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

ZHIYONG MENG AND FUQING ZHANG

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

ABSTRACT

In Part I of this two-part work, the feasibility of using an ensemble Kalman filter (EnKF) for mesoscale and regional-scale data assimilation through various observing system simulation experiments was demonstrated assuming a perfect forecast model for a winter snowstorm event that occurred on 24–26 January 2000. The current study seeks to explore the performance of the EnKF for the same event in the presence of significant model errors due to physical parameterizations by assimilating synthetic sounding and surface observations with typical temporal and spatial resolutions. The EnKF performance with imperfect models is also examined for a warm-season mesoscale convective vortex (MCV) event that occurred on 10–13 June 2003. The significance of model error in both warm- and cold-season events is demonstrated when the use of different cumulus parameterization schemes within different ensembles results in significantly different forecasts in terms of both ensemble mean and spread. Nevertheless, the EnKF performed reasonably well in most experiments with the imperfect model assumption (though its performance can sometimes be significantly degraded). As in Part I, where the perfect model assumption was utilized, most analysis error reduction comes from larger scales. Results show that using a combination of different physical parameterization schemes in the ensemble forecast can significantly improve filter performance. A multischeme ensemble has the potential to provide better background error covariance estimation and a smaller ensemble bias. There are noticeable differences in the performance of the EnKF for different flow regimes. In the imperfect scenarios considered, the improvement over the reference ensembles (pure ensemble forecasts without data assimilation) after 24 h of assimilation for the winter snowstorm event ranges from 36% to 67%. This is higher than the 26%–45% improvement noted after 36 h of assimilation for the warm-season MCV event. Scale- and flow-dependent error growth dynamics and predictability are possible causes for the differences in improvement. Compared to the power spectrum analyses for the snowstorm, it is found that forecast errors and ensemble spreads in the warm-season MCV event have relatively smaller power at larger scales and an overall smaller growth rate.

1. Introduction

In the past few years, ensemble-based data assimilation has drawn increasing attention from the data assimilation community due to its prevailing advantages such as flow-dependent background error covariance, ease of implementation, and its use of a fully nonlinear model. Since first proposed by Evensen (1994), ensemble-based data assimilation has been implemented in numerous models to various realistic extents for different scales of interest (Houtekamer and Mitchell 1998; Hamill and Snyder 2000; Keppenne 2000; Anderson 2001; Mitchell et al. 2002; Keppenne and Rienecker 2003; Zhang and Anderson 2003; Snyder and Zhang 2003; Houtekamer et al. 2005; Whitaker et al. 2004; Dowell et al. 2004; Zhang et al. 2004, 2006a, hereafter Part I; Tong and Xue 2005; Aksoy et al. 2005). More background on ensemble-based data assimilation can be found in recent reviews of Evensen (2003), Lorenc (2003), and Hamill (2006).

While Hamill and Snyder (2000), Whitaker and Hamill (2002), and Anderson (2001) showed that using an ensemble Kalman filter (EnKF) in the context of a perfect model (i.e., both the truth and ensemble propagate with the same model) can significantly reduce er-

Corresponding author address: Dr. Fuqing Zhang, Department of Atmospheric Sciences, Texas A&M University, College Station, TX 77845-3150.
E-mail: fzhang@tamu.edu

ror and outperform competitive data assimilation methods such as three-dimensional variational data assimilation (3DVAR), the perfect model assumption must be dropped in real-world studies where model error is caused by inadequate parameterization of subgrid physical processes, numerical inaccuracy, truncation error, and other random errors. The presence of model error can often result in both a large bias of the ensemble mean and too little spread and can ultimately cause the ensemble forecast to fail. Fortunately, studies (e.g., Houtekamer et al. 1996; Houtekamer and Lefaivre 1997) show that including model error in an ensemble can lead to a more realistic spread of the forecast solution. Despite this, model error, especially that at the mesoscale, is generally hard to identify and to deal with due to the chaotic nature of the atmosphere, its flow-dependent characteristics, and the lack of sufficiently dense observations for verification (e.g., Orrell et al. 2001; Orrell 2002; Simmons and Hollingsworth 2002; Stensrud et al. 2000).

There have been several different approaches for including model error in ensemble forecasts. One popular (yet ad hoc) approach involves the use of different forecast models (e.g., Evans et al. 2000; Krishnamurti et al. 2000) or different physical parameterization schemes (e.g., Stensrud et al. 2000). Other ways to include model error are to apply statistical adjustment to ensemble forecasts (Hamill and Whitaker 2005) or to use stochastic forecast models and/or stochastic physical parameterizations (e.g., Palmer 2001; Grell and Devenyi 2002).

Mitchell et al. (2002), Hansen (2002), Keppenne and Rienecker (2003), Hamill and Whitaker (2005), and Houtekamer et al. (2005) have all discussed explicit treatment of model error in ensemble-based data assimilation. For example, Keppenne and Rienecker (2003) obtained encouraging results using covariance inflation (first proposed in Anderson 2001) with an oceanic general circulation model and real data. In a study that showed ensemble data assimilation can outperform 3DVAR, Whitaker et al. (2004) also used the covariance inflation method to reanalyze the past atmospheric state using a long series of available surface pressure observations. Despite these successes, covariance inflation can cause a model to become unstable due to excessive spread in data-sparse regions (Hamill and Whitaker 2005). The additive error method (Hamill and Whitaker 2005; Houtekamer et al. 2005) and the covariance relaxation method of Zhang et al. (2004) have recently been proposed as alternatives to covariance inflation. The performance of certain additive error methods was found to be superior to covariance inflation for the treatment of model truncation

error caused by lack of interaction with smaller-scale motions, and additive error methods might outperform a simulated 3DVAR method (Hamill and Whitaker 2005). Meanwhile, Houtekamer et al. (2005) used a medium-resolution, primitive equation model with physical parameterizations and similarly parameterized model error by adding noise consistent in structure with 3DVAR background error covariance. The EnKF performed similarly to the 3DVAR method implemented in the same forecast system.

Most of the aforementioned studies that included an explicit treatment of model error used global models. To the best of our knowledge, the impacts of model error on ensemble data assimilation with a mesoscale model have rarely been addressed in literature. Applications of an EnKF to the mesoscale have only recently begun with simulated observations (Snyder and Zhang 2003; Zhang et al. 2004; Tong and Xue 2005; Caya et al. 2005; Part I) and with real data (Dowell et al. 2004; Dirren et al. 2007). In Part I, the authors examined the performance of an EnKF implemented in a mesoscale model through various observing system simulation experiments (OSSEs) assuming a perfect model. It is found that the EnKF with 40 members works very effectively in keeping the analysis close to the truth simulation. The result that most error reduction comes from large scales is consistent with Daley and Menard (1993). Furthermore, the EnKF performance differs among variables; it is least effective for vertical motion and moisture due to their relatively strong smaller-scale power, and it is most effective for pressure because of its relatively strong larger-scale power.

As the second part of a two-part study, this paper examines the performance of the same EnKF in the presence of significant model error due mainly to physical parameterizations. Past studies (e.g., Stensrud et al. 2000) suggested that a considerable part of model error comes from parameterization of subscale physical processes. The "surprise" snowstorm of 24–26 January 2000 that was examined in Part I is also examined here.

In next section, we describe the methodology, experimental design, and ensemble and model configurations. A synoptic overview and the control experiment results are described in section 3. Section 4 demonstrates the sensitivity of the EnKF to model error due to physical parameterizations. The EnKF performance in another case with a distinguishably different flow regime [the long-lived warm-season mesoscale convective vortex (MCV) event that occurred on 10–13 June 2003] is then examined in section 5 to address the impact of flow-dependent predictability. Finally, section 6 gives our conclusions and a discussion.
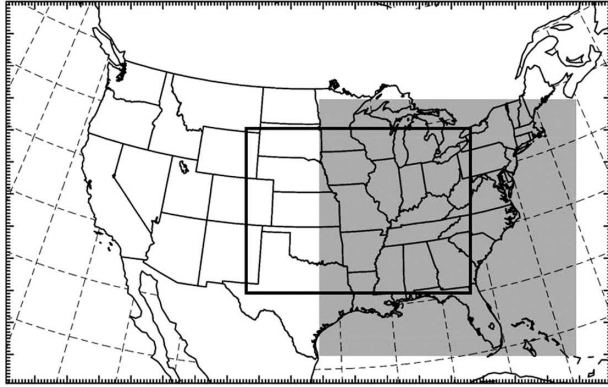
FIG. 1. Map of the model domain. Observations are extracted only from the area inside the shaded (solid) box for the snowstorm (MCV) case.

TABLE 1. Model configurations of various experiments

	Physical parameterization schemes		
Expt	Cumulus parameterization	PBL	Microphysics
CNTL	Grell	Eta	Graupel Reisner
KFens	KF	Eta	Graupel Reisner
KF2ens	KF2	Eta	Graupel Reisner
BMens	BM	Eta	Graupel Reisner
KUOens	KUO	Eta	Graupel Reisner
Multi1	Grell, BM, KUO, KF	Eta	Graupel Reisner
Multi2	KF2, BM, KUO, KF	Eta	Graupel Reisner
KF3ens	KF	MRF	Graupel GSFC
Multi3	KF2, BM, KUO, KF	MRF	Graupel GSFC
Multi4	Refer to Table 2		

2. Methodology and experimental design

Unless otherwise specified, the EnKF system used here is the same as that employed in Part I (section 2). It is a square root EnKF with 40 ensemble members that uses covariance relaxation [Zhang et al. 2004, their Eq. (5) where $\alpha = 0.5$] to inflate the background error covariance. The Gaspari and Cohn (1999) fifth-order correlation function with a radius of influence of 30 grid points (i.e., 900 km in horizontal directions and 30 sigma levels in vertical domain) is used for covariance localization.

The third version of the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5; Dudhia 1993) is used herein with 190×120 horizontal grid points and 30-km grid spacing to cover the continental United States (Fig. 1; a slightly newer update of MM5 version 3 is used here than was used in Part I). The model setup also includes 27 layers in the terrain-following vertical coordinate with the model top at 100 hPa, and a smaller vertical spacing within the boundary layer. The National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis data are used to create the initial and boundary conditions.

Various experiments are performed with different model configurations (Table 1) to explore the sensitivity of the EnKF to the uncertainties in physical parameterizations. Serving as a benchmark, the control experiment "CNTL" is performed under the assumption of a perfect forecast model in the same manner as in Part I (section 3) and it utilizes the Grell cumulus parameterization scheme, the Reisner microphysics scheme with graupel, and the Mellor–Yamada (Eta) planetary boundary layer (PBL) scheme [refer to Grell et al. (1994) and Wang and Seaman (1997) for a de-

scription of different parameterization schemes]. Aside from the control experiment, sensitivity experiments are conducted with different cumulus parameterizations (Kfens, KF2ens, BMens, KUOens, Multi1, Multi2, KF3ens, Multi3, and Multi4) and are described in Tables 1 and 2.

The initial conditions for both the truth simulation and the ensemble are generated with the MM5 3DVAR method (Barker et al. 2004; Part I). The perturbation standard deviations are approximately 1 m s^{-1} for wind components u and v, 0.5 K for temperature T, 0.4 hPa for pressure perturbation p', and 0.2 g kg^{-1} for water vapor mixing ratio q. Other prognostic variables (vertical velocity w, mixing ratios for cloud water q_c, rainwater q_r, snow q_s, and graupel q_g) are not perturbed. The 3DVAR perturbations are added to the NCEP reanalysis at 0000 UTC 24 January 2000 to form an initial ensemble that is then integrated for 12 h to develop a realistic, flow-dependent error covariance structure before the first data is assimilated. A relaxation inflow–outflow boundary condition is adopted for both the truth simulation and the ensemble.

As in Part I, the tendencies in lateral boundaries are not perturbed. Instead, the analysis step of the EnKF is implemented only upon an inner area far from the inflow boundary as shown by the shaded box in Fig. 1. Since the reference ensemble forecast has no apparent decrease of variance in the inner (assimilation) domain, the lack of boundary perturbations is assumed to have minimal effects upon the experiments. Also, examination of both the EnKF analyses and subsequent ensemble forecasts reveals no apparent inconsistencies at and near the boundaries of the inner (assimilation) domain.

