**11.2**

# Flow-Dependent Background Error Covariance and Mesoscale Predictability Estimation through Ensemble Forecasting

*Fuqing Zhang, Joy Ham and Altug Aksoy*
*Department of Atmospheric Sciences, Texas A&M University, College Station, TX 77843-3150*

## 1. Introduction

Over the past decade, ensemble forecasting has emerged as a powerful tool for numerical weather prediction. Not only does it produce the best estimates of the state, it also provides uncertainties associated with the best estimate and the predictability of a certain event, which also provides invaluable information to estimate the background error covariance for data assimilation. In this study, random perturbations have been used to initialize a mesoscale ensemble forecast of the 24-25 January 2000 “surprise” snowstorm that occurred along the East Coast of the United States. Our previous studies of this storm found that the mesoscale predictability can be seriously limited by model grid resolution and strong upscale growth of small-scale small-amplitude initial error in the presence of moist convection (Zhang et al. 2002a, 2002b). However, it remains unexplained that the 36-h forecast difference using two independent initial analyses (operational Eta analysis and the ECMWF analysis, larger initial difference) is considerably larger than difference with/without individual soundings or idealized small-amplitude initial perturbations. The ensemble forecasts were initiated with *realistic* initial uncertainties with error magnitude comparable to the difference between operational Eta and ECMWF analyses. Short-term (24- to 36-h) forecast sensitivity of the snowstorm, error growth characteristics (predictability) in the presence of larger-amplitude initial errors, and the flow-dependent background error covariance will be investigated through these ensemble forecasts.

## 2. Experimental design

NCAR/PSU mesoscale model MM5 version 2 is used for this study. The 30-km resolution model domain covers the whole continental United States and its configurations are the same as those used in the real-time forecast system running at NCAR (http://rain.mmm.ucar.edu/mm5). The reference initial analyses at 12Z 23 January and 00Z 24 January 2000 are both generated using the real-time operational Eta model as the first guess and then reanalyzed with observations using the standard procedure in MM5. Real-time Eta-forecasts from these initial analyses are then used as boundary conditions for all the simulations discussed herein. Details of the reference simulation setup and physics configurations can be found in Zhang et al. (2002a).

