P 2.2

ENSEMBLE BASED DATA ASSIMILATION OF OBSERVATIONS OF HURRICANE HUMBERTO

Brian J. Etherton[1] and Sim D. Aberson[2]

1-University of North Carolina at Charlotte, 2-NOAA/AOML/HRD

1. INTRODUCTION

During 2001, NOAA and NASA conducted the joint Hurricanes at Landfall (HAL) and CAMEX-IV. The most comprehensive datasets obtained were during two consecutive days during Hurricane Humberto (Feuer et al., 2002). During these two sets of missions, centered on nominal times of 0000 UTC 24 and 25 September, operational synoptic surveillance missions were conducted with the G-IV. The dropwindsonde data gathered during these missions led to substantial track forecast improvements in both of the leading operational dynamical forecast models (AVN and GFDL) at the nominal times of the missions. In addition to improving forecasts from these operational models, the data from the surveillance missions provided an opportunity to test an ensemble-based Kalman filter data assimilation technique.

The Ensemble Transform Kalman Filter (ET KF) (Bishop et al., 2001) is currently used at the National Centers for Environmental Prediction (NCEP) to select flight tracks for dropwindsonde equipped weather reconnaissance aircraft in winter storms (Szunyogh et al., 2000, 2002, Majumdar et al. 2002), and is being tested in tropical cyclone surveillance flight track selection. In order to select target sites, the ET KF calculates an estimate of uncertainty of the model forecast. The ET KF estimates error covariance matrices in terms of the outer product of a matrix containing transformed ensemble perturbations. The transformation coefficients are obtained by repeatedly solving the Kalman filter error statistics equations within the ensemble subspace (Bishop et al., 2001).

Corresponding Author Address:
Brian Etherton, Department of Geography and Earth Sciences, The University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC 28223
e-mail: betherto@uncc.edu

Data assimilation, the optimal blending of observations with a model forecast field to produce an analysis of the atmosphere, requires error statistics for both the observational data as well as the model first guess field. The same error statistics used for targeting, which measure the uncertainty of the model forecast, can be applied to data assimilation.

Ensemble-based error statistics have flow dependent characteristics, such as one-point error correlation patterns that stretch out along fronts. In contrast, the time-invariant one-point error correlation patterns used for data assimilation at operational centers such as NCEP (Parrish and Derber, 1992) and (until recently) the European Center for Medium range Weather Forecasting, ECMWF (Courtier et al., 1998, which was superceded by Rabier et al. 2000) and those run by the National Aeronautic and Space Administration, NASA (Cohn et al., 1998) and by the U.S. Navy (Daley and Barker, 2001), are generally horizontally isotropic and temporally invariant. Given that tropical cyclones are not systems with a discrete, isotropic structure, flow dependent error statistics have the potential to perform better than time-invariant, spatially isotropic error statistics.

2. METHODS

The Ensemble Kalman Filter (EnKF) is an approach to data assimilation that uses ensembles to determine the error statistics associated with a model first guess field. Approaches for ensemble-based data assimilation include those presented in Evensen (1994), Evensen and van Leeuwen (1996), Burgers et al. (1998), Houtekamer and Mitchell (1998), Anderson and Anderson (1999), Hamill and Snyder (2000), Keppenne (2000), Anderson(2001), Pham (2001), and Whitaker and Hamill (2002) and Mitchell et al. (2002). The ET KF (Bishop et al., 2001) uses the differences between ensemble members to calculate the correlations of errors as a covariance matrix, and then produces an eigenvector decomposition of the ensemble-based covariance matrix.

As given in Cohn (1997), the minimum analysis error variance increment

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

where y are the observations, $\mathbf{x}^f$ is the first guess field, $\mathbf{x}^a$ is the new analysis, $\mathbf{H}$ is the observation operator translating variables from observation to model space, $\mathbf{R}$ is the observation error covariance matrix providing the error statistics of the observational data, and $\mathbf{P}^f$ is the prediction error covariance matrix providing information about how errors of each analysis variable at each grid point correlate with all others. The innovation vector, ($\mathbf{y}$-$\mathbf{H}\mathbf{x}^f$), is equal to the difference between the observation and the first guess at the observation locations. If observational errors are assumed to be uncorrelated, $\mathbf{R}$ is diagonal, with the diagonal elements equal to the error variance associated with each observation. Equation 1 is solved so that the analysis error variance is minimized.

