Development of a Mesoscale Ensemble Data Assimilation System at the Naval Research Laboratory

QINGYUN ZHAO

Marine Meteorology Division, Naval Research Laboratory, Monterey, California

FUQING ZHANG

The Pennsylvania State University, University Park, Pennsylvania

TEDDY HOLT AND CRAIG H. BISHOP

Marine Meteorology Division, Naval Research Laboratory, Monterey, California

QIN XU

National Severe Storms Laboratory, Norman, Oklahoma

(Manuscript received 18 January 2013, in final form 15 July 2013)

ABSTRACT

An ensemble Kalman filter (EnKF) has been adopted and implemented at the Naval Research Laboratory (NRL) for mesoscale and storm-scale data assimilation to study the impact of ensemble assimilation of high-resolution observations, including those from Doppler radars, on storm prediction. The system has been improved during its implementation at NRL to further enhance its capability of assimilating various types of meteorological data. A parallel algorithm was also developed to increase the system's computational efficiency on multiprocessor computers. The EnKF has been integrated into the NRL mesoscale data assimilation system and extensively tested to ensure that the system works appropriately with new observational data stream and forecast systems. An innovative procedure was developed to evaluate the impact of assimilated observations on ensemble analyses with no need to exclude any observations for independent validation (as required by the conventional evaluation based on data-denying experiments). The procedure was employed in this study to examine the impacts of ensemble size and localization on data assimilation and the results reveal a very interesting relationship between the ensemble size and the localization length scale. All the tests conducted in this study demonstrate the capabilities of the EnKF as a research tool for mesoscale and storm-scale data assimilation with potential operational applications.

1. Introduction

An ensemble Kalman filter (EnKF; Evensen 1994) was recently developed at the Naval Research Laboratory (NRL) for use as an advanced high-resolution data assimilation system for the U.S. Navy's Coupled Ocean–Atmosphere Mesoscale Prediction System (COAMPS;[1] Hodur 1997). The objectives of the EnKF development at NRL are twofold: (i) to investigate the impact of flow-dependent background error covariance on mesoscale and storm-scale data assimilation, especially when applied to high-resolution nonconventional sensor data, in the presence of rapid and/or complex changes in storm structure and intensity, and (ii) to provide a useful tool for research and system development for assimilating meteorological observations from nonconventional sensors such as Doppler radars into numerical weather prediction (NWP) models to improve forecasts of high-impact weather events over oceans and in remote areas. The parameters observed by these nonconventional sensors are often nonlinear functions of the state variables. In dynamically and statistically estimating the

[1] COAMPS is a registered trademark of the Naval Research Laboratory.

Corresponding author address: Dr. Qingyun Zhao, Naval Research Laboratory, 7 Grace Hopper Ave., Mail Stop II, Monterey, CA 93943.
E-mail: allen.zhao@nrlmry.navy.mil

DOI: 10.1175/WAF-D-13-00015.1

error correlations between such observed parameters and the model state variables, the EnKF provides a relatively straightforward and simple procedure with which to estimate these parameters (Aksoy et al. 2009; Tong and Xue 2005, 2008).

The COAMPS-based ensemble square root filter (EnSRF; Whitaker and Hamill 2002) developed and tested in this study has its origins in a research version of the EnKF (Zhang et al. 2006, 2009; Meng and Zhang 2007, 2008a,b) originally developed for the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (Penn State–NCAR) Mesoscale Model (MM5; Grell et al. 1994) and the Weather Research and Forecasting (WRF) model (Skamarock et al. 2005). This predecessor regional-scale EnKF had been extensively tested with both simulated and real-data observations that exhibited skill over a three-dimensional variational data assimilation (3DVAR) system for the WRF model (Zhang et al. 2006, 2009; Meng and Zhang 2007, 2008a,b).

Several major changes and improvements have been made to the EnKF to make it more suitable for COAMPS applied both in an operational environment and for scientific research at NRL. First, interfaces have been developed to connect the EnKF to COAMPS ensemble forecasts. The EnKF was enhanced to have the same capabilities as COAMPS in its grid nesting, dynamical grid configuration, and global relocatability. Second, Navy Operational Global Atmospheric Prediction System (NOGAPS; Hogan and Rosmond 1991) ensemble forecasts are used as boundary conditions for COAMPS ensemble forecasts and first guesses for cold starts of the EnKF ensemble analyses. This setup avoids the need to randomly generate the initial and lateral boundary perturbations for COAMPS ensemble members and, most importantly, enables the global ensemble to transmit flow-dependent boundary condition uncertainty information to the regional ensemble. This approach is useful for operational centers where global ensemble forecasts are available. Details of the ensemble transform (ET) method used to create the global NOGAPS ensemble are given in McLay et al. (2008, 2010). For the experiments reported in this paper a special 128-member NOGAPS ET ensemble was run. Third, the EnKF shares the same observational data, data processing and quality control algorithms, and observational forward operators with the NRL Atmospheric Variational Data Assimilation System (NAVDAS) 3D- and 4DVAR systems, except for radar observations. This setup makes it easy to compare the ensemble-based data assimilation algorithm (when radar observations are not assimilated) with variational approaches. More importantly, this provides a stepping stone to a future hybrid system that couples the ensemble-based and variational data assimilation methods at NRL. The most significant change to the EnKF is in the development of a parallelization algorithm to convert the EnKF from a single-processor version to a parallelized version using the message-passing interface (MPI) technique to improve the system's computational efficiency on a massive central processing unit (CPU) computer. The most important change to the EnKF is the addition of NRL radar data processing, quality control, and assimilation algorithms to further enhance the system with the capability of assimilating high-resolution Doppler radar observations.

This paper focuses on COAMPS EnKF development and testing at NRL. In this paper, we will present and discuss some interesting results from our experiments that show how the system performs with different ensemble sizes and localizations. We will also compare the COAMPS EnKF with the current operational NAVDAS 3DVAR system to better understand the advantages and issues that the current EnKF data assimilation system has. The overall goal of this study is to ensure that the COAMPS EnKF is a mature system for mesoscale and storm-scale data assimilation study. Section 2 gives a general description of the current EnKF with a focus on the improvements to the EnKF after the system was adopted at NRL. Details of the independent validation procedure to examine the impact of EnKF data assimilation are also given in this section. In section 3, results from experiments are presented along with discussions about the ensemble size, covariance localization, ensemble spread, and their impacts on ensemble analyses. In section 4, evaluations of the COAMPS ensemble forecasts initialized by the EnKF are given. The impacts of ensemble size and localization length scale on ensemble forecasts are also discussed. Comparison between the EnKF and the NAVDAS 3DVAR for a period of about a week is examined in section 5 to demonstrate the capability and performance of the ensemble data assimilation system. Section 6 summarizes the study and discusses future work.

2. System description

The EnKF implemented for the COAMPS model and tested in this study is an EnSRF serial analysis algorithm (Whitaker and Hamill 2002; Tippett et al. 2003). The analysis equations for the EnSRF are given by

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\overline{\mathbf{x}^f}) \quad \text{and} \tag{1}$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f, \tag{2}$$

where

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \quad (3)$$

$\overline{\mathbf{x}^f}$ is the mean state vector from the ensemble forecasts with sample covariance $\mathbf{P}^f$, $\mathbf{y}^o$ is the observation vector, $\mathbf{R}$ the observational error covariance, and $\mathbf{H}$ the observation forward operator. The superscripts *a* and *f* denote analysis and forecast, respectively. The readers are referred to Whitaker and Hamill (2002) for the description and explanation of the EnSRF serial algorithm. The algorithms and procedures for ensemble localization and for updating the ensemble mean and perturbations in the EnKF remain unchanged during the implementation for COAMPS. Further refinements to the algorithm applications can be found in Snyder and Zhang (2003), Zhang et al. (2004, 2006, 2009), and Meng and Zhang (2008a,b).

Although the EnKF was developed primarily for radar data assimilation at NRL, it has the capability to assimilate all conventional observations (including those from rawinsondes and pilot balloons, land surface stations, ships and buoys, and commercial aircraft) and satellite products (most satellite winds and derived temperature profiles) that the operational NAVDAS 3DVAR assimilates into COAMPS. In this study, we will test the EnKF with conventional data and satellite products, instead of using radar observations, based on two considerations: (i) the conventional and satellite data are appropriately quality controlled and well-studied and (ii) this makes it possible to directly compare the results from the EnKF with those from the NAVDAS 3DVAR. The EnKF has the capability to update all model state variables. In all the data assimilation experiments we conducted in this study, however, the EnKF only updates the model fields of potential temperature (θ), horizontal winds (u, v), and water vapor mixing ratio (q_v). Vertical wind speed (w) is not updated because the observations we assimilated in this study cannot observe storm structures while w is basically storm related. The pressure field in a nonhydrostatic model is partially related to upward motion. We found in our current storm-scale radar data assimilation study that better results are achieved by updating the pressure and w fields concurrently than when just updating the pressure alone (the results will be published separately).

a. Development of a parallel technique for the EnKF

When an EnSRF serial analysis algorithm is applied to real data assimilation problems with a reasonable number of observations over the analysis domain of a reasonable size (such as the contiguous United States or CONUS, for example), the computational efficiency of the EnKF becomes a concern, especially for operational applications. To address this issue, a parallel algorithm has been developed for the EnKF at NRL to increase the computational efficiency by using the MPI technique running on a computer with multiple CPUs.

