NOAA THORPEX ANNUAL PROGRESS REPORT

July 3, 2006

Project title: *Impact of fundamental assumptions of probabilistic data assimilation/ensemble forecasting: Conditional mode vs. conditional mean*

PI: *Milija Zupanski, Colorado State University*

SUMMARY

During the second year of the project, the efforts were directed toward development and testing of the Maximum Likelihood Ensemble Filter (MLEF) system with NCEP Global Forecasting System (GFS) spectral model and operational observations. The work was conducted in collaboration with Arif Albayrak from CIRA/Colorado State University, who is also a member of the CSU THORPEX group. At present, the observations include only standard observations, and the GFS resolution is T62L28. In the next year the observations will additionally include satellite radiances, and the GFS model will be in the resolution T126L55 for the control and T62L28 for ensemble members, as part of the dual-resolution MLEF algorithm. Important algorithmic note is that the MLEF is developed to include the NCEP operational codes as modules. In addition to the GFS operational script and code, the MLEF is also using the Spectral Statistical Interpolation (SSI) data assimilation script and code in interfacing with observations. All this was developed on the NCEP IBM computers, and is available to other interested researchers. Main accomplishments achieved during the second year of the project are:

- The MLEF with GFS T62L28 and operational observations (standard only) is developed and tested on a 3-day interval in January 2004,
- Modular MLEF code which includes NCEP operational infrastructure is developed,
- The model bias capability is included and tested,
- Strategies for the initiation of ensembles in the MLEF are being developed.
- Preliminary results are presented at the NOAA THORPEX PI meeting at NCEP in January 17-19, 2006.

To illustrate some of the issues, we present the results under three major topics:

1) Strategies for the MLEF ensemble initiation,
2) Other strategies for improving the MLEF performance,
3) Modular code development.

1. Strategies for the MLEF ensemble initiation

This addresses an issue discovered earlier within MLEF applications, namely that its performance is impacted by the choice of the initial ensemble perturbations (e.g., Zupanski et al. 2005). It was found that it is beneficial to introduce correlations to initial random perturbations. Note that this applies only to the first cycle; in all subsequent cycles the ensembles are created using the MLEF algorithm. The problem with spectral model applications is that correlations between spectral coefficients do not necessarily transform into desired correlations between variables in physical space. We tried various options related to the structure of initial ensemble perturbations. In conclusion, we found that there are better ways to create initial perturbations, and one of those methods is used to present the results here. In addition, we found that using an information content mask does improve the robustness of the MLEF. The idea is similar to what is done elsewhere, i.e. assimilating only the observations with an increased information measure.

Preliminary experiments are performed with 200 ensembles. This number of ensembles is planned only for the warm-up cycles (up to 15-20 cycles) while only about 100 ensembles will be used for the remaining data assimilation cycles. The model resolution is T62L28, and only standard observations are assimilated (excluding satellite radiances and radar observations, which will be used in the next phase). The experiments are run for the 3-day period (Jan 7-10, 2004). In Fig.1 the root-mean-squared (RMS) errors of the 48-hour temperature forecast are shown, started from the MLEF and the SSI analyses. The panels are showing the RMS errors in tropics, the northern hemisphere and in the southern hemisphere extra-tropics. The RMS errors are calculated at observation locations and essentially represent the RMS of innovation vectors (e.g., observation minus first guess), calculated from the MLEF and SSI control forecasts. One can see that the MLEF and SSI produce comparable results in the tropics, while in the extra-tropics the SSI shows better results, except in the southern hemisphere upper troposphere. This is in contrast to the results from other ensemble methods which generally show a relatively better performance in the extra-tropics and a relatively worse performance in the tropics.

