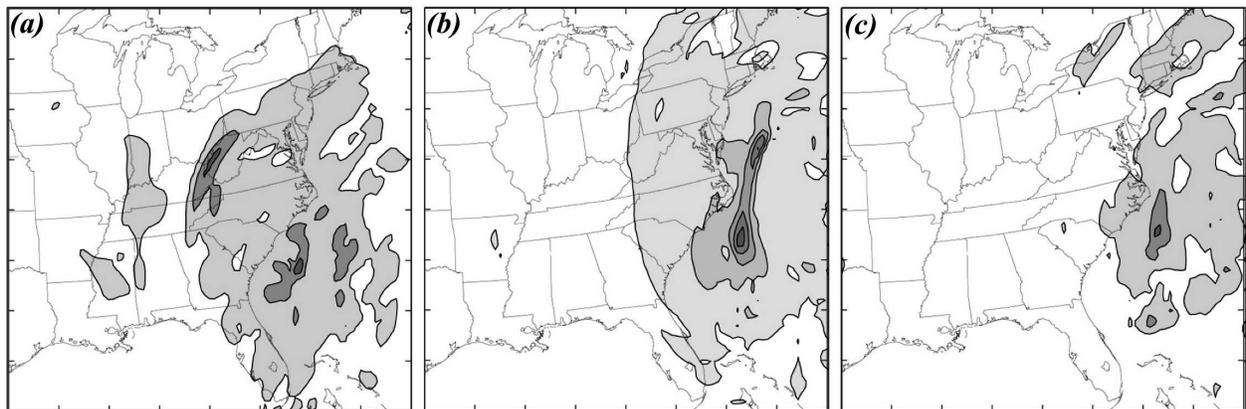


FIG. 10. As in Fig. 6 but from ensemble forecasts for (a) EF12H at 24 h, (b) EF12H at 36 h, and (c) EF24H at 36 h.