Simulated soundings and surface observations of u, v, and T are extracted from within the assimilation domains of the truth simulations. The soundings are spaced every 300 km horizontally and sounding obser-

TABLE 2. Model configuration of experiment Multi4.

No. of ensemble members	Cumulus scheme	No. of ensemble members	Cloud microphysics scheme	No. of ensemble members	PBL scheme
10	Grell	5	Graupel Reisner	3	Eta
				2	MRF
		5	Graupel GSFC	3	MRF
				2	Eta
10	KUO	5	Graupel Reisner	3	Eta
				2	MRF
		5	Graupel GSFC	3	MRF
				2	Eta
10	KF	5	Graupel Reisner	3	Eta
				2	MRF
		5	Graupel GSFC	3	MRF
				2	Eta
10	BM	5	Graupel Reisner	3	MRF
				2	Eta
		5	Graupel GSFC	3	Eta
				2	MRF

vations are taken at every sigma level. Surface observations are spaced every 60 km at the lowest model level (approximately 36 m above the ground). Assimilation of real surface observations could be more problematic due to the representative error, strong gradients, and fluxes near the surface. We assume that all observations have independent Gaussian errors with zero mean and a standard deviation of 2 m s^{-1} for u and v, and 1 K for T. Sounding and surface observations are assimilated every 12 and 3 h, respectively. Starting from the 12th hour into the integration, data assimilation continues for 24 h. Only the state variables inside the shaded box are updated and analyzed.

3. Overview of the event and the control experiment

a. Synoptic overview

The case that we investigate is an intense winter storm that occurred during 24–26 January 2000 off the southeastern coast of the United States and brought heavy snowfall from the Carolinas through the Washington, DC, area and into New England. Snow associated with this storm fell across North Carolina and the Raleigh–Durham area reported a record snowfall total of over 50 cm (Zhang et al. 2002). The system developed as an upper-level short wave embedded in a broad synoptic trough over the eastern United States moved southeastward across the southeast states. A 300-hPa low formed around 0000 UTC 25 January near the coasts of Georgia and South Carolina and moved northward along the coast. The upper-level low reached southeastern North Carolina by 1200 UTC 25

January (Zhang et al. 2002, their Fig. 2), and the storm produced the most intense snowfall in this area. The minimum mean sea level pressure (MSLP) associated with the surface cyclone rapidly dropped from 1005 hPa at 1200 UTC 24 January to 983 hPa at 1200 UTC 25 January 2000. The surface low then gradually weakened as it followed the northward-moving upper low along the coast.

b. The control EnKF experiment

The control EnKF experiment (CNTL) utilizes the perfect model scenario of Part I in which the truth and the ensemble are simulated with the same forecast model physics configuration. The truth simulation is the ensemble member that most accurately simulates the observed location and intensity of the surface and 300-hPa cyclones and simulated reflectivity (Fig. 2). The reference ensemble forecast uses the same initial conditions as the ensemble in CNTL but is a pure ensemble forecast without any data assimilation, demonstrates rapid error growth in terms of both MSLP and surface wind forecast error (Figs. 3a,b) and in terms of the square root of column-averaged (mean) difference total energy (RM-DTE; Figs. 3d,e). The DTE is defined as

$$\text{DTE} = 0.5(u'u' + v'v' + kT'T'), \qquad (1)$$

where the prime denotes the difference between the truth and the ensemble mean or between any two realizations, $k = C_p/T_r$, $C_p = 1004.7$ J kg^{-1} K^{-1} and the reference temperature $T_r = 270$ K. Large increases can be seen in the maximum forecast error of different variables from 12 to 36 h. For example, the error increases from 1.5 to 8.5 hPa with MSLP (Figs. 3a,b, respec-
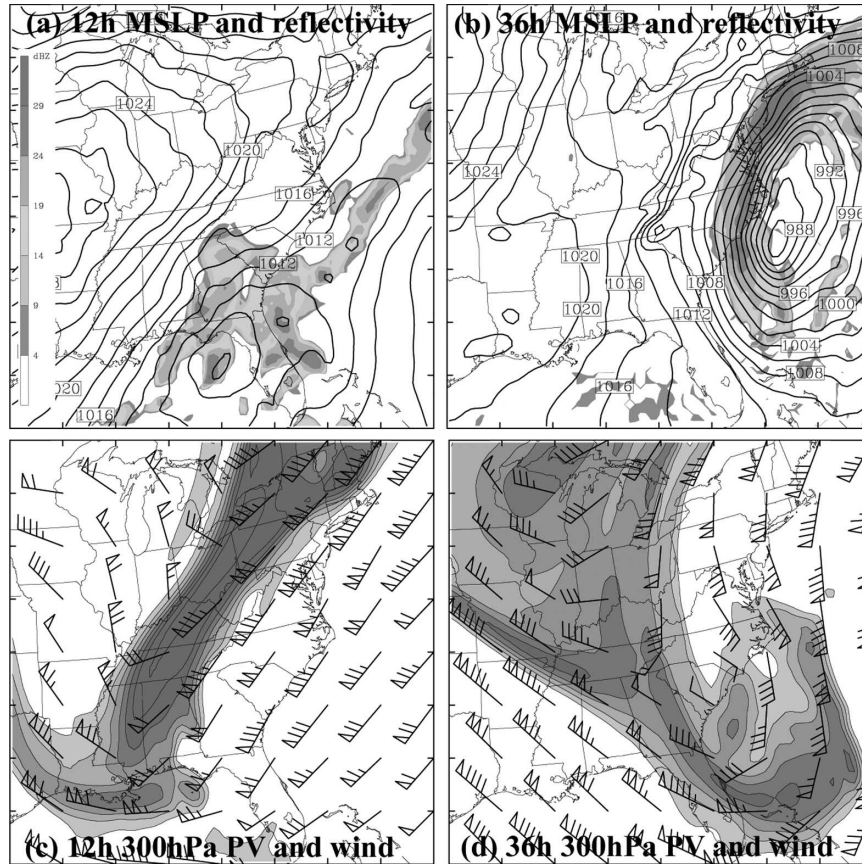
FIG. 2. The MSLP (every 2 hPa) and simulated reflectivity (shaded) valid at (a) 12 and (b) 36 h from the truth simulation for the snowstorm case. (c), (d) Same as in (a), (b) but for the potential vorticity (every 1 PVU) and wind vectors (full barb 5 m s^{-1}) at 300 hPa.

tively), from 2.5 to 12.5 m s^{-1} with the surface wind, and from 1.2 to 16 m s^{-1} for the column-averaged RM-DTE (Figs. 3d,e). Note that large errors generally occur near the surface cyclone, the upper-level shortwave trough, and the associated fronts and moist processes. These results are consistent with Part I, Zhang et al. (2002, 2003) and Zhang (2005).

After 24 h of assimilation with the EnKF, the analysis error (defined as the difference between the posterior ensemble mean and the truth) decreases significantly for all variables of interest. The EnKF analysis of MSLP (Fig. 3c) and the column-averaged RM-DTE (Fig. 3f) are almost indistinguishable from those of the truth simulation. Relative error reduction [(RER) as in Part I] will be used to verify the performance of the EnKF. RER is defined as

$$\text{RER} = \frac{E_f - E_a}{E_f} \times 100\%, \qquad (2)$$

where E_f denotes the root-mean-square error of the reference ensemble forecast of an arbitrary variable in

an experiment, and E_a denotes the root-mean-square error of the corresponding analysis of the same variable. As shown in the time evolution of forecast and analysis error for different variables including u, v, T, p', w, and q (Fig. 8 in Part I), the EnKF reduces the analysis error by as much as 85% for pressure perturbation, 80% for horizontal wind and temperature, 45% for water vapor mixing ratio, and 30% for vertical velocity. The largest improvement is obtained when both sounding and surface observations are assimilated (Part I).

The effectiveness of the EnKF in CNTL is shown in Figs. 4 and 5. For example, there is no apparent filter divergence because the ensemble spread (dotted line in Fig. 4a) and analysis errors (solid thick dark-gray line in Fig. 4a) are quite close to each other. Also, compared to that in the reference ensemble forecast without data assimilation (dashed thick dark-gray line in Fig. 4a), RM-DTE is reduced by 73% (to 1.1 m s^{-1}) after the 24-h assimilation period (solid thick dark-gray line in Fig. 4a). In fact, the RM-DTE value after the assimila-

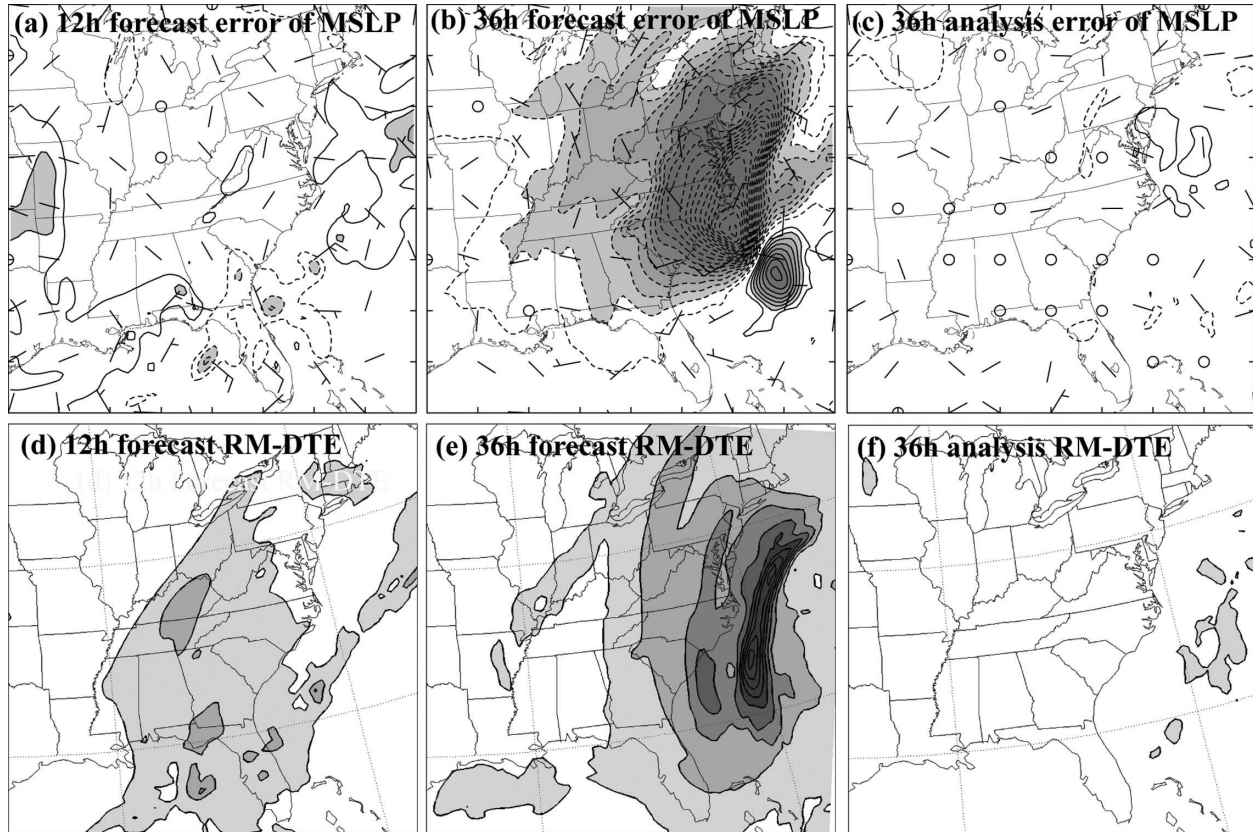FIG. 3. Forecast errors of surface wind vectors (full bard 5 m s^{-1}) and MSLP (every 0.5 hPa) at (a) 12 and (b) 36 h for the snowstorm case. (c) The analysis error of the same fields at 36 h. (d)–(f) Same as in (a)–(c) but for the column-averaged RM-DTE (every 2 m s^{-1}).

tion period is less than that of typically specified observation errors. The vertical profile of horizontally averaged RM-DTE at 36 h (Fig. 4d, thick dark-gray lines) suggests that the largest improvement occurs where the reference ensemble forecast has the largest error. Moreover, the power spectrum analysis of DTE at 36 h (Fig. 5a, solid thick dark-gray line for analysis and dotted line for reference forecast) demonstrates that the EnKF is very efficient at decreasing the error at larger scales where the covariance is most reliable. The EnKF less effectively reduces error at smaller, marginally resolvable scales. This is possibly due to the poor representation of background error covariance, faster error growth at smaller scales, and/or insufficient observation information (Part I).