The parallel algorithm consists of two parts: the input/output (I/O) part and the analysis part. The I/O part takes care of the input and output of the ensemble forecasts and analyses from and to each member of the COAMPS ensemble. This part was efficiently designed so that the algorithm makes assignment of the ensemble members to each of the selected CPUs as even as possible. For example, when the EnKF reads in COAMPS ensemble forecasts from 16 members with 4 CPUs, each CPU is assigned with 4 COAMPS members. When 6 CPUs are selected, however, the first 2 CPUs are assigned with 2 COAMPS members each while the last 4 CPUs are assigned with 3 members each. If 32 CPUs are selected, the last 16 CPUs have to wait until the first 16 CPUs finish reading in the COAMPS forecasts. This algorithm is also applied to the calculation of the prior ensemble perturbations at observation locations from the forecasts of each ensemble member during the assimilation.

The major challenge in parallelizing the system is in the analysis. In a serial analysis algorithm, the ensemble state vector is updated immediately after the assimilation of the current observation and then is used to compute the prior ensemble perturbations for the next observation (see Fig. 1 in Anderson and Collins 2007). This makes it very inconvenient and difficult to parallelize the algorithm (by assigning different groups of observations to different CPUs that run in parallel), especially if we want to guarantee that the analyses from the same ensemble forecasts and observations but using different numbers of CPUs are identical. While an innovative algorithm for the EnSRF serial analysis algorithm that parallelizes the observations while requiring that the analyses are completely independent of the number of CPUs used for the data assimilation is still under development at NRL, the current parallel algorithm for the analysis part in this study is relatively straightforward (we can call this an intermediate step of parallelizing the serial analysis algorithm). In the EnSRF serial analysis algorithm, the analysis produced by each observation is masked by a monotonic function decaying with distance of separation localized within an ellipsoid contained within a three-dimensional (3D) cube. The parallel algorithm developed in this paper divides the 3D cube, as evenly as possible (in the same way as in the I/O part), into 3D small boxes of (NX $\times$ NY $\times$ NZ), where NX, NY, and NZ are the numbers of CPUs selected in the x, y, and z directions, respectively. Figure 1 gives an example of the divided 3D cube with NX = 3, NY = 3, and NZ = 3, with the total number of

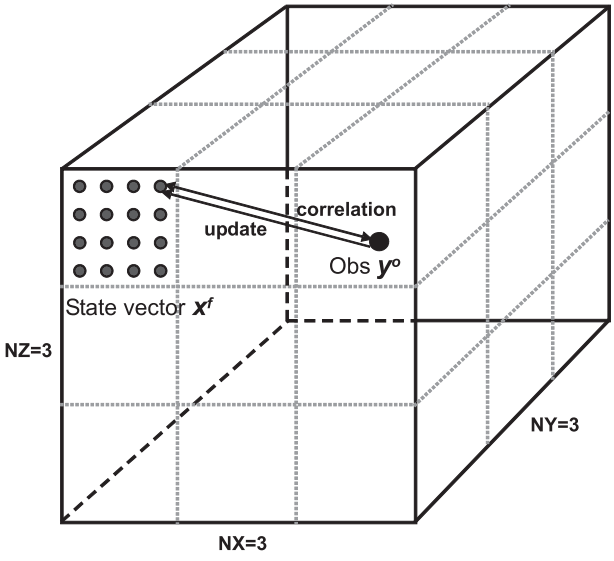

FIG. 1. Illustration of the EnKF analysis parallelization. The three-dimensional cube that contains the EnKF localization ellipsoid for one observation (located at the center of the cube) is divided into NX × NY × NZ subdomains. Each subdomain is for one CPU.

selected CPUs being 27. In this case, the 3D cube is divided into 27 small boxes. Each of the 27 CPUs calculates the correlations between the prior ensemble approximation at the current observation location and the ensemble values at each of the grid points within the small box and then updates the ensemble mean and perturbations at these grid points based on the correlations and the innovation at the observation location. These CPUs do the calculations separately and simultaneously and, then, communicate to each other once the assimilation of the observation is done. Next, they prepare for next observation assimilation, and so on.

Preliminary tests show that the current parallel algorithm works well with a reasonable ensemble size and CPU number. Analyses from experiments with the same observations and same ensemble size but different CPU numbers yield exactly the same results. Table 1 gives an example of wall-clock times (normalized by setting the wall-clock time for one CPU to 1.0) from a test with an ensemble size of 32 and the numbers of CPUs being 1, 16, and 32, respectively. Results show that the algorithm is reasonably scalable. The real wall-clock time for the EnKF to finish one analysis depends on individual computer speed, the number of observations, the length scale of localization, the model domain size and grid resolution, and the number of CPUs used. Our recent results from experiments show that for an analysis domain that covers the CONUS and parts of the Atlantic and Pacific Oceans with three nested grids with grid resolutions of

TABLE 1. An example of normalized wall-clock time (by setting the wall-clock time for one CPU to 1.0) as a function of CPU number from the EnKF analysis parallelization algorithm. The ensemble size is 32.

No. of CPUs	Normalized wall-clock time
1	1.0
16	0.1638
32	0.1034

45, 15, and 5 km and grid numbers of 160 × 160, 160 × 192, and 192 × 192, respectively, the EnKF assimilates about 5×10^4 observations within about 1 h of wall-clock time with 32 CPUs for an ensemble size of 32.

The main advantage of the EnKF parallel algorithm is its complete independence of the parallel setup of the NWP model. Thus, the EnKF can be run separately from the model, at a different time, or even on a different machine. Another major advantage of the algorithm is the independence of the analyses on the number of CPUs used in the data assimilation. The algorithm, however, also has some weaknesses. One of the weaknesses is the use of the serial algorithm. In cases with the observation number reaching 10^6–10^8, the wall-clock time required for the EnKF could become a concern for most research and operational applications. We also found that when the ensemble size and the number of CPUs are greater than 100, the scaling of the algorithm degrades because of the heavy overhead charge caused by the intercommunications among the processors.

b. A procedure for data assimilation impact evaluation

Usually it is not easy or may even be impossible to directly evaluate the impact of data assimilation on the final analysis (without making the subsequent forecasts and evaluating them with future observations), especially if all available observations are assimilated and thus no observation can be used for an independent verification. In a serial algorithm, however, observations are assimilated one after another. The model fields are updated immediately after the assimilation of one observation and then the updated model fields are used as a first guess for the next observation assimilation. This procedure makes it possible to use any of these observations (except the first one) to independently evaluate the impact of those previously assimilated observations whose influence areas cover the current observation right before the current observation is assimilated. Below is a procedure developed at NRL for the EnKF to study the data assimilation impact.

Let d_i^b denote the innovation computed from the ith observation y_i by subtracting its associated original

background value (not updated by any observations for the current assimilation cycle):

$$d_i^b = [\, y_i - \overline{\mathbf{H}_i(\mathbf{X}_b)}\,], \tag{4}$$

where $\mathbf{H}_i$ is the forward operator for the ith observation and $\mathbf{X}_b$ is the state vector from the ensemble forecasts without assimilating any observations for the current assimilation cycle. The overbar indicates the ensemble mean. Similarly, the innovation computed from the same ith observation by subtracting its associated updated background value (after the previous i–1 observations, $y_1, y_2, \ldots, y_{i-1}$, are assimilated) can be expressed as

$$d_i^a = [\, y_i - \overline{\mathbf{H}_i(\mathbf{X}_a)}\,], \tag{5}$$

where $\mathbf{X}_a$ is the state vector after the previous $i - 1$ observations, $y_1, y_2, \ldots, y_{i-1}$, are assimilated. Then, the innovation reduction evaluated at the ith observation due to the assimilation of the observations $(y_1, y_2, \ldots, y_{i-1})$ can be defined as

$$D_i = |d_i^a| - |d_i^b|. \tag{6}$$

By definition, a negative value of D_i means a positive impact of the previously assimilated data. This verification procedure is henceforth referred as the innovation reduction procedure.