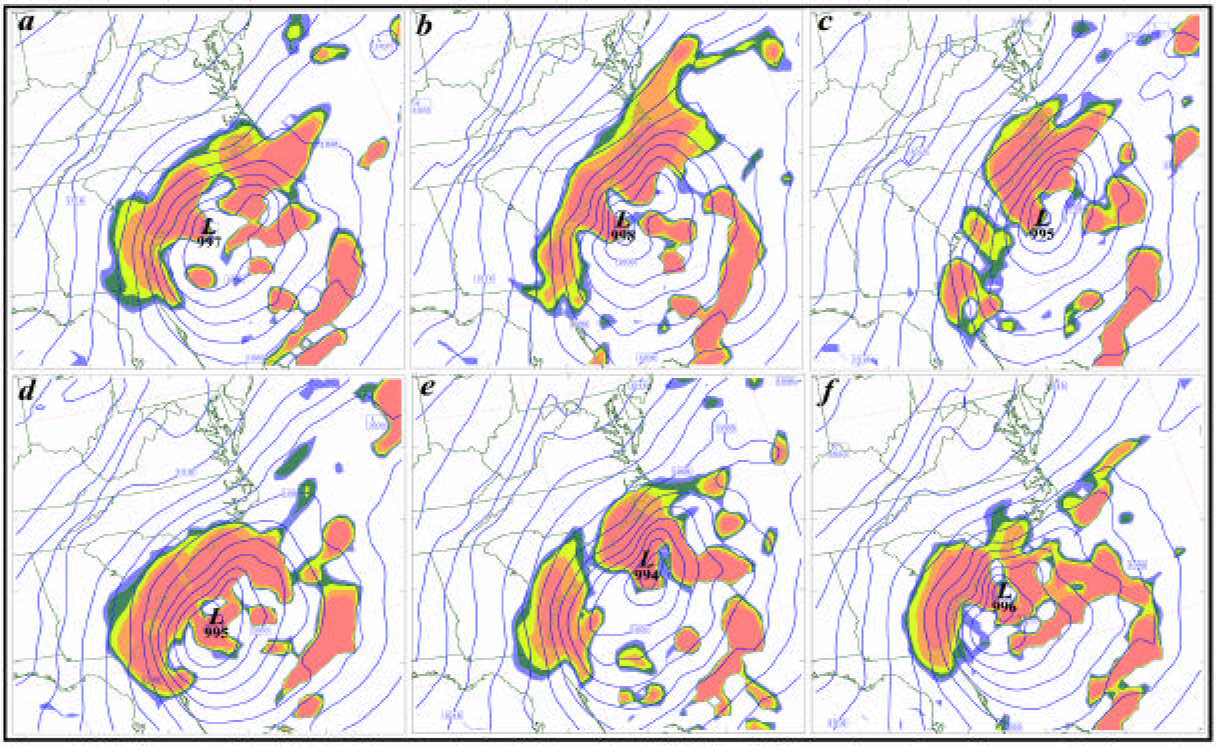

Figure 1. 24-h forecast of the MSLP (D=2mb) and reflectivity valid at 00Z 25 Jan 2000 from 6 members of the ensemble forecast.

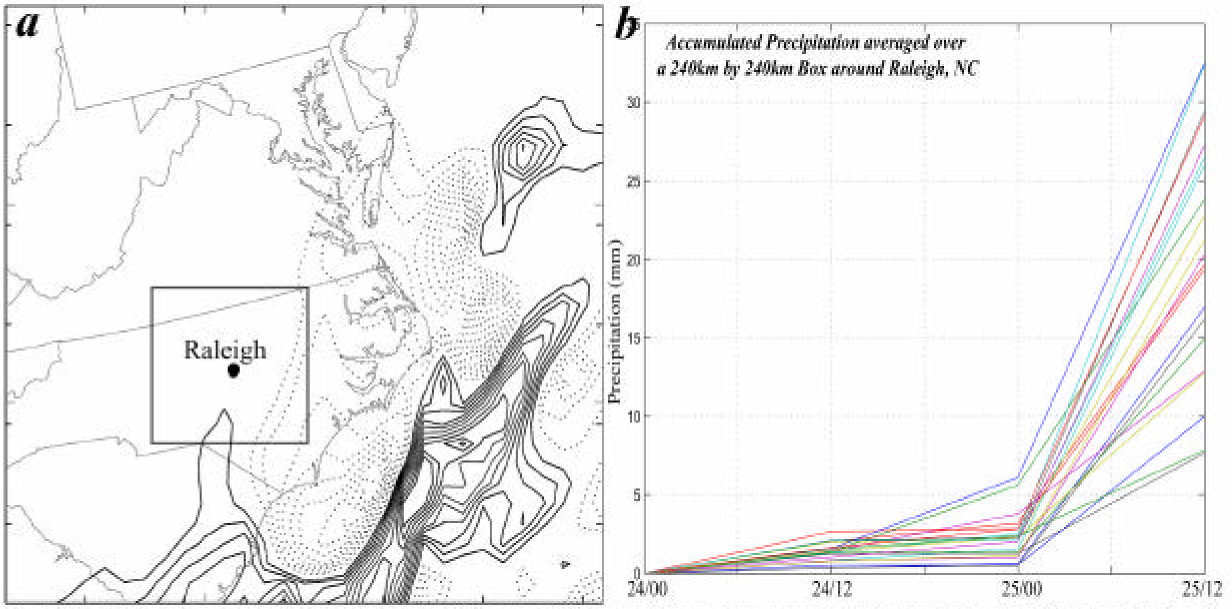

Figure 2. (a) 36-h accumulated precipitation difference (D=10mm) between two ensemble members. (b) Time evolution of the accumulated precipitation (mm) averaged over a 240km by 240km box around Raleigh, NC shown in panel (a).

First, "grid-point" random perturbations with standard deviation of 3 m/s and 3 K are added to the reference analyses at all model grid points at 12Z 23 January to generate 10 perturbed initial conditions. Each of these initial states is then integrated for 12 hours. Since these initial perturbations are totally uncorrelated, after 12-h simulations, the difference total energy [DTE=0.5*(U*U+V*V)+k*T*T] between any two simulations has greatly decreased (to ~20% of its original value). However, these isotropic initial perturbations do not decrease by the same degree across the domain; coherent structures begin to develop from these random perturbations after 12-h model integration. The variations are maximized over the region of moist activity as well as in the vicinity of the jet-front systems where strong gradients exist; the perturbations are still largely uncorrelated and mostly in smaller scales (not shown). We then linearly rescale the 12-h forecast difference of all prognostic variables between any two ensemble members according to the difference between the operational Eta analysis and ECMWF gridded analysis (in terms of DTE) valid at 00Z 24 January. These rescaled perturbations are then added to the reference MM5 analysis valid at 00Z 24 January to generate the initial conditions of a 20-member ensemble forecast. The ensemble forecast is integrated forward for 36 h. The ensemble generation used is similar to the "breeding method" used at NCEP (Toth and Kalnay 1993) except that "grid-point" random perturbations were used initially and there was only one "breeding" cycle applied.

