6.2

DISTANCE-DEPENDENT FILTERING OF BACKGROUND ERROR COVARIANCE ESTIMATES IN AN ENSEMBLE KALMAN FILTER

Thomas M. Hamill and Jeffrey S. Whitaker

NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado

Chris Snyder

National Center for Atmospheric Research, Boulder, Colorado*

1. INTRODUCTION

Many groups are experimenting with data assimilation schemes for complex numerical weather and oceanographic prediction models where background forecast error covariances are estimated using an ensemble. Much of this experimentation is based on approach is known as the ensemble Kalman filter, or "EnKF" (Evensen 1994, Evensen and van Leeuwen 1996, Houtekamer and Mitchell 1998, 2001, Hamill and Snyder 2000, etc.). The EnKF uses a specifically constructed ensemble of parallel short-term forecasts and data assimilation cycles. Statistics derived from the short-term forecasts are used to estimate background error covariances during the subsequent data assimilation step. Ensemble members are updated to the new observations in parallel data assimilation cycles.

Houtekamer and Mitchell (1998) noted that the EnKF analysis could be improved by excluding observations greatly distant from the grid point being analyzed. They concluded that this was because background error covariance estimates generated from a small ensemble often produced spuriously large magnitude background error covariance estimates between greatly separated grid points; estimates from a larger ensemble showed that the true covariances were generally small. These large covariances resulted in unduly large corrections to the analyses far from the observation location. Hence, the analyses were more accurate when the observations were excluded than when they were included and assimilated with degraded background error statistics. Houtekamer and Mitchell (2000) have since experimented with filtering covariance estimates produced by the ensemble using a "Schur product," whereby the ensemble-based covariance estimates are multiplied element by element with a distance-dependent correlation function that varies from 1.0 at the observation location to 0.0 at some prespecified radial distance. They have found that the analysis errors are substantially improved when the Schur product with a correlation function is incorporated. The analyses also had other desirable properties such as improved smoothness.

This research continues the exploration into the distance-dependent filtering of covariance estimates generated by a finite ensemble. Our goal is to understand why such filtering may be beneficial, how much improvement may be expected from filtering, and how this may change with the observational data density and the size of the ensemble. An extended version of this preprint has been accepted by *Mon. Wea. Rev.*, (Hamill et al. 2001). Contact the author for a copy if you are interested in details not described here.

2. SIMPLE PROPERTIES OF COVARIANCE MATRICES FROM RANDOM SAMPLES

Let us start by trying to understand one of the most basic effects that an error in the specification of covariances in the background error will have on data assimilation. Namely, we consider how an error in the covariance will affect the analysis at a grid point away from the observation location. We consider the simplest system possible, a 2-dimensional model state with a single observation. We will use the nomenclature of Bayesian statistics; for example, in this section, capital letters will denote continuous random variables, and lowercase letters the actual values. Assume we have a random vector $\mathbf{X}^{\mathrm{T}} = (X_1^{\mathrm{T}}, X_2^{\mathrm{T}})$ representing the unknown true state of the model. We have a sample forecast $\mathbf{x}^{\mathrm{b}} = (x_1^{\mathrm{b}}, x_2^{\mathrm{b}})$, denoting the background, or "first guess" forecast sample of the true state, with background error covariance matrix $\mathbf{P}^{\mathrm{b}}$ defined by

$$\mathbf{P}^{\mathrm{b}} = \begin{pmatrix} \sigma_1^2 & c_{12} \\ c_{12} & \sigma_2^2 \end{pmatrix}. \qquad (1)$$

Thus, in the absence of new observations, we have a prior probability distribution $\pi(\mathbf{X}^{\mathrm{T}}) \sim N(\mathbf{x}^{\mathrm{b}}, \mathbf{P}^{\mathrm{b}})$, where $N \sim (\mathbf{a}, \mathbf{B})$ indicates the distribution is normal with expected value $\mathbf{a}$ and variance/covariance $\mathbf{B}$. Assume a new observation then becomes available. Y is a scalar random variable denoting the observation, and

---

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

Corresponding author address: Dr. Thomas M. Hamill, NOAA-CIRES CDC R/CDC 1, 325 Broadway, Boulder, CO 80303-3328. hamill@cdc.noaa.gov ; (303) 497-3060 ; telefax (303) 497-7013

the actual observation is y, taken at the location of the first component of the state vector. Errors ϵ_0 for the observation are defined by $\epsilon_0 \sim N(0, \sigma_0^2)$.