3. Testing of the EnKF system

The predecessor of the current EnKF implemented in the MM5 and WRF models has been tested extensively with both simulated and real-data observations with promising results (Zhang et al. 2006, 2009; Meng and Zhang 2007, 2008a,b). The current study presents some exemplary results from extensive testing of the EnKF implemented in COAMPS. Such testing and reevaluation are necessary to ensure that the filter has been integrated properly with a new forecast model given changes in the model state variables, observational data feed, observation forward operators, and the source for the initial and boundary perturbations for the ensemble.

A procedure has been designed to test the EnKF, in which the NOGAPS ensemble is used at the very beginning as the initial conditions to cold start the COAMPS ensemble (no spinup is needed since the NOGAPS ensemble has been cycled for a period of several days). The 12-h COAMPS ensemble forecast is then used as the background for the EnKF data assimilation. After that, COAMPS ensemble forecast is warm started using the EnKF ensemble analyses as initial conditions. The cycled EnKF data assimilation–COAMPS ensemble forecast continues until the end of the test period. During the test, NOGAPS ensemble forecasts are also used as boundary conditions for the COAMPS ensemble. All the results shown hereafter are from warm starts.

One of the major advantages of EnKF is its capability of dynamically estimating the background error covariance from ensemble forecasts. To ensure that the adopted EnKF has the capability to appropriately estimate the background error covariance and also to study the impact of ensemble size on the covariance estimation, experiments were performed in which correlations of the perturbation state variables between the observation location and the model grid points from forecasts of different ensemble sizes are examined. Figure 2 gives examples of perturbation temperature autocorrelations at the $Z = 5800$ m model level with respect to a temperature observation point at 500 hPa at Salt Lake City, Utah (near the center of the domain in Fig. 2), computed with the ensemble sizes set to 16, 32, 64, and 128, respectively. The model horizontal grid spacing in these examples is 45 km. There are two important features in Fig. 2. First, the correlations computed from different ensemble sizes look similar (Figs. 2b–d) except for the one from the ensemble with 16 members, which has apparent spuriously high ensemble correlations near the Four Corners area of the U.S. southwest (Fig. 2a). An ensemble size of 16 is apparently statistically insufficient for accurate estimation for the background error covariance. It is also interesting to note that as the ensemble size increases, the correlation becomes smoother and the maximum value of the correlation decreases slightly. Second, the correlations appear to have some dependence on the horizontal winds at the same vertical level. The possible dynamics underlying such a flow dependence is still yet to be explored.

Cross correlations are also examined. Examples of cross correlations of temperature perturbation T' (at the same observation point as that in Fig. 2) to horizontal wind perturbations (u', v') and water vapor mixing ratio perturbation q_v' (at the same model level as shown in Fig. 2) are given in Fig. 3 for 64-member ensembles. The cross correlation between T' and q_v' shows some scattered patterns in Fig. 3c that could plausibly reflect the thermodynamic relationship between these two perturbation variables over that localized area. Note that the general patterns of the cross correlations between T' and (u', v') in Figs. 3a,b (at 5800 m, which is above the localized T') show some basic geostrophic relationship between these two fields. This is an indication, as expected, that the estimated cross correlations are constrained by the approximate geostrophy and its associated thermal–wind relationship over the synoptic and large scales (on the 45-km grid).

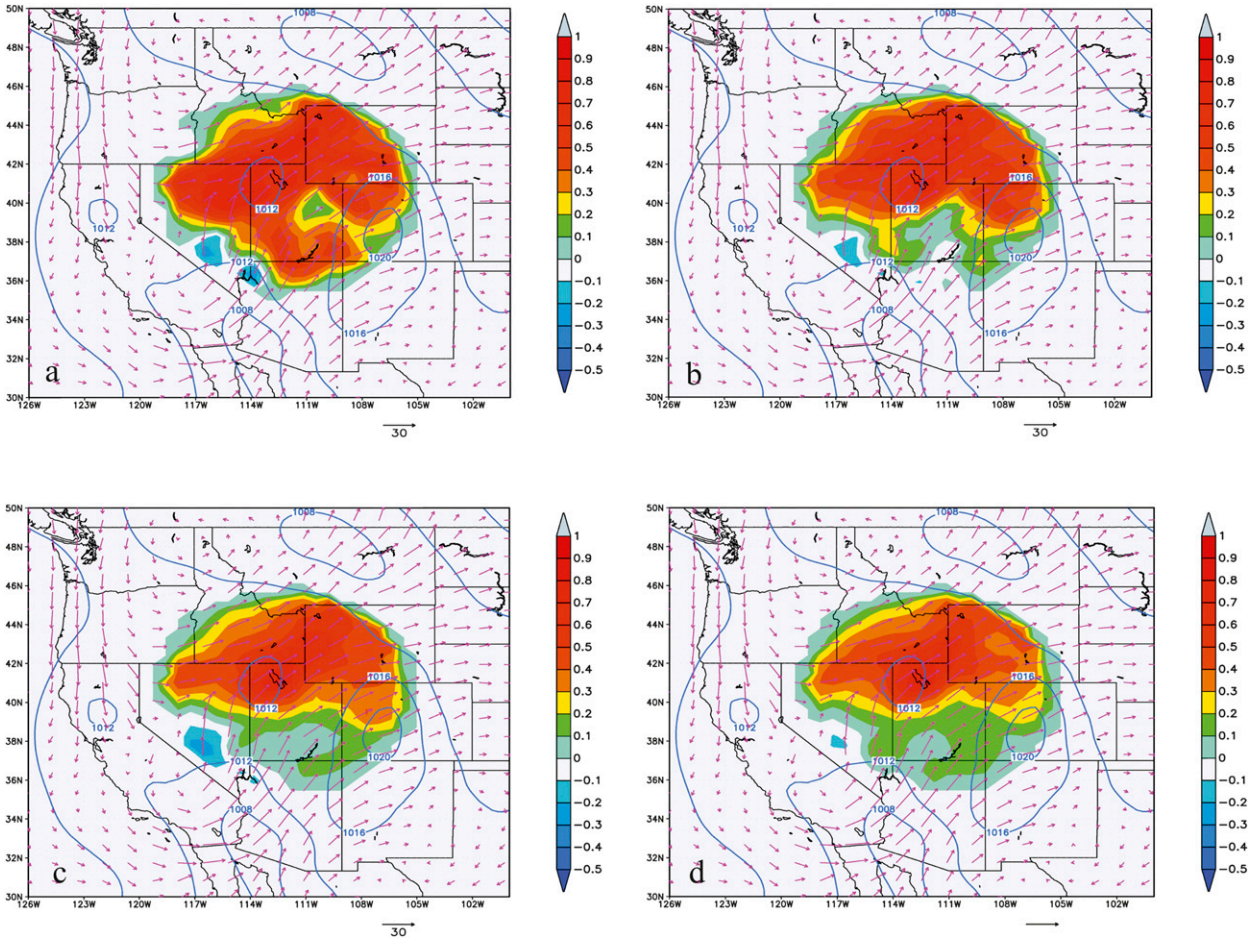

FIG. 2. Perturbation temperature autocorrelations (colored areas) at $Z = 5800$ m with respect to a temperature observation point at 500 hPa at Salt Lake City and the ensemble mean of horizontal wind analyses (arrows) at the same level. The temperature observation was taken at 1200 UTC 28 Jun 2005. The contours are the ensemble mean of sea level pressure (hPa). The ensemble sizes are (a) 16, (b) 32, (c) 64, and (d) 128.

The impact of ensemble size on ensemble analyses has also been studied. The innovation reduction procedure described in section 2b is used here to examine how effectively the EnKF updates the ensemble analysis mean with different ensemble sizes. For this purpose, a set of about 30 900 real observational data points inside the analysis domain from rawinsondes, pilot balloons, land surface stations, ships and buoys, commercial aircrafts, and satellite products (mostly satellite winds and derived temperature profiles) were selected to test the system.

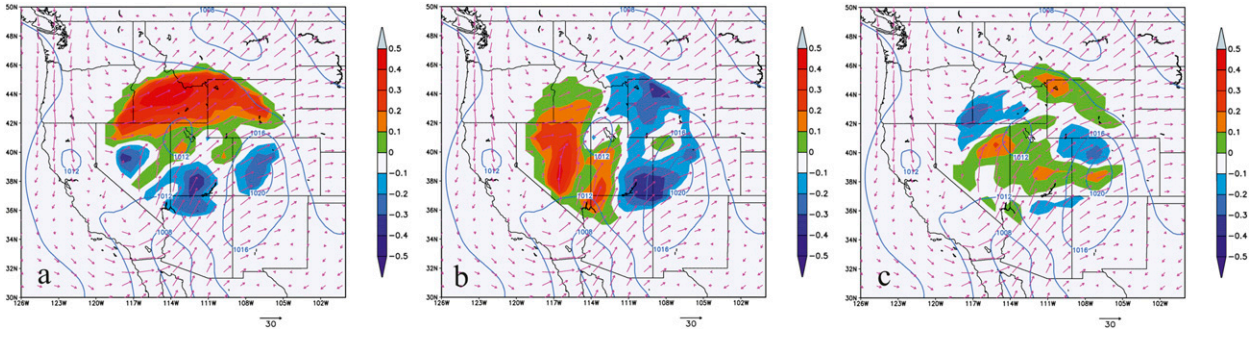

FIG. 3. Cross correlations (colored areas) of T' (at the same temperature observation point as that in Fig. 2) to (a) u', (b) v', and (c) q'_v (at the same model level as shown in Fig. 2) for a 64-member ensemble. The arrows and contours are the same wind and sea level pressure analyses as in Fig. 2.

