12.1 A COMPARISON BETWEEN THE 4D-VAR AND THE ENSEMBLE FILTER TECHNIQUES FOR RADAR DATA ASSIMILATION

A. Caya, J. Sun and C. Snyder

National Center for Atmospheric Research,*
Boulder, Colorado

1. INTRODUCTION

A complete atmospheric state estimate is required for numerical weather prediction (NWP). A prior state estimate (background) is usually available and can be improved by using newly obtained observations of the atmosphere. The four dimensional variational data assimilation (4D-Var) and the ensemble Kalman filter techniques are two family of methods for combining the background and the observations in an analysis suitable for NWP.

The ability of 4D-Var has been proven in some NWP centers. For radar data assimilation, the 4D-Var was also used successfully (Sun and Crook, 1997). Recently, ensemble Kalman filter techniques have been developed and tested on simple models (for examples, see Whitaker and Hamill, 2002). The application of ensemble Kalman filtering with more realistic atmospheric models is yet to be fully explored. In this paper, the ensemble Kalman filter is applied to the assimilation of radar data in a cloud-resolving model and compared with the results of 4D-Var.

2. ASSIMILATION ALGORITHMS

In the theoretical context of perfect model and linear dynamics, the 4D-Var and the Kalman filter give identical results at the end of the assimilation (Lorenc, 1986). Because of model nonlinearity, sampling error, and computational procedure, results from 4D-Var and the ensemble Kalman filter are expected to differ.

2.1 *The 4D-Var algorithm*

The 4D-Var analysis $\mathbf{x}^a$ is obtained through the minimization of a cost function J that measures the misfit between the model trajectory $\mathbf{Hx}$ and the observations $\mathbf{y}$ over a time window:

$$J = (\mathbf{x}-\mathbf{x}^f)^T\mathbf{P}^{f\,-1}(\mathbf{x}-\mathbf{x}^f)+(\mathbf{Hx}-\mathbf{y})^T\mathbf{R}^{-1}(\mathbf{Hx}-\mathbf{y}), \quad (1)$$

where $\mathbf{H}$ is the generalized observational operator and $\mathbf{R}$ is the observational-error covariance matrix. The minimization of the cost function is achieved by adjusting the initial condition $\mathbf{x}$ of the model at the beginning of the assimilation window while remaining close to the background state $\mathbf{x}^f$, which is usually a forecast. This closeness is specified by the forecast-error covariance matrix $\mathbf{P}^f$. The minimization itself is performed by integrating the forecast model through the end of the assimilation window to evaluate the cost function. Then the adjoint model is integrated backward in time to give the variation of the cost function with respect to the initial condition. This information is used iteratively in a descent algorithm until a satisfactory solution is found. The solution at the end of the assimilation window is used as background and first guess for the next assimilation cycle.

2.2 *An ensemble Kalman filter*

Unlike the 4D-Var algorithm, the Kalman filter assimilates the observations sequentially. First, the model state $\mathbf{x}^f$ and the forecast-error covariance matrix $\mathbf{P}^f$ are propagated forward in time until the next observational time. Second, a new state estimate $\mathbf{x}^a$ is provided by minimizing J, which is now a 3D-Var problem for the Kalman filter since only the data at that time are considered. Third, the forecast-error covariances are reduced by the assimilation of the observations. This procedure is repeated until all data in the assimilation window are processed.

Unfortunately, the full representation and evolution of the forecast-error covariance matrix $\mathbf{P}^f$ required by the Kalman filter are not feasible at present for NWP model. The ensemble Kalman filter has been developed to circumvent this problem (Evensen, 1994). The forecast-error covariance matrix $\mathbf{P}^f$ is represented with an ensemble of equally possible states $\{\mathbf{x}_i^f,\, i = 1, \ldots, n_e\}$ by:

$$\mathbf{P}^f = \frac{1}{n_e - 1}\mathbf{X}^f\mathbf{X}^{f\,T}, \quad (2)$$

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Alain Caya; NCAR, P.O. Box 3000, Boulder, CO 80307-3000; caya@ucar.edu.

where $\mathbf{X}^f$ is the matrix whose columns are the deviations from the ensemble mean of $\mathbf{x}_i^f$ and is also a square root (to a constant) of $\mathbf{P}^f$. By using a limited number of ensemble members n_e, the method can be as tractable as 4D-Var.

An ensemble square-root filter (EnSRF), which does not require perturbed observations (Whitaker and Hamill, 2002), is used in this study. When observational-error correlations can be neglected (i.e. $\mathbf{R}$ can be reduced to a diagonal matrix), the observations can be assimilated one at a time. This greatly simplifies the computation in the analysis step in the EnSRF and is exploited in this work.