## 3. Mesoscale predictability

With realistic initial uncertainty, strong variability of all aspects of the snowstorm has been found in the 24- to 36-h ensemble forecast, echoing the massive failure of the short-range deterministic prediction by most of the operational models in real time. Figure 1 shows the 24-h forecast of the mean sea-level pressure (MSLP) and simulated Radar reflectivity of the 20-member ensemble forecast valid at 00 UTC 25 January 2000. This is the time when Raleigh began to have the solid precipitation in the real time. From this six members selected, we can see that the low pressure of the surface cyclone varies from 994 to 998 hPa; the location of the cyclone center can easily be separated by a distance of 300km (Fig.1d, e); and most importantly, the instantaneous precipitation bands indicated by the simulated reflectivity can be totally dislocated. For example, even though the surface cyclones (low pressure and location) are very close to each other in two simulations (Fig.1c, f), dramatic difference in the precipitation patterns occurred between forecasts from these two members. For member 3, most of the inland precipitation is along the coast of North Carolina but little in South Carolina (Fig.1c); for member 6, most of the inland precipitation is along the coast of South Carolina and Georgia with little precipitation on the coast of North Carolina (Fig. 1e). To compare against the forecast sensitivity to individual soundings shown in Fig. 15 of Zhang et al. (2002a, page 1629), the 36-h accumulated precipitation difference between two members (1 and 2) and the precipitation evolution of all members in a 240 km by 240 km grid box around Raleigh, North Carolina are plotted in Fig. 2. In Zhang et al. (2002a), maximum precipitation difference is as large as 40mm over the Atlantic Ocean between experiments with and without the Little Rock, AK sounding in the initial analysis and the precipitation averaged over Raleigh can be altered by 40% among the 10 individual sounding experiments. With realistic initial uncertainties and considerably larger magnitude of initial difference, maximum 36-h precipitation difference is over 100mm (between member 1 and member 2) and the averaged precipitation over Raleigh can easily be altered by 200% (Fig. 2). The greater short-term forecast sensitivity in Figs. 1-2 is consistent with the larger forecast difference between two forecasts initialized with the Eta analysis and ECMWF analysis, respectively, found in Fig. 11b-c of Zhang et al. (2002a, page 1626). A new set of ensemble forecasts using exactly the same initial conditions are performed exactly the same as above except that the latent heating/cooling is turned off throughout the 36-h hour integration (the "fake dry" experiments). Variability of the short-term forecast, especially in the lower troposphere has been greatly decreased in the dry environment starting from the same initial difference. The 24-h surface cyclone forecasts from all 20 ensemble members stayed within 50-km in positioning and 2 hPa in magnitude (not shown). This is consistent with the findings from Zhang et al. (2002a, 2002b) that moist convection strongly impacts mesoscale predictability regardless of the initial error amplitude. The 24-h forecasted mean and standard deviation of the potential vorticity (PV) at 300 and 600 hPa from both the "fully-moist" and "fake-dry" sets of ensemble forecast are plotted in Fig.3. We can see that moist processes not only modified the balanced dynamics of the baroclinic system by curving the southern end of the PV filament more anti-cyclonically and stretching the tropopause more to the middle-lower troposphere, they also changed the variability (thus predictability) and the vertical distribution of the forecast uncertainty. Variance of the PV (as well as the winds and temperature, not shown) in the dry environment stayed mostly in the upper troposphere along the upper front; variance in the moist environment has the same magnitude but broader area in the upper troposphere but also has significant middle-lower tropopheric maximum above the surface cyclone and coastal front. This suggests that cautions should be taken to generalize theories on ensemble generation or data assimilation achieved from lower-dimensional dry models to more complex prediction systems simulating realistic atmosphere with moist dynamics.

