10.7 **Impacts of Initial Analyses and Observations on the**

Convective-Scale Data Assimilation with an Ensemble-Kalman Filter

Fuqing Zhang[1], Chris Snyder[2] and Juanzhen Sun[2]

[1]Department of Atmospheric Sciences, Texas A&M University, College Station, Texas
[2]National Center for Atmospheric Research, Boulder, Colorado

1. Introduction

The key problem in covective-scale data assimilation is to infer other variables given observations of radial velocity and reflectivity. This problem has been addressed using a number of retrieval techniques (e.g., Shapiro et al. 1995 and references) and, more recently, with four-dimensional variational assimilation schemes (4DVar; Sun and Crook 1997). The ensemble Kalman filter (EnKF) is an alternative approach; it uses an ensemble of short-range forecasts to estimate the flow-dependent background error covariances required in data assimilation. This method has attracted considerable recent interest in the fields of atmospheric and oceanic sciences and hydrology because of a number of appealing properties: it does not require adjoints of either the forecast model or observation operators, it integrates data assimilation and ensemble forecasting and thus produces estimates of forecast uncertainty at no extra cost, it is highly parallel, and it is largely independent of the forecast model (Evensen 1994; Houtekamer and Mitchell 1998; Hamill and Snyder 2000; Anderson 2002). In perfect-model experiments using simulated observations of radial velocity from a supercell storm, Snyder and Zhang (2002) demonstrated the feasibility of the EnKF for convective-scale data assimilation. The present study seeks to explore further the potential and behavior of the EnKF at convective scales by considering more realistic initial analyses and variations in the availability and quality of the radar observations. A detailed report of this research is in Zhang et al. (2002).

In a general sense, both the EnKF and 4DVar utilize the time history of the flow to gain information on unobserved variables. While 4DVar accomplishes this by fitting a solution of the forecast model to the observations over an interval of time, the EnKF summarizes the effects of previous observations and of past growth of forecast errors in terms of $\mathbf{B}^f$, the forecast-error covariance matrix. Although their role may seem obscure, these covariances provide direct information on unobserved variables; if we know, say, the correlation in our background (or first guess) forecast between errors in the components of the velocity, then an observation of one component can be used to estimate the others.

The update equation for the Kalman filter is,

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \qquad (1)$$

where the flow-dependent estimate of background error covariance **B** is estimated through the ensemble forecast (Evensen 1994):

$$\mathbf{B} \approx \mathbf{N}_e^{-1}\Sigma(\mathbf{x}_i^f - \underline{\mathbf{x}})(\mathbf{x}_j^f - \underline{\mathbf{x}})^T \qquad (2)$$

2. The control experiment

The numerical model used is that of Sun and Crook (1997). The forecast model was assumed to be perfect; that is, the numerical model produces the forecasts and the reference simulation from which observations are taken. The environmental profile is based on the 00Z 5/25/97 Oklahoma City sounding which was taken just before the development of a supercell thunderstorm. We have added a mean wind to the sounding so that main cell has small propagation speed. The forecast model employs a domain of 70 km by 70 km horizontally with 2-km grid spacing and 18 km vertically with 0.5-km resolution. The initial warm bubble of 2.0-K was added to the liquid water potential temperature field to initiates a convective cell. Figure 1 shows the 5-km AGL zonal winds, vertical velocity and total liquid water at 40, 60 and 80 min, respectively. The storm spits between 60 to 80 minutes after the warm bubble initiation; the right moving storm is dominant while the weaker left storm move faster relative to the basic flow.

Simulated Doppler-radar wind observations are taken from the reference simulation. The radar is located at the southwest corner of the computational domain; it measures the radial velocity in a spherical coordinate system centered on the radar; the observations have a independent, Gaussian random errors of zero mean and variance of 1m/s; and the radial velocity is observed only where the radar reflectivity is greater than 15 dBZ. The observations are thus related to the reference state by observing the radial velocity and the dependence on the fall speed of rain has been neglected for simplicity.

The assimilation experiments begin at 40 min. Observation sets, consisting of at all points with exceeding the threshold given above, are available at 45 min and every 5 min thereafter. The only additional information used in the assimilation is the environmental sounding. Thus, each ensemble member is initialized 5 min prior to the first observations by adding realizations of Gaussian to the environmental sounding. This noise is independent at each grid point and has variance of 3.0 m/s for each component of velocity and 3.0 K for the liquid-water potential temperature. Water vapor is initialized using the environmental sounding at each level. No cloud water and rain water is allowed in the initial ensembles.

