P6.1 TESTS OF AN ENSEMBLE KALMAN FILTER AT CONVECTIVE SCALES

C. Snyder, F. Zhang, J. Sun and A. Crook

National Center for Atmospheric Research,*
Boulder, Colorado

1. INTRODUCTION

Numerical forecasts at convective scales require an accurate estimate of the atmospheric state for initialization. Accurate state estimates at convective scales in turn require the inference of unobserved 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. Although it remains largely unproven for atmospheric flows, the EnKF has 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. The EnKF and its variants have also shown substantial promise in simple test problems (Houtekamer and Mitchell 1998; Hamill and Snyder 2000; Anderson 2001). This preprint presents preliminary results from our evaluation of the EnKF at convective scales.

2. BACKGROUND

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 $\mathbf{P}^{\mathrm{f}}$, 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.

Of course, attempts to estimate and evolve each of the $n_x(n_x + 1)/2$ distinct entries of $\mathbf{P}^{\mathrm{f}}$ (where n_x is the number of degrees of freedom of the forecast model) are doomed to failure, as present computers do not possess even sufficient storage. The EnKF overcomes this difficulty by representing $\mathbf{P}^{\mathrm{f}}$ and other required covariances in terms of a small ensemble (a few tens or hundreds of members) of forecasts.

More specifically, suppose we have both an ensemble of forecasts $\{\mathbf{x}_i^{\mathrm{f}}, i = 1, \ldots, n_e\}$ valid at some time t, where x represents the state vector of the forecast model and the superscript f denotes a forecast quantity, and a set of observations $\mathbf{y} = \mathbf{H}\mathbf{x}^{\mathrm{t}} + \epsilon$ valid at the same time, where $\mathbf{x}^{\mathrm{t}}$ is the true state and $\epsilon \sim N(0, \mathbf{R})$. Let $\mathbf{X}^{\mathrm{f}}$ be the matrix whose columns are the ensemble deviations (that is, the difference of each member from the ensemble mean, $\mathbf{x}^{\mathrm{f}}$) multiplied by $(n_e - 1)^{-1/2}$. The EnKF then uses the Kalman filter update for the ensemble mean,

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

but replaces the Kalman gain matrix,

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

by

$$\mathbf{K} = \mathbf{X}^{\mathrm{f}}(\mathbf{H}\mathbf{X}^{\mathrm{f}})^{\mathrm{T}}\left((\mathbf{H}\mathbf{X}^{\mathrm{f}})(\mathbf{H}\mathbf{X}^{\mathrm{f}})^{\mathrm{T}} + \mathbf{R}\right)^{-1}. \tag{3}$$

When the observations and the the true state are related nonlinearly by $\mathbf{y} = h(\mathbf{x}^{\mathrm{t}}) + \epsilon$, $\mathbf{H}\mathbf{X}^{\mathrm{f}}$ is replaced by the matrix whose columns are the difference between $h(\mathbf{x}_i^{\mathrm{f}})$ and the ensemble mean.

In addition to updating the ensemble mean, it is also necessary to update each ensemble member given the observations, so that the sample covariance based on the ensemble approximates the analysis error covariance. Here we use the simplest method in which each member is updated as in (1) and (3) using the observations $\mathbf{y}$ contaminated by a realization of the observational error ϵ (Houtekamer and Mitchell 1998).

The accuracy of the update (1) can be increased, and its computational cost reduced, by

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

Corresponding author address: Chris Snyder; NCAR, P.O. Box 3000, Boulder, CO 80307-3000; chriss@ucar.edu.

explicitly accounting for the fact that correlations between elements of the state variable decrease with distance, so that observations at a sufficient distance have little influence on the analysis at any specific model grid point. For the results shown here, we enforce this property by only considering observations within a radius of a few kilometers of a given analysis point; more sophisticated approaches are possible (Houtekamer and Mitchell 2001).

3. RESULTS

Our experiments use the anelastic model of Sun and Crook (1997). This model is used both to produce a reference simulation of a supercell (that is, $\mathbf{x}^t$) and to assimilate simulated observations taken from the reference solution. Thus, these initial experiments assume a perfect forecast model. The reference simulation of a supercell is the same as that used in Sun and Crook (2001, in this preprint volume); further details may be found there.

3.1 *Ensemble evolution*

Since the size of the state vector is large compared to the likely number of ensemble members ($n_x \gg n_e$), the skeptical reader will remark that the sampling error in (3) could be prohibitively large. This is certainly true if the errors in different state variables are uncorrelated, but it need not be true in the opposite case that the forecast errors are highly correlated. In nonlinear dynamical systems (such as the cloud model considered here), small perturbations typically grow in only a few directions in phase space, while decaying, often rapidly, in most others; thus, errors that are initially uncorrelated will develop correlations during a forecast. The feasibility of the EnKF (with a limited ensemble) depends on this behavior and its time scale.

To evaluate the time scale over which the dynamics alters small errors, we have perturbed the reference solution at t = 40 min with 50 realizations of Gaussian noise in velocity and temperature. This noise is independent for each grid point and each variable, with 1 ms^{-1} and 1 K standard deviations for winds and temperature, respectively. In the following, we will denote error variables by primes.