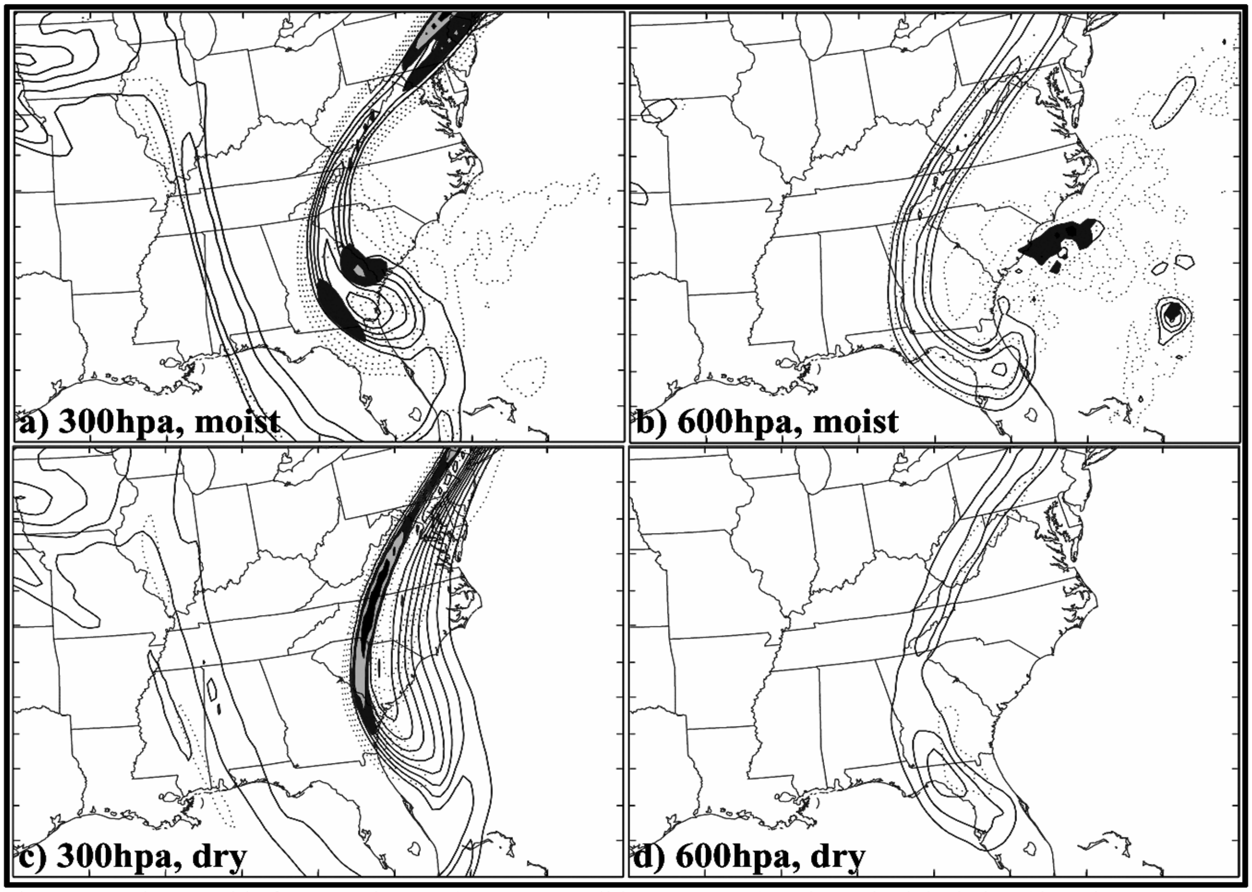

Figure 3. 24-h ensemble forecasted mean (>1.5PVU, solid, D=1.0PVU) and standard deviation (dotted, D=0.25PVU, >1.0 PVU colored) of PV at 300 and 600 hpa from both the "fully-moist" and "fake-dry" sets of ensemble forecasts.