We seek the posterior probability distribution for the analyzed state conditional on (updated to) the new observation, $\pi(\mathbf{X}^{\mathrm{T}} = \mathbf{x}|Y = y) = \mathbf{X}^{\mathrm{a}} = (\mathrm{X}_1^{\mathrm{a}}, \mathrm{X}_2^{\mathrm{a}})$. It can be shown that $\mathbf{X}^{\mathrm{a}} \sim N(\mathbf{x}^{\mathrm{a}}, \mathbf{P}^{\mathrm{a}})$, where $\mathbf{x}^{\mathrm{a}} = (\mathrm{x}_1^{\mathrm{a}}, \mathrm{x}_2^{\mathrm{a}})$ and

$$\mathbf{P}^{\mathrm{a}} = \begin{pmatrix} Var(\mathrm{X}_1^{\mathrm{a}}) & Cov(\mathrm{X}_1^{\mathrm{a}}, \mathrm{X}_2^{\mathrm{a}}) \\ Cov(\mathrm{X}_1^{\mathrm{a}}, \mathrm{X}_2^{\mathrm{a}}) & Var(\mathrm{X}_2^{\mathrm{a}}) \end{pmatrix}. \qquad (2)$$

Thus, the analyzed values and the expected analysis error variance obtained by updating the background are

$$\mathrm{x}_1^{\mathrm{a}} = \mathrm{x}_1^{\mathrm{b}} + \frac{\sigma_1^2}{\sigma_0^2 + \sigma_1^2}\left(\mathrm{y} - \mathrm{x}_1^{\mathrm{b}}\right), \quad Var(\mathrm{X}_1^{\mathrm{a}}) = \sigma_1^2\left(1 - \frac{\sigma_1^2}{\sigma_0^2 + \sigma_1^2}\right)$$

$$\mathrm{x}_2^{\mathrm{a}} = \mathrm{x}_2^{\mathrm{b}} + \frac{\mathrm{c}_{12}}{\sigma_0^2 + \sigma_1^2}\left(\mathrm{y} - \mathrm{x}_1^{\mathrm{b}}\right), \quad Var(\mathrm{X}_2^{\mathrm{a}}) = \sigma_2^2 - \frac{\mathrm{c}_{12}^2}{\sigma_0^2 + \sigma_1^2}.$$

$$Cov(\mathrm{X}_1^{\mathrm{a}}, \mathrm{X}_2^{\mathrm{a}}) = \frac{\mathrm{c}_{12}\sigma_0^2}{\sigma_0^2 + \sigma_1^2} \qquad (3)$$

Now, suppose we have an inaccurate estimate $\hat{\mathbf{P}}^{\mathrm{b}}$ of the covariance matrix $\mathbf{P}^{\mathrm{b}}$, where variances are correctly specified but the covariance has an error, or "noise" $\epsilon_c \sim N(0, \tau_{12})$:

$$\hat{\mathbf{P}}^{\mathrm{b}} = \begin{pmatrix} \sigma_1^2 & \mathrm{c}_{12} + \epsilon_c \\ \mathrm{c}_{12} + \epsilon_c & \sigma_2^2 \end{pmatrix} \qquad (4)$$

We seek to understand the effect on the quality of the analysis for $\mathrm{x}_2^{\mathrm{a}}$. If the error ϵ_c is uncorrelated with errors in y, $\mathrm{x}_1^{\mathrm{b}}$, and $\mathrm{x}_2^{\mathrm{b}}$, it can be shown that

$$Var(\mathrm{X}_2^{\mathrm{a}}) = \sigma_2^2 - \frac{\mathrm{c}_{12}^2}{\sigma_0^2 + \sigma_1^2}\left[1 - \left(\frac{\tau_{12}}{\mathrm{c}_{12}}\right)^2\right]. \qquad (5)$$

Let us denote $\frac{\tau_{12}}{\mathrm{c}_{12}}$ the "relative error" in the covariance, a measure of noise relative to signal. Notice that when the relative error is greater than one, the analysis of $\mathrm{x}_2^{\mathrm{a}}$ is typically *degraded* by assimilating the observation y. Notice also that the amount of improvement or degradation will be proportional to the square of the covariance c_{12}. That is, for a given relative error > 1.0, the degradation will be worse for larger covariances.

