J10.1

MODEL ERRORS IN ENSEMBLE FORECASTS: THE STRUCTURE OF ERRORS FROM UNREPRESENTED SCALES

Thomas M. Hamill and Jeffrey S. Whitaker

NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado

1. INTRODUCTION

Ensemble weather forecasts have been conducted operationally in the U.S. and Europe since 1993 (Toth and Kalnay 1993, Molteni et al. 1996, Houtekamer et al. 1996). Until recently, these operational ensembles consisted solely of parallel forecasts using the same forecast model but different initial conditions. Typically, the spread of forecasts in the ensemble was too small, presumably due in large part to imperfections in the forecast model. Ideally, the ensemble forecast system ought to include ways of properly including the additional uncertainty introduced by an imperfect model.

Estimating model uncertainty is a difficult problem. Buizza et al. (1999) proposed a simple scheme to introduce stochastic noise to simulate uncertainties in model parameterizations. In this scheme, the parameterized tendencies were modeled stochastically by multiplying them by a random number between 0.5 and 1.5. Many groups have suggested that individual ensemble member forecasts be integrated from different models or use different parameterizations. For ensemble data assimilation (Evensen 1994, Houtekamer and Mitchell 1998, Hamill 2004) purposes, Mitchell and Houtekamer (2000) proposed to estimate the correlation length scale and the vertical correlations of model error using data assimilation innovation statistics and an approach suggested by Dee (1995). Random samples of this model error distribution were then added to each first guess field before estimating the covariances from the ensemble.

The general approach of using innovation statistics to estimate model-error statistics is sensible; practically, the true state of the atmosphere is never known, so a comparison of observations and model forecasts provides trustworthy data to make such comparisons. As we are still learning about model errors, however, it may prove useful to use an alternative approach, constructing a toy system where the true state is known. In this study we will use a comparatively high-resolution model integration (T126) to examine the errors introduced in a T31 model as a consequence of unresolved scales. We are interested in some of the general properties of these errors. Are they large or small? Are they random or flow dependent? To what extent are they temporally and spatially correlated? Though these experiments will be conducted with relatively simple models, the general framework could be extended to real NWP models, e.g., a synthetic T2000 simulation could be used to understand the model-error characteristics of a T300 simulation. The knowledge gained here may teach us how to more effectively parameterize model uncertainty for ensemble forecast and ensemble-based data assimilation purposes.

Corresponding author address: Dr. Thomas M. Hamill, NOAA-CIRES CDC, R/CDC 1, 325 Broadway, Boulder, CO 80305-3328. e-mail: tom.hamill@noaa.gov

2. EXPERIMENT DESIGN

A dry, spectral primitive equation model with no terrain is used in this experiment. The model is essentially the same as the model described more completely in Hamill et al. (2003). The model is forced by relaxation of temperature to a prescribed zonal mean state, which is similar but not identical to the state described by Held and Suarez (1994). The only difference in this version is a more stable tropical troposphere to lessen the incidence of tropical superadiabatic lapse rates. In these simulations the model was spun up for a period of 200 days from a random perturbation imposed on a resting state. After 200 days, the high-resolution model solution (T126 L28) was used as the true model state, and forecasts at reduced resolution were conducted.

The general methodology for exploring model errors is illustrated in Figure 1. Panel (a) shows a T126 initial condition, regarded as the true state at time t=0. Surface temperature is denoted by colors and sea-level pressure by the black contours. Land/sea boundaries are overlaid to provide perspective as to scale, but are essentially meaningless. Panel (b) shows the true state 12 h later, generated by making a 12-h forecast from the T126 initial condition using the T126 forecast model. Suppose we are concerned with an accurate forecast for scales T31 and below, the "resolved scales" (as opposed to T32-T126 being the "unresolved scales"). Truncating the 12-h true state to T31, we generate panel (c), the true depiction of the resolved scales. If a forecast is made using a T31 model, then the initial condition at T31, panel (d), is generated from a truncation of (a), and then a 12-h forecast is made forward, shown in panel (e). Comparing (c) and (e), we thus can assess the errors in the resolved scales contributed by the lack of interaction with the unresolved scales. Note that this simulation concept was also recently utilized in a paper by Tribbia and Baumhefner (2003).