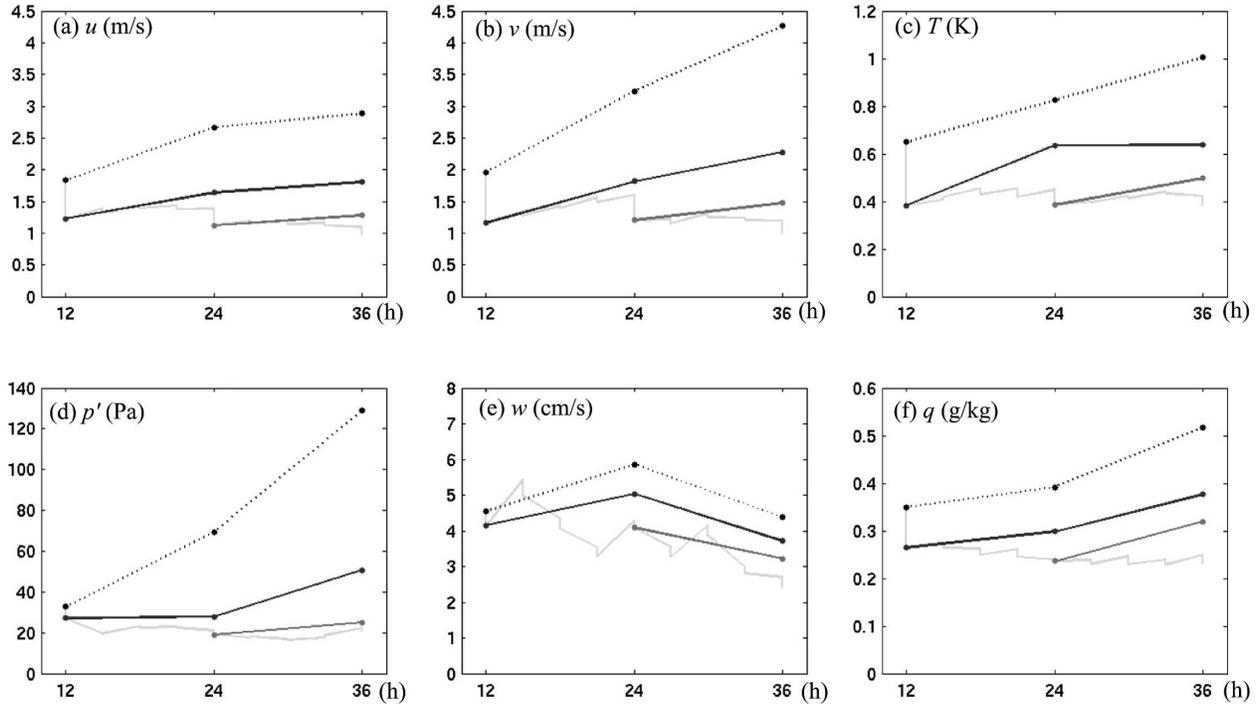


FIG. 11. Time evolution of domain-averaged root-mean-square errors from the reference forecast ensemble (dotted), EF12H (solid black; from 12 h), EF24H (solid black; from 24 h), and the CNTL analysis (gray) for six prognostic variables (u , v , T , p' , w , q).

is used (not shown). Thus, the application of the variance relaxation from Zhang et al. (2004) helps prevent filter divergence, which occurs when small ensembles are used. We also tested different implementations of the variance inflation method (e.g., applying the variance inflation either before or after the EnKF analysis) used in Anderson (2001) using 40-member ensembles, with the inflation factors from 1.05 to 1.5, either applied before or after the EnKF analysis. None of these additional experiments (not shown) exhibited satisfactory performances comparable to that from the control experiment.

The performance of the EnKF assimilation is also very sensitive to covariance localization. Apparent degradation of EnKF performance and the lack of ensemble spread compared to analysis error (possible filter divergence) are seen in the EnKF experiment (“IR60DX”; Fig. 12c) in which the three-dimensional distance used in the Schur-product is set to be too large (60 rather than 30 grid points in the CNTL, which is equivalent to 1800 km versus 900 km in terms of purely horizontal distance). When a 450-km cutoff radius of influence is used in another EnKF experiment (not

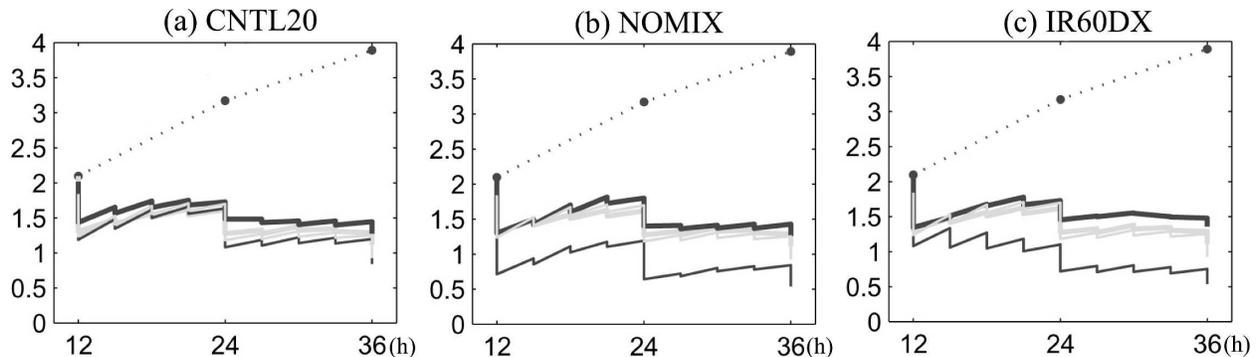


FIG. 12. Time evolution of the domain-averaged RM_DTE (m s^{-1} , black, thick) and standard deviation (m s^{-1} , black, thin) from sensitivity experiments (a) CNTL20, (b) NOMIX, and (c) IR60DX. RM_DTE and standard deviation from the control experiment and the RM_DTE from the reference forecast ensemble are also displayed in thick gray, thin gray, and dotted curves, respectively.