4. Sensitivity to model error in physical parameterizations

As mentioned in the introduction, model error can result in bias of the ensemble mean and insufficient ensemble spread due to its smaller projection onto the correct error growth direction. In numerical models,

those processes that cannot be explicitly resolved have to be approximated through different parameterization schemes that are major sources of model error. To test the performance of the EnKF in the presence of model error caused by physical parameterization schemes, we assume that the Grell cumulus scheme, the Eta PBL, and the Reisner microphysics with graupel, which are employed to generate the truth simulation, are perfect. The ensemble forecast in the sensitivity experiments is then performed with either one or multiple parameterization schemes that differ from the truth simulation.

a. Impact of cumulus parameterization under perfect PBL and microphysics schemes

Cumulus parameterization, the problem of formulating the statistical effects of moist convection to obtain a closed system for predicting weather and climate (Arakawa 2004), has greater uncertainty than any other aspect of mesoscale numerical prediction (Molinari and Dudek 1992). Cumulus parameterization generally improves precipitation forecasts when it is utilized in a global–synoptic-scale model with a grid spacing of about 100 km or larger (Molinari and Corsetti 1985).
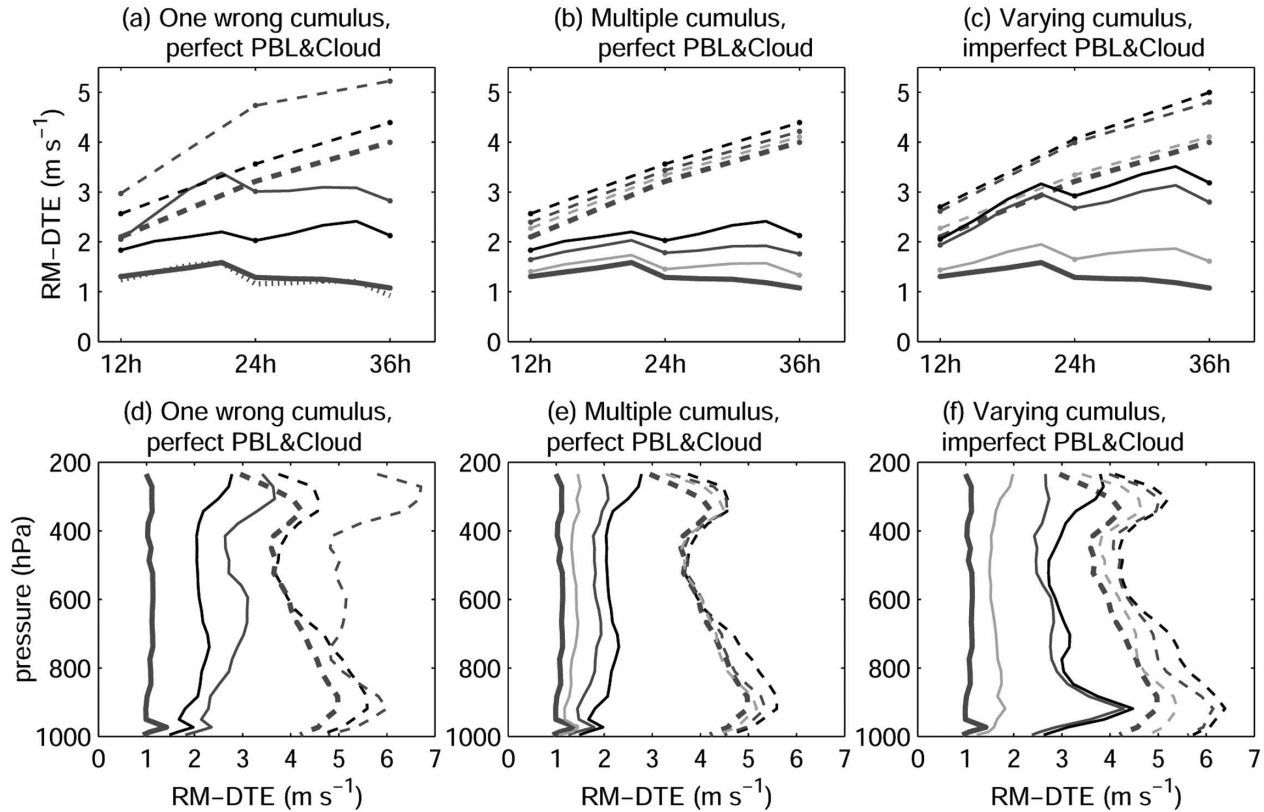
FIG. 4. (a)–(c) Time evolution of domain-averaged RM-DTE for different experiments and (d)–(f) the vertical distribution of horizontally averaged RM-DTE of the EnKF analysis (solid lines) and corresponding reference forecast (dashed lines) for the snowstorm case at 36 h. (a), (d) One wrong cumulus parameterization scheme and perfect PBL and microphysics. This includes experiments KFens (thin black), KUOens (thin dark gray), and CNTL (thick dark gray). The dotted line in (a) denotes the standard deviation of the EnKF analysis ensemble in CNTL in terms of RM-DTE. (b), (e) Same as (a), (d) but for multiple cumulus schemes, including Multi1 (thin gray), Multi2 (thin dark gray), KFens (black), and CNTL (thick dark gray). (c), (f) Same as in (a), (d) but for varying cumulus and imperfect PBL and microphysics including experiments KF3 (black), Multi3 (thin dark gray), Multi4 (thin gray), and CNTL (thick dark gray).

Problems arise when the grid spacing reduces to below 50 km and initially irresolvable clouds turn into resolvable mesoscale circulations at later times. The lack of a power gap between cloud scale and mesoscale renders the conceptual basis of cumulus parameterization ill-posed for smaller grid spacings (Cotton and Anthes 1989).

Cumulus parameterization schemes generally contain convective initiation, a closure assumption, and a cloud model. Different approaches to these three elements form different parameterization schemes. For example, seven cumulus parameterization schemes are available with MM5 (refer to Grell et al. 1994 for descriptions of individual schemes). Here we choose two convective adjustment methods that do not explicitly formulate the convective process [the Anthes–Kuo scheme (KUO) and the Betts–Miller scheme (BM)] and three mass flux methods [the original Kain–Fritch scheme (KF), the revised Kain–Fritsch scheme with shallow convection (KF2), and the Grell scheme],

which include a cloud model to directly simulate the convective process.

1) THE USE OF A SINGLE BUT WRONG CUMULUS PARAMETERIZATION (SINGLE-SCHEME ENSEMBLE)

Four experiments named KUOens, KFens, KF2ens, and BMens are executed to evaluate the EnKF performance with the use of a single wrong cumulus parameterization scheme in the ensemble forecast ("wrong" implies a difference from rather than inferiority to the scheme used for the truth). In these experiments, the truth simulation is generated using the Grell scheme (as in CNTL), but each ensemble forecast for the EnKF uses one of the four different cumulus parameterization schemes [i.e., KUO, KF, KF2, and BM (Table 1)]. Because different physical parameterizations rely on significantly different underlying assumptions, the use of any scheme in the ensemble other than that used to
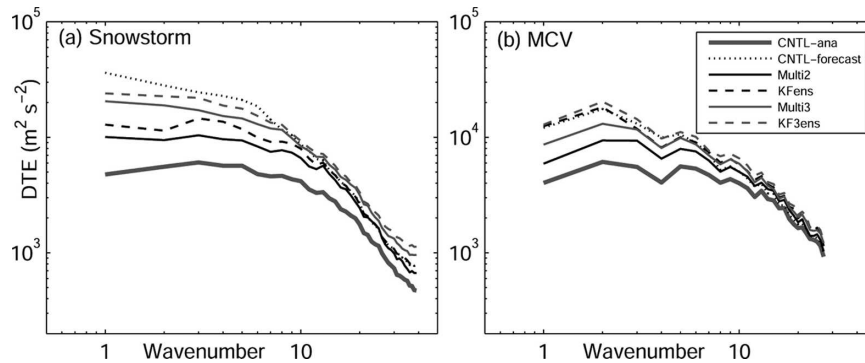
FIG. 5. Power spectrum of DTE for (a) the snowstorm at 36 h and (b) the MCV event at 48 h. The minimum (maximum) wavenumber 1 (40) in (a) and 1 (28) in (b) correspond to a horizontal wavelength of 2400 (60) km in (a) and 1680 (60) km in (b).

generate the truth simulation will unavoidably incur model error. This is also true when using any single-scheme ensemble to assimilate real-world observations.

To simplify subsequent discussions, we define "bias" as the difference between the ensemble means of the reference ensemble forecast with the perfect physics and the one with imperfect scheme(s) in terms of root-mean difference total energy [RM-DTE, defined in Eq. (1)]. The biases of the four ensemble means (Figs. 6a,c) are found to be significantly different from each other (the mean sampling error in bias estimation is less than 0.1 K for temperature and less than 0.2 m s^{-1} for u and v). KFens (dashed black in Fig. 6a) and KUOens (dashed gray), respectively, have the smallest and largest biases, while those of BMens and KF2ens (not shown) are between the two extremes. This suggests that the Kain–Fritsch and Grell (truth) schemes, which are significantly different from the other two convective adjustment schemes, perform similarly to one another in the winter season. Also, the magnitude of the bias is very different among the ensembles at the altitude of its two primary vertical peaks (dashed black and gray lines in Fig. 6c). These peaks are located at around 850 and 300 hPa and are associated with moist convection and upper-level fronts, respectively.

The spectral analysis of bias in the above experiments indicates that it exists mostly at large scales and that it is noticeably different among different schemes (not shown). For example, BMens exhibits a similar bias to KUOens at large scales but has a relatively smaller bias at smaller scales. The bias of KFens is consistently smaller at all scales than that in KUOens and BMens. Moreover, differences are also observed in the domain-averaged reference ensemble spread at 36 h (Fig. 6b), with the smallest spread in KUOens due to its smaller spread at lower levels (dashed gray line in Fig. 6d). The aforementioned differences in the error

growth structure will have profound impacts on the performance of the EnKF.

Figures 7 and 4a demonstrate degraded EnKF performance in the single wrong cumulus parameterization experiments (their ~50% error reduction is significantly less than the 73% error reduction in CNTL). The decreased performance is possibly a result of the worsened error covariance structure and bias of the ensemble mean. In general, the larger the mean bias of the reference forecast (model error) or the smaller the ensemble spread, the larger the degradation of the EnKF performance. This is demonstrated among the four single-scheme experiments, where KUOens shows the least improvement (46%), while KFens, KF2ens, and BMens show error reductions of 52%, 48%, and 55%, respectively (black bins in Fig. 7a). Similarly, the absolute analysis error measured in terms of the domain-averaged RM-DTE after the 24-h EnKF assimilation is 2.8, 2.0, 2.1, and 2.3 m s^{-1} for KUOens, BMens, KFens, and KF2ens (black bins in Fig. 7b), respectively. This analysis error is comparable in magnitude to the observational error specified. In addition, most of the error reduction comes from larger scales and the maximum error decrease is obtained in the lower troposphere (Fig. 4d).