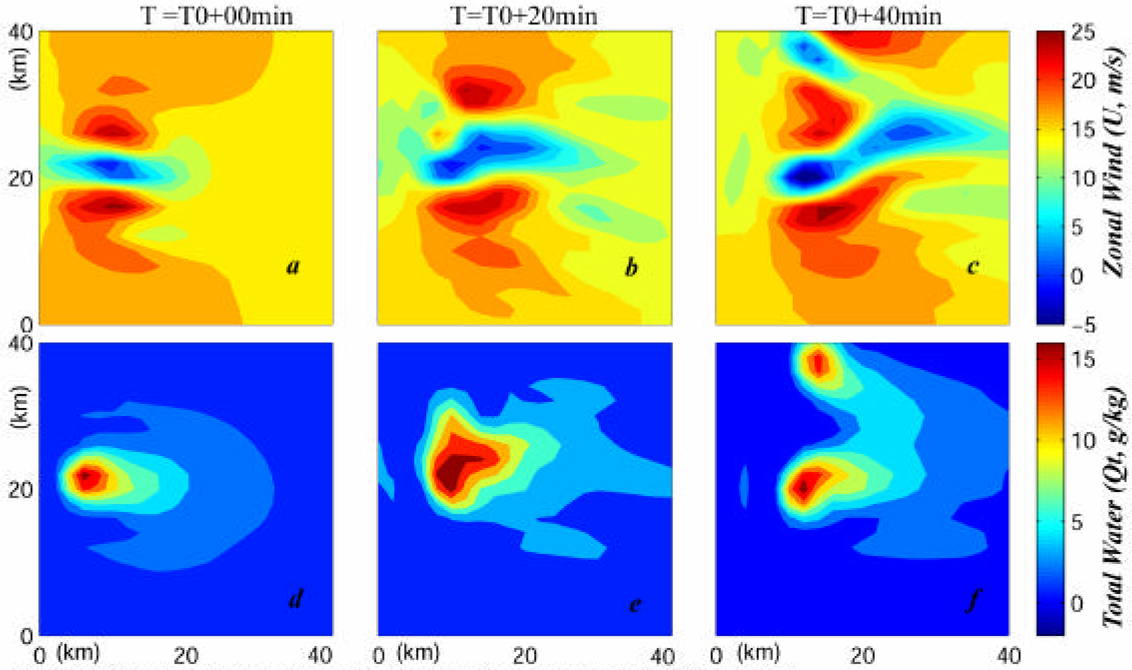

Figure 1. Zonal wind and total water in the the reference simulation (T0=40 minutes).