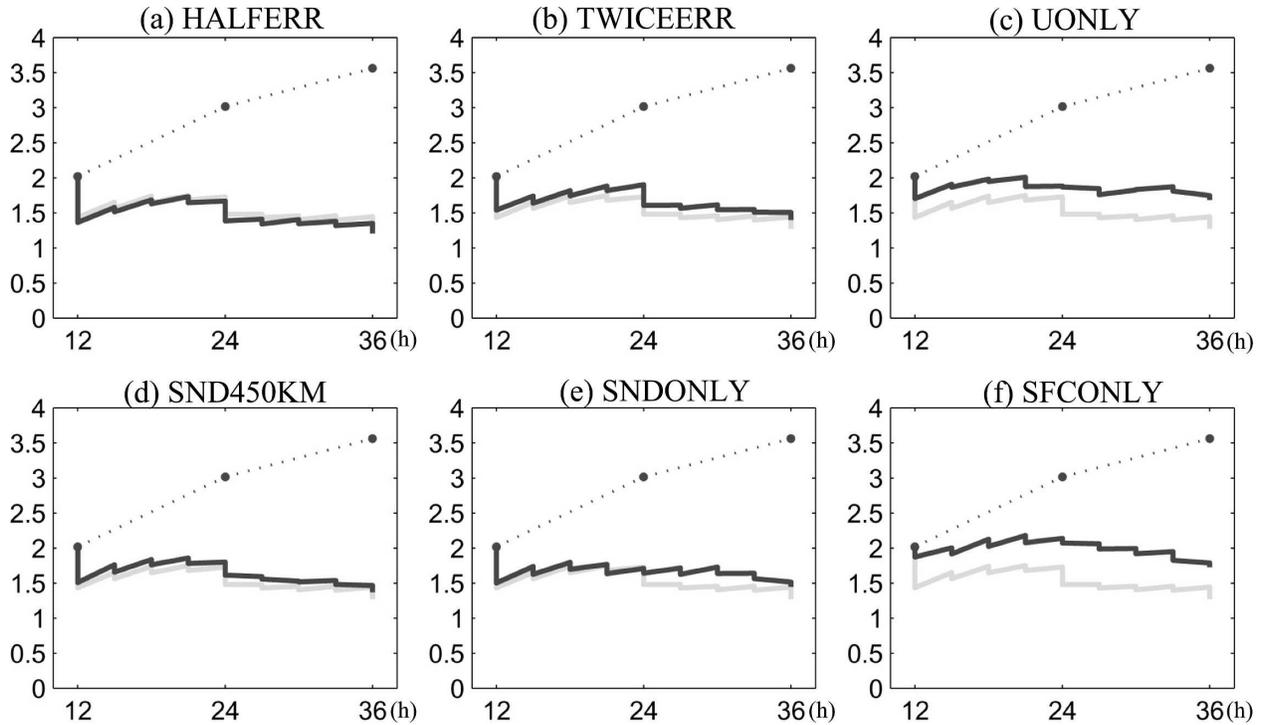


FIG. 13. Time evolution of the domain-averaged RM_DTE (m s^{-1} , thick solid curve) from sensitivity experiments (a) HALFERR, (b) TWICEERR, (c) UONLY, (d) SND450KM, (e) SNDONLY, and (f) SFCONLY. RM_DTE from experiment CNTL20 and the 20-member reference forecast ensemble are also displayed in gray and dotted curves, respectively.

shown), the overall performance is similar to CNTL but the ensemble spread is bigger than the analysis error throughout the assimilation period. Similar sensitivity was also reported in the EnKF experiments in Houtekamer and Mitchell (2001, their Fig. 4). Since the best value for the radius of influence is not known a priori, the current EnKF configuration may unavoidably need to be “tuned” for different weather systems for best performance.

These sensitivity experiments demonstrate that the ratio of the rms error of the ensemble forecast and EnKF analysis to the STD of ensemble variance, as a common index of filter divergence, is a complex function of ensemble size, the cutoff radius of influence, and variance inflation. Bigger ensemble size, smaller cutoff radius, and the implementation of variance relaxation method lead to larger ensemble spread, potentially preventing severe filter divergence.

b. Observation quality and availability

The ensemble Kalman filter combines information from the initial estimate, the dynamics of the forecast model and the observations to get the best estimate and the associated uncertainty. The quality and availability (coverage, resolution, and accuracy) of sounding and

surface observations is different from case to case, which could impact the ability to estimate the true state. In this subsection, various possible observational scenarios are tested using the EnKF, some of which follow closely those of Zhang et al. (2004). We use a 20-member EnKF with the same truth simulation and the same initial ensemble as those in CNTL20 for all the sensitivity experiments investigated in this subsection since the difference between CNTL and CNTL20 is rather insignificant (Fig. 12a).