The sampling error associated with the limited ensemble size means that the small correlations between widely separated state variables are difficult to estimate (Hamill et al., 2001). For this reason, each observation is allowed to influence only state variables located within a certain cut-off radius of the observation.

3. METHODOLOGY

The cloud model of Sun and Crook (1997) is used in this study to generate a control simulation of a supercell storm as well as the assimilation model for 4D-Var and the EnSRF. Thus, the two methods are compared here in the context of perfect model. The prognostic variables are the velocity components (u, v, w), the liquid water potential temperature (θ_l), the rain water mixing ratio (q_r), and the total water mixing ratio (q_t). The supercell storm is produced by initializing the model with the 00Z 5/25/97 Oklahoma City sounding on which a warm bubble is superimposed. The warm bubble is centered at 1.25 km altitude, is 16 km wide and 2 km deep, and is 1 degree warmer than the environment. The model has 2 km resolution in the horizontal and 500 m between vertical levels. The domain is 140x140x17.5 km^3 and is large enough as to mitigate the effect of the boundaries on the simulated storm. The model equations are integrated with a time step of 5 s.

3.1 *Simulated data*

The radial velocity observations are simulated at each 5 minutes from the control simulation where the rain water mixing ratio q_r exceeds 0.13 g kg^{-1}, or approximately 12 dBZ, as seen by a single radar located at the South-West corner of the domain. A 1 m s^{-1} random Gaussian noise is added to the radial velocities before the assimilation.

In the 4D-Var experiments, q_r is also assimilated at those points where it exceeds 0.13 g kg^{-1}. Assimilation of q_r does not improve the EnSRF analyses at present and so the results shown for the EnSRF are based only on assimilation of radial velocity.

3.2 *Background and forecast-error covariance*

The true rain water mixing ratio is used in the background and the liquid water potential temperature θ_l is modified accordingly following the definition

$$\theta_l = \theta \left(1 - \frac{L_v}{c_p T}(q_c + q_r)\right). \qquad (3)$$

The cloud water mixing ratio q_c is set to zero. The remaining of the model variables are set to the values of the sounding.

The forecast-error correlations are neglected in $\mathbf{P}^f$ of the 4D-Var cost function. The forecast-error covariance matrix is thus modeled in 4D-Var as diag(σ_i^2), that is, as a diagonal matrix of the error variances σ_i^2 for each state variable. Forecast-error correlations are also ignored in the initial EnSRF ensemble; each member is initialized with the environmental sounding plus Gaussian noise with zero mean and covariance diag(σ_i^2). Table 1 shows the corresponding standard deviations for both 4D-Var and the initial EnSRF ensemble. Note that all the negative values produced during the addition of the noise are changed to zero for q_r and q_t.

Table 1. Forecast-error standard deviations.

	$\sigma_{u,v,w}$ (m s^{-1})	σ_T (K)	σ_{q_r,q_t} (g kg^{-1})
4D-Var	1.0	1.0	0.1
EnSRF	3.0	3.0	0.1

A 100-member EnSRF is initialized at t=35 minutes into the control simulation. The cut-off radius limits the correlation length to 6 km. The 4D-Var uses a 10-minute assimilation window. The data are assimilated from t=35 minutes, when the precipitating cell is organized, and the assimilation is stopped after 65 minutes. From there 2-hour forecasts are issued from 4D-Var and EnSRF analyses.

4. RESULTS

Figure 1 shows the rain water mixing ratio at the lowest model level (250 m) with the horizontal wind vectors at 65 minutes (the end of the analysis cycles) in the control simulation, and the corresponding 4D-Var and EnSRF analyses. Both 4D-Var and EnSRF analyses kept track of the single cell that the warm bubble produced but it is better represented in the 4D-Var analysis. Figure