2) THE USE OF MULTIPLE CUMULUS PARAMETERIZATION SCHEMES (MULTISCHEME ENSEMBLE)

In practice, it is hard to determine a priori which cumulus parameterization scheme is the most suitable to predict certain kinds of weather systems in different flow regimes. For example, the above single-scheme experiments demonstrate that model error due to the use of a single wrong cumulus scheme can degrade the EnKF performance to different degrees. Also, Wang and Seaman (1997) compared the performance of four different cumulus parameterizations (i.e., the KUO,
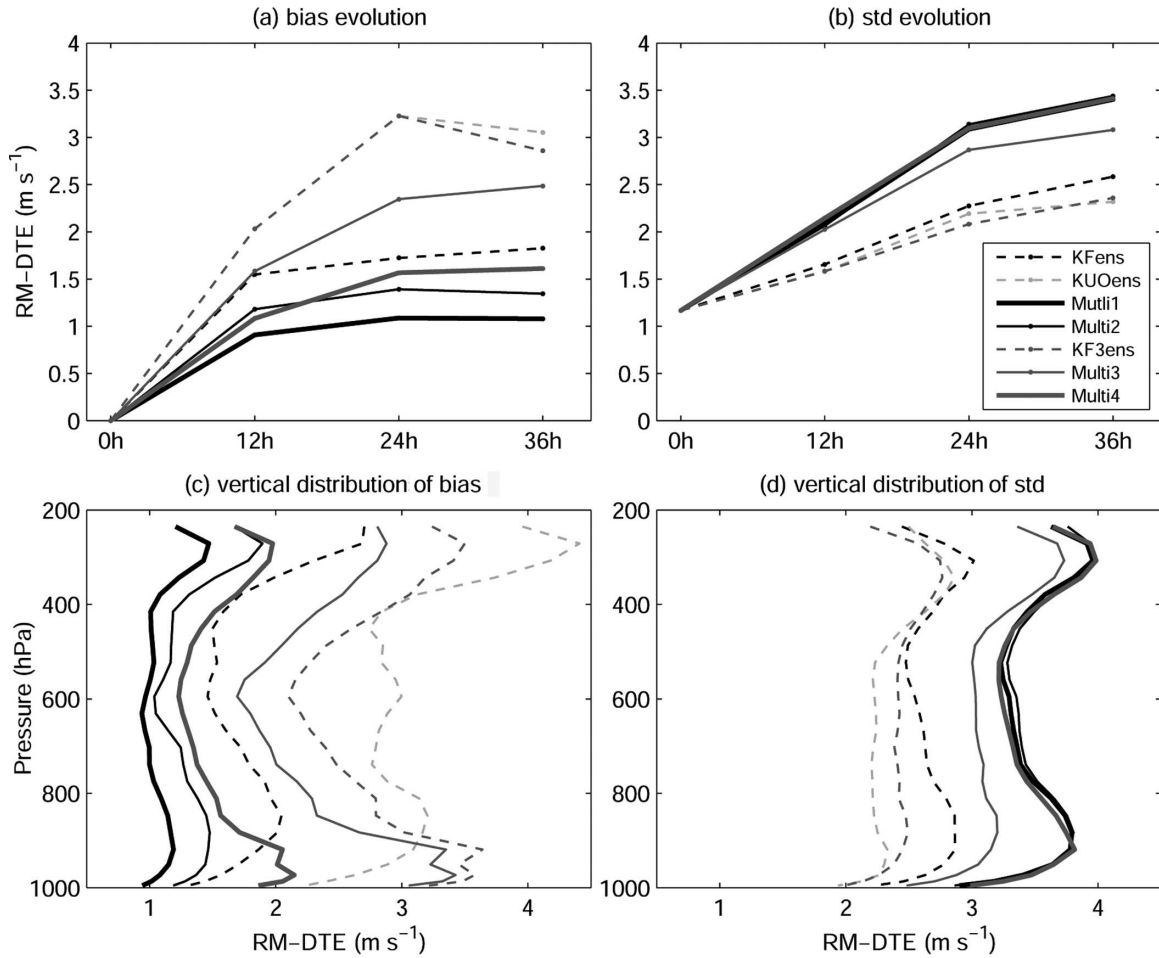
FIG. 6. Time evolution of (a) the bias (the rms difference between the imperfect-experiments' reference ensemble mean and the CNTL reference ensemble mean) in terms of RM-DTE and (b) the corresponding reference ensemble spreads (std dev or std) of RM-DTE for the snowstorm case. (c), (d) Same as in (a), (b) but for the vertical distribution at 36 h. The dashed lines denote one-scheme ensembles with black for KFens, gray for KUOens, and dark gray for KF3ens. The solid lines represent multischeme ensembles including Multi1 (thick black), Multi2 (thin black), Multi3 (thin darkngray), and Multi4 (thick dark gray).

BM, KF, and Grell schemes) in MM5 and showed that none of them demonstrates consistently better results than others.

A very natural treatment to account for model error from cumulus parameterization is thus to integrate an ensemble using a combination of different cumulus parameterization schemes (Stensrud et al. 2000; Grell and Devenyi 2002). Through the use of different closure assumptions, cloud models, and convection triggering mechanisms, a multischeme ensemble may provide a better estimate of the background error covariance by including both initial condition and model uncertainties. In this context, experiment Multi1 (Table 1) is constructed by adopting four different cumulus parameterization schemes including the Grell, KF, BM, and KUO schemes in the ensemble forecast (which implies that part of the cumulus parameterizations used in the mul-

tischeme ensemble are perfect). These four schemes are each assigned to a 10-member subset of the 40-member ensemble. Our use of a multischeme ensemble was motivated by a recent study using real-data EnKF experiments (Fujita et al. 2005).

The reference ensemble forecast of Multi1 shown in the solid thick black lines in Fig. 6 has significantly smaller bias (solid thick black line in Fig. 6a) and bigger spread (solid thick black line in Fig. 6b) at each level than do any of the single-scheme ensembles (Figs. 6c,d). As expected, the multischeme ensemble contributes to larger error reduction than do the single-wrong-scheme ensembles in the EnKF data assimilation. The domain-averaged RM-DTE and the vertical distribution of horizontally averaged RM-DTE after the 24-h data assimilation are plotted in Figs. 4b,e (thin gray lines). For direct comparison, KFens (which has aver-
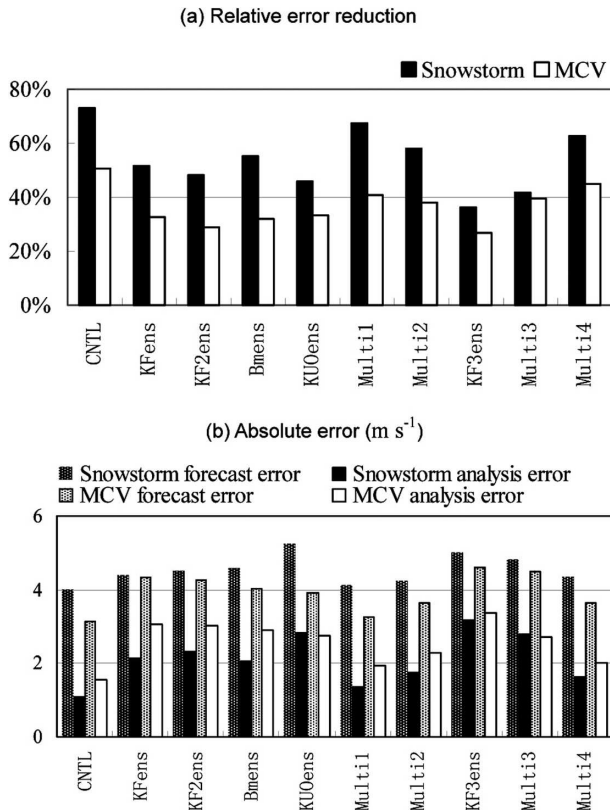
(a) Relative error reduction

(b) Absolute error (m s⁻¹)

FIG. 7. (a) Relative error reduction and (b) absolute forecast/analysis errors (m s^{-1}) in terms of domain-averaged RM-DTE at the final analysis time for the snowstorm case at 36 h (black bins) and the MCV case at 48 h (white bins). The experiments are labeled on the x coordinate.

age performance) is repeated here to represent the single-wrong-scheme experiments. Compared to the 52% improvement in KFens, nearly 67% error reduction is achieved in Multi1. The 1.3 m s^{-1} RM-DTE in Multi1 is also smaller than any of the single-wrong-scheme experiments (Fig. 7b). Again, the largest improvement occurs in the lower troposphere (Fig. 4e).

Because a quarter of the ensemble members in Multi1 still use a perfect (the Grell) scheme, which is unrealistic, experiment Multi2 replaces the Grell scheme in Multi1 with the KF2 so that all cumulus schemes used in the ensemble are different from the truth (and thus imperfect, see Table 1). The reference ensemble forecast bias in Multi2 (solid thin black line in Figs. 6a,c) is systematically larger than that in Multi1 but smaller than the bias in KFens. The relative error reduction in Multi2 is about 58% and its absolute RM-DTE is 1.8 m s^{-1} at 36 h. Though it reduces error less than Multi1, Multi2 systematically outperforms any of the single-scheme experiments (Fig. 7a). Compared to KFens (dashed black line in Fig. 5a), most of the im-

provement in Multi2 comes from larger scales (solid thin black line in Fig. 5a). The horizontal distribution of column-averaged RM-DTE also shows consistent improvement over KFens, and the greatest error reduction is in the vicinity of the surface cyclone (Figs. 8a–c).

b. Impact of cumulus parameterization under imperfect PBL and microphysics schemes

Not only does forecast error come from cumulus parameterization, but it also comes from parameterization of other subgrid-scale processes such as microphysics and PBL processes. This section explores the impact of model error from cumulus parameterization with imperfect PBL and microphysics schemes.

To account for the possibility of error from parameterization of multiple subgrid-scale processes, the ensemble in experiment KF3ens uses all imperfect schemes including the KF cumulus scheme, the MRF PBL scheme and the Goddard microphysics scheme with graupel. This ensemble performs significantly worse than any aforementioned experiment and exhibits relative error reduction of only 36% and absolute analysis error of 3.2 m s^{-1} (thin black lines in Figs. 4c,f and 7). With additional model error from PBL and cloud microphysics, the reference ensemble of KF3ens has a large bias but a small spread (the largest bias is in the lower levels among all experiments as shown in dashed dark-gray line in Fig. 6).

Experiment Multi3 expands on KF3ens by using the same combination of four (imperfect) cumulus parameterization schemes (i.e., KF, KF2, BM, and KUO) as Multi2 and the same imperfect PBL and microphysics schemes as KF3ens (Table 1). Even in the presence of model error from PBL and microphysics parameterizations, the use of the multiple-cumulus-scheme ensemble still helps to decrease the bias and increase the spread significantly at all levels (solid thin dark-gray line in Fig. 6) compared to KF3ens. Consequently, the EnKF performs better in Multi3 than in KF3ens by reducing the relative error by 42% and the absolute analysis error to 2.8 m s^{-1} at 36 h (Figs. 7 and 4c). Also, most of the improvement occurs at large scales (solid thin dark-gray line in Fig. 5a) and at middle to upper levels (solid thin dark-gray line in Fig. 4f).

Experiment Multi4 accounts for the possibility that some schemes may be nearly perfect under certain flow regimes since all parameterization schemes are developed to represent real physical processes. To do this, Multi4 uses a combination of different cumulus, PBL and microphysics schemes, each of which includes some of the same schemes as in the truth (Table 2). Specifically, each 10-member subset of Multi1 is further divided into four subsets. Among the 10 members of each
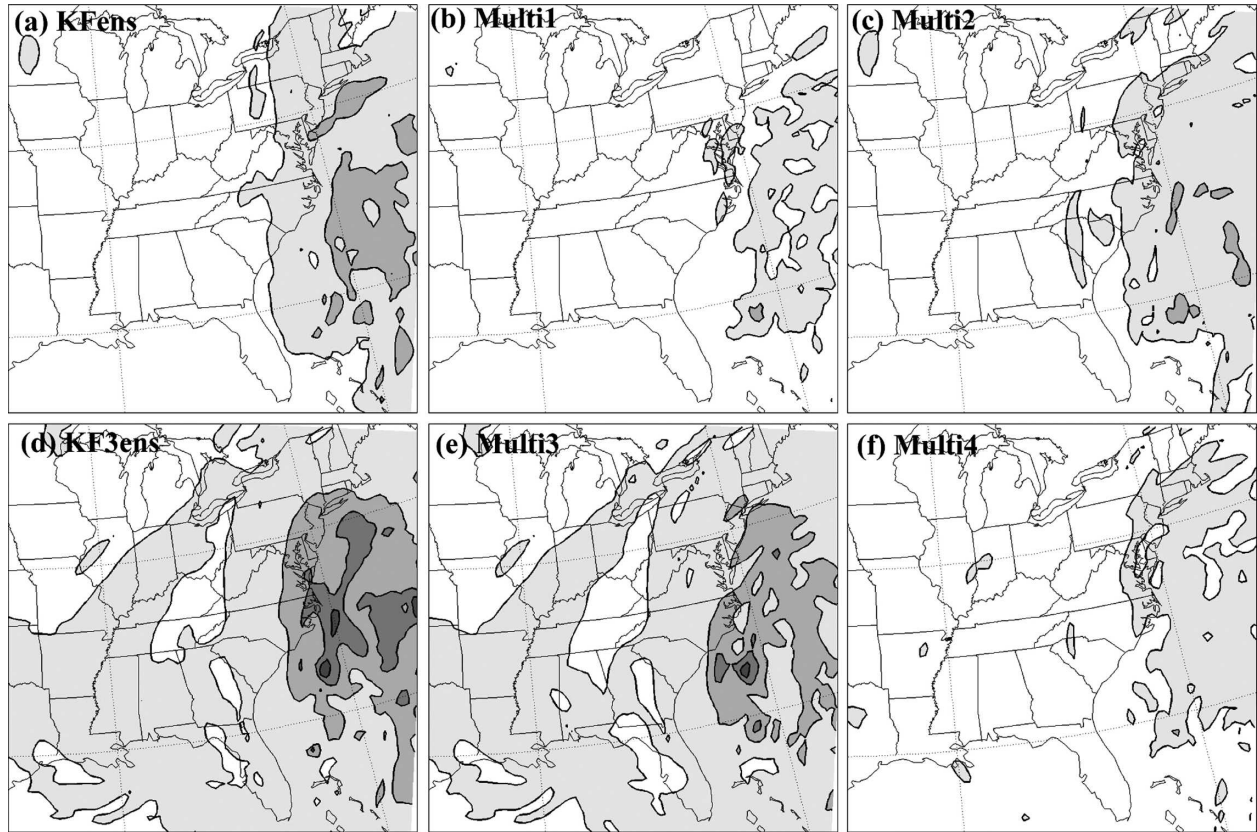
FIG. 8. Horizontal distribution of column-averaged RM-DTE (every 2 m s^{-1}) at 36 h for the snowstorm case for (a) KFens, (b) Multi1, (c) Multi2, (d) KF3ens, (e) Multi3, and (f) Multi4.