One drawback of this approach is that the model climate of the resolved scales in the T126 model and the T31 model are not identical. Figure 2 shows a power spectrum from 100-day nature runs of both models. Note that there is less energy in the shorter resolved scales in the T31 model than in the T126. Hence, a T31

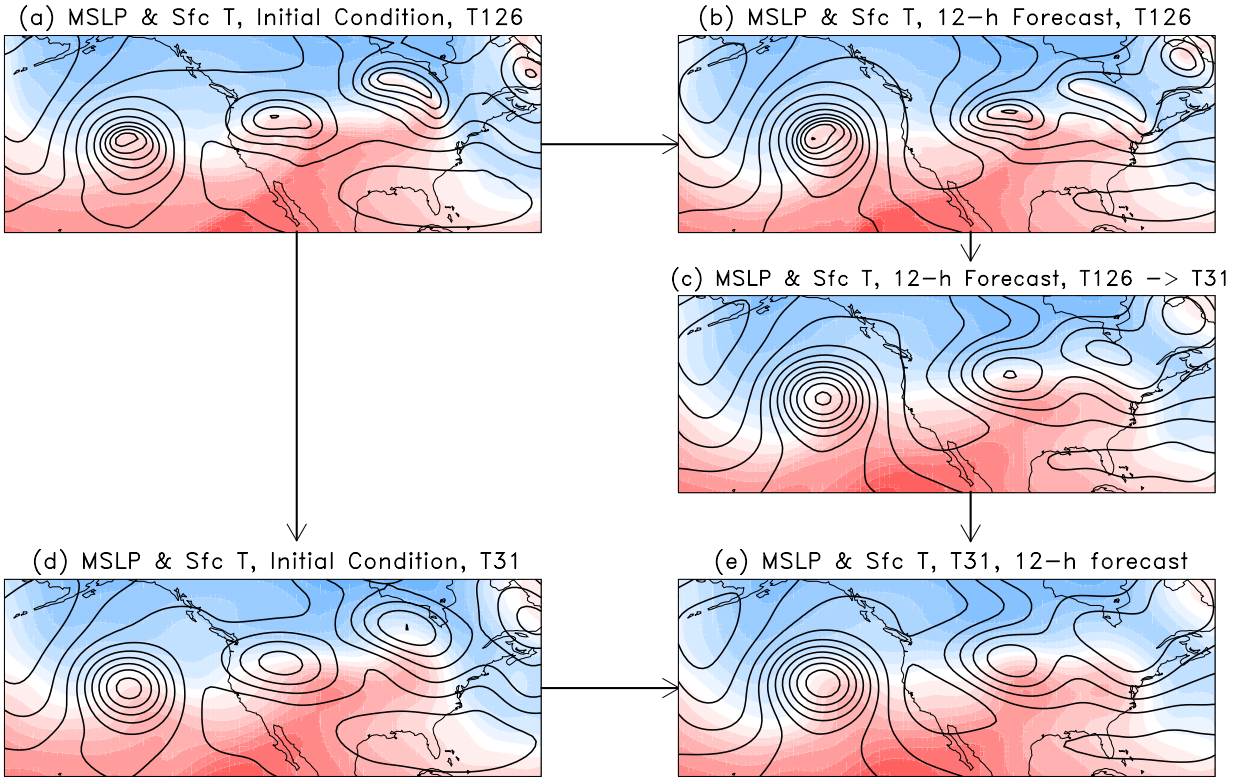

Figure 1. Illustration of process for examining 12-h error contributed by unresolved scales in T31 forecast. (a) Snapshot of a T126 initial condition, regarded as the true state at time t=0. Surface temperature is denoted by colors and sea-level pressure by the black countours. (b) T126 true state 12 h later. (c) Truncation of the 12-h true state to T31, the true depiction of the resolved scales. (d) T31 initial condition at T31, generated from a truncation of (a). (e) 12-h forecast of T31 model. Forecast errors of resolved scales determined by comparing (c) and (e).

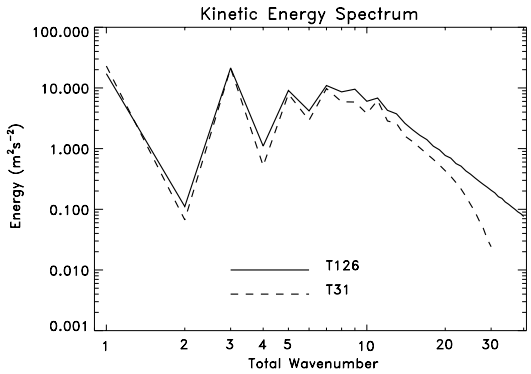

Figure 2. Kinetic energy spectrum of nature run simulations at T126 and T31.

forecast initialized with the resolved scales of the T126 initial condition will have errors from two sources, the model error from the lack of interaction with unresolved scales and the systematic drift toward the different T31 attractor.