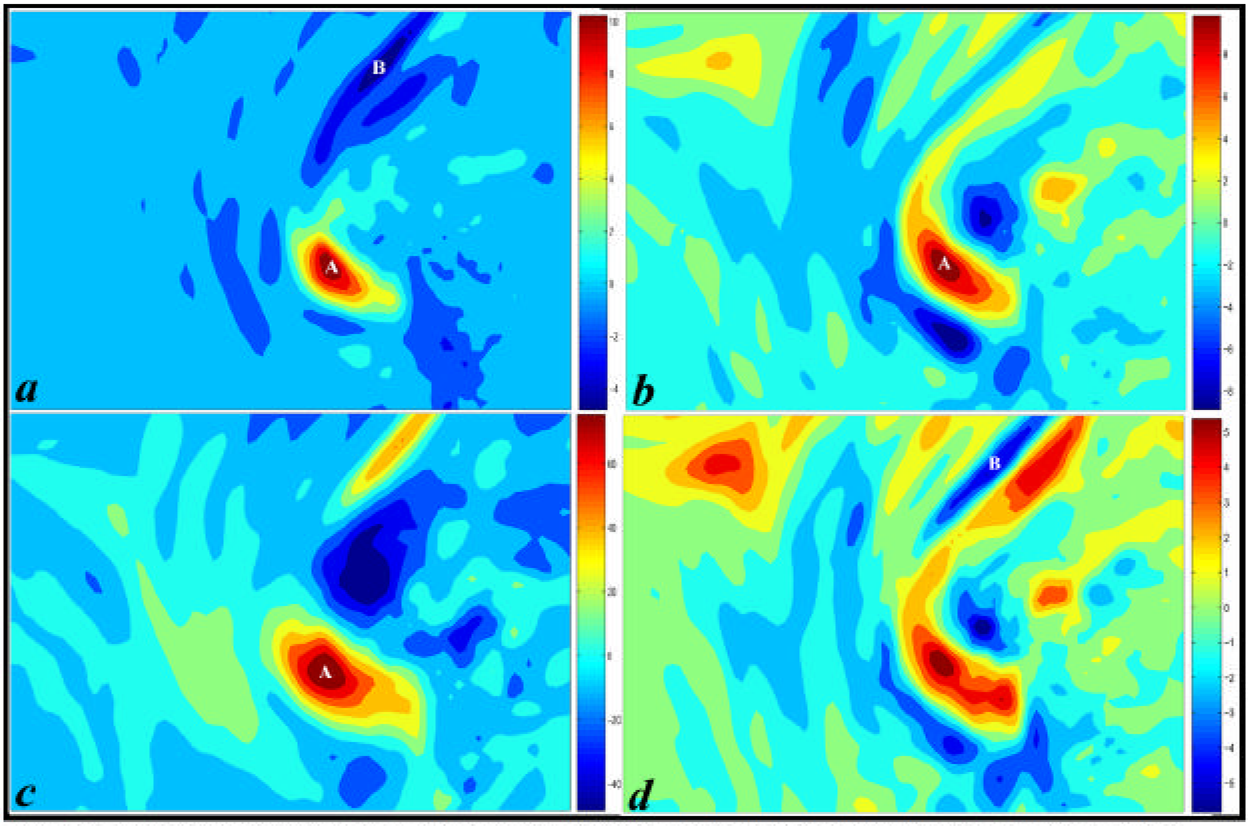

Figure 4. 24-h ensemble forecasted background error covariances at 300hPa for (a) u't', (b) $u'_A t'$, (c) $u'_A u'$, and (d) $u'_B t'$. The units are Km/s for panels (a) (b) (d) and $(m/s)^2$ for (c). The domain is the same as in Figure 3.

## 4. Background error covariance

Our previous studies found that, even with nearly-perfect initial analysis, mesoscale predictability of the record-breaking snowstorm can ultimately be limited by model resolutions (or model error in general) and strong nonlinear upscale growth of uncontrollable small-scale small-amplitude initial errors in the presence of moist convection (Zhang et al. 2002a, 2002b). However, in the near term, given the uncertainty in our operational analysis are still huge, significant prediction skills can be gained with better data assimilation techniques (to reduce the amplitude of the initial errors), probabilistic or ensemble forecast and the combination of both (i.e., the ensemble-based state estimation such as ensemble-Kalman filter or EnKF).

One of the key issues for data assimilation is the treatment of background error covariance. Currently, one-point horizontally-isotropic and time-independent error correlation is commonly used for data assimilation at most operational centers such as NCEP. This is in strong contrast to the flow-dependent predictability and variance distribution discussed above. The flow-dependent background error covariance is thus examined in detail using these ensemble forecasts. Figure 4 shows part of the background error covariance matrix. We can see that, after 24-h integration, the initially uncorrelated random errors develop strong spatial correlation not only among the same variable (auto-covariance) but also between different forecast variables (cross-covariance) especially over the region of strong cyclogenesis and along the upper front. When an observation is taken, these auto- and cross-correlations (covariances) developed from the short-term ensemble forecast can wisely spread the information to both observed and unobserved variables. The ensemble forecast can also be used to determine where the optimum observations ("target observation") should be taken by maximizing the Kalman gain (Bishop et al. 2001).

Our ultimate goal is to use the background error covariances estimated from ensemble forecasts to develop an ensemble-based data assimilation system (EnKF) for meso-/regional scale data assimilation.

**Acknowledgements.** This research is supported by NSF/ITR 02-05599 and also by NSF/NCAR USWRP.

## Reference

Bishop, C. H. et al. 2001: *MWR*, **129**, 420.
Toth, Z. and E. Kalnay, 1993: *Wea. Forecasting*, **12**, 140.
Zhang, F., C. Snyder and R. Rotunno, 2002a: *MWR*, **130**, 1617.
Zhang, F., C. Snyder and R. Rotunno, 2002b: *JAS*, in review.

Corresponding author: FZ (fzhang@tamu.edu)