subset, five use the Reisner-graupel microphysics scheme while the other five adopt the (Goddard Space Flight Center) GSFC-graupel scheme. The five-member subsets using the Reisner-graupel scheme are further divided into two groups of three and two members employing the Eta and MRF PBL schemes, respectively. The other five members with the GSFC-graupel scheme are treated similarly except that the two PBL schemes are switched between the three- and two-member groups. This particular configuration is used to make sure that any of the three categories of the physical parameterization schemes are evenly distributed among the 40 ensemble members.

The reference ensemble forecast (without the EnKF assimilation) of Multi4 (solid thick dark-gray line in Fig. 6) has smaller bias and larger spread than those of both KF3ens and Multi3 during the whole integration period. Figure 7 also shows that Multi4 performs better than nearly all other imperfect-model experiments (except Multi1, which also includes the same schemes as in the truth). The relative error reduction for Multi4 is 63%, and absolute analysis error of 1.6 m s^{-1} is observed in this experiment. This reduction is evident in

both the column average (Figs. 8d–f) and vertical distribution (solid thin gray line in Fig. 4f) of RM-DTE. Though they might be overly optimistic, experiments Multi1 and Multi4 suggest that better EnKF performance can be achieved if parts of the parameterizations used in the multischeme ensemble are perfect.

The large differences observed between KFens and KF3ens and between Multi2 and Multi3 demonstrate that the use of imperfect PBL and microphysics schemes (in addition to imperfect cumulus parameterizations) can significantly degrade the EnKF performance (Fig. 5a). However, due to the limited availability of microphysics and PBL parameterization schemes in MM5, we cannot examine the impact of using multischeme ensembles in which none of the schemes in PBL or microphysics parameterizations is perfect (this is partially due to limited choices of usable PBL or microphysics schemes in MM5).

c. Comparison of error covariance between single- and multischeme ensembles

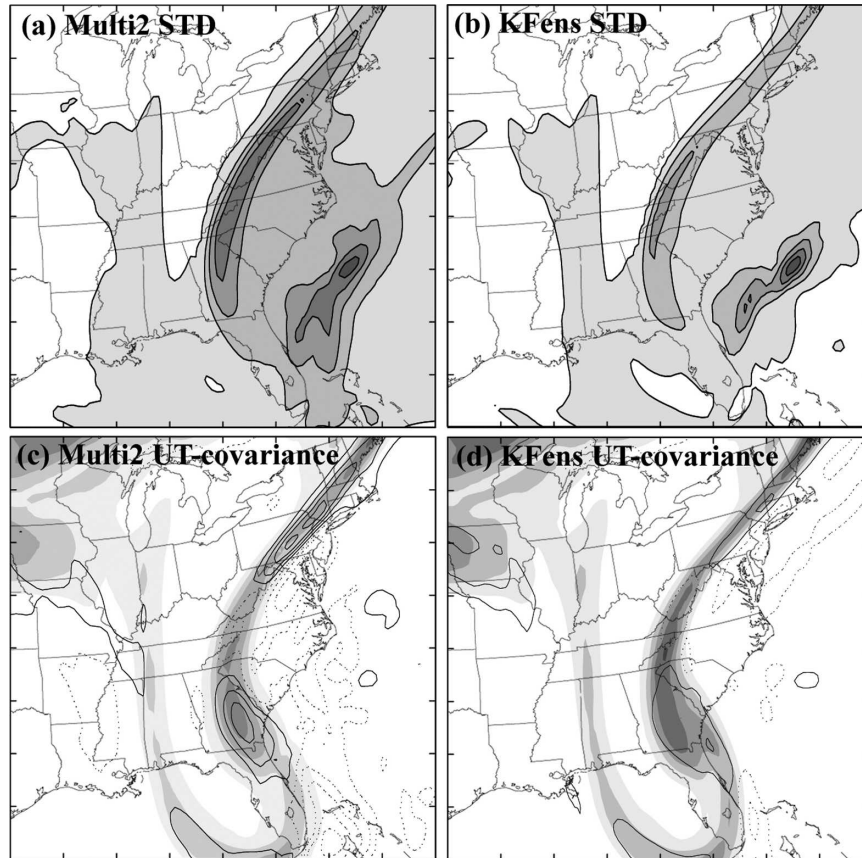This section further investigates the reasons why the EnKF performs better with a multischeme ensemble

FIG. 9. Horizontal distribution of the standard deviation of column-averaged RM-DTE (every 2 m s^{-1}) for (a) Multi2 and (b) KFens at 24 h for the snowstorm case. (c), (d) Same as in (a), (b) but for the covariance between u and T on 300 hPa (every 2 K m s^{-1}; negative, dotted). The shading in (c), (d) is PV at 300 hPa every 1 PVU.

than with a single-wrong-scheme ensemble. For example, the previous subsections showed that while the EnKF is quite effective at reducing the analysis error in the presence of significant model uncertainties, the analysis error in the imperfect-model experiments is noticeably larger than that of CNTL. This indicates that the EnKF performance can be degraded to different extents with different physical parameterizations (Fig. 7). Such difference in the EnKF performance might be due to the ensemble mean error (bias) and/or insufficient ensemble spread resulting from the use of an imperfect model.

The horizontal distributions in Figs. 9a,b show that the reference ensemble forecast of Multi2 has a significantly larger standard deviation of column-averaged RM-DTE than does KFens at 24 h. Zhang (2005), a previous study of this snowstorm, observed similar large-scale, balanced features that evolved from initially uncorrelated, small-scale, unbalanced errors in a period of 12–24 h. The maximum error growth in the

disturbances is associated with the upper trough and the surface cyclone and is collocated with the strongest PV gradient. The spectral analysis of the ensemble spread also shows a much larger difference between Multi2 and KFens at larger scales (i.e., wavenumber < 10 or wavelength > 240 km) than at smaller scales (not shown). The differences between balanced disturbances of Multi2 and KFens have implications when using the EnKF because the EnKF is most effective at correcting errors at larger scales (as shown in Part I).

To further illustrate the differences between the large-scale error structures of KFens and Multi2, the cross-covariance between u and T at 300 hPa at 24 h is also examined for each ensemble (Figs. 9c,d). While Multi2 and KFens exhibit similar covariance structures with increased covariance in the vicinity of strong PV gradients, the magnitude of the covariance in Multi2 is noticeably larger due to its relatively larger ensemble spread (Figs. 9a,b). When the ensemble spread is significantly smaller than the error of the ensemble mean,

increase of the ensemble spread could improve the performance of the EnKF. A larger spread in the multischeme ensembles may increase the likelihood of keeping the truth within the uncertainties spanned by the imperfect ensemble, and a large covariance has the potential to propagate observational information more efficiently between variables. This is consistent with Fujita et al. (2005), a recent real-data study that partially motivated the use of multischeme ensembles in the current study.

To understand whether or not the covariance structure developed in one of the above ensembles (i.e., KFens or Multi2) is systematically better than the covariance structure of the other, four static EnKF experiments (i.e., Pmulti2-Mmulti2, Pkf-Mmulti2, Pmulti2-Mkf, and Pkf-Mkf) are conducted. These experiments are "static" in the sense that observations are assimilated at only one selected time without subsequent forecast and analysis cycles. The naming convention is as follows: "M . . ." refers to the reference ensemble mean, "P . . ." refers to perturbations/deviations from the mean and ". . ." refers to the experiments in previous subsections. For example, Pmulti2-Mmulti2 and Pkf-Mkf use the (unaltered) reference ensemble forecast of Multi2 and KFens, respectively, to estimate the background error covariance for the EnKF. Pkf-Mmulti2 and Pmulti2-Mkf are performed by switching the ensemble means of Multi2 and KFens so that the perturbations of Multi2 are added to the mean of KFens, and the perturbations of KFens are added to the mean of Multi2. Because any two experiments formed using the same ensemble mean have the same forecast error (e.g., Pkf-Mkf and Pmulti2-Mkf), the quality of the covariance structure associated with each ensemble can be ascertained by the differences in error between the same two experiments after the assimilation cycle (i.e., the analysis error).

The results in Table 3 show that a systematically smaller analysis error can be achieved by using the background error covariance estimated from the multischeme ensemble (Multi2) rather than the single-wrong-scheme ensemble (KFens). Similar results are also obtained for KF3ens and Multi3 (see Table 3) and for different reference forecast times (not shown). Using a multischeme ensemble is also found to be beneficial in a warm-season MCV event for both the continuously evolving and static EnKF assimilation experiments (detailed in section 5).

While KFens also has the problem that its ensemble spread (solid thin gray line in Fig. 10a) is noticeably smaller than its analysis error (solid thick gray line in Fig. 10a), the potential for filter divergence with this ensemble may be alleviated with covariance inflation.

TABLE 3. Domain-averaged RM-DTE for one-time data assimilation experiments valid at 36 (48) h for the snowstorm (MCV) case that switch perturbations between the single-scheme KFens and the multischeme EnKF experiments. EF means the reference ensemble forecast.

		RM-DTE (m s^{-1})	
Ensemble mean	Expt	Snowstorm	MCV
Multi2	EF of Multi2	4.22	3.66
	Pmulti2-Mmulti2	2.50	2.43
	Pkf-Mmulti2	2.64	2.62
KFens	EF of KFens	4.39	4.34
	Pmulti2-Mkf	2.49	2.47
	Pkf-Mkf	2.79	3.08
Multi3	EF of Multi3	4.80	4.47
	Pmulti3-Mmulti3	3.12	2.85
	Pkf3-Mmulti3	3.31	3.09
KF3ens	EF of KFens	5.00	4.61
	Pmulti3-Mkf3	3.10	2.86
	Pkf3-Mkf3	3.44	3.33

Experiment KFens_0.7 is conducted by changing the weighting coefficient α in the relaxation method [Zhang et al. 2004; their Eq. (5)] from 0.5 to 0.7 to give more weight to prior perturbations. The use of a larger weight for the prior estimate as an alternative for covariance inflation (e.g., Anderson 2001) consequently leads to systematically larger ensemble spreads (though still insufficient, solid thin black line in Fig. 10a) and slightly improved the EnKF performance over 24 h of data assimilation (solid thick black line in Fig. 10a).