3. RESULTS

a. Error growth characteristics

How large are the errors contributed by the unresolved scales, and how do these errors change with time? Following the general procedure outlined in Fig. 1, we initialized a T31 forecasts with the resolved scales from the T126. The domain mean specific kinetic energy for a given lead time was calculated over the nt=180 separate cases according to

$$KE = 0.5 \frac{\sum_{j=1}^{nt} \sum_{i=1}^{np} \alpha(i) \left(u'_j(i)^2 + v'_j(i)^2 \right)}{\left(nt \sum_{i=1}^{np} \alpha(i) \right)} \qquad (1)$$

where $u'_j(i)$ and $v'_j(i)$ are the $u-$ and v-velocity differences between the truncated T126 and the T31 forecasts at the ith grid point, and α_i is the grid area associated with the ith grid point of np grid points. The global growth of the magnitude of a typical error, i.e., $\sqrt{2\,KE}$ was then examined as a function of lead time (Fig. 3). Errors grow somewhat more quickly in the first 12 h, but steadily thereafter.

Examining the global kinetic energy error growth as a function of total wavenumber (Fig. 4), the errors are initially largest at the smallest scales but grow

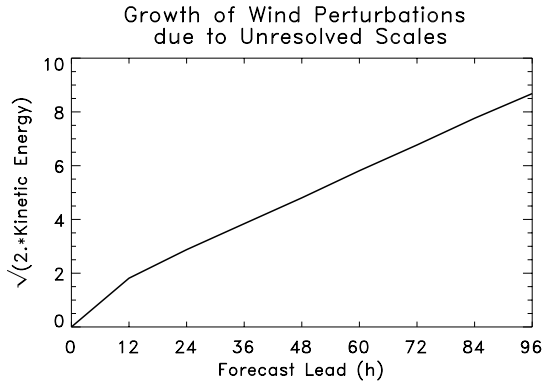

Figure 3. Growth of wind errors due to unresolved scales as function of forecast lead time. Data averaged over 180 cases separated by 12 h.

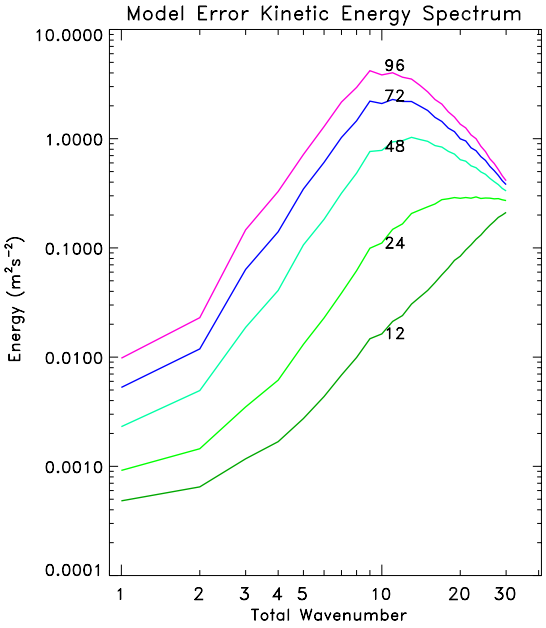

Figure 4. Kinetic energy error spectrum due to unresolved scales as a function of total wavenumber and lead time in hours.

upscale so that by 48h they are peaked at the synoptic scales. Thereafter, the errors grow relatively uniformly at the synoptic and planetary scales. These results are consistent with Tribbia and Baumhefner (2003).

b. Temporal continuity of short-term errors

We now consider whether the errors from the unrepresented scales have flow dependence and temporal continuity. Suppose the process illustrated in Fig. 1 is repeated every 12 h, continually re-initialized so that the resolved scales have no error at the start of a 12-h forecast and whatever error develops in the subsequent 12 h is a consequence of those unrepresented scales. Figure 5 shows a time series of four such sequential 12 h forecast errors, here of 1000 hPa temperature. Note that the 12-h errors are generally small in scale, consistent with Fig. 4, and over much of the domain they do not appear to have much temporal consistency. However, for one synoptic feature on the "west coast" of the U.S. (again, geography is fictional) and another in the western Atlantic Ocean, the errors for each 12-h period appear to be flow-dependent and to propagate with the short wave. This illustrates that in some special circumstances, perhaps in more baroclinically active situations, the errors of the unrepresented scales are closely tied to the flow of the day.