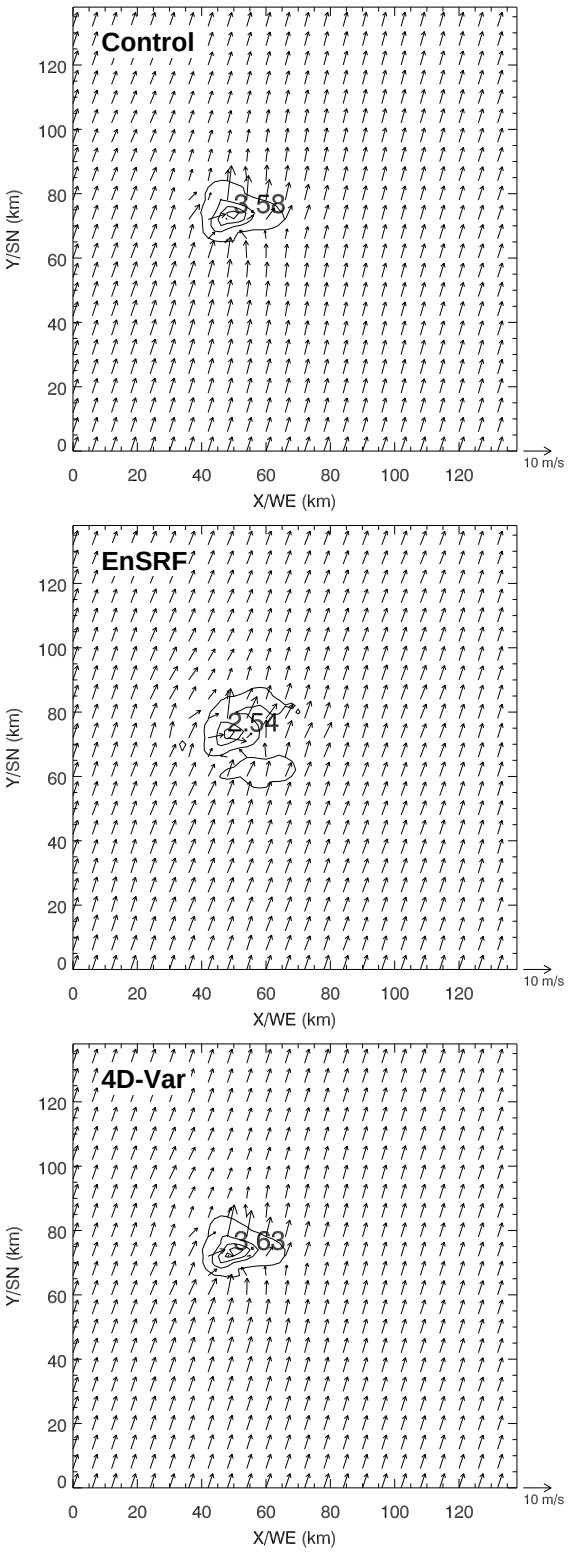

Figure 1. Rain water mixing ratio at 250 m and 65 minutes into the simulation (end of analysis cycles). Contour interval is 0.5 g kg^{-1} starting at 0.13 g kg^{-1}. Arrows are horizontal wind vectors.

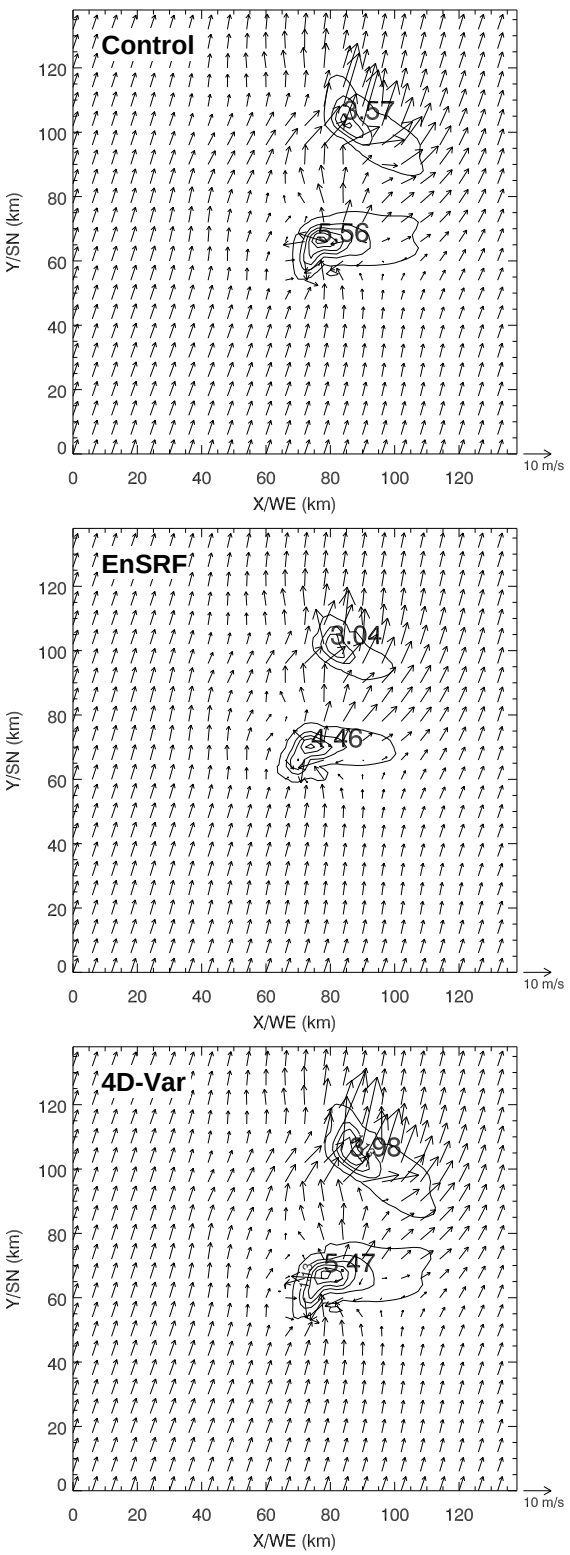

Figure 2. Same as Figure 1. but at 125 minutes into the simulation (1 hour forecast).