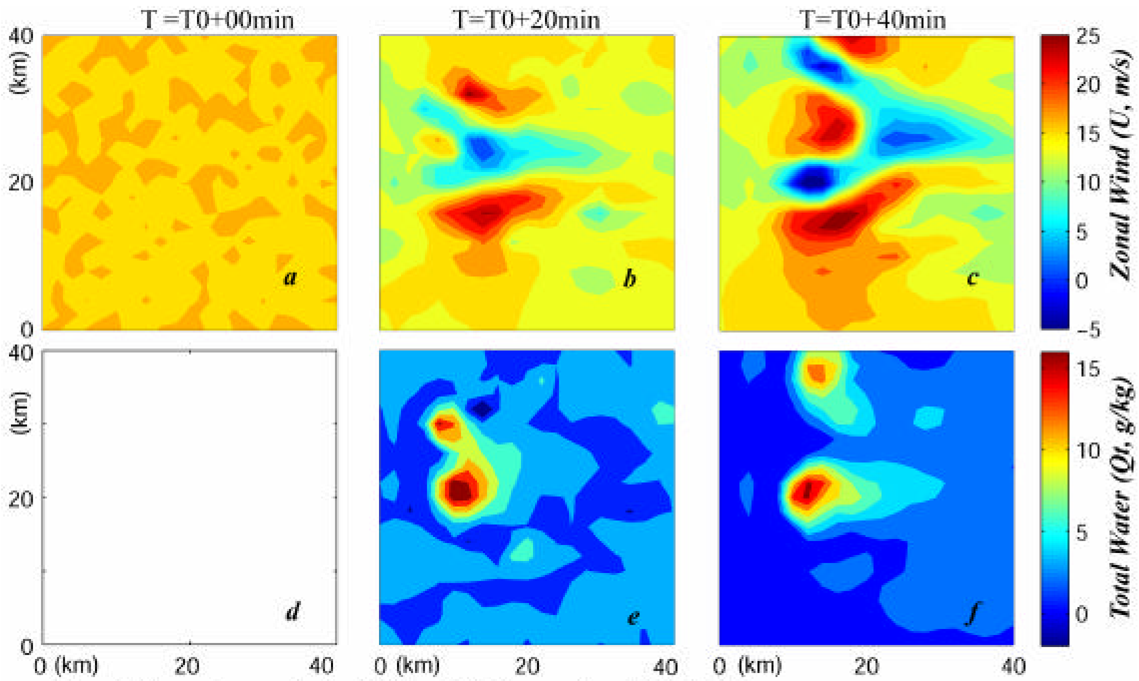

Figure 2. Ensemble mean in the CNTL assimilation experiment (T0=40min).

Only 20 members were used for the ensemble forecasts. Each observation was assimilated sequentially with the radius of influence set to be 3 km. The posterior and prior background error covariances were averaged after each assimilation cycle. The 5-km AGL initial ensemble means of U, and Qt at 40 min are shown in Fig.2a-c, which contains no information of the initial storm shown as in Fig.1a-c. After 20 minutes (4 cycles) of assimilation of radial velocity observations where dBZ>15 with EnKF, strong indication of the existence of the storm is presented not only in the wind fields (Fig.2b) which are partially observed variables), but also in the total water field (Fig.2e) and in the temperature field as well, even though the assimilated storm is less organized and still significantly different to the reference storm at this time. After another 20 min, the EnKF has captured completely the splitting supercell with comparable right structure and strength in almost every aspect of the storm (Fig.2d-f vs. Fig.1d-f). There is very small difference in the updraft of the left moving storm and also in the rain shafts between two cells. Domain-averaged root-mean square error (RMSE) between the assimilated ensemble mean and the reference run of all six prognostic model variables after every assimilation cycle for 90 min is shown in Fig.3. Except for the first several cycles, by taking only the radial velocity observation when there is significant radar reflectivity (dBZ>15), we can see the RMSE is consistently and rapidly dropping for all variables. The RMSE of U and V winds drops below the observational error bar (1.0 m/s) after 40 min and continuously drops to ~0.5 m/s after 80-90 min and appears to stabilize afterwards. Similar convergence of the updated/assimilated ensemble mean to the reference state is also true for the unobserved prognostic variables (Fig.3b).

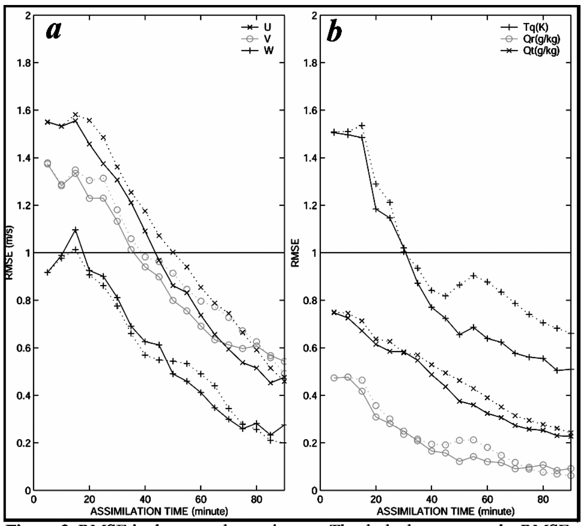

Figure 3. RMSE in the control experiments. The dashed curves are the RMSE without mixing of the error covariance after each assimilation cycle.