In subsequent work it may be desirable to generate random model-error states that have the same spatial and temporal characteristics as short-range forecast errors due to model truncation. This may be useful in a simulation experiment to test whether such additive model error (e.g., Mitchell and Houtekamer 2000) improves ensemble-based data assimilations. This research has not yet proceeded to the point of testing a T31 ensemble-based assimilation method, assimilating observations from a T126 nature run and adding random samples of simulated model errors due to truncation. However, one can make several qualitative conclusions about the characteristics of additive errors and reasonable ways to simulate them. First, as indicated by Fig. 4, the additive noise that one would add will have a very different structure depending on how often observational data is assimilated. If observations are infrequent, the model-error structures will be dominated by the baroclinic scales. If assimilation cycles are separated by only a few hours, a more realistic assumption, the noise due to truncation errors will primarily be small in scale. This suggests that the model errors contributed by the unresolved scales will be small in scale, which is different, for example, than the model error parameterization suggested by Mitchell and Houtekamer (2000) [though their model error parameterization accounts for more than just interactions with unresolved scales].

We have developed a general framework for creating random samples that simulate short-range forecast errors due to truncation. Let's assume we want to simulate a time series of 12-h errors. With a set of T126 and T31 short-range forecasts generated as in Fig. 1, we form a matrix $\mathbf{X} = (\mathbf{x}_1^{'}, \ \ldots \ , \mathbf{x}_n^{'})$ whose ith column is the ith time's state vector of 12-h forecast differences due to model truncation. A singular value decomposition (SVD) of $\tilde{\mathbf{X}} = \mathbf{EX}$ is then performed, where $\mathbf{E}$ is a diagonal matrix of weights that (1) accounts for the geographical variations in grid box size associated with each state vector element, and (2) weights temperatures, winds, and pressures according to the total-energy norm (Hamill et al. 2003). Specifically, a weight of $(c_p/T_{ref})^{-1/2}$ is applied to temperature components, where c_p is the specific heat capacity of dry air at constant pressure (1004 $J\ kg^{-1}\ K^{-1}$), and