Given that large relative errors in the magnitude of background error covariances may degrade the analysis, we shift focus to understand what can cause such errors when they are estimated from an ensemble. We examine this question through some simple experiments with 2 × 2 sample covariance matrices. Again, assume we have an ensemble of vector background values $\mathbf{x}^{\mathrm{b}} = (\mathrm{x}_1^{\mathrm{b}}, \mathrm{x}_2^{\mathrm{b}})$ sampled from $\mathbf{X}^{\mathrm{T}}$. Here $\mathrm{x}_1^{\mathrm{b}}$ represents the value at the observation location, and $\mathrm{x}_2^{\mathrm{b}}$ is the value at some distance from the observation.

It can be shown that given the true covariance matrix $\mathbf{P}^{\mathrm{b}}$ with variances $\sigma_1^2 = 1$, $\sigma_2^2 = \eta^2$, true correlation $\rho = \mathrm{Corr}(\mathbf{X}_1^{\mathrm{T}}, \mathbf{X}_2^{\mathrm{T}})$, (and hence true covariance $\mathrm{c}_{12} = \rho\sigma_1\sigma_2 = \rho\eta$), then the variance τ_{12} of the error in the calculation of the covariance from a sample ensemble of $\mathbf{x}^{\mathrm{b}}$ is approximately

$$Var(\epsilon_c) = \tau_{12} \simeq \frac{1}{n}(1 + \rho^2)\eta^2 \qquad (6)$$

for large enough sample sizes n.

How do errors change as the true correlation and the ensemble size changes? Figure 1 shows the corresponding relative error of the covariance, τ_{12}/c_{12}. Relative error increases greatly as ρ decreases and as the ensemble size decreases. Since ρ typically decreases with increasing distance from the observation, in a numerical weather prediction model, the noise-to-signal ratio would thus be expected to typically increase with increasing distance from the observation (this is, on average, the case; sometimes there may be large magnitude true correlations over long distances. Section 4 provides some evidence that this is quite uncommon, however.) We conducted two sets of tests, a set of single observation experiments designed to illuminate the characteristics of signal and noise in the ensemble, and a test of analysis accuracy for different observational networks, different sized ensembles, and different filter characteristics. For both, a 90-day set of analyses were computed, updated with new observations every 12 h.

3. DESIGN OF THE EXPERIMENT

a. Forecast Model

Results in the rest of the paper will be based on a dry two-layer PE model, described in Zou et al.

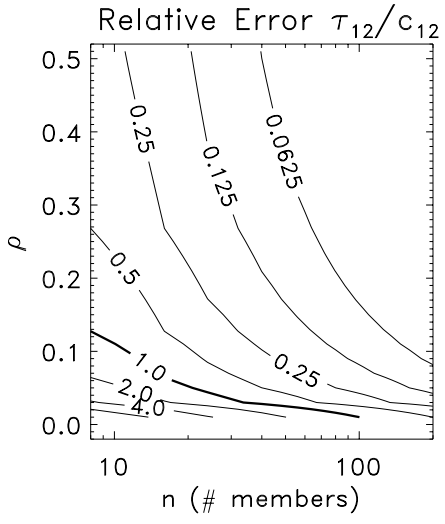

Figure 1. (a) Relative errors (sample error variance divided by the magnitude of the true covariance)as a function of correlation ρ and ensemble size.

(1993). The model state vector consists of vorticity and divergence spectra at two levels as well as Exner function π at the lower surface and at an interface. The model is spectral with a T31 triangular truncation. There is a simple, wavenumber 2 terrain, but there are no land/water interfaces. A fourth-order Runge-Kutta scheme is used for the numerical integration, there is ∇^8 diffusion, and the model is forced by damping the interface π toward an equilibrium state.

b. Observations

Two observational networks with approximately uniform data density over the globe were tested, one with 46 and another with 126 raobs, with the raobs located on a geodesic grid (see map with locations in Hamill et al. 2001). We observed u and v components of the wind at both model levels and Exner function π at the lower surface and interface. Wind component error variances were assumed to be $(3.0 \cos\phi)^2$ m^2 s^{-2}. Lower boundary π variances are assumed to be 0.09 J^2 kg^{-2} K^{-2}, or about 1 hPa^2 pressure error variance. Interface π variances were set to 9.0 J^2 kg^{-2} K^{-2}, which corresponds to about 1 K^2 temperature error variance. These same observation-error covariances were used both to generate random observation errors and were those assumed by the data assimilation scheme. Observations and new analyses were generated every 12 h, followed by a 12-h forecast with the PE model that served as background at the next analysis time.