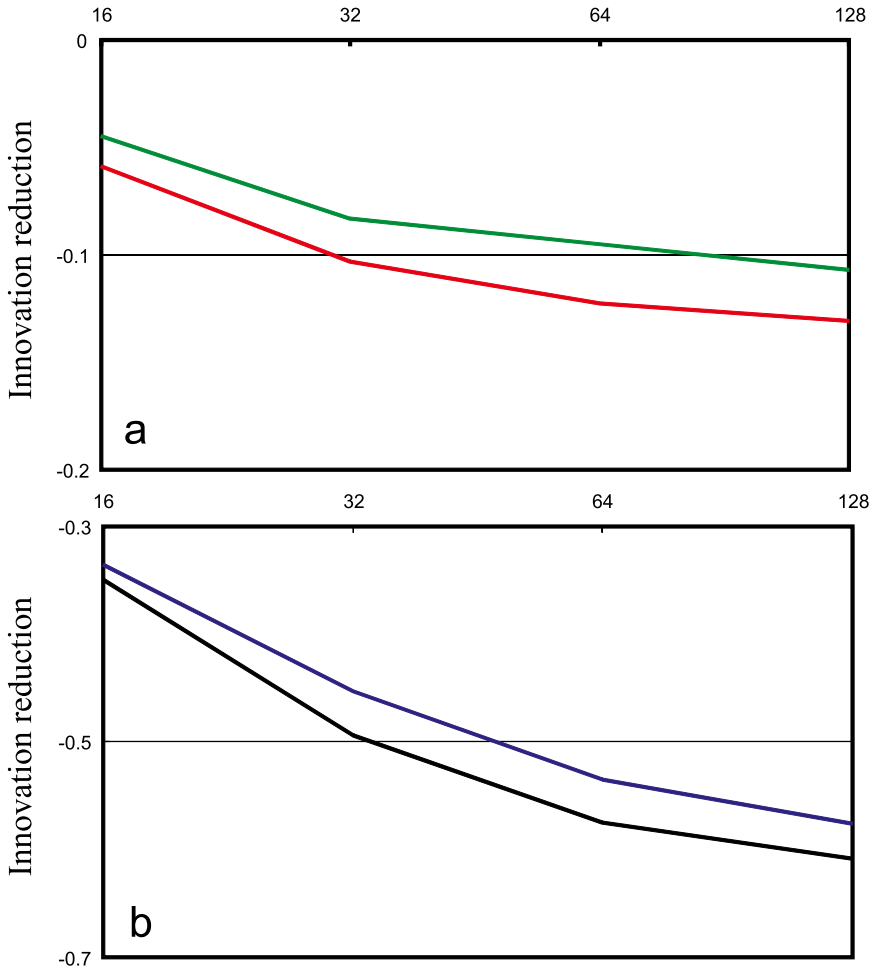

FIG. 4. Domain-averaged innovation reductions plotted as functions of ensemble size for (a) T (°) and q_v (green curve, g kg^{-1}), and (b) u (blue curve, m s^{-1}) and v (black curve, m s^{-1}). A horizontal localization length scale of 675 km is used in these experiments.

A horizontal localization length scale of 675 km was used for the test. Equations (5) and (6) are used to compute the innovation reduction at each observation location right before the observation at this location is assimilated. Then, the domain-averaged innovation reductions for the same observation parameters are calculated using

$$\overline{D} = \frac{1}{M}\sum_{i=1}^{M} D_i, \tag{7}$$

where M is the number of observations of the same observation parameter. Four experiments were conducted with ensemble sizes of 16, 32, 64, and 128, respectively. Figure 4 gives the domain-averaged innovation reductions for each of the four observation parameters (T, u, v, and q_v) as a function of ensemble size. It is evident in Fig. 4 that the ensemble size affects the data assimilation impact. For an ensemble size of 16, the data impact on the model fields is the smallest. This could be due to the undersampling issue caused by the insufficient ensemble number. The largest increase in data assimilation impact occurs when the ensemble size increases from 16 to 32. When the ensemble size increases from 32 to 64, the increase in data assimilation impact is notable but smaller than that between 16 and 32. The increase in the data assimilation impact from ensemble 64 to ensemble 128 is the smallest. It is not clear though from Fig. 4 whether there is a saturation point, after which increasing ensemble size provides no help in improving data assimilation impact. Unfortunately, we cannot run ensembles of more than 128 members due to the lack of a NOGAPS ensemble for COAMPS ensemble boundary and initial conditions.

There is no doubt that the shape of the curves in Fig. 4 is affected by the order of observations in the serial algorithm simply because, as mentioned earlier,

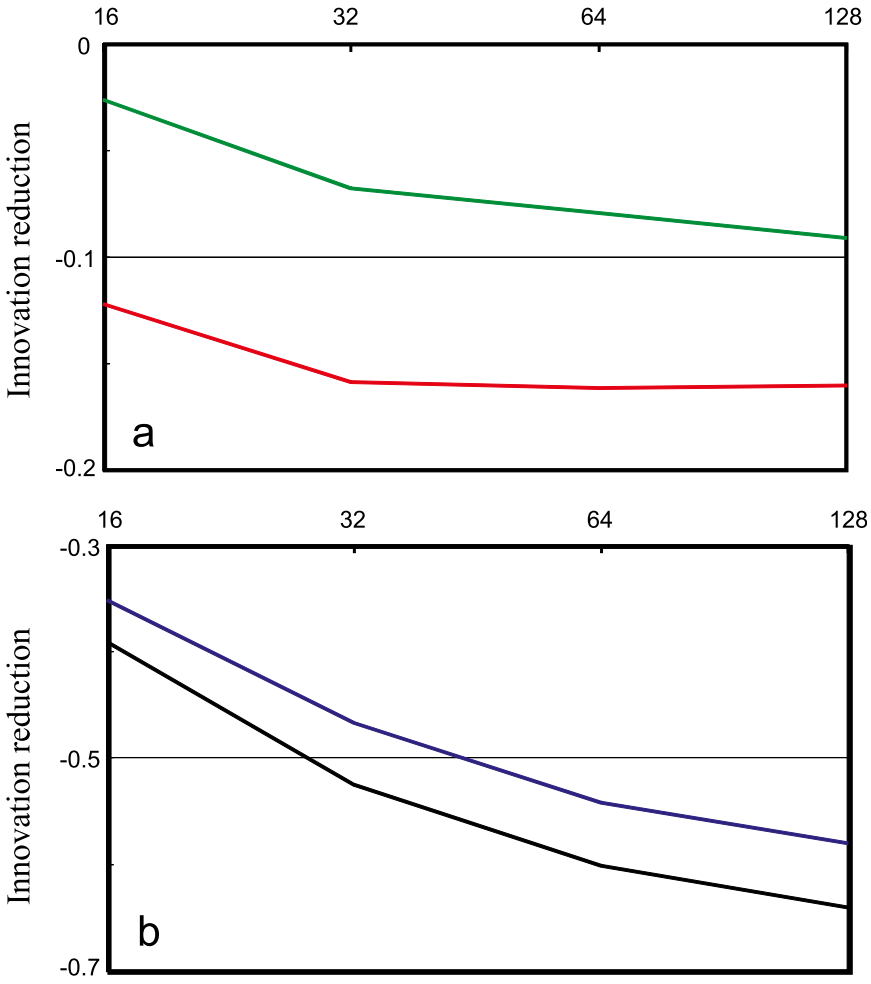

FIG. 5. As in Fig. 4, except that the order of observations was reversed in the serial algorithm.

the innovation reduction at the *i*th observation is a result of the assimilations of previous observations whose areas of influence cover the *i*th observation. The qualitative aspect of the relationship between the ensemble size and the data assimilation impact, however, should not be significantly impacted by the data assimilation sequence. To verify this, experiments have been carried out in which the observations and ensemble sizes are exactly the same as those in Fig. 4 but the order of the observations used in the serial algorithm is reversed. The last observation is assimilated first and the first observation is assimilated last. The resulting domain-averaged innovation reductions are presented in Fig. 5. Apparently, the relationship between the ensemble size and the data assimilation impact is very similar to that in Fig. 4. Although this is only one case, it suggests that the innovation reduction procedure developed here can be a useful measure for tuning the ensemble size and the horizontal localization length as well. The results shown in Figs. 4 and 5 can also be used to determine the minimum ensemble size for a given computational resource. In the above example, an ensemble size of 16 is probably too small.