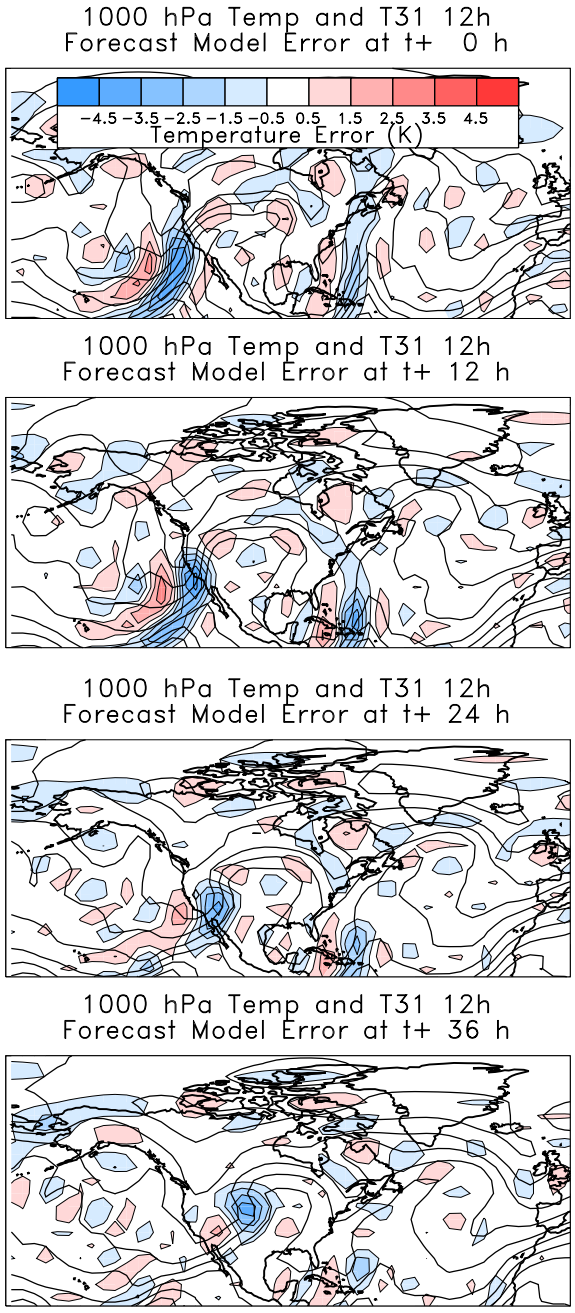

Figure 5. Illustration of flow dependence and temporal correlation of errors due to truncation of unrepresented scales. Following procedure outlined in Fig. 1, a set of four sequential 12-h forecasts of 1000 hPa temperature (solid lines, contours every 5 K) and temperature error (colors) are shown.

T_{ref} is a reference temperature (300 K). A weight of $(R\,T_{ref}/P_{ref})^{-1/2}$ is applied to pressure components, where P_{ref} = 1000 hPa, and R is the gas constant for dry air (287 $J\ kg^{-1}\ K^{-1}$). In any case, the SVD of $\tilde{\mathbf{X}}$ is $\tilde{\mathbf{X}} = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^{\mathrm{T}}$. Hereafter we will focus on the properties of the nsv right singular vectors $\mathbf{U} = \mathbf{E}^{-1}\tilde{\mathbf{U}} = (\mathbf{u}_1, \ldots, \mathbf{u}_{nsv})$.

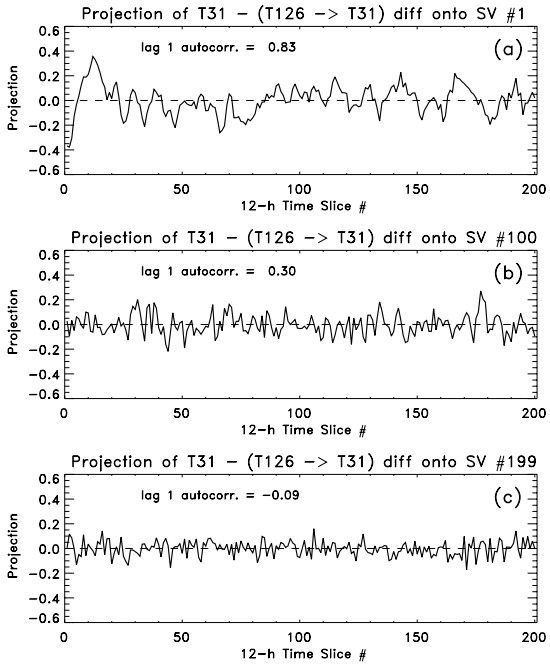

Figure 6. Projection of time series of 12-h model errors onto singular vectors of 12-h model errors. (a) Leading singular vector, (b) 100th singular vector, (c) 200th singular vector.

Figure 6 shows the projection of the time series of 12-h model errors onto the 1st, 100th, and 199th ordered singular vectors. As indicated, the lag-1 (12-h) autocorrelation is 0.85, 0.41, and -0.14, respectively (and though not shown, the leading singular vector is much larger in scale and amplitude than the trailing singular vector). Now, suppose one wished to develop a perturbation method that attempts to simulate random model errors with the correct, time-averaged spatial and temporal characteristics. Let $\mathbf{x}'(t)$ denote a synthetic model-error perturbation to the state at time t. We assume that such a perturbation can be generated as a linear combination of the singular vectors:

$$\mathbf{x}'(t) = \sum_{i=1}^{nsv} p_i(t)\mathbf{u}_i \qquad (2)$$

where $p_i(t)$ is a projection coefficient for the ith singular vector at time t. These coefficients are simulated with a first-order autoregressive model (Wilks 1995):

$$p_i(t) = p_i(t-1)\ \phi\ (i) + r_i\ \sigma_e(i). \qquad (3)$$

Here, $\phi\ (i)$ is the lag-1 correlation coefficient, r_i is a random, $N(0,1)$ number, and $\sigma_e^2(i)$ is the white-noise variance of the ith singular vector. This will be modeled (ibid, p. 306) using the variances of the projection coefficients $\sigma_x^2(i)$:

$$\sigma_e^2(i) = (1 - \phi^2(i)\ )\ \sigma_x^2(i). \qquad (4)$$

As shown in Fig. 7, it a simple model for $\phi(i)$ and $\sigma_x(i)$ is possible to estimate as a function of the singular vector number.