c. Ensemble Kalman filter data assimilation system

The EnKF presupposes an ensemble of background states are available to generate background covariance estimates. We started with an ensemble of n analyses at some time t_0 generated in the manner described in Hamill and Snyder (2000). These perturbed analyses were generated by adding random spatially correlated noise to the truth analysis. We then repeated the following three-step process for each data assimilation cycle: (1) Make n forecasts to the next analysis time, here, 12 h hence. These forecasts will be used as background fields for n parallel analyses. (2) Given the already imperfect observations at this next analysis time (hereafter called the "control" observations), generate $i = 1, \ldots, n$ independent sets of *perturbed* observations $\mathbf{y}_i^{\mathrm{o}}$ by adding random noise to the control observations $\mathbf{y}^{\mathrm{o}}$. The noise is drawn from the same distributions as the observation errors (see section 3b), and the noise is constructed to ensure that the mean of the perturbed observations is equal to the control observation. (3) Perform n objective analyses, updating each of the n background forecasts using the associated set of perturbed observations. The analysis equation for the ith member is

$$\mathbf{x}_i^{\mathrm{a}} = \mathbf{x}_i^{\mathrm{b}} + \hat{\mathbf{P}}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}\left[\mathbf{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right]^{-1}\left(\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}\mathbf{x}_i^{\mathrm{b}}\right). \qquad (7)$$

$\mathbf{x}_i^{\mathrm{b}}$ is the m-dimensional model state vector for the ith member background forecast of an n-member ensemble, comprised of gridded u and v wind components at the two model levels as well as lower surface and interface π. $\mathbf{x}_i^{\mathrm{a}}$ is the subsequently analyzed state for the ith member. $\mathbf{H}$ (here assumed linear) is an operator that converts the model state to the observation type and location. $\mathbf{R}$ is the $n_o \times n_o$ measurement error covariance matrix. $\hat{\mathbf{P}}^{\mathrm{b}}$ is now an approximation of the background error covariances generated from the collection of background forecasts. In its most simple form in the EnKF, $\hat{\mathbf{P}}^{\mathrm{b}}$ is approximated by

$$\hat{\mathbf{P}}^{\mathrm{b}} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{x}_i^{\mathrm{b}} - \overline{\mathbf{x}^{\mathrm{b}}}\right)\left(\mathbf{x}_i^{\mathrm{b}} - \overline{\mathbf{x}^{\mathrm{b}}}\right)^{\mathrm{T}}, \qquad (8)$$

where $\overline{\mathbf{x}^{\mathrm{b}}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{\mathrm{b}}$ is the ensemble mean state.

Here, as in Evensen (1994) and Houtekamer and Mitchell (1998, 2001), $\mathbf{P}^{\mathrm{b}}$ is not computed explicitly by itself but as the product $\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}$, and as in Houtekamer and Mitchell (2001), a covariance localization is applied. Define

$$\overline{\mathbf{H}\mathbf{x}^{\mathrm{b}}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{H}\mathbf{x}_i^{\mathrm{b}},$$

which represents the mean of ensemble states converted to the observation variable's type and location, respectively. Then

$$\hat{\mathbf{P}}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} = \rho_S \circ \left[\frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{x}_i^{\mathrm{b}} - \overline{\mathbf{x}^{\mathrm{b}}}\right)\left(\mathbf{H}\mathbf{x}_i^{\mathrm{b}} - \overline{\mathbf{H}\mathbf{x}^{\mathrm{b}}}\right)^{\mathrm{T}}\right]. \qquad (9)$$

The operation $\rho_S \circ$ in (9) denotes a Schur product (an element-by-element multiplication) of a correlation matrix $\mathbf{S}$ with the covariance model generated by the ensemble. The Schur product of matrices $\mathbf{A}$ and $\mathbf{B}$ is a matrix $\mathbf{C}$ of the same dimension, where $C_{ij} = A_{ij}B_{ij}$. For sequential data assimilation, the function $\mathbf{S}$ depends upon the observation location; it is a maximum of 1.0 at the observation location and typically decreases montonically to zero at some finite distance from the observation.

To define the correlation matrix $\mathbf{S}$, we used a 5th-order function of Gaspari and Cohn (1999), which is similar in shape to the Gaussian function, but tapers to zero at a finite distance. The controlling parameter is l_c, a correlation length scale for the function. See Hamill et al. (2001) for more description.