Covariance localization has been used by many ensemble filters as a means to reduce the spurious correlation between an observation and model state variables due to sampling error when small ensembles are used. The Schur-product-based localization algorithm (Gaspari and Cohn 1999) is used by the EnKF, in which the localization length scale (LLS, the radius in the Schur-product-based localization) is empirically determined. Doubtlessly, LLS is critical in the estimation of background error covariance, and ensemble size is an important factor in determining LLS values for effective data assimilation. To quantify the relationship between LLS and ensemble size, domain-averaged innovation reductions are computed (for the same set of observations above) with LLS values of 225, 450, 675, 900, 1125, 1350,

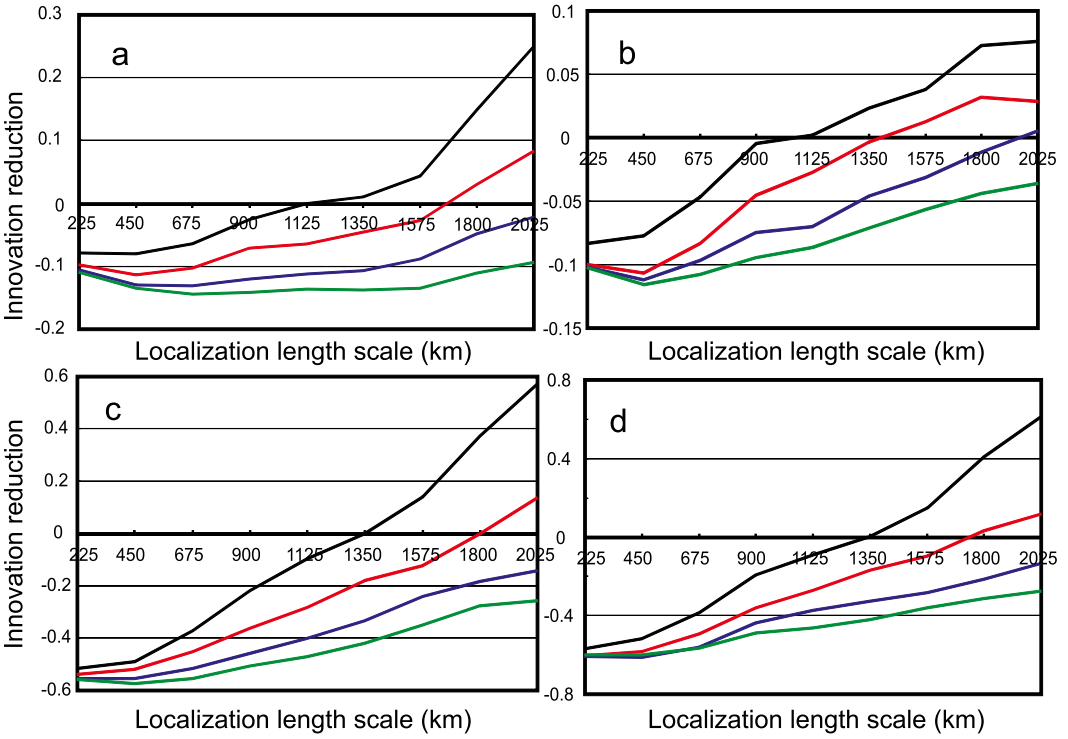

FIG. 6. Domain-averaged innovation reductions at observation locations of (a) T (°), (b) q_v (g kg^{-1}), (c) u (m s^{-1}), and (d) v (m s^{-1}), respectively, for different values of LLS and ensemble sizes. The black, red, blue, and green curves are for ensemble sizes of 16, 32, 64, and 128, respectively.

1575, 1800, and 2025 km, respectively, and again with ensemble sizes of 16, 32, 64, and 128. The results are given in Fig. 6. Several interesting issues can be identified from the results. First, as expected, LLS has a remarkable impact on innovation reductions, especially for small ensemble sizes. Small ensemble sizes favor small length scales in localization. This is because a large value of LLS cannot effectively eliminate spurious background covariance–correlations (see Fig. 2a) due to the insufficient ensemble members in the statistical calculations. The issue here is that the small values of LLS required by small ensemble sizes will limit the influences of observations on updating the model state variables and, as a result, will limit the effectiveness of the ensemble data assimilation system. But the good news from Fig. 6 is that with an appropriate selection of LLS, satisfactory data assimilation impacts can still be achieved by small ensemble sizes when computational resources are limited. Second, as the ensemble size increases, the analyses become less sensitive to LLS. This can also be seen in Fig. 2, where the background error correlations become smoother and less noisy as the ensemble size increases. This supports the theoretical expectation that the necessity and importance of localization should diminish as the ensemble size increases boundlessly. Finally, Fig. 6 suggests that wind and q_v observations favor slightly smaller LLSs than does the observations of T. This raises a question as to whether we should use different values of LLS for different observation parameters. More studies will be needed to address this question.

We should also mention that vertical localization is important, especially for observations inside storms where the vertical mixing is dynamic and very different from the large-scale environment. In this study, however, we used the same types of observations as those used by the current operational NAVDAS 3DVAR system. So we simply adopted the length scales of the vertical background error correlations, which have been well predefined and tested (Daley and Barker 2001), from the 3DVAR as the vertical localization length scales for the types of observations used in the test of the EnKF.

4. Evaluation of the ensemble forecasts

The previous section examined the major features of the EnKF and how the ensemble size and LLS impacted the effectiveness of the EnKF in assimilating observations into the analyses. The other question that remains

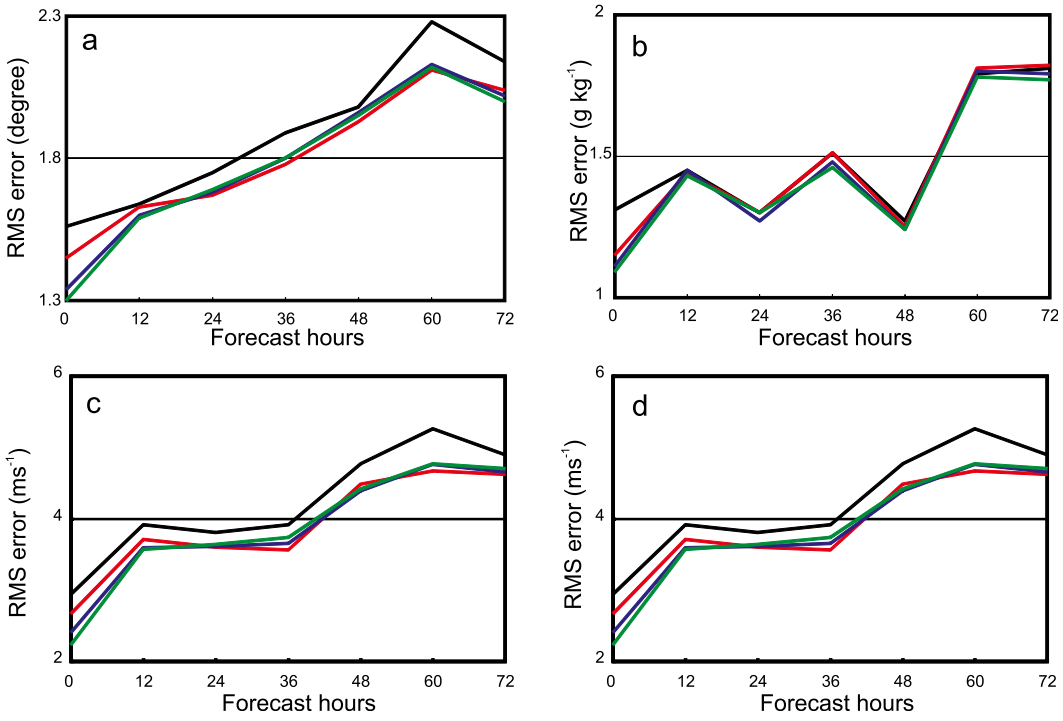

FIG. 7. RMSEs of (a) T (°), (b) q_v (g kg^{-1}), (c) u (m s^{-1}), and (d) v (m s^{-1}) of the ensemble forecast mean verified against radiosonde observations. The black, red, blue, and green curves are for ensemble sizes of 16, 32, 64, and 128, respectively. LLS is equal to 675 km.

is how these factors affect the ensemble forecasts initialized by the EnKF analyses. To answer this question, COAMPS ensemble forecasts are launched using the ensemble analyses generated from the experiments in the previous section.