2 depicts the same fields after 1 hour into the forecast. Both forecasts have correctly produced the splitting of the cell in two but the intensity of precipitation in the forecast from the EnSRF analysis remains too weak and the 4D-Var analysis results in a better forecast.

Figure 3a shows the r.m.s errors in the 3D domain of the three wind components

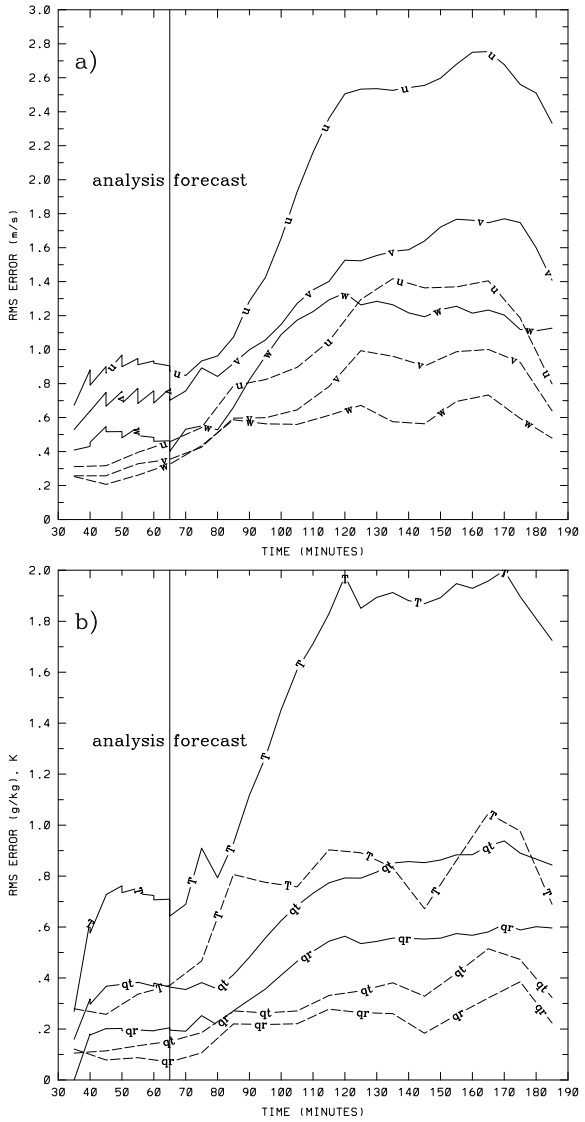

Figure 3. The r.m.s. error of the model variables during the assimilation from 35 to 65 minutes and the 2-hour forecast for the ensemble mean of the EnSRF (full line) and the 4D-Var analysis (dashed line): a) the three wind components and b) the liquid water potential temperature θ_l, the rain water q_r and the total water contents q_t.

during the assimilation and the forecast from the 4D-Var and EnSRF analyses. Figure 3b is for the r.m.s errors of the liquid water potential temperature, and rain water and total water mixing ratios. The 4D-Var is continuously producing better results for this case. We emphasize, however, that the r.m.s errors of the 4D-Var analyses generally increase at later times while those of the EnSRF tend to decrease after the third analysis cycle (at 50 minutes). By 90 minutes, the two methods give similar r.m.s errors (not shown).

5. FUTURE WORK

There are three changes that could improve the performance of the EnSRF. First, the sample covariances should be localized with a compactly supported correlation function (Hamill et al. 2001) rather than simply set to zero beyond a given radius. Second, instead of using uncorrelated noise over the entire domain to initialize the EnSRF, spatially correlated noise localized where there are radar echoes may be more suitable. This should reduce the standard deviations required in the initial ensemble and enable a better comparison with 4D-Var. Third, radar reflectivity observations are undoubtedly an important source of information and should be assimilated, although effective techniques for this remain to be developed for the EnSRF.

We also plan to examine the sensitivity of the methods to the presence of model errors and correlated observational errors. The EnSRF and 4D-Var methods are also applied on a real data case at this conference in presentations 13.9 and 4B.1, respectively.

REFERENCES

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**(C5), 10143–10162.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distant-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.

Sun, J., and N. A. Crook, 1997: Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. *J. Atmos. Sci.*, **54**, 1642–1661.

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