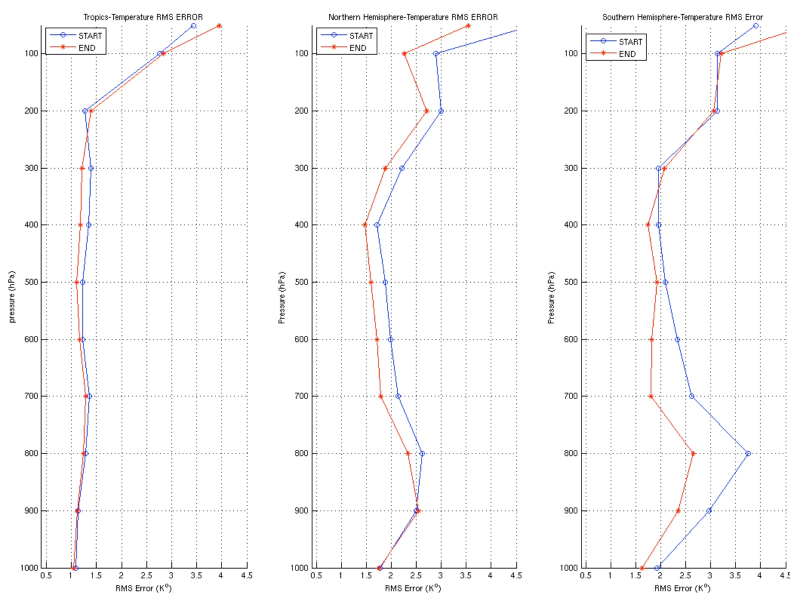

Fig.1. Vertical distribution of a 48-hour forecast root-mean-squared temperature error calculated from the first 9 data assimilation cycles, valid from 18 UTC, January 7, 2004, until 12 UTC, January 9, 2004. The left panel is for the tropics $(20^\circ S - 20^\circ N)$, the middle panel for the northern hemisphere $(20^\circ N - 90^\circ N)$, and the right panel is for the southern hemisphere $(20^\circ S - 90^\circ S)$. The red line (dots) represents the forecast from the SSI, and the blue line (open circles) represents the forecast from the MLEF.

In order to better understand the reasons for a relatively good performance of the MLEF in the tropics, the differences between various ensemble members and the control are calculated. In

Fig.2, a typical surface pressure difference between an ensemble member and the control is shown. In particular, the results are shown for the ensemble member 75 in a randomly chosen cycle. One can immediately notice that most of the perturbations are confined to the tropics, where the MLEF has an improved performance.

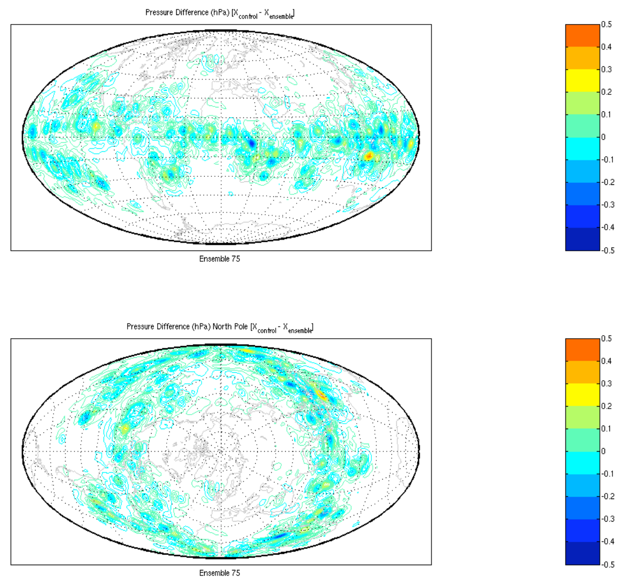

Fig.2. Horizontal map of a 6-hour surface pressure forecast difference (hPa) between the ensemble member 75 and the control forecast (first guess). The upper panel represents a side-view from the equator and the bottom panel shows a top-view from the North Pole.

The results shown in Figs.1 and 2 suggest that there is a need to change the initial ensemble perturbation method in the MLEF, in such way to allow realistic perturbations in the extra-tropics. These experiments are now under way, and will be reported in near future. We believe that the new approach will allow more efficient use of ensemble members, eventually improving the overall MLEF performance.