The first issue to address is the impact of ensemble size on the performance of COAMPS ensemble forecasts. The case of a storm associated with a frontal system along the east coast of the United States stretching from New England all the way down to the Gulf of Mexico on 28 June 2005 is selected for this study since the large-scale storm system lasted basically for the whole forecast period of 72 h. Four ensemble forecasts are launched at 1200 UTC 28 June 2005 using the ensemble analyses generated from the experiments in the previous section with ensemble sizes of 16, 32, 64, and 128, respectively. For this study, only one nested COAMPS grid is used, with a horizontal grid resolution of 45 km and 30 vertical levels. An LLS value of 675 km is used. These forecasts are verified against radiosonde observations and their root-mean-square (RMS) differences are computed as a function of forecast hours. The computed RMS difference for each variable estimates the square root of the sum of the true forecast and observation error variances at each forecast hour (except for 0 h, since the analyzed initial field is not independent of the observations). Since the observation error variance is estimated by a constant value for each variable, the time or case variation of the computed RMS difference reflects the time or case variation of the forecast RMS error for each variable (after 0 h). Figure 7 gives the results for model state variables of T, u, v, and q_v, respectively. As one can see, the performance of these ensemble forecasts (in terms of RMS errors) from 0 to 12 h basically follow the same order of the ensemble sizes beginning with the best results from the largest ensemble. This is not surprising since the ensemble analyses that are used for initializing these ensemble forecasts also show the same impact of ensemble size (in terms of innovation reductions; see Fig. 4). It is interesting, however, to note that the differences in the RMS errors among the ensemble forecasts of ensemble sizes of 32, 64, and 128 decrease quickly with forecast time and become barely notable at about 24 forecast hours. After that, all the ensemble forecasts basically have the same RMS errors except for the one with 16 ensemble members whose RMS errors remain larger than the others for essentially the entire forecast period. This suggests again that the ensemble size of 16 is too small (at least for this case) to reliably estimate the forecast error statistics. Increasing the ensemble size to 32 is still insufficient for the ensemble analyses, as discussed earlier, but it appears to

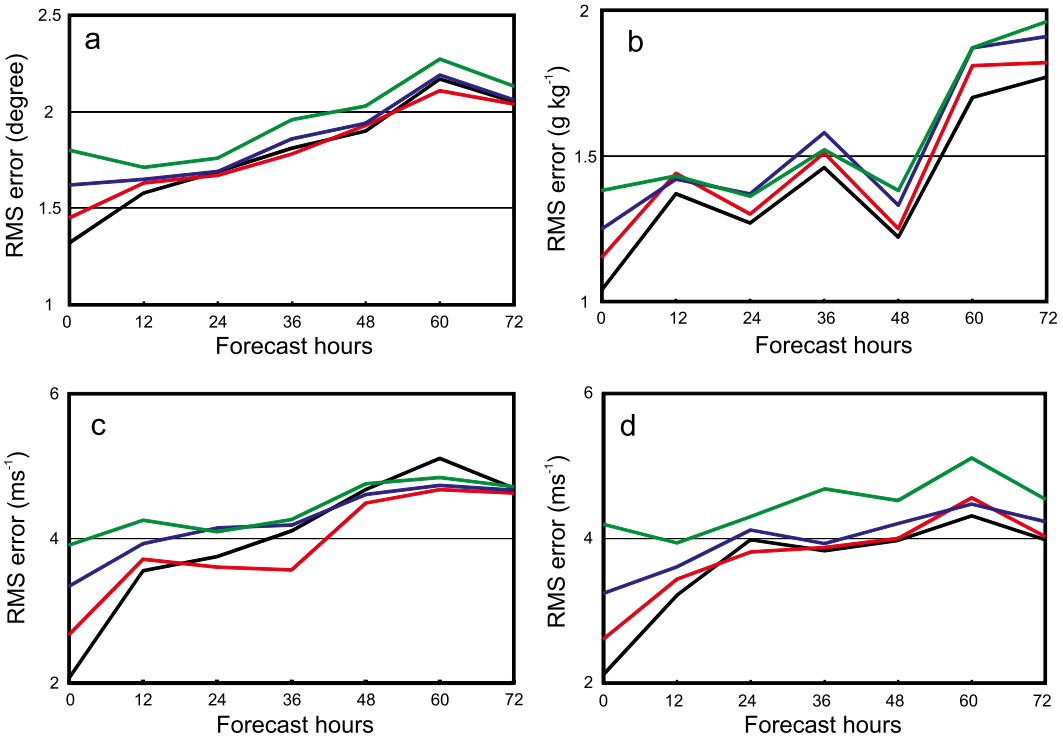

FIG. 8. As in Fig. 7, but for different LLSs in the EnKF data assimilation. The black, red, blue, and green curves are for LLS values of 225, 675, 1125, and 2025 km, respectively. Ensemble size is 32.

be significant and marginally adequate for reducing the forecast errors. The results in Fig. 7 also appear to indicate that fewer ensemble members may be actually needed for ensemble forecasts of longer than 24 h (in this case), although a large number of ensemble members is required for near-term forecasts (up to 12 h in this case) to produce the background ensemble for subsequent ensemble analyses. Please note that the above discussions are based on only one case study. While the results are encouraging, they could be limited to this particular case. More studies are still needed to further study the general impact of ensemble size on ensemble forecast.

The second issue is the impact of localization on model forecasts. The same storm case above is selected again for this investigation. As seen in Fig. 6, at ensemble size of 32, the EnKF is sensitive to LLS. But with an appropriate selection of the LLS value, the EnKF is able to produce innovation reductions comparable to those from large ensemble sizes. It is also seen in Fig. 7 that after about 24 h of model integration, the ensemble forecast of 32 members has about the same RMS error as those from the experiments with larger ensemble sizes. Therefore, the ensemble size of 32 is selected to study the impact of LLS on ensemble forecasts.

Figure 8 gives the RMS differences of the ensemble forecasts of T, u, v, and q_v with respect to radiosonde observations as functions of forecast hours. A couple of interesting features can be noted in Fig. 8. First, LLS affects the ensemble forecasts (with an ensemble size of 32) significantly. Therefore, the selection of the LLS value is important for ensemble forecasts with small ensemble sizes. Second, the differences among the EnKF analyses from different LLS values again decrease quickly with the forecast time during the first 12 h of model integration. After that, they still remain notable. Apparently for small ensemble sizes, weak localizations with LLS values of 1025 and 2025 km lead to large forecast errors basically throughout the forecast period. The strongest localization with LLS = 225 km produces the smallest forecast errors during about the first 12 h. But after that, it has mixed impacts on the ensemble forecasts compared to those from the experiment with LLS = 675 km. The latter overall is probably the best choice for this particular ensemble size.

5. Comparison to NAVDAS 3DVAR

NAVDAS 3DVAR (Daley and Barker 2001) has been the operational data assimilation system for COAMPS since October 2006. Currently, it supports the operational COAMPS On-Scene (OS), which is running at many locations around the world to support U.S. Navy operations,

