ENSEMBLE-BASED DATA ASSIMILATION

BY FUQING ZHANG AND CHRIS SNYDER

Ensemble-based data assimilation is a collection of flexible-state estimation techniques that use ensemble forecasts to estimate flow-dependent background error covariance or other probabilistic aspects of the background forecast. The b`est-known form of ensemble-based assimilation is the ensemble Kalman filter (EnKF) and its variant; a concise introduction to the EnKF can be found in Snyder and Zhang (2003), and Evensen (2003) and Hamill (2006) provide the most recent review.

At this 2.5-day workshop, more than 40 participants from seven universities, the National Center for Atmospheric Research (NCAR), and the National Oceanic and Atmospheric Administration (NOAA) discussed recent progress and emerging trends in ensemble-based data assimilation. This was a follow up to an informal workshop with more than 20 participants held at NCAR in September 2003. Both workshops were motivated by the rapid development and growth of research in this area. For example, the total number of publications with "ensemble Kalman filter" in the title or abstract has grown from 1 paper each year in 1996 and 1997 to 10 papers in 2001, and then to 40 papers for the first 8 months of 2006 [according to the Science Citation Index (SCI)].

This growth was reflected at the workshop, where presentations covered a variety of applications from convective to global scales and employed numerous different forecast models. Perhaps most striking was that all of the research was carried out by individual investigators, postdoctoral scientists, and graduate students rather than by the large teams more common in data assimilation research. Regardless of whether its operational use continues to expand, we believe ensemble-based assimilation already allows the research community new access to a wide variety of assimilation problems.

In addition to the workshop summary we provide here, a detailed workshop agenda and all presentations are available online at www.met.tamu.edu/people/faculty/fzhang/EnDA2006/.

WORKSHOP ON ENSEMBLE-BASED DATA ASSIMILATION

WHAT: Researchers assembled to discuss advances in ensemble-based data assimilation for state and parameter estimation in numerical weather prediction models ranging from convective to global in scale.
WHEN: 10–12 April 2006
WHERE: Marble Falls, Texas (40 miles northwest of Austin)

PROGRESS AND HIGHLIGHTS. Numerous presentations at the workshop demonstrated the effectiveness of the EnKF in experiments with real observations.

AFFILIATIONS: ZHANG—Texas A&M University, College Station, Texas; SNYDER—National Center for Atmospheric Research, Boulder, Colorado
CORRESPONDING AUTHOR: Fuqing Zhang, Department of Atmospheric Sciences, Texas A&M University, 1204 Eller O&M Building, College Station, TX 77843-3150
E–mail: fzhang@tamu.edu

DOI:10.1175/BAMS-88-4-565

In final form 8 January 2007

At the regional scale, a real-time EnKF implemented in a regional-scale model [Weather Research and Forecasting (WRF)] running continuously for nearly 2 yr performed comparable to the operational models for the northwest region of the United States, even though far fewer observations were assimilated in the EnKF than in the operational models. Promising results, again using limited-area models, were also shown for case studies of severe storm environments, mesoscale convective systems, and tropical cyclones (Fig. 1).

At convective scales, assimilation of radar radial velocity and reflectivity with the EnKF has advanced from observing system simulation experiments to several encouraging case studies with real Doppler observations (Fig. 2). These case studies highlight the feasibility of EnKF assimilation at convective scales. But, the studies also emphasize the uncertainties in both microphysical and planetary boundary layer parameterizations for existing convective-scale models and the difficulty with evaluating convective-scale analyses and forecasts owing to a lack of observations, especially of thermodynamic variables.

At the global scale, two different groups presented results from real-data experiments using the EnKF with a reduced-resolution version of the operational National Centers for Environmental Prediction (NCEP) Global Forecasting System (GFS) and a subset of the fully operational observation set. The EnKF performed comparable to or better than the operational three-dimensional variational system, especially over data-sparse regions where observation spacing is comparable to or greater than characteristic correlation lengths and/or some state variables are unobserved. Ongoing operational testing at Environment Canada supports this same view.

The EnKF has also been used with simulated observations for parameter estimation and offers hope to treat different sources of parametric model error. Through simultaneous state and parameter estimation, the EnKF was shown in experiments with simulated observations to be capable of retrieving some critical parameters in microphysics parameterizations and boundary layer schemes.

Presentations at the workshop further demonstrated that the EnKF is not only capable of assimilating standard surface, sounding, and profiler observations, but also remote sensing observations derived from satellites such as cloud-drift winds and radio occultation measurements from GPS. Moreover, it can also easily accommodate the assimilation of the position of tropical cyclones (Figs. 1b–d), which is extremely difficult in typical variational data assimilation systems. Promising results were also reported in coupling the EnKF with variational methods using simulated observations.