4. SIGNAL AND NOISE ESTIMATES FROM ENSEMBLES

Hamill et al. (2001) shows how a very large ensemble (n=400 members) can be used to estimate the characteristics of signals and noise in background error covariances estimated from a much smaller (n=25) ensemble. Generally, it is shown that the ratio of noise to signal in the small ensemble gets larger the further from the observation, i.e., as the true covariance progressively decreases, the noise in the covariance estimate tends

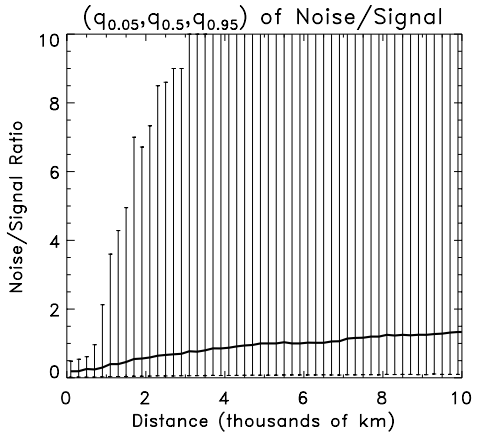

Figure 2. 5th, 50th (solid line) and 95th percentiles of distribution of noise/signal ratio as a function of distance from the observation.

to overwhelm the signal at finite distance from the observation (around 4000 km, in this experiment) (Fig. 2). This is in accordance with the theory outlined in section 2. See Hamill et al. (2001) for more details.

5. ANALYSIS ERRORS WITH FILTERED COVARIANCES

We applied the 5th-order function in Gaspari and Cohn (1999) as discussed in section 3c to an EnKF with 25, 100, and 400 members. Forecasts and analyses were cycled for 90 days, with updates every 12 h. We shall examine the analysis error characteristics of interface π averaged over the last 60 days of the integration.

a. Analysis errors as function of filter length scale

Figures 3 a-b present the time-average ensemble mean error for the sparse network (46 observation locations; Fig. 3a) and the denser network (126 observation locations; Fig. 3b). To the right of the dots plotted for a given correlation length scale in Fig. 3, filter divergence occurred for tested larger length scales and the analyses were useless. Figure 3 suggests some interesting characteristics of the EnKF coupled with the localization of covariances. First, as expected, the analyses were significantly improved by using more observations. Note that the optimal length scale is a function of the size of the ensemble. Smaller ensembles had a smaller optimal length scale than for larger ensembles, indicating that noise in the covariance estimates overwhelms signal at relatively short distances from the observations when the ensemble size is small, but for larger ensembles, noise doesn't overwelm signal until much further from the observation. This is similar to a result Houtekamer and Mitchell (1998) found using a cutoff radius to eliminate observations.

We also generated rank histograms (Hamill 2001 and references therein) as a way of measuring the reliability of the ensemble. Ideally, a sample of forecast

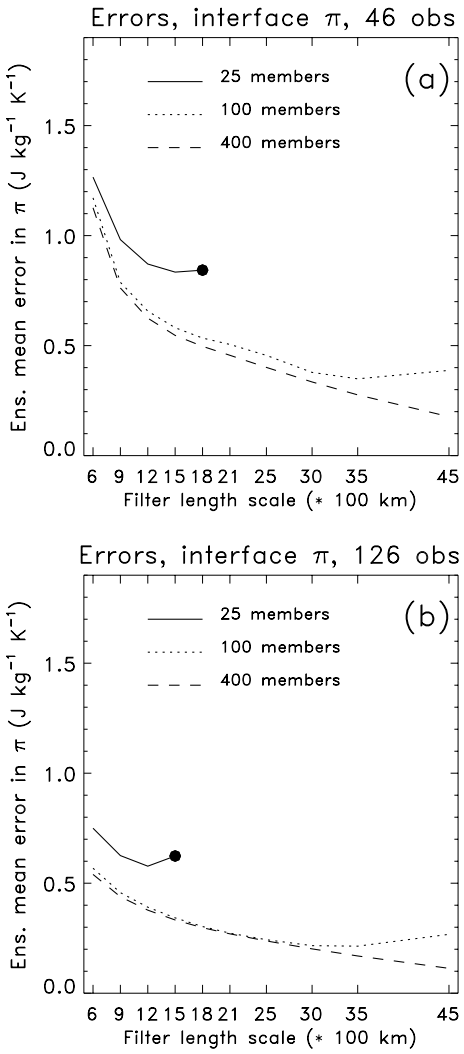

Figure 3. (a) Time averaged ensemble mean error in interface π for 46-observation network as function of correlation length scale of the filter. (b) As in (a), but for the 126-observation network.