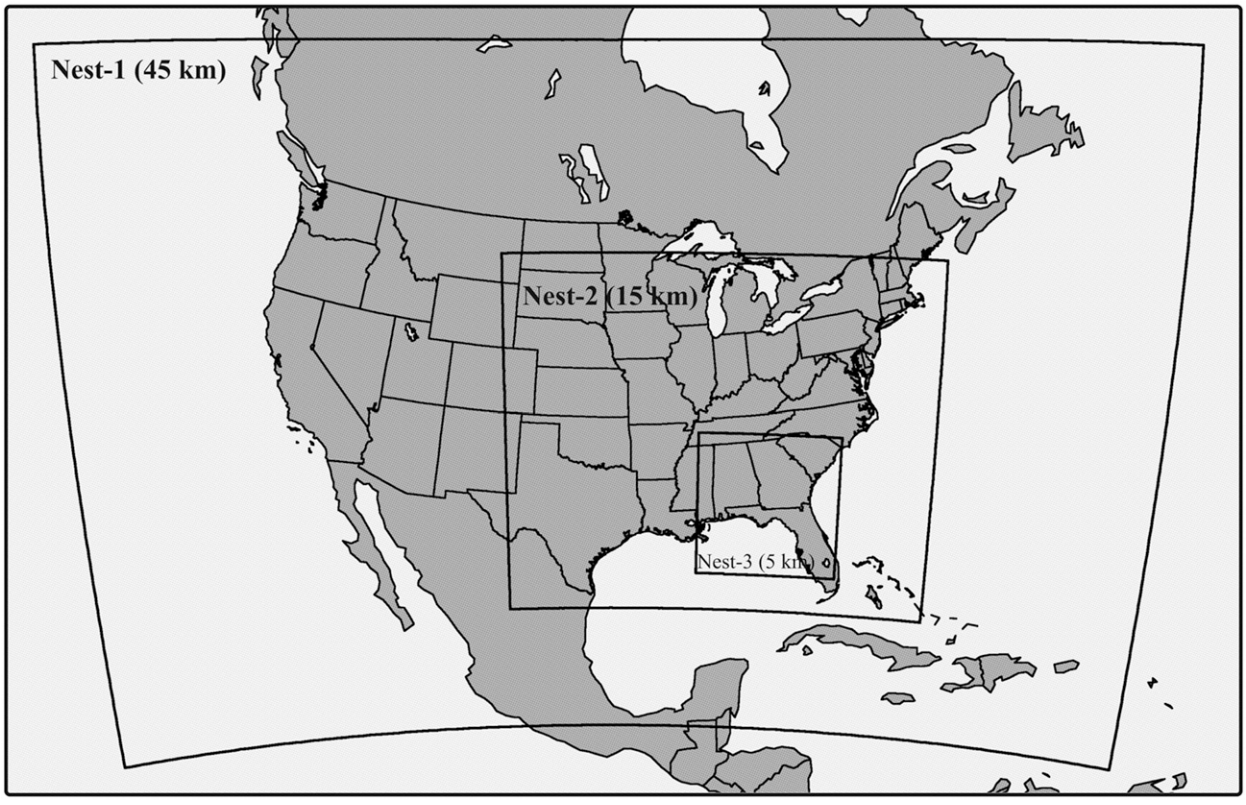

FIG. 9. COAMPS domains of three nested grids with grid spacings of 45, 15, and 5 km.

and the COAMPS Tropical-Cyclone (TC) model for TC prediction with highly regarded performance. The best way to demonstrate that the EnKF is a mature data assimilation system for real storm prediction is to compare the COAMPS ensemble forecasts from the EnKF with the control run that is initialized by the NAVDAS 3DVAR with exactly the same observational data assimilated. For this purpose, two experiments have been conducted. One is the control run initialized by the NAVDAS 3DVAR (the experiment is called CNTL-3DVAR hereafter) and another is an ensemble experiment with 32 members initialized by the EnKF (Ens32-EnKF). The COAMPS model uses three nested grids at 45-, 15-, and 5-km spacing. Figure 9 gives the model domains for the three nested grids. Both the CNTL-3DVAR and the Ens32-EnKF were cold started at 0000 UTC 22 June 2005 and initialized by a single NOGAPS forecast and a 32-member NOGAPS ensemble, respectively. After the cold start, both experiments were warm started with a 12-h update cycle starting at 1200 UTC 22 June 2005. Exactly the same observational data available to the operational NAVDAS 3DVAR at that time were assimilated at each update cycle into COAMPS by both the 3DVAR and EnKF schemes (no radar data were assimilated). After each update, 72-h COAMPS forecasts were launched from both the CNTL-3DVAR and the Ens32-EnKF. NOGAPS single and ensemble forecasts were again used as boundary conditions for COAMPS. After 10 continuous analysis–forecast update cycles, the experiments ended at 1200 UTC 27 June 2005. Note that we decided to omit a spinup period for the COAMPS ensemble because the NOGAPS ensemble used as the COAMPS cold start had already been running for more than 20 days.

The model forecasts of T, u, v, and q_v from the CNTL-3DVAR deterministic runs and the mean values of those from the Ens32-EnKF ensemble runs during this experiment period were verified against raob observations. RMS errors were computed for each of the three nested grid domains. Figure 10 gives the averaged RMS errors over the 3D domain for the experimental period as a function of forecast lead time. To examine how the EnKF performs compared to NAVDAS 3DVAR at both large and storm scales, only the scores for the 45- and 5-km domains are displayed. Before 24 h of forecast lead time, both the CNTL-3DVAR and the Ens32-EnKF simulation have basically similar performance at both large and storm scales. After that, the Ens32-EnKF gradually shows better forecasts than the CNTL-3DVAR, especially for the 5-km domain. The forecast

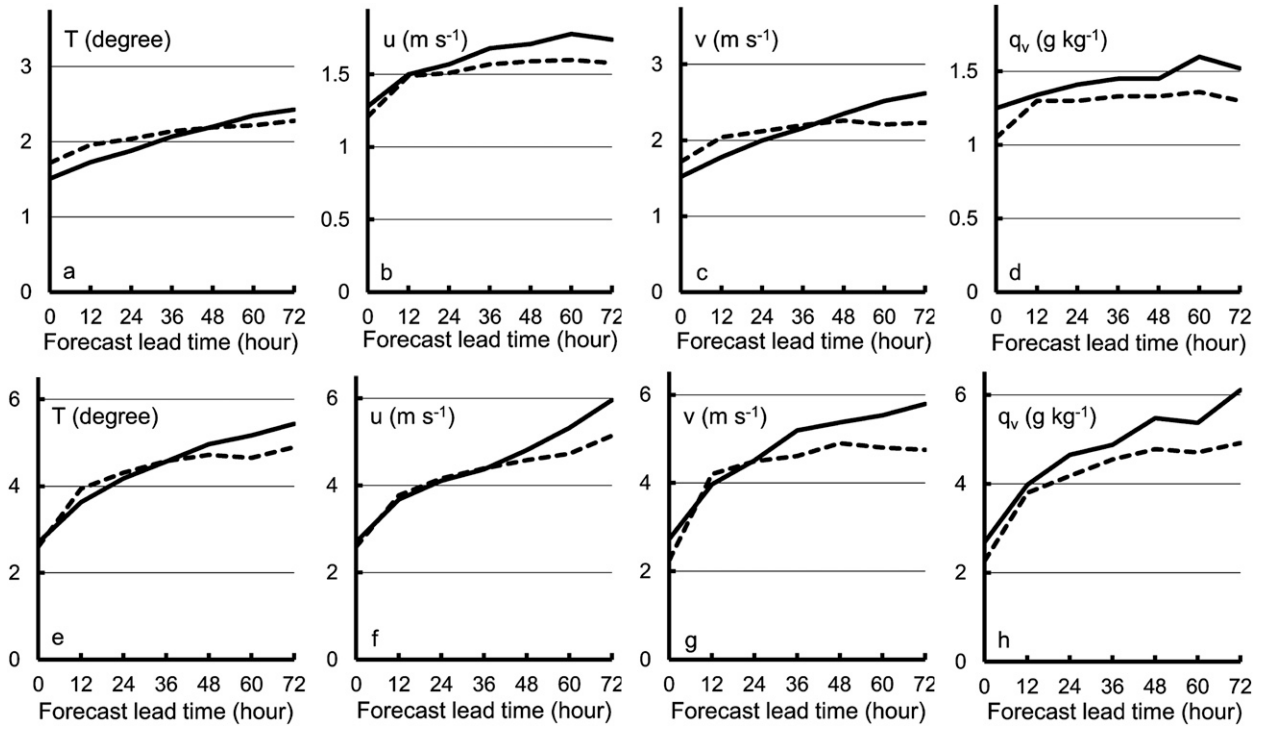

FIG. 10. RMSEs of model forecasts of T (°), u ($\mathrm{m\,s^{-1}}$), v ($\mathrm{m\,s^{-1}}$), and q_v ($\mathrm{g\,kg^{-1}}$) from CNTL-3DVAR (solid lines) and Ens32-EnKF (dashed lines) as a function of forecast lead time for the (a)–(d) 45- and (e)–(h) 5-km domains, for the forecast period of 1200 UTC 22 Jun–1200 UTC 27 Jun 2005 verified against raob observations.

improvement by Ens32-EnKF reaches a maximum at 72 h. If we look at different state variables, we will find that Ens32-EnKF has basically the same forecast performance as CNTL-3DVAR for temperature, is slightly better for winds, and is significantly better for the water vapor mixing ratio. But overall, Ens32-EnKF outperforms CNTL-3DVAR, which is consistent with recent studies of Meng and Zhang (2008a,b) based on the WRF model. Vertical profiles of the averaged RMS errors of the four state variables over the experiment period at forecast lead times of 72 h from both experiments are also calculated and given in Fig. 11. As we can see, the improvement in the model forecasts by Ens32-EnKF at this forecast hour is basically through the whole column of the atmosphere rather than at particular model levels. This further demonstrates that the EnKF data assimilation system developed at NRL has shown strong potential to become a mature data assimilation system comparable or possibly superior to the operational NAVDAS 3DVAR.

6. Summary

An EnKF has been adopted and implemented at NRL for COAMPS ensemble data assimilation and forecasts. The implemented EnKF has gone through extensive tests with real observational data to examine the key features of the system to ensure that the system is working appropriately in a new environment. The results reported here show that the system has the capability of usefully estimating background error covariance when the ensemble reaches a certain size (it appears to be 32 in Fig. 3). As the ensemble size increases, the covariance looks smoother. As the flow-dependent error covariance is estimated from the ensemble perturbations, the synoptic and large-scale structures were found to be, as expected, qualitatively consistent with the geostrophic relationship between temperature and wind in the atmosphere.