ISSUES AND CHALLENGES. During the presentations and discussions, existing and foreseeable issues with the EnKF were discussed, with each demanding further research. For example, the performance of the EnKF is often found to be sensitive to, and can be thus substantially compromised

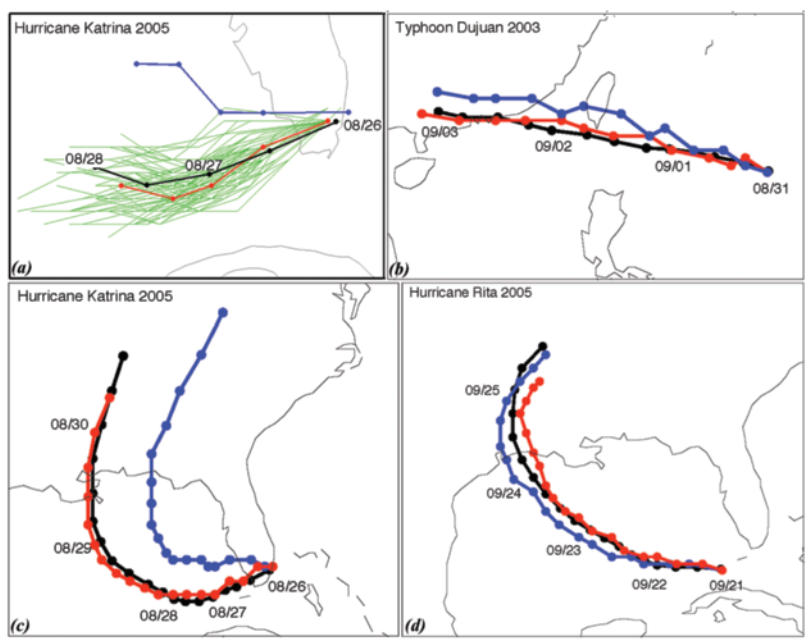

Fig. 1. (a) The WRK–EnKF ensemble forecast tracks (members, green; mean, red) of Hurricane Katrina (2005) in comparison to the National Hurricane Center (NHC) best track (black) and NCEP/GFS forecast (blue). The ensemble was initiated with the WRF–EnKF analysis at 0000 UTC 26 Aug 2005 through assimilation of all conventional observations. (Courtesy of R. Torn and G. Hakim at the University of Washington). The NHC best track (black) and the WRF forecast tracks of Typhoon Dujuan (2003), Hurricane Katrina (2005), and Hurricane Rita (2005) initialized from the EnKF mean (red) and NCEP/GFS (blue) analyses at 0000 UTC (b) 31 Aug 2003, (c) 26 Aug 2005, and (d) 21 Sep 2005, with a different WRF–EnKF system assimilating hurricane position and intensity, and satellite cloud-drift winds. (Courtesy of Y. Chen at NCAR.)

by, inaccuracies in the sampling of the background error covariance and errors in the forecast model and observation operator.

Sampling error can appear in the form of spurious, weak spatial autocorrelations between the same or different distant variables and/or spurious, weak cross correlations between different neighboring variables. The error is ultimately due to the limited affordable number of ensemble members. Several methods used to alleviate this problem include covariance localization, inflation, or a hybrid of both, but all current methods are ad hoc in nature. One obvious challenge of covariance localization occurs when the physical processes of interest contain a hierarchy of scales and typical observations contain multiscale information. Sampling error may also come from ensemble initiation resulting from incomplete and/or inaccurate representation of initial uncertainties that begin the EnKF cycles. This is exacerbated in convection (or even regional) -scale data assimilation in which the initial deficiency may persist through the limited life cycles of the processes of interest. Lateral boundary conditions are another source of sampling error for limited area applications.