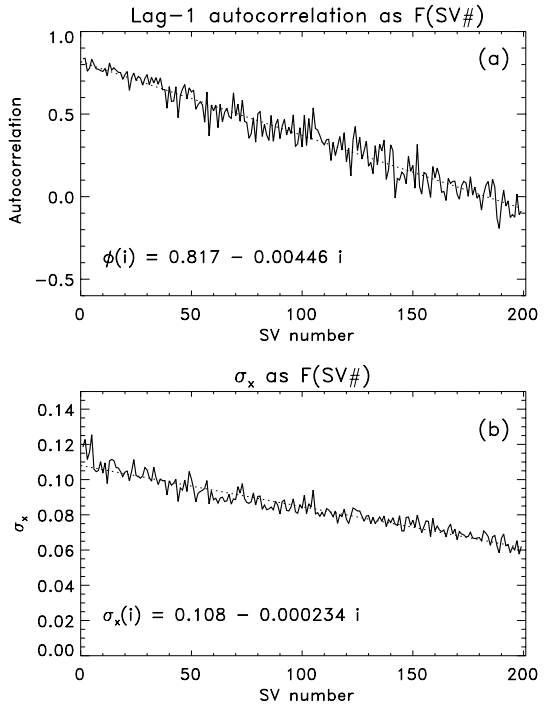

Figure 7. (a) Lag-1 (12-h) autocorrelation, derived, as in Fig. 6, from a time series of 12-h projection coefficients onto the ordered singular vectors. Dashed line indicates linear regression model fit as a function of the singular vector number; regression equation shown in lower left. (b) Standard deviation of time series of projection coefficients as a function of singular vector number. Fitted linear regression model again shown.

An examination of the characteristics of such synthetic time series is just beginning. Preliminary examinations have shown that the time series does not exhibit the flow-dependent, coherent structures shown in Fig. 5, and aspects such as the balance of these perturbations and their efficacy in data assimilation is yet to be explored. We hope to present an extension of this work at the conference.

4. CONCLUSIONS

In this preprint we examined the characteristics of errors in a global T31 forecast model. T31 forecasts were initialized with the resolved scales (T31 and below) from a T126 forecast model, identical in every respect except for the higher resolution. In this manner, it is possible to explore the contributions of the unresolved scales of motion (T32 and above) on the evolution of the resolved scales. This provides an analogue for understanding some of the error characteristics of operational weather forecast models when their resolution must be arbitrarily truncated.

We found that errors due to unresolved scales begin small in scale but are eventually dominated by the baroclinic scales after $\sim$ 48 h. Growth of the magnitude of a typical perturbation is relatively uniform through 96 h. In some instances, a time series of short-term forecast errors can be temporally coherent and propagate with a significant feature of the flow.

A method for generating additive noise that simulates short-term model errors due to the contribution of the unresolved scales was proposed. We plan a subsequent simulation experiment whereby a T126 simulation will be used as a nature run, for verification and for generating observations to be assimilated into a T31 ensemble-based assimilation system. The additive noise model developed here will be used in the ensemble assimilation system and compared against other parameterizations of model error.

5. ACKNOWLEDGMENTS

This research was partially funded by NSF ATM-0205612 and NSF ATM-0112715.

REFERENCES

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic simulation of model uncertainty in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887-2908.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, , **123**, 1128-1145.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99 (C5)**, 10143-10162.

Hamill, T. M., C. Snyder, and J. S. Whitaker, 2003: Ensemble forecasts and the properties of flow-dependent analysis-error covariance singular vectors. *Mon. Wea. Rev.*, **131**, 1741-1757.

Hamill, T. M., 2004: Ensemble-based data assimilation: a review. *Mon. Wea. Rev.*, submitted. Available from http://www.cdc.noaa.gov/∼hamill .

Houtekamer, P. L., L. Lefaivre, and J. Derome, 1996: The RPN ensemble prediction system. *Proc. ECMWF Seminar on Predictability, Vol II*, Reading, United Kingdom, 121-146. [Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom].

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

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

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73-119.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317-2330.

—— , and —— , 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

Tribbia, J. J., and D. P. Baumhefner, 2003: Scale interactions and atmospheric predictability: an updated perspective. *Mon. Wea. Rev.*, in press.