One approach to creating $\mathbf{P}^f$ commonly used in operational centers in 3D- and 4D-Var schemes is to produce a parameterized, time-invariant matrix in which the impact of each observation decays isotropically away from the observation location. One example is given in Daley (1991, Eq. 4.3.21):

$$f(r) = \exp(\ln[0.1](r/D)^2), \qquad (2)$$

where r is the distance between the analysis and observation locations, and D is a correlation length scale at which the impact of the observation is 0.1 of the impact at the observation location. This correlation function has been used in Bishop et al. (2001) and Etherton and Bishop (2004) to construct covariance matrices for use assimilating data into a simple barotropic vorticity model.

An alternative formulation of the error covariance matrix associated with the model first guess field is an ensemble generated covariance matrix. Using an EnKF, the matrix $\mathbf{P}^f$ is represented as the outer product of a matrix $\mathbf{Z}$, where each column of $\mathbf{Z}$ is a perturbation from the ensemble mean:

$$\mathbf{P}^f = \mathbf{Z}\mathbf{Z}^T \qquad (3)$$

A third alternative for the prediction error covariance matrix is a hybrid scheme that approximates the forecast error covariance matrix $\mathbf{P}^f$ with a mix of parameterized covariances, $\mathbf{B}^f$, such as given in Eq. 2, and flow dependent, ensemble-based covariances, $\mathbf{F}^f$, such as that in Eq. 3. The forecast error covariance matrix $\mathbf{P}^f$ from Hamill and Snyder (2000) is given by:

$$\mathbf{P}^f = (1-\alpha)\mathbf{F}^f + \alpha\mathbf{B}^f, \qquad (4)$$

With α equal to 1, $\mathbf{P}^f$ is $\mathbf{B}^f$, the 3D-Var correlation matrix. With α between zero and one, $\mathbf{P}^f$ is a mixture of flow dependent and time invariant error statistics. For a further discussion of ensemble-based versus isotropic covariance matrices, see Bishop et al. (2001).

The first guess field $\mathbf{x}^f$, (equation 1) was the VICBAR ensemble mean forecast at the nominal observation time at 1 degree latitude/longitude resolution within a domain spanning from 5 to 45 degrees north, 105 to 45 degrees west. The value of D (equation 2) is set so that the distance at which correlations decay to 0.1 of the value at the observing site is 5 degrees latitude. To produce an increment to the first guess field, covariance matrices $\mathbf{P}^f$ and $\mathbf{R}$ were computed. Since 0.5 ms^{-1} is the expected error of a dropwindsonde wind observation (Hock and Franklin 1999) the diagonal elements of R were set to 0.25 m^2s^{-2}. The average error of the first guess field was assumed to be 2 ms^{-1}, so the diagonal elements of $\mathbf{B}^f$ for the 3D-Var technique were set to 4 m^2s^{-2}. $\mathbf{F}^f$ was calculated taking the product of the matrix of VICBAR ensemble perturbations as per equation 3. The traces of the covariance matrices $\mathbf{B}^f$ and F^f are re-scaled using equation 5. Increments to the ensemble mean were calculated for alpha values of 0.0, 0.5, and 1.0, a 3D-Var, hybrid, and ET KF increment. Adding the ensemble mean to each of these increments produced 3 distinct initial conditions for the VICBAR model. These fields were truncated to a 2.5 degree resolution.

3. RESULTS

The assimilation of the dropwindsonde data at nominal time 0000 UTC 24 September illustrates the differences between ensemble-based and conventional data assimilation. Figure 1a shows the ensemble mean 12-h forecast VICBAR DLM wind field valid at the nominal time. This wind field serves as the first guess state of the atmosphere for the data assimilation. Figure 1b shows the vector difference between the DLM dropwindsonde observation and the first guess field at the observation sites. The largest differences, or innovations, are to the north and east of Humberto. Figures 1c and 1d show that the largest increments to the first guess are also in this region. However, the use of the different error statistics results in greatly different increments.

A circulation center in the subtropical ridge is evident in the first guess field (Fig. 1a) near 32ºN 59ºW. The 3D-Var data assimilation does not significantly modify this feature since it is far from observation locations (Figs. 1c, 1d). The ensemble-based scheme weakens this feature and moves it eastward. Other important differences are seen in the anticyclonic/cyclonic pair to the southwest of Humberto, and over Florida. These features combined to provide an improved VICBAR forecast than in the more conventional data assimilation (Fig. 2).