3. Impacts of initial estimate

As true to all Kalman Filter-type of state estimation which combines the prior estimate (first guess or initial estimate) and the observation as well as the associated uncertainties, the quality of the initial estimate sometime can be key to the final best estimate. We first begin with the assimilation experiment with good initial guess at 40min right before the filter starts. In “Good0Guess”, there exists a supercell storm in the initial ensemble mean; each wind component of the initial ensemble mean differs from the reference storm by a random error of 3m/s and liquid water potential temperature by a random error of 3 K, unlike in CNTL where there is no prior information of the supercell except for the background sounding. Though the initial root-mean-square (rms) error is bigger than the control run by design (not shown), the ensemble mean will be quickly drawn closely to the truth run after four-assimilation cycles. The ensemble mean is continuously improving to a high accuracy after 40 minutes. The benefit of Good0Guess over CNTL vanishes after 50-60 min (Fig.4a). In the forecast experiments with the good initial estimate as in Good0Guess but without EnKF assimilations, the RMSE error of the pure ensemble forecast mean relative to the reference simulation can grow to a magnitude of 3m/s for velocities and 3 K for Tq (not shown). The EnKF is intelligently and continuously drawing information from the observations to keep the solution from diverging.

The initial estimate of most data assimilation system usually comes from the previous short-term forecast; there is strong possibility that a storm may exist in the initial estimate but in the wrong location. In “Bad0Guess”, the initial supercell was completely dislocated from the reference simulation by a distance of 10 km. For the first several assimilation cycles, the assimilation is struggling to develop a supercell in the right location and to destroy the incorrectly located initial storm by taking Vr observations in the vicinity of the truth storm location. After 20 minutes, both the newly assimilated storm and the “false” initial storm coexist in the ensemble mean; the RMSE of all prognostic variables is considerably larger than that in the CNTL. However, after 40 min of assimilation of Vr observations, the EnKF filter has successfully assimilated the supercell, with the same accuracy as those in CNTL (Fig.4b) indicates that the EnKF is resilient under different initial estimates.

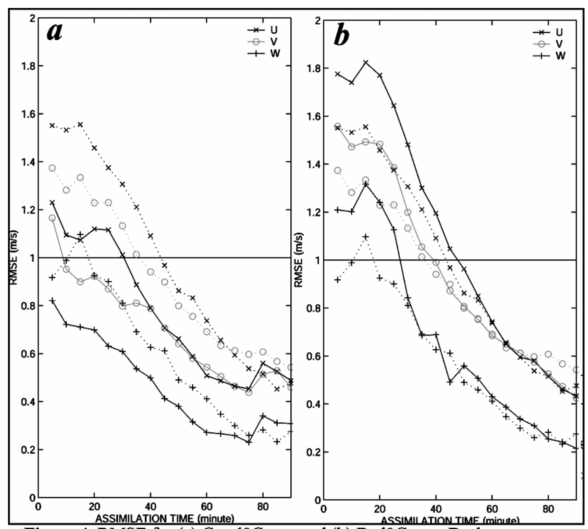

Figure 4. RMSE for (a) Good0Guess and (b) Bad0Guess. Dashes curves are the RMSE from the CNTL experiment for direct comparison.

4. Impacts of Observational coverage

In real event, the convective storms are often at a distance to the Radar site and thus the lowest boundary layer of the storm is likely missing from the Radar observations. Three different experiments have been designed. Experiments “No2kmPBL" and “No4km1” are performed exactly the same as CNTL with the same initial estimate except that there are no observations taken below 2 km and 4 km, respectively. In "No4kmPBL2", a dense (every 4 km by 4 km) surface mesonet with U, V, Tq observations was added; the standard deviation of the surface observational error is 1m/s for U and V and 1K for Tq. As expected, the EnKF in No2kmPBL will initially converge toward the truth slower than that in the CNTL; after 4 or 5 assimilation cycles, the structure and characteristics in the assimilated ensemble mean will match closely those in the truth and the EnKF captures the storm in the ensemble mean after 40-min assimilation, though the RMSE is slightly larger than from the CNTL throughout the whole 90-min assimilation (Fig.5a). No4km1 has the difficulty to assimilate the storm with the right strength after 40 min even though the assimilated splitting supercells developed in the right locations. The assimilation will not fully recover the storm even after an hour; however, the assimilated ensemble mean does eventually approach the true state after 60-80 minutes. Significant improvement of the EnKF performance can be found when a surface mesonet of wind and temperature observations are added, as can be seen in No4km2. In this experiment, the radius of influence of each surface observation is set to 5 km. Even though the EnKF still has the difficulty assimilating the left-moving storm (which is the weaker one and has relatively little reflectivity above 4km, it captures the dominant cell after 40 min of assimilation. The domain averaged RMSE in No4km2 fall close to that of the CNTL after an hour while remains above or equal to the observational error bar for U and V throughout the assimilation (Fig.5b). Experiment No4km2 certainly suggests the addition of a hypothetical surface mesonet can be very beneficial in convective-scale data assimilation.