2. Other strategies for improving the MLEF performance

In addition to the ensemble initiation strategies, we have been testing few other strategies likely to improve the performance of ensemble data assimilation (EnsDA).

a. Model bias

The MLEF experiments including model bias correction were performed for the same period (i.e. Jan 7-10, 2004). The bias correction strategy follows the method of Zupanski and Zupanski (2006). Following common assumption that larger scales of model error will make a stronger impact on the longer forecasts, the model bias was defined only for the T21 (i.e. spectral coefficients up to T21 were adjusted for the bias), while the remaining coefficients were not changed. The results, however, did not show significant improvement, and are not reported here.

We will continue these experiments by including all spectral coefficients, and devising other strategies.

b. Digital filtering

Another possible way to improve the EnsDA performance is to improve control of gravity waves in the analysis. In order to address this issue, an experiment with digital filter initialization was conducted. The results indicated that there was a short-lived improvement, up to the second cycle (12 hours of data assimilation). After that, the results did not differ noticeably from the experiment without digital filter. These experiments indicate that the MLEF algorithm is not producing an excess of gravity waves. Since the use of digital filter significantly increases the computational cost (e.g., about 30 %), these experiments were discontinued.

c. Constrained forecast error covariance

Most of EnsDA systems utilize covariance localization, which has an implicit property of increasing the number of the degrees of freedom in the analysis. The idea is that with more degrees of freedom a better fit to the observations, and therefore a better analysis, can be achieved. The adverse aspect of covariance localization is that it introduces additional free parameters to be specified, requires a notable code change, and increases the overall computational cost. Regarding the forecast error covariance matrix, localization acts as a constraint, in such way that the originally unrestricted, non-diagonal ensemble matrix is changed to a block-diagonal form, with off-diagonal blocks equal to zero.

In order to test the potential benefit of the block-diagonal forecast error covariance and thus the impact of an increased number of the degrees of freedom in the analysis, without an explicit localization, however, a new approach was taken. A block-diagonal matrix was created by neglecting the cross-correlations between control variables. Although these cross-correlations are important, the idea is that through the cycling of data assimilation, all variables will be eventually correlated. Note that this approach is less restrictive, and because of that it has fewer degrees of freedom, than the localization approach. Since there are four control variables in our experiment (logarithm of surface pressure, temperature, vorticity and divergence), the effective degrees of freedom were increased four times. In these experiments we used 100 ensemble members, thus there are 400 degrees of freedom in the analysis. As expected, the fit to observations was dramatically improved (i.e. a cost function decrease from 10 % to 25 %). However, the 48-hour forecast RMS error, used as a measure of success, was not significantly affected. It appears that additional degrees of freedom were not relevant for the uncertainty of the dynamics. We will look more into this issue, and try to understand better why the error covariance localization may, or may not be beneficial within MLEF.

3. Modular code development

One of the characteristics of the MLEF code is that it follows a modular approach. This means that all major parts of the algorithm (forecast model, observation operator, MLEF matrix calculations) are almost independent, having only interfaces to connect them. In particular, the MLEF exploits the codes and the scripts available for the GFS forecast model and for the SSI observation operators. This approach allows the MLEF to take advantage of an upgrade of the forecast model or observation operator in an algorithmically efficient fashion. This may be important when changing the model resolution, or when introducing new observations that may become available through the operational SSI (or GSI) data assimilation. Especially important is that this approach allows the quality control of data to be same as the NCEP operational quality control.

In conclusion, the work on this THORPEX project is closely following the plan for the second year. The next year work will focus on including the assimilation of satellite radiances, and evaluating the dual-resolution MLEF algorithm. We collaborate with NCEP scientists Yucheng Song and Mozheng Wei, as well as with the research groups at Univ. of Maryland (Istvan Szunyogh, Eugenia Kalnay), NOAA/ESRL (Tom Hamill, Jeff Whitaker), NRL Monterey (Craig Bishop), and others.