A new serial ''innovation reduction'' method for assessing the accuracy of the ensemble-based error covariance models has been introduced. The technique provides a computationally inexpensive means of tuning localization length scales and assessing the potential benefit of increased ensemble size. Using this approach, we demonstrated that the ability of the ensemble-based error covariance model to reduce analysis errors increases monotonically as the ensemble size increases from 16 to 64, but the change becomes rather small as the ensemble size further grows from 64 to 128. The approach was also used to find the most effective localization length scales for various ensemble sizes. It was found that for relatively small ensemble sizes of 16–32

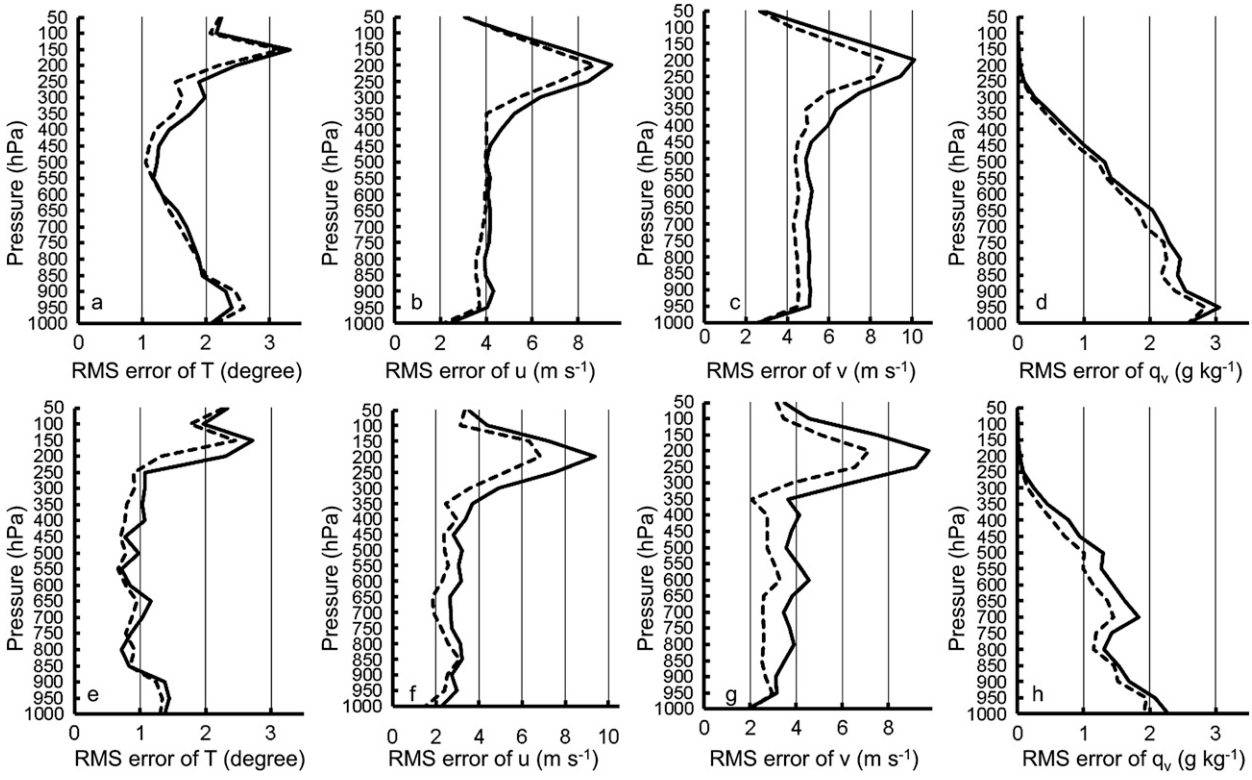

FIG. 11. Vertical profiles of RMS errors of model forecasts of T (°), u ($m\,s^{-1}$), v ($m\,s^{-1}$), and q_v ($g\,kg^{-1}$) from CNTL-3DVAR (solid lines) and Ens32-EnKF (dashed lines) at the forecast lead time of 72 h for the (a)–(d) 45- and (e)–(h) 5-km domains, for the forecast period of 1200 UTC 22 Jun–1200 UTC 27 Jun 2005 verified against raob observations.

members, relatively tight localization length scales of 200–600 km were required for effective performance while for ensemble sizes of 64 and 128 a larger localization radius was more effective. The performance of the EnKF was much less sensitive to the localization radius at large ensemble sizes.

COAMPS ensemble forecasts initialized by the EnKF analyses from the above experiments are also evaluated. The impacts of ensemble size and LLS on ensemble forecasts are examined by looking at the RMS errors of the ensemble forecasts verified against radiosonde observations. It is found that both ensemble size and LLS affect the forecasts. The effect of ensemble size, however, decreases rapidly with forecast time and becomes barely visible at about 24 forecast hours for ensemble sizes larger than a certain number (32 in this case study). For a small ensemble size of 16, the large forecast errors basically remain throughout the whole forecast period. The impact of LLS also decreases quickly during the first 12 h of model integration for an ensemble size of 32, and then remains small but notable during the subsequent forecast period.

An extensive study has been performed to compare the COAMPS EnKF with the current operational NAVDAS 3DVAR for COAMPS at NRL. It has been found from a week-long test that the ensemble forecast from the EnKF shows slightly faster error growth than the control forecast initialized by the NAVDAS 3DVAR during the first 12 h of model integration. After that, the error growth in the ensemble forecast slows down and becomes significantly slower than that from the control forecast. The EnKF outperforms the 3DVAR in the moisture analyses and the subsequent forecasts at both large and storm scales. For the temperature and wind fields, however, the results are mixed for the first 36 h. During this period, the 3DVAR seems to perform better than the EnKF mainly at large scales. After the 36 h of model integration, the EnKF steadily improves the forecast all the way to 72 h where the improvement reaches a maximum. Vertical profiles of RMS errors from the ensemble and control forecasts at 72 h also show that the improvement in the model forecasts by the EnKF is basically across the whole depth of the atmosphere. This further demonstrates that the COAMPS EnKF has the potential to become a mature data assimilation system for mesoscale and storm-scale data assimilation. Doppler radar data assimilation into the EnKF is currently being investigated. The preliminary results show further

improvement of the ensemble analyses and forecasts with radar observations. More studies will be performed and the results will be presented in follow-up papers.

Acknowledgments. The authors thank the On-Demand Section led by Mr. John Cook, the Mesoscale Modeling Section led by Dr. Jim Doyle, and the Data Assimilation Section led by Dr. Nancy Baker of the Marine Meteorology Division, Naval Research Laboratory, for their strong support and help during this study. Special thanks go to Dr. Keith Sashegyi of the Data Assimilation Section for his generous help in providing the data processing and quality control algorithms as well as observational forward operators from NRL NAVDAS. This research was supported by the Office of Naval Research (ONR) under PE 0602435N, N000140910526, and other ONR-sponsored collaboration project.

REFERENCES

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.,* **137,** 1805–1824.

Anderson, J. L., and N. Collins, 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.,* **24,** 1452–1463.

Daley, R., and E. Barker, 2001: NAVDAS source book 2001. Naval Research Laboratory Rep. NRL/PU/7530-01-441, 161 pp.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

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

Grell, G. A., J. Dudhia, and D. R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR Tech. Note NCAR/TN-3981STR, 138 pp.

Hodur, R. M., 1997: The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Mon. Wea. Rev.,* **125,** 1414–1430.

Hogan, T. F., and T. E. Rosmond, 1991: The description of the Navy Operational Global Atmospheric Prediction System's spectral forecast model. *Mon. Wea. Rev.,* **119,** 1786–1815.

McLay, J. G., C. H. Bishop, and C. A. Reynolds, 2008: Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.,* **136,** 1093–1108.

——, ——, and ——, 2010: A local formulation of the ensemble transform (ET) analysis perturbation scheme. *Wea. Forecasting,* **25,** 985–993.

Meng, Z., and F. Zhang, 2007: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. *Mon. Wea. Rev.,* **135,** 1403–1423.

——, and ——, 2008a: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study. *Mon. Wea. Rev.,* **136,** 552–540.

——, and ——, 2008b: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with 3DVAR in a month-long experiment. *Mon. Wea. Rev.,* **136,** 3671–3682.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRF version 2. NCAR Tech. Note NCAR/TN-468STR, 88 pp.

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

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.,* **131,** 1485–1490.

Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.,* **133,** 1789–1807.

——, and ——, 2008: Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part II: Parameter estimation experiments. *Mon. Wea. Rev.,* **136,** 1649–1668.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.,* **130,** 1913–1924.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.,* **132,** 1238–1253.

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

——, Y. Weng, J. Sippel, Z. Meng, and C. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.,* **137,** 2105–2125.