When covariance inflation is applied to other ensembles for which the ensemble spread is not too small, the results are worsened somewhat. For example, when the relaxation coefficient in Multi2 is modified from 0.5 to 0.7 in experiment Multi2_0.7, the analysis ensemble spread (solid thin black line in Fig. 10b) quickly becomes larger than the analysis error ("overinflation") and the EnKF performance worsens (solid thick black line in Fig. 10b). The ensemble spread eventually gets closer to or slightly smaller than the error and draws the analysis error back to that of Multi2 at 36 h. This negative impact of overinflation is more apparent when the relaxation coefficient changes from 0.5 to 0.7 in Multi4 (Fig. 10c) because the initial spread is already comparable to the error. The larger ensemble spread results in consistently larger errors during the whole period.

d. Other experiments

Various experiments using the conventional covariance inflation of Anderson (2001) and additive error method of Hamill and Whitaker (2005) are also performed to account for model error from physical parameterizations. None of these experiments with differ-
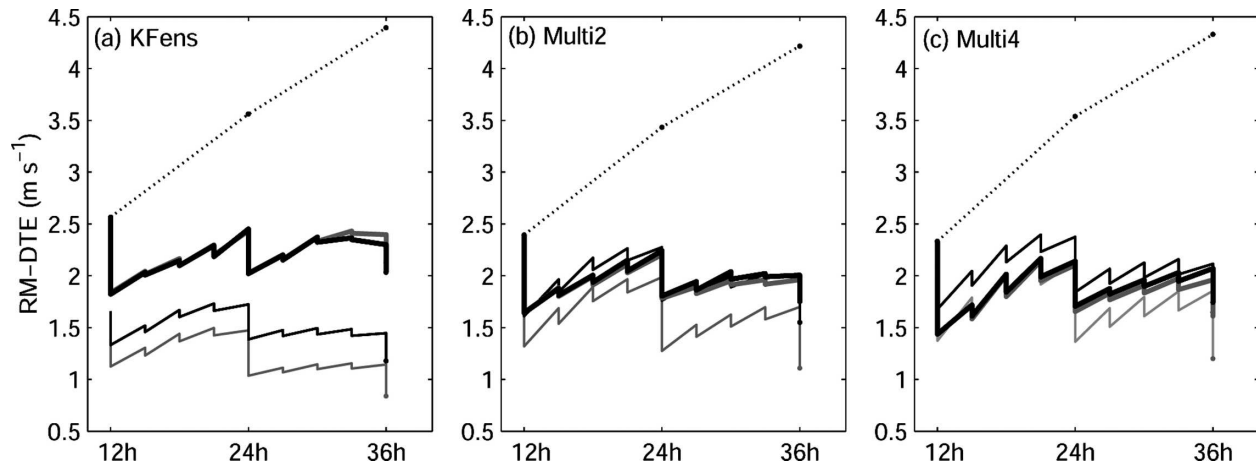
FIG. 10. The domain-averaged RM-DTE (thick solid lines) and analysis ensemble spread of RM-DTE (thin solid lines) with different weights (α) of prior perturbations in the covariance inflation (mixing) method for experiments (a) KFens, (b) Multi2, and (c) Multi4 for the snowstorm case. The black lines are for $\alpha = 0.7$, gray lines for $\alpha = 0.5$. The reference ensemble forecast errors are also plotted in dotted lines.

ent covariance inflation factors or different additive error gives acceptable EnKF performance (not shown). The traditional inflation leads to spuriously large ensemble spread in data-sparse areas. For the additive error experiments, the additive error covariance sampled from the differences between different cumulus parameterization schemes (at different times) fails to increase the ensemble spread in desired regions where there is active parameterized convection at analysis times. This result is in strong contrast to the success of using similar additive error methods to account for model truncation error (Hamill and Whitaker 2005), which is likely to be less flow dependent.

5. Impact of flow-dependent error growth dynamics

In this section, we investigate the performance of the EnKF for a vastly different flow regime than in previous sections. Since weather systems under different flow regimes may have different error growth dynamics and mesoscale predictability, and the EnKF performance is significantly scale and dynamic dependent (Part I), the EnKF is likely to behave differently in different regimes. The particular case examined is a long-lived warm-season MCV event that occurred on 10–13 June 2003. A recent study (Hawblitzel et al. 2007) shows that the predictability of this MCV event is very limited due to its extreme sensitivity to convection. This result is not surprising given that past studies (e.g., Wang and Seaman 1997; Zhang et al. 2006b) suggest that model error, especially that from cumulus param-

eterization, can be more detrimental to warm-season forecasts than to winter events.

a. Overview of the MCV event and the EnKF configuration

This MCV event occurred during an intense observation period (IOP8) of the Bow Echo and Mesoscale Convective Vortex Experiment (BAMEX) conducted from 18 May to 7 July 2003 over the central United States. At 0000 UTC 10 June 2003, a disturbance embedded in the subtropical jet triggered convection over eastern New Mexico and western Texas. An MCV developed from the remnants of this convection over central Okalahoma at 0600 UTC 11 June 2003, and matured by 1800 UTC 11 June 2003 as it traveled northeastward to Missouri and Arkansas. The MCV transitioned into an extratropical baroclinic system after 0000 UTC 12 June 2003.

The EnKF configuration is the same as for the winter snowstorm event except that a 15-point (450 km) radius of influence here due to the relatively smaller scale of the weather system. The assimilated data and the updated grid points are constrained to within the solid box of Fig. 1. Because of the longevity of the MCV, a 36-h data assimilation is performed from 1200 UTC 10 June to 0000 UTC 12 June 2003. The assimilation follows a 12-h ensemble forecast that starts at 0000 UTC 10 June 2003. Employing the same method used for the winter case, synthetic soundings are assimilated at 12-h intervals and synthetic surface observations are assimilated every 3 h. The ensemble member with the 48-h forecast being closest to the observed MCV is adopted as the
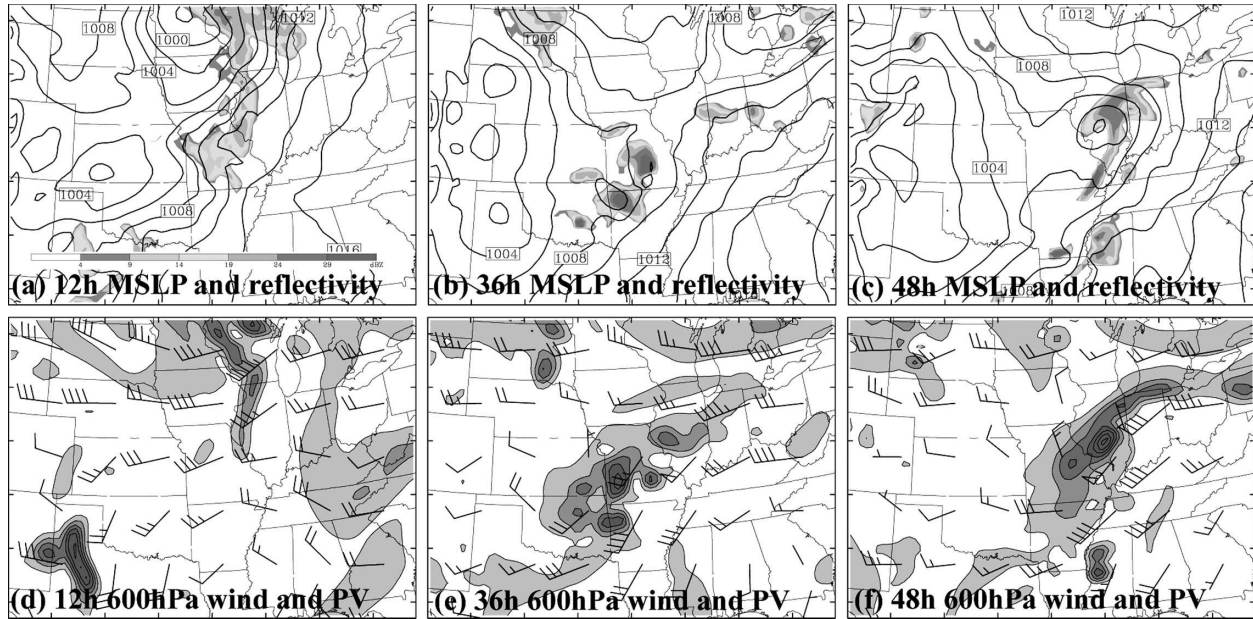
FIG. 11. The MSLP (every 2 hPa) and simulated reflectivity (shaded) valid at (a) 12, (b) 36, and (c) 48 h and the potential vorticity (every 1 PVU) and wind vectors (full barb 5 m s^{-1}) at 600 hPa valid at (d) 12, (e) 36, and (f) 48 h from the truth simulation for the MCV case.

truth from which the observations are extracted (Fig. 11).

b. The control EnKF experiment for the MCV event

The control experiment for this MCV event, which is also conducted under a perfect model assumption using the Grell scheme (as in the snowstorm simulation), reveals that the largest errors are strongly associated with the MCV dynamics. The reference ensemble forecast error in terms of both the MSLP and the surface wind at 12 and 48 h and the column-averaged RM-DTE are shown in Fig. 12. Comparison of Figs. 12 and 3 reveals that the overall error amplitude in this MCV event at 36 h (as well as 48 h) is significantly smaller than that in the snowstorm event. Spectral analysis of the reference ensemble forecast error shows that the MCV event has relatively larger smaller-scale error but smaller larger-scale error (dotted line in Fig. 5b) compared to the snowstorm event (dotted line in Fig. 5a). The smaller-scale error in the MCV event initially grows faster and quickly saturates while the larger-scale error grows slowly.

Despite the apparent difference in error, spectral composition, and growth rate between the MCV event and the snowstorm event, the control EnKF (CNTL) performs reasonably well for the MCV event. After the 36-h data assimilation in CNTL, the maximum MSLP error is reduced from 4 to 1 hPa while the area of error

larger than 0.5 hPa also decreases significantly (Fig. 12c). Error reduction in the surface wind field is also apparent as the maximum error value reduces from approximately 7.5 to 5 m s^{-1} (Fig. 12c). Significant error reduction is also exhibited in column-averaged RM-DTE for the entire assimilation domain, especially where the MCV is located. Furthermore, the maximum RM-DTE value decreases from 8 to 2 m s^{-1} (Fig. 12f). At 600 hPa, the maximum PV error reduces from 2.5 to 1 PVU and the maximum velocity error decreases from 10 to 2.5 m s^{-1} in the vicinity of the MCV (not shown).

Figure 13 shows that the evolution of domain-averaged root-mean-square analysis and forecast error and the analysis ensemble spread for u, v, T, p', w, and q for the CNTL of the MCV event are similar to those of the snowstorm case (see Fig. 8 in Part I). As with the winter case, the ratio of the analysis error to the ensemble spread is very close to 1.0 (except for p' and w), suggesting no apparent filter divergence for the warm-season event. After the 36-h data assimilation, the relative error reduction of the observed variables u, v, and T is about 40%–60%. Pressure perturbation (p') still has the largest error reduction of about 60%, but its reduction is still less than that with the snowstorm event. Also, about 40% improvement is obtained in the moisture field. Again, the least improvement (about 37%) is observed with vertical velocity. In terms of column-averaged RM-DTE, the overall error reduction at 48 h is about 51% (Fig. 7a and thick dark-gray lines
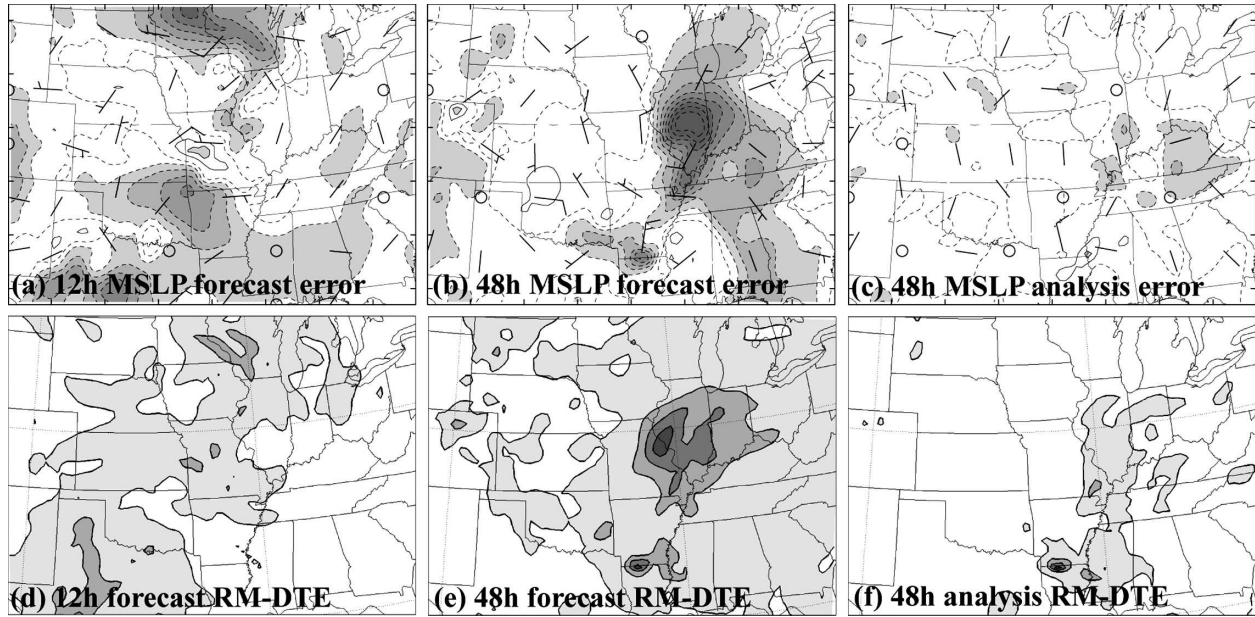
FIG. 12. Forecast errors of surface wind vectors (full bard 5 m s^{-1}) and MSLP (every 0.5 hPa) at (a) 12 and (b) 48 h for the MCV case and (c) analysis error of the same fields at 48 h. (d)–(f) Same as in (a)–(c) but for the column-averaged RM-DTE (every 2 m s^{-1}).