4. CONCLUSIONS

The data from the surveillance missions provided an opportunity to test an ensemble-based Kalman filter data assimilation technique in a barotropic hurricane track forecast model, VICBAR, using a 41-member modified bred mode ensemble forecasting system. An isotropic assimilation scheme such as 3D-Var is only able to impact the initial conditions in a region surrounding each data point. However, the ETKF is able to make modifications based upon the unique “flow of the day” such that meaningful initial increments to the first guess field are created. This has led to further improvements to track forecasts in the model than possible with currently operational data assimilation schemes.

Clearly, further tests of both these targeting and data assimilation schemes must be made before any meaningful and statistically significant results may be found. Such tests are currently being conducted, and may potentially be used in the upcoming NOAA/NASA experiment in the Eastern Pacific basin. The current results suggest that the ETKF has promise as a targeting and data assimilation technique for tropical cyclones.

5. REFERENCES

Anderson, J. L., 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.,* **129**, 2884–2903.

Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Mon. Wea. Rev.*, **127**, 2741-2758.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Mon. Wea. Rev*., **129**, 420-436.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis Scheme in the Ensemble Kalman Filter. *Mon. Wea. Rev*., **126**, 1719-1724.

Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan.,* **75**, 257–288.

Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, C. Lamich, 1998: Assessing the Effects of Data Selection with the DAO Physical-Space Statistical Analysis System. *Mon. Wea. Rev.,* **126**, 2913-2926.

Courtier, P, E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier and M. Fisher, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation, **124**, 1783-1808.

Daley, R. and E. Barker, 2001: NAVDAS: Formulation and Diagnostics. *Mon. Wea. Rev.*, **129**, 869-883.

Dee, D. P., 1995: On-line Estimation of Error Covariance Parameters for Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **123**, 1128-1145.

Etherton, B. J., and C. H. Bishop, 2004: The resilience of hybrid ensemble/3D-Var analysis schemes to model error and ensemble covariance error. *Mon. Wea. Rev.,* **132**, 1065-1080.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99**, C510143–10162.

Evensen, G., 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.

Feuer, S. E., J. F. Gamache, M. L. Black, and F. D. Marks, 2002: A multiple aircraft experiment in Hurricane Humberto 2001). Part I: Wind Fields. *Preprints*, American Meteorological Society, 25th Conf. Hurr. Trop. Met., 206-207.

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

Hock, T. F., and J. L. Franklin, 1999: The NCAR GPS dropwindsonde. *Bull. Amer. Meteor. Soc*., **80**, 407-420.

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

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **129**, 123–137.

Keppenne, C. L., 2000: Data Assimilation into a Primitive-Equation Model with a Parallel Ensemble Kalman Filter. *Mon. Wea. Rev*., **128**, 1971-1981.

Majumdar, S. J., C. H. Bishop, B. J. Etherton, and Z. Toth, 2002: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part II: Field Program Implementation. *Mon. Wea. Rev*., **130**, 1356-1369.

Mitchell, H. L., P. L. Houtekamer, and G. Pelerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.,* **130**, 2791–2808.

Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.,* **129**, 1194–1207.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Mon. Wea. Rev.*, **120**, 1747-1763.

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.,* **126**, 1143–1170.

Szunyogh, I., Z. Toth, A. V. Zimin, S. J. Majumdar and A. Persson, 2002: Propagation of the Effect of Targeted Observations: The 2000 Winter Storm Reconnaissance Program. *Mon. Wea. Rev*., **130**, 1144-1165.

Szunyogh, I., Z. Toth, R. E. Morss, S. J. Majumdar, B. J. Etherton, and C. H. Bishop, 2000: The Effect of Targeted Dropsonde Observations during the 1999 Winter Storm Reconnaissance Program. *Mon. Wea. Rev*., **128**, 3520-3537.