Figure 1 shows x-z cross sections after 10 min of the variance for v', the correlation of u' with $v'(x_0, z_0)$ [where (x_0, z_0) is the location of max(Var(v'))], and the correlation of liquid-water potential temperature θ_l' with $v'(x_0, z_0)$; all of these are calculated as sample quantities, that is, by averaging over the 50 realizations of the error. Note that, while the initial fields of variance for the errors are uniform (up to variations produced by sampling error), the variance of v' after 10 min of forecast exhibits substantial spatial variation, with maxima associated with the updraft of the supercell. Similarly, there are also significant covariances among variables after 10 min even though those covariances are zero initially (again, to within sampling error). Thus, the dynamical evolution of the error fields introduces structure into the error statistics on a time scale of 10 minutes, which is comparable to that for the evolution of the reference simulation.

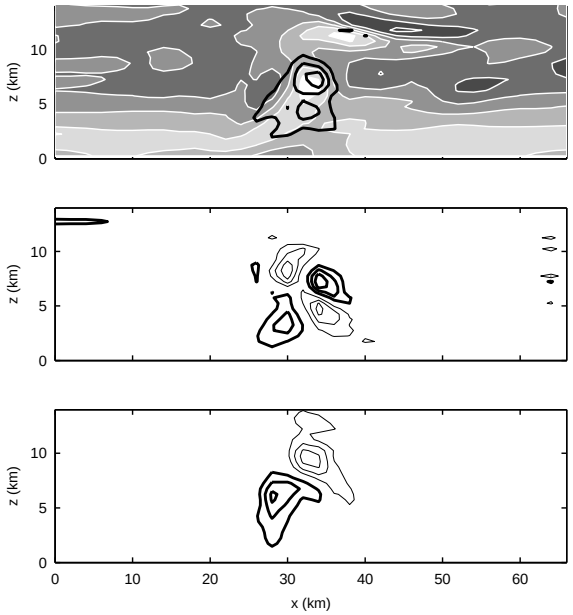

Figure 1. Cross sections through the center of the supercell of (left) the sample variance for v' (upper, contour interval 3 $\mathrm{m}^2\mathrm{s}^{-2}$); the covariance of u' with $v'(x_0, z_0)$, where (x_0, z_0) is the location of max(Var(v')) (middle, contour interval 1.5 $\mathrm{m}^{-2}\mathrm{s}^{-2}$; and the covariance of liquid-water potential temperature θ_l' with $v'(x_0, z_0)$ (lower, contour interval 1 mKs^{-1}). Positive contours are shown as thick lines, negative as thin. Also shown shaded in the upper panel is v from the reference simulation; the contour interval 2.5 ms^{-1}, with lowest contour value (and lightest color) corresponding to $-5\,\mathrm{ms}^{-1}$.

3.2 *Assimilation of simulated observations*

We have performed initial tests of the EnKF using a 50-member ensemble and simulated observations of each of u, v, w, and θ_l at one of every 8 gridpoints. Observational errors are independent for each variable and at each location with standard deviations of 1 ms^{-1} for wind and 1 K for θ_l, and are available every 5 minutes. The assimilation begins at t = 40 min of the reference simulation and uses a prior estimate at that time that consists of the reference solution perturbed with a

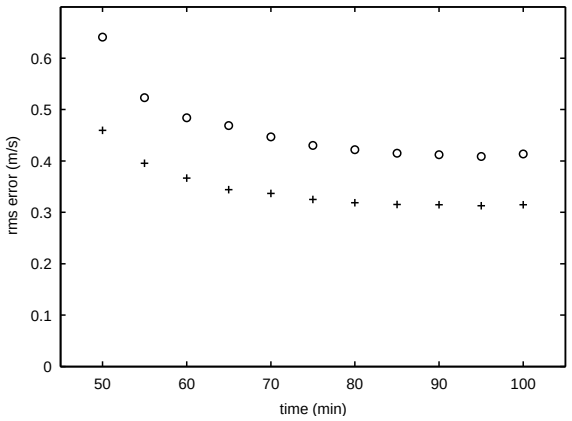

Figure 2. The r.m.s. wind error of the ensemble mean in the background forecast (open circles) and the analysis (plus signs) from the EnKF.

realization of the observational error and ensemble members that in turn differ from that mean by independent realizations of the observational error.

The EnKF employs a simple, serial scheme to compute the update (3); the scheme processes observations one by one, updating each member after each observation and requiring only the inversion of a scalar in (3). More computationally efficient implementations are possible.

The r.m.s. error of the ensemble mean decreases steadily over 10 cycles of the EnKF; the wind error for both the analysis and the background forecast are shown in Fig. 2. This initial test indicates that the filter is at least stable and well behaved for an ensemble of 50 members. Results of further (and more realistic) tests, such as using observations of radial wind alone, will be presented at the conference.

REFERENCES

Anderson, J. L., 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.*, accepted.

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

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

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Shapiro, A., S. Ellis and J. Shaw, 1995: Single-Doppler velocity retrievals with Phoenix II data: Clear air and microburst wind retrievals in the planetary boundary layer. *J. Atmos. Sci.*, **52**, 1265–1287.

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.