in Fig. 14a). As with the snowstorm case, most of the error reduction comes from larger scales (solid thick dark-gray line in Fig. 5b) and is maximized at lower levels (thick dark-gray lines in Fig. 14d). Both the

analysis error after the control EnKF assimilation (solid thick dark-gray line in Fig. 5b) and the reference ensemble forecast error have a larger smaller-scale component than does the snowstorm event error (solid
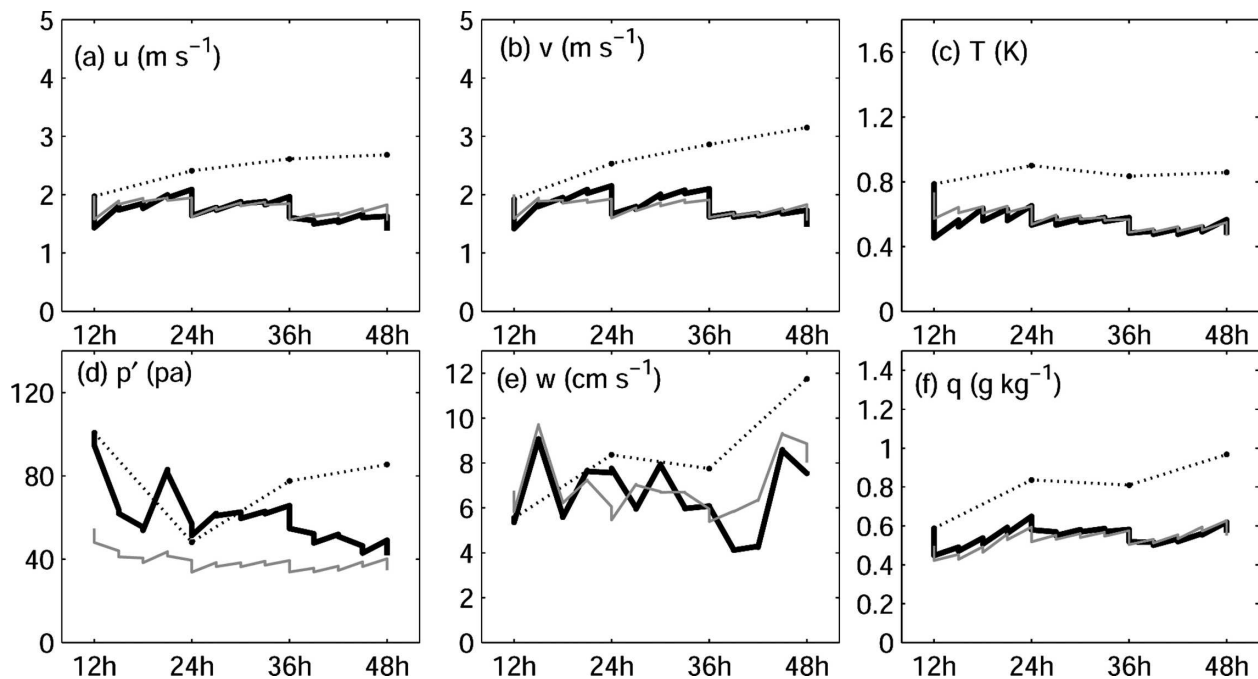

FIG. 13. Time evolution of the domain-averaged root-mean-square errors of (a) u, (b) v, (c) T, (d) p', (e) w, and (f) q for the EnKF analysis (solid black) and the reference ensemble forecast (dotted black, computed every 12 h) of CNTL in the MCV case. The gray lines are the std dev of analysis ensemble.
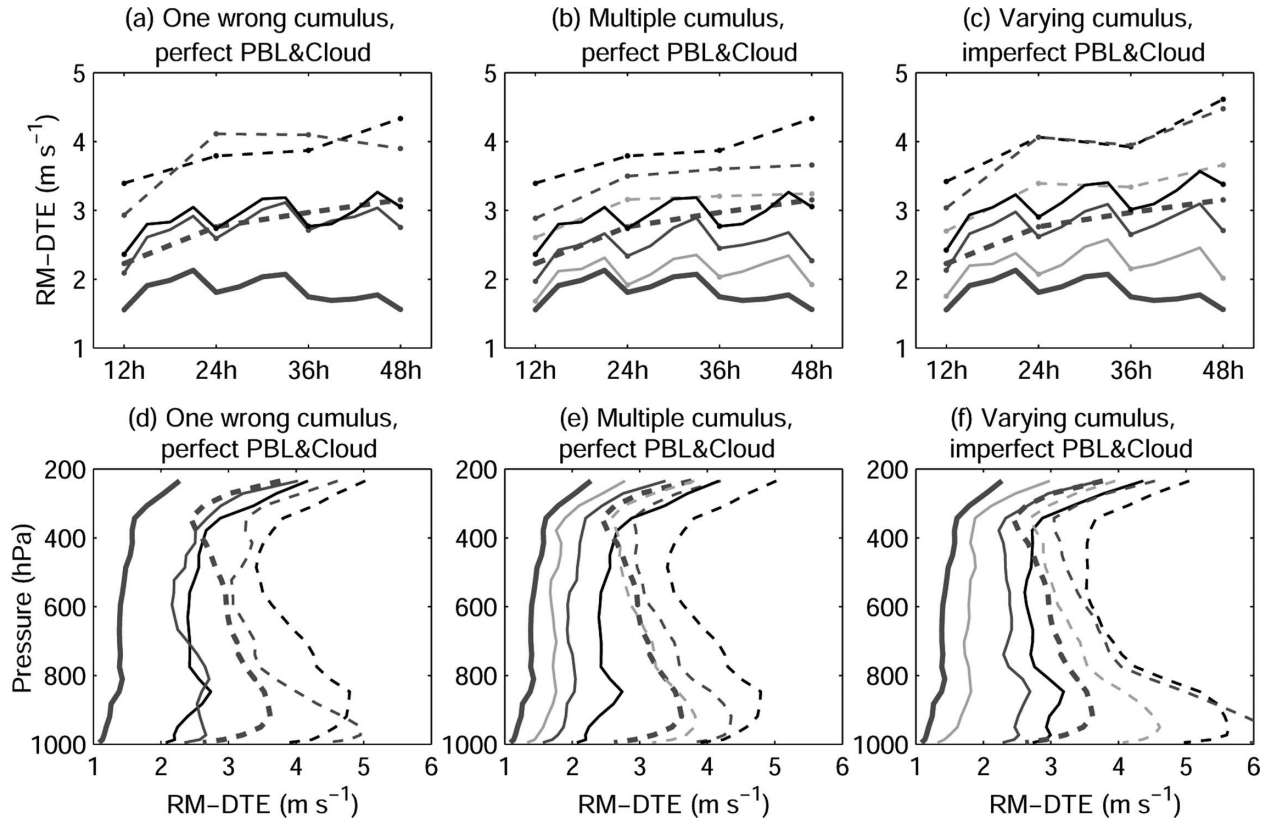
FIG. 14. Same as in Fig. 4 but for the MCV case with (d)–(f) valid at 48 h.

thick dark-gray line in Fig. 5a). Since the EnKF is less effective for small, marginally resolvable scales (Part I), the overall relative error reduction for the MCV event is smaller than that for the snowstorm event.

c. Impact of model error for the MCV event

The difference between the forecast ensemble mean in the control experiment and various sensitivity experiments using different physical parameterization schemes (bias) in this warm-season case evolves differently from that in the winter case (Fig. 15 versus Fig. 6). The multiple-cumulus-scheme ensemble biases are much closer to each other than are those in the snowstorm case. The largest bias after 48 h of integration is observed in KF2ens (not shown) and the smallest bias is observed with KUOens (dashed gray in Fig. 15). The biases of KFens (dashed black in Fig. 15) and BMens (not shown) fall between the two extremes. The ensemble spreads of these experiments are also quite close to each other (dashed lines in Fig. 15b). The vertical profiles of the biases (dashed lines in Fig. 15c) and spreads (dashed lines in Fig. 15d) exhibit a two-peak pattern similar to the winter case. The higher upper peaks in the MCV case than that in winter case are due

to the higher tropopause and upper-level fronts in the summer. The 950-hPa bias peaks in the MCV case are at a slightly lower altitude and are stronger in magnitude than the ~900-hPa bias peaks of the snowstorm case (Fig. 15b). However, the lower peaks of the ensemble spread of the MCV case are at similar altitudes to those in the snowstorm case (around 900 hPa, dashed black and gray lines in Figs. 15d and 6d).

When the EnKF is used with the above ensembles, the error reduction is smaller than with the snowstorm case and the filter performance is very similar among the experiments KFens, BMens, KF2ens, and KUOens (Figs. 7 and 14a,d). These similarities are not surprising given the similarity between reference ensembles. One possible culprit for the roughly similar results is the observed fast error saturation.