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

Figure 1:

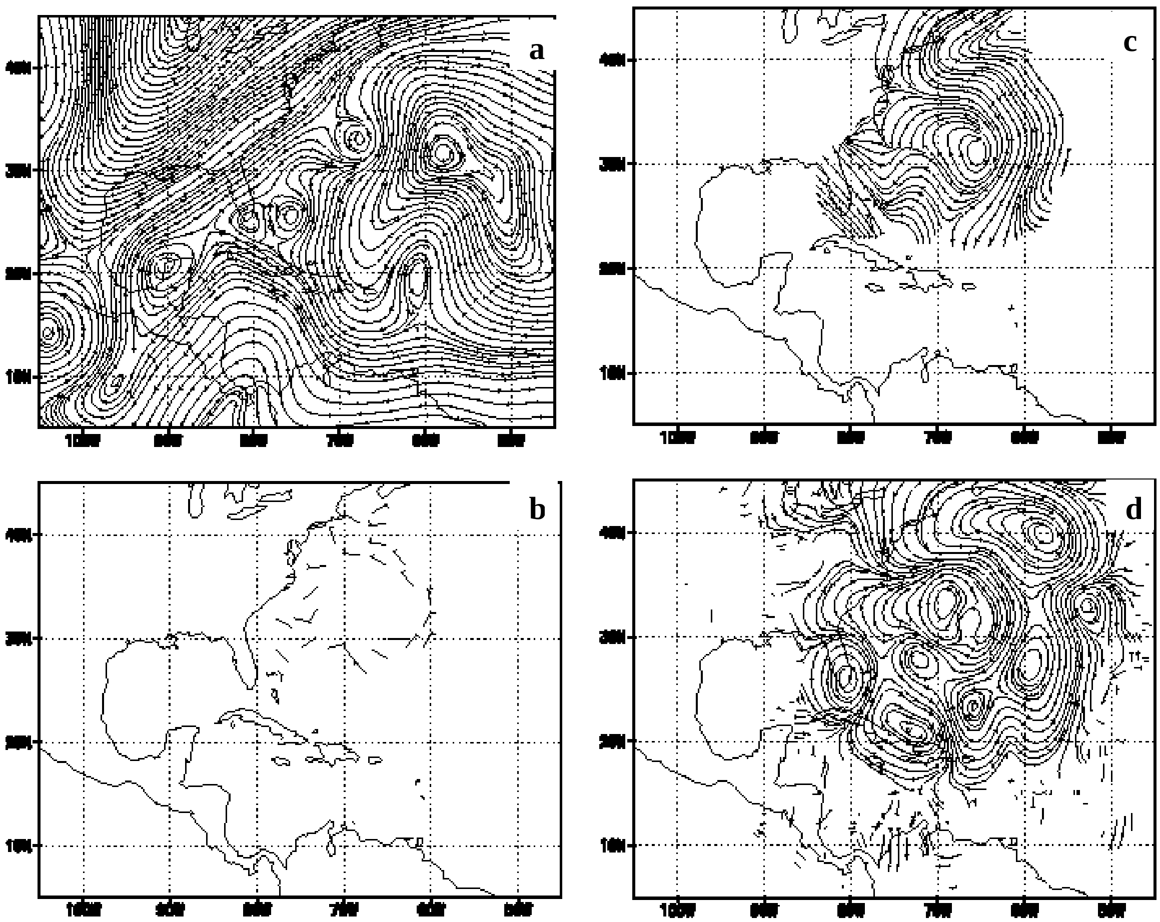

Figure 1. The VICBAR DLM wind (a) ensemble mean and (b) innovations (differences between the observations and the first guess) for 00Z 24 September 2001. The increments to the VICBAR DLM wind field first guess obtained using (c) isotropic and (d) ensemble-based error statistics at 00Z 24 September 2001.

Figure 2:

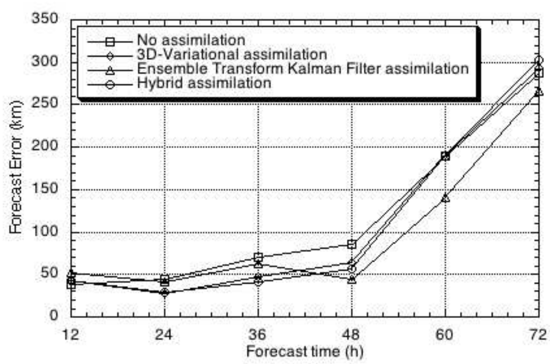

Figure 2. Track forecast errors for the 0000 UTC 24 September 2001 VICBAR runs with the various data assimilation schemes. Errors are in units of kilometers.