Experiment “HALFERR” (“TWICEERR”) differs from CNTL20 in that the observational errors of the observed variables (u , v , and T) are reduced (increased) to half (twice) of those used in CNTL20. The rms of the DTE in HALFERR (TWICEERR), albeit slightly (<5% smaller (larger)), shows very similar convergence toward the reference solution in comparison to that of CNTL20 (gray curves; Figs. 13a,b). These two experiments demonstrate that, as long as the observational errors are uncorrelated, assimilation with the ensemble filter is rather insensitive to the observational accuracy given the typical range of observational errors for sounding and surface observations, consistent with those convective scale experiments in Zhang et al. (2004).

Experiment “UONLY” differs from CNTL20 in that

only the zonal wind from the sounding observations is assimilated, which is similar to a case if we use radar radial velocity instead of sounding observations. Again, the EnKF analysis converges well toward the truth simulation over the 24-h assimilation; the RM_DTE at 36 h is only $\sim 10\%$ – 20% larger than that in the CNTL20 (Fig. 13c). In another experiment similar to CNTL20 but with the addition of pressure perturbation and moisture observations in the soundings, there is no significant improvement in the EnKF analysis compared to CNTL20 for all prognostic variables including p' and q (not shown). Filter performance is also nearly unchanged when the horizontal spacing of the sounding network changed from 300 km in CNTL20 to 450 km in the experiment "SND450KM" (Fig. 13d).

Experiments "SNDONLY" and "SFCONLY" differ from CNTL20 in that only either sounding or surface observations are assimilated every 3 h. For the first 12-h assimilation period of SNDONLY, the analysis follows closely that of the CNTL20 but the loss of surface observations cannot be corrected by more frequent sounding observations for the final 12-h assimilation period (Fig. 13e). Consistent with Whitaker et al. (2004), it is very encouraging to notice that the filter also converges well to the truth simulation when only surface observations are assimilated (Fig. 13f), even though the advantage of sounding observations is clearly seen when compared to CNTL20.

c. Different truth simulations

We also performed several additional experiments with the same set of initial ensembles as in CNTL20 but using different realizations of the 3DVAR perturbations to generate the truth simulation. Quantitatively similar performance (to the CNTL20) has been achieved in all of these EnKF experiments (not shown). Another experiment with the same truth simulation as in CNTL20 but a different set of 20 ensemble members behaves in a similar manner (not shown).

7. Summary and discussions

Through various observing system simulation experiments, this study exploits the potential of using the ensemble Kalman filter (EnKF), which estimates error covariances through an ensemble of short-term forecasts, for mesoscale and regional-scale data assimilation. The EnKF is implemented in the nonhydrostatic MM5 to assimilate simulated sounding and surface observations derived from truth simulations of the "surprise" snowstorm of January 2000. This is an explosive east coast cyclogenesis event with strong error growth at all scales as a result of interactions between convective-, meso-, and subsynoptic-scale dynamics.

It is found that the EnKF is very effective in keeping the analysis close to the truth simulation. In the control experiment (CNTL), a 24-h continuous EnKF assimilation of sounding and surface observations with realistic temporal and spatial resolutions can have an error reduction of as much as 80% for horizontal winds and temperature, 85% for pressure perturbation, and 45% for water vapor mixing ratio in comparison to the reference forecast ensemble.

Error growth characteristics in the ensemble forecast with and without the EnKF, including the scale, structure, and evolution of the forecast and analysis errors of different variables are also examined. It is found the EnKF is most effective in reducing larger-scale errors but less effective in reducing errors at smaller, marginally resolvable scales. This is consistent with the analysis of spectral characteristics of Kalman filter systems by Daley (1991) and Daley and Menard (1993). The scale-dependent error reduction may also be due to the faster error saturation (thus shorter predictability) and thus poorer quality of the prior estimate and background error covariance at the smaller, marginally resolvable scales. It could also arise from observational information that is insufficient to allow for a good estimate at smaller scales. There are also apparent improvements in the forecast initiated with EnKF analysis. Since error grows at all scales but saturates quicker at smaller scales, error growth in the ensemble forecasts may be dominated by initial errors at larger scales.