values from the ensemble and the true state ought to be able to be considered random samples from the same probability distribution. If this is true, then when the rank of the true state is compared to an n-member ensemble sorted from lowest to highest, the rank of the true state should be equally likely to occur in any of the $n + 1$ possible ranks. A histogram of the rank of the truth tallied over many points provides evidence of the reliability of the ensemble. A U-shaped rank histogram (excessive population at the lowest and highest ranks) indicates insufficient spread or bias in the ensemble. An excess population at the middle ranks indicates too much spread.

Figures 4 a-b show rank histograms for the 46- and 126-observation networks, respectively. Rank histograms for the 100- and 400-member ensembles were generated by taking a subset of 25 of the members,

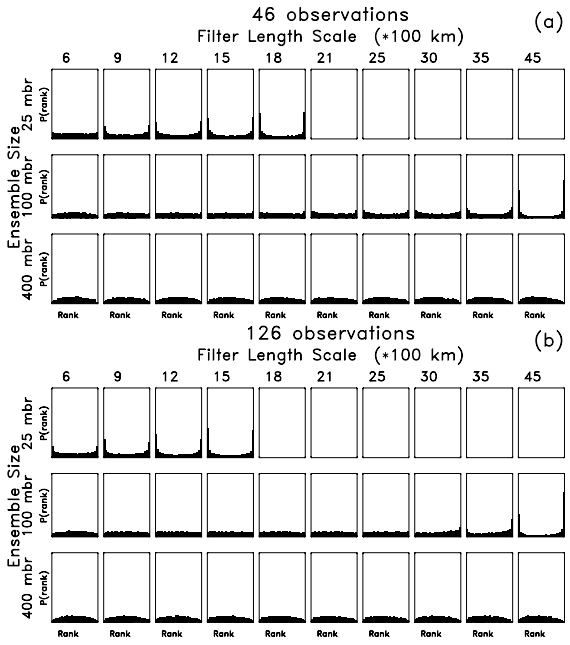

Figure 4. (a) Rank histograms for the 46-observation network as a function of the ensemble size and the filter correlation length. Where rank histograms are not plotted, filter divergence occurred, (b) As in (a), but for the 126-observation ensemble.

so that comparisons with the 25-member ensemble could be facilitated. For the 25-member ensembles, at all but the shortest tested length scale, rank histograms are consistently overpopulated at the extreme ranks. This result suggests that the small ensemble may not be able to correctly specify error variances over the full range of growing directions in the ensemble. For the 25- and 100-member ensembles, there is a trend toward more population at the extreme ranks as the filter length scale increases. This change from underpopulation to overpopulation as filter length increases is a primarily reflection of differing amounts of variance reduction associated with different filter lengths. With strong filtering (a short l_c), only grid points very near the observations are adjusted during the assimilation, and at the rest, the original variance in the background is preserved in the ensemble of analyses and propagated forward to the next cycle. Thus, when filter length is shorter than appropriate for a given sized ensemble, the background covariances estimated from the ensemble are reduced too much in magnitude, undercorrecting the analysis far away from observation locations.

As the length scale of the filter increases, the more the background error covariances estimated from the ensemble are trusted; hence more and bigger corrections to the analysis are possible far from the observation location. If the covariances are very noisy, though, as shown before, the corrections are inappropriate, and the result is an overly adjusted, variance-deficient ensemble. In the extreme, for very long correlation lengths, this can induce filter divergence. This can be noticed in the the rank histograms, which become increasingly U-shaped as correlation length is increased.

6. CONCLUSIONS

Ensemble-based data assimilation approaches show much potential for improving the quality of initial conditions. The essence of ensemble approaches lies in being able to effectively weight observations and background forecasts using the ensemble to define the background error statistics. We have shown here that these statistics are subject to errors that can be improved through the application of a distance-dependent filtering algorithm.

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.

—— , and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85-96.

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

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

—— , J. S. Whitaker, and —— , 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, accepted.

—— , 2000: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550-560.

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.

Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416-433.

Zou, X., A. Barcilon, I. M. Navon, J. Whitaker, and D. G. Cacuci, 1993: An adjoint sensitivity study of blocking in a two-layer isentropic model. *Mon. Wea. Rev.*, **121**, 2833-2857.