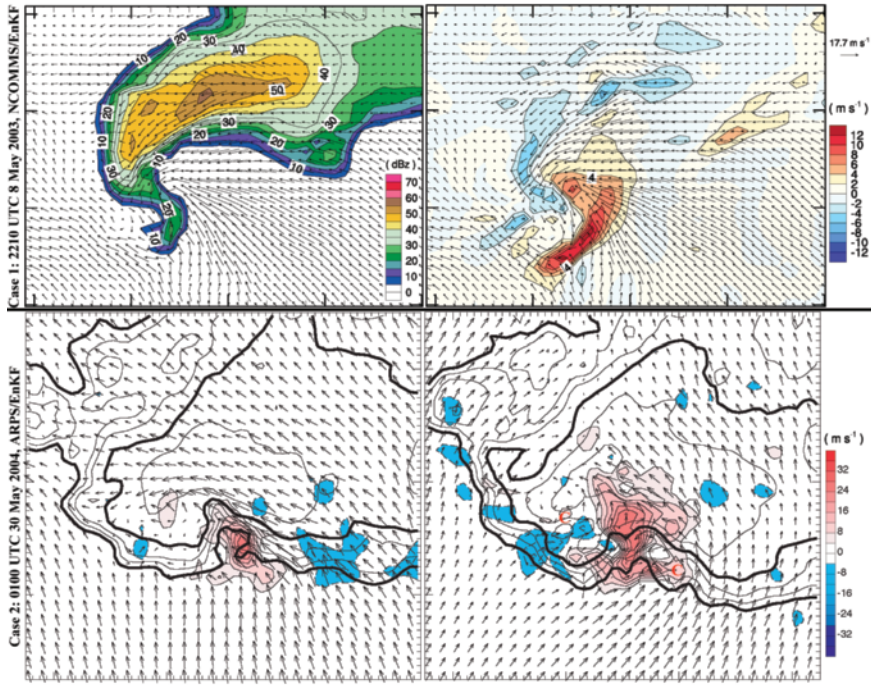

Fig. 2. (top left) Reflectivity (Δ=5 dBZ), (top right) vertical velocity (Δ=2 m s^{-1}), and horizontal storm-relative winds (vectors, plotted every 1 km) at 1.0 km AGL at 2210 UTC 8 May 2003, as the Oklahoma City, OK, supercell was entering its tornadic phase, which later produced an F4 tornado. The EnKF analysis was produced by assimilating Doppler velocity and reflectivity observations from the KOUN National Severe Storms Laboratory (NSSL) Collaborative Model for Multiscale Atmospheric Simulation for 85 min. (Courtesy of D. Dowell at NCAR). (bottom) Vertical velocity (contours and color shading), horizontal storm-relative winds (vectors, plotted every 2 km) and reflectivity (Δ=10 dBZ, with the 10- and 40-dBZ contours highlighted) from the ensemble mean analysis at (left) 1.5 and (right) 4 km MSL at 0100 UTC 30 May 2004. The EnKF analysis was produced by assimilating Doppler velocity and reflectivity observations from both KLTX and KVNX radars into the Advanced Regional Prediction System for 55 min. This 29–30 May northern Oklahoma City thunderstorm event lasted 9 h and produced 16 tornadoes, the strongest being an F3. (Courtesy of M. Xue and M. Tong at University of Oklahoma.)

Model error is a challenging problem for almost all aspects of numerical weather prediction (NWP), including ensemble-based data assimilation, which relies on the model forecast to faithfully propagate the time- and space-dependent error covariance from one assimilation cycle to the next. Limited temporal and spatial resolutions and parameterization of subgrid-scale and stochastic physics processes are two primary sources of model error. Limited, application-dependent successes have been reported through the use of additive noise, covariance inflation, hybrids of the EnKF with variational methods, and multiple physics schemes in the same ensemble. However, as is the case with sampling error, these treatments are still largely ad hoc in nature. Also, systematic treatment of model bias, which is routinely done in many operational NWP models, has yet to be implemented and evaluated in the current EnKF approaches. At convective scales, there are not enough observations to define model bias in the thermodynamic fields (e.g., temperature). Simultaneous state and parameter estimation through state augmentation offers hope to treat parametric model error (and perhaps also model bias), although this method may be more vulnerable to spatial and temporal variation of "optimal" parameters and the sampling error between the parameters and the observed state variables. Another potential problem for this approach is the nonlinear state response in the presence of multiple unknown parameters.

As is common to all data assimilation methods, observation quantity, quality, and error covariance may also impact the performance of the EnKF. Some observations, such as satellite radiance, radar reflectivity, and polarimetric measurements, require complex observation operators that are often hard to assimilate when tangled with model error. In cases of too many often correlated and redundant observations, such as satellite and radar measurements, it may be necessary to perform data thinning and/or superobservations (i.e., synthesizing/combining multiple observations into one) to reduce the computational cost and limit the effects of sampling error.

CONCLUDING REMARKS. The results presented at this workshop showed that the EnKF is a maturing assimilation technique for NWP across a range of scales. Beyond the evident scientific progress, the workshop was also remarkable for the breadth of applications and forecast models considered, and because all of the EnKF systems discussed at the workshop were developed by small research groups.

We expect continued scientific progress in the next few years as research on ensemble-based data assimilation has clearly reached a critical mass in the United States. The flexibility and ease of development of the EnKF will also likely encourage further data assimilation research in the university community, providing stimuli for data assimilation research and education as called for in the U.S. Weather Research Program (Fritsch and Carbone 2004).

ACKNOWLEDGMENTS. We acknowledge all participants for their invaluable input during the workshop and for their comments on earlier versions of this summary. This workshop is partially supported by National Science Foundation Grants 0205655 and 0205599.

REFERENCES

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53,** 343–367.

Fritsch, J. M., and R. E. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season: A USWRP research and development strategy. *Bull. Amer. Meteor. Soc.*, **85,** 955–965.

Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation. *Predictability of Weather and Climate*, T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 124–156.

Snyder, C., and F. Zhang, 2003: Tests of an ensemble Kalman filter for convective-scale data assimilation. *Mon. Wea. Rev.*, **131,** 1663–1677.