Error growth characteristics and the quality of initial estimate and background error covariance also differ greatly from variable to variable, resulting in different degrees of error reduction for different variables. The EnKF is least effective on the vertical motion and moisture fields, which have more energy in smaller scales while pressure perturbation in general enjoys the biggest error reduction because it has the strongest larger-scale component among all variables. Different error growth from different variables also results in inconsistency between the analysis error and ensemble spread of different variables when the same localization or error inflation/relaxation is used for all variables.

It is also found that the ratio of the root-mean-square analysis/forecast error to the standard deviation of the ensemble variance, as a common index of filter divergence, is a complex function of ensemble size, the cutoff radius of influence (localization), and variance relaxation (inflation). Consistent with past studies, it is found that bigger ensemble size, smaller cutoff radius, and the implementation of the variance relaxation method all lead to larger ensemble spread and potentially prevent filter divergence.

Various experiments are also performed to test the sensitivity of the EnKF to the number of observed variables and the density and accuracy of sounding and surface observations. The EnKF is found to be quite resilient in most of the realistic observational scenarios tested.

The above conclusions on the mesoscale data assimilation with the EnKF are drawn from observation system simulation experiments under the perfect model assumption. Such a strong EnKF performance should not be readily expected in real-world situations where the forecast model unavoidably has errors and the initial ensemble statistics may be far from perfect (refer to Houtekamer et al. 2005). The EnKF performance under various imperfect-model scenarios will be explored in Part II.

Acknowledgments. The authors are grateful to Chris Snyder, Jeff Anderson, Dale Barker, Tom Hamill, Wei Huang, Mei Xue, Wei Wang, John Nielsen-Gammon, and Amy Stuart for their help and comments on forecast model, ensemble initiation, and filter design. Snyder, Nielsen-Gammon, and Stuart also provided thorough reviews of an earlier version of the manuscript. Thanks are also due to two anonymous reviewers for their insightful comments. This research is sponsored by the NSF Grant ATM0205599 and by the Office of Navy Research under Grant N000140410471.

REFERENCES

- Aksoy, A., F. Zhang, J. W. Nielsen-Gammon, and C. C. Epifanio, 2005: Ensemble-based data assimilation for thermally-forced circulations. *J. Geophys. Res.*, **110**, D16105, doi:10.1029/JD005728.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Barker, D. M., W. Huang, Y.-R. Guo, and A. J. Bourgeois, 2003: A three-dimensional variational (3DVAR) data assimilation system for use with MM5. NCAR Tech. Note NCAR/TN-453+STR, 68 pp.
- , —, —, —, and Q. N. Xiao, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897–914.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- , and R. Menard, 1993: Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Wea. Rev.*, **121**, 1554–1565.
- 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.
- Dudhia, J., 1993: A nonhydrostatic version of the Penn State–NCAR Mesoscale Model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.*, **121**, 1493–1513.
- 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.
- 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.
- , —, and R. E. Morss, 2002: Analysis-error statistics of a quasigeostrophic model using three-dimensional variational assimilation. *Mon. Wea. Rev.*, **130**, 2777–2791.
- , —, and J. S. Whitaker, 2003: Ensemble forecasts and the properties of flow-dependent analysis-error covariance singular vectors. *Mon. Wea. Rev.*, **131**, 1741–1758.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- , and —, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.
- , —, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620.
- Keppenne, C. L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971–1981.
- , and M. M. Rienecker, 2002: Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.*, **130**, 2951–2965.
- Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 2791–2808.
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677.
- , T. M. Hamill, and S. B. Trier, 2003: Linear evolution of error covariances in a quasigeostrophic model. *Mon. Wea. Rev.*, **131**, 189–205.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- , G. P. Compo, X. Wei, and T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**, 1190–1200.
- Zhang, F., 2005: Dynamics and structure of mesoscale error covariance of a winter cyclone estimated through short-range ensemble forecasts. *Mon. Wea. Rev.*, **133**, 2876–2893.
- , C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the “surprise” snowstorm of 24–25 January 2000. *Mon. Wea. Rev.*, **130**, 1617–1632.
- , —, and —, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185.
- , —, 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.
- Zhang, S., and J. L. Anderson, 2003: Impact of spatially and temporally varying estimates of error covariance on assimilation in a simple atmospheric model. *Tellus*, **55A**, 126–147.