Mobile Doppler radar or Doppler on wheels (DOW) has become an increasingly promising tool to observe and study the convective storms. Many of these DOWs observe only the lower part of the storms with higher accuracy. Experiment “Only4kmPBL” is designed to assimilate Vr observations only below 4 km, mostly in the boundary layer. With only the lower-layer observations, even though the filter begins to develop storms in the ensemble mean after 20 minutes similar to those in the CNTL at 5-km AGL, the domain-averaged RMSE of all prognostic variables is significantly larger; most of the difference exists in the upper troposphere where no observations are taken. After 40 min, the filter appears to have captured the splitting storms in the right location and strength though the zonal wind field has the most difficulty from both the 5-km AGL fields and the RMSE (not shown), which improves greatly after 60 min.

5. Concluding remarks

The ensemble Kalman filter (EnKF) uses short-range ensemble forecasts to estimate the flow-dependent background error covariances required in data assimilation. Here we further demonstrated the feasibility of the EnKF for convective-scale data assimilation in consideration of impacts of initial analyses and observations. EnKF using radial-velocity observations and 20 ensemble members can be successful in most realistic observational scenarios for supercell thunderstorms, although because of our assumption of a perfect forecast model these are undoubtedly upper bounds on the performance to be expected with real observations and an imperfect model. Even though the filter converges toward the truth simulation faster from a better initial estimate, an accurate estimate of the storm can be achieved within an hour even the initial storm was totally misplaced. Similarly, radial-velocity observations below 2km are certainly beneficial but in their absence the assimilation scheme can still achieve a comparably accurate estimate of the state of the storm given a slightly longer assimilation period. The addition of a dense network of wind and temperature observations at the surface, the EnKF can again provide an accurate estimate of the storm with no radial velocity observation below 4km. It can also successfully assimilate the storm in the case of radar observations only below 4 km.

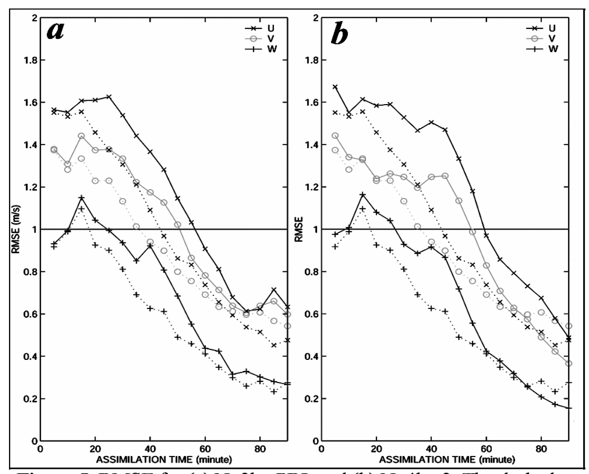

Figure 5. RMSE for (a) No2kmPBL and (b) No4km2. The dashed curves are the RMSE from the CNTL for direct comparison.

6. Reference

Anderson, J. L., 2001: *MWR,* **127**, 2741.

Evensen, G., 1994: *JGR*, **99**(**C5**), 10143.

Hamill, T. M., and C. Snyder, 2000: *MWR*, **128**, 2905.

Houtekamer, P. L., and H. L. Mitchell, 1998: *MWR*, **126**, 796.

Houtekamer, P. L., and H. L. Mitchell, 2001: *MWR*, **29**, 123.

Shapiro, A., S. Ellis and J. Shaw, 1995: *JAS*, **52,** 1265–1287.

Snyder, C., and F. Zhang, 2002: Assimilation of simulated Doppler radar observations with an ensemble-Kalman Filter. *MWR*, to be submitted.

Sun, J., and N. A. Crook, 1997: *JAS*, **54**, 1642.

Zhang, F., C. Snyder, and J. Sun, 2002: Impacts of initial estimate and observations on the convective-scale data assimilation with an ensemble Kalman filter. *MWR*, to be submitted.

Acknowledgements. This research is supported by NSF/ITR 02-05599 and also by NSF/NCAR USWRP.

Corresponding author: FZ (fzhang@tamu.edu)