As with the snowstorm event, experiments using multischeme ensembles for this MCV event show improvement over those using single-scheme ensembles. Multi1 and Multi2, the perfect PBL and microphysics multischeme experiments, have smaller bias and larger spread than KFens (Figs. 15a,b); this result is similar to that of the snowstorm case. A systematically larger bias is observed for all experiments in the MCV case than in
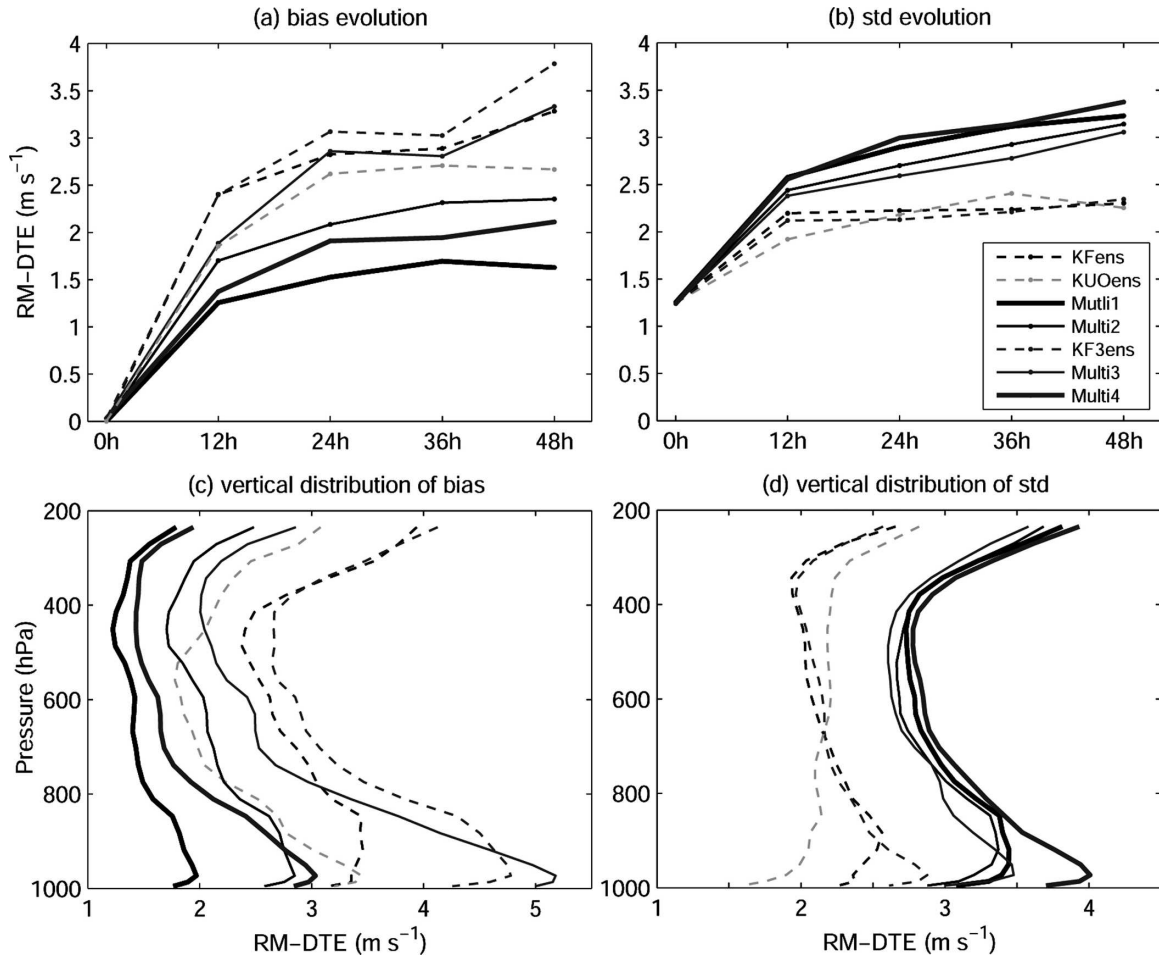
FIG. 15. Same as in Fig. 6 but for the MCV case with (c) and (d) valid at 48 h.

the winter case, and this suggests a larger impact of physical parameterizations for the warm-season event. The covariance between u and T at 600 hPa after a 36-h integration is larger in Multi2 than in KFens, but it is weaker in both experiments when compared to the covariance in the winter case. One possible reason for this is the low predictability of smaller-scale convective activity. After 36-h data assimilation, the relative error reduction for KFens, Multi2, and Multi1 is 33%, 38%, and 41%, respectively, and the absolute error is respectively 3.0, 2.3, and 1.9 m s^{-1} (Figs. 14b and 7). There is thus consistent improvement when a multischeme ensemble is adopted. Power spectrum analysis also shows that the improvement of Multi2 over KFens comes mainly from the large scales (Fig. 5b).

Similar improvement in multischeme ensembles over single-wrong-scheme ensembles is also observed under imperfect PBL and microphysics parameterizations in KF3ens, Multi3, and Multi4. Vertical distribution of the ensemble spread shows that the lower peaks of the

spreads of these three experiments are at slightly lower levels and are larger than the corresponding peaks in the snowstorm case. This indicates that PBL processes may have a larger impact on error growth in the MCV than the snowstorm case (Fig. 15d versus Fig. 6d). The EnKF result shows significant improvement of Multi3 over KF3ens (Fig. 14c) at large scales (Fig. 5b) and on each level (Fig. 14f), suggesting the multicumulus ensemble can decrease PBL error more than the winter case where very small differences are seen at lower levels between KF3ens and Multi3. Similarly, Multi4 consistently reduces error during the whole period at all levels (gray line in Figs. 14c,f).

Experiments in this MCV event further demonstrate that a multischeme ensemble is capable of providing better estimation of the background error covariance than a single-wrong-scheme ensemble. The significance of improving error covariance by using a multischeme ensemble is also demonstrated through static EnKF experiments by switching the means of the reference en-

semble forecast for KFens and Multi2 and for KF3ens and Multi3 (Table 3) in a similar way to that discussed in section 4c for the snowstorm event.

6. Conclusions and discussion

Through various observing system simulation experiments, the performance of an ensemble Kalman filter is explored in the presence of significant model error caused by physical parameterization. The EnKF is implemented in the mesoscale model MM5 to assimilate synthetic sounding and surface data derived from the truth simulations at typical temporal and spatial resolutions for the cold-season snowstorm event that occurred on 24–26 January 2000 and the warm-season MCV event that occurred on 10–13 June 2003.

Results show that although the performance of the EnKF is degraded by different degrees when a perfect model is not used, the EnKF can work fairly well in different kinds of imperfect scenario experiments. A 36%–67% overall relative error reduction (improvement over the reference ensemble forecast) is found in each imperfect scenario for the snowstorm event. In both the perfect and imperfect scenarios, most of the error reduction comes from larger scales and it is maximized in the lower troposphere.

The performance of the EnKF was tested and found to be very sensitive to model error introduced by different cumulus parameterizations. Sensitivity experiments herein used ensembles with either single or multiple imperfect cumulus parameterizations with and without model error from PBL and microphysics. The results demonstrate that using a combination of different cumulus parameterization schemes can significantly improve the EnKF performance over experiments using a single inaccurate parameterization scheme. Our results suggest that the improvement comes from a smaller bias and from a better background error covariance structure developed from the multischeme ensemble. This is consistent with a recent real-data EnKF study of Fujita et al. (2005). Model uncertainties from PBL and microphysics processes also have significant impacts on the EnKF performance.

The EnKF performance depends strongly on the scales and dynamics of the flow of interest. Comparison of the EnKF performance in the two events with distinguishably different flow regimes exemplifies the impacts of flow-dependent predictability. It is found that the EnKF behaves consistently in corresponding experiments examining the two events, but the relative error reduction over the reference ensemble forecast is 10%–15% smaller in the warm-season event. The growth of reference ensemble forecast error is much slower in the MCV event than in the snowstorm case. Slower error growth and the relatively smaller scale of the MCV circulation may be responsible for a smaller error reduction and also for less bias when using different cumulus parameterization schemes in the ensemble forecast (Fig. 7). Impact of PBL and microphysics processes seems to be more significant for the warm-season case than for the winter case.

As a pretest for assimilating real data, this study is aimed at examining the impact of model error on an ensemble-based mesoscale data assimilation system. Apart from the errors explored here, there are other sources of uncertainty such as those from ensemble initialization, truncation error, lateral boundary, and surface processes. In real data assimilation, model error could potentially be more detrimental than considered in this study. We not only need to understand the impact of various model errors on the EnKF, but we also need to design effective ways to treat them such as with parameterization of model error (e.g., Hamill and Whitaker 2005) and simultaneous estimation of parametric model error (e.g., Aksoy et al. 2006a,b).

REFERENCES

Aksoy, A., F. Zhang, J. W. Nielsen-Gammon, and C. C. Epifanio, 2005: Data assimilation with the ensemble Kalman filter for thermally forced circulations. *J. Geophys. Res.,* **110,** D16105, doi:10.1029/2004JD005718.

——, ——, and ——, 2006a: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Mon. Wea. Rev.,* **134,** 2951–2970.

——, ——, and ——, 2006b: Ensemble-based simultaneous state and parameter estimation with MM5. *Geophys. Res. Lett.,* **33,** L12801, doi:10.1029/2006GL026186.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.,* **129,** 2884–2903.

Arakawa, A., 2004: The cumulus parameterization problem: Past, present and future. *J. Climate,* **17,** 2493–2525.

Barker, D. M., W. Huang, Y.-R. Guo, A. J. Bourgeois, and Q. N. Xiao, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.,* **132,** 897–914.

Caya, A., J. Sun, and C. Snyder, 2005: A comparison between the

4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.,* **133,** 3081–3094.

Cotton, W. R., and R. A. Anthes, 1989: *Storm and Cloud Dynamics.* Academic Press, 883 pp.

Daley, R., and R. Menard, 1993: Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Wea. Rev.,* **121,** 1554–1565.

Dirren, S., R. D. Torn, and G. J. Hakim, 2007: A data assimilation case study using a limited-area ensemble Kalman filter. *Mon. Wea. Rev.,* **135,** 1455–1473.

Dowell, D., 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.

Evans, R. E., M. S. J. Harrison, R. J. Graham, and K. R. Mylne, 2000: Joint medium-range ensembles from the Met Office and ECMWF systems. *Mon. Wea. Rev.,* **128,** 3104–3127.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99,** 10 143–10 162.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.,* **53,** 343–367.

Fujita, T., D. J. Stensrud, and D. C. Dowell, 2005: Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. Preprints, *11th Conf. on Mesoscale Processes,* Alberqueque, NM, Amer. Meteor. Soc., CD-ROM, 1M.3.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.,* **125,** 723–757.

Grell, G. A., and D. Devenyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.,* **29,** 1693, doi:10.1029/2002GL015311.

——, J. Dudhia, and D. R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR Tech. Note NCAR/TN-398+STR, 117 pp.

Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation. *Predictability of Weather and Climate,* T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 124–156.

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

——, and J. S. Whitaker, 2005: Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Mon. Wea. Rev.,* **133,** 3132–3147.

Hansen, J. A., 2002: Accounting for model error in ensemble-based state estimation and forecasting. *Mon. Wea. Rev.,* **130,** 2373–2391.

Hawblitzel, D. P., F. Zhang, Z. Meng, and C. A. Davis, 2007: Probabilistic evaluation of the dynamics and predictability of the meoscale convective vortex of 10–13 June 2003. *Mon. Wea. Rev.,* **135,** 1544–1563.

Houtekamer, P. L., and L. Lefaivre, 1997: Using ensemble forecasts for model validation. *Mon. Wea. Rev.,* **125,** 2416–2426.

——, and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.,* **126,** 796–811.

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

——, H. L. Mitchell, G. Pellerin, M. Buehner, 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, 2003: Assimilation of temperature into an isopycnal ocean general circulation model using a parallel Ensemble Kalman Filter. *J. Mar. Syst.,* **40–41,** 363–380.

Krishnamurti, T. N., C. M. Krishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate,* **13,** 4196–4216.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—A comparison with 4Dvar. *Quart. J. Roy. Meteor. Soc.,* **129,** 3183–3203.

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.

Molinari, J., and T. Corsetti, 1985: Incorporation of cloud-scale and mesoscale downdrafts into a cumulus parameterization: Results of one- and three-dimensional integrations. *Mon. Wea. Rev.,* **113,** 485–501.

——, and M. Dudek, 1992: Parameterization of convective precipitation in mesoscale numerical models: A critical review. *Mon. Wea. Rev.,* **120,** 326–344.

Orrell, D., 2002: Role of the metric in forecast error growth: How chaotic is the weather? *Tellus,* **54A,** 350–362.

——, L. A. Smith, J. Barkmeijer, and T. N. Palmer, 2001: Model error in weather forecasting. *Nonlinear Proc. Geophys.,* **8,** 357–371.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parameterization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.,* **127,** 279–304.

Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.,* **128,** 647–677.

Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.,* **131,** 1663–1677.

Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.,* **128,** 2077–2107.

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.

Wang, W., and N. L. Seaman, 1997: A comparison study of convective parameterization schemes in a mesoscale model. *Mon. Wea. Rev.,* **125,** 252–278.

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 co-

variance 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 meso-scale predictability. *J. Atmos. Sci.,* **60,** 1173–1185.

——, ——, and J. Sun, 2004: Impacts of initial estimate and obser-vation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.,* **132,** 1238–1253.

——, Z. Meng, and A. Aksoy, 2006a: Tests of an ensemble Kal-man filter for mesoscale and regional-scale data assimilation. Part I: Perfect model experiments. *Mon. Wea. Rev.,* **134,** 722–736.

——, A. Odins, and J. W. Nielsen-Gammon, 2006b: Mesoscale predictability of an extreme warm-season precipitation event. *Wea. Forecasting,* **21,** 149–166.

Zhang, S., and J. L. Anderson, 2003: Impact of spatially and tem-porally varying estimates of error covariance on assimilation in a simple atmospheric model. *Tellus,* **55A,** 126–147.