15B.6 Impact of Model Error and Imperfect Initial Condition Perturbations on Ensemble-Based Probabilistic Forecasts: UNPREDICTABLE SPOTS

Jun Du*

Environmental Modeling Center/NCEP/NOAA and SAIC, Washington DC

Abstract

The ultimate goal of ensemble forecasting is to reliably estimate the time-evolution of a probabilistic density function (PDF) of meteorological fields by quantifying forecast uncertainties both at the initial time and over the entire model integration. In this study, using the NCEP short-range ensemble forecasting (SREF) system, two "perfect model" experiments were conducted to address the following three issues: (1) given a near-perfect Ensemble Prediction System (EPS), how well can PDF be predicted? (2) how can bad PDF forecast regions be identified (referring to "unpredictable spots")? and (3) what is the relative importance between model error and imperfect initial condition (IC) perturbations over the evolution of the PDF?

Although a good ensemble system could produce good mean, spread and probability forecasts at the majority of model grid points, it's almost certain that it also generates extremely bad and misleading forecasts at some locations which are defined as *"unpredictable spots"*. As long as the model used is imperfect, "unpredictable spots" will never diminish even if the IC perturbations used in an EPS is perfect. Identifying the location of "unpredictable spots" is important for forecast calibration, but it's not an easy task because those spots are not well correlated with ensemble spread (or predictability) in general, *i.e.*, ensemble spread alone might not be a good indicator for identifying them. Our results further indicate that the correctness of the model physics might be more important than that of the IC perturbations in order to have a correct PDF forecast, at least in the big picture (but it is not conclusive at this point). Only if given a perfect model and very realistic IC perturbations could an EPS produce good (but still not perfect) forecasts over nearly the entire model domain. In a word, the task of correctly predicting the probability distribution or PDF using ensembles is extremely challenging if not impossible in reality.

Key words: ensemble forecasting, probability distribution, initial condition perturbation, model error, unpredictable spots

1 Introduction

Given the existence of intrinsic uncertainties in both initial conditions (IC) including boundary forcing and the model (including physics, dynamics and numerical calculations), an ensemble approach, a group of model forecasts started using slightly different but equally-likely ICs and various versions of model physics, including stochastic physics (Leith, 1974; Mullen *et al.*, 1999; Palmer *et al.*, 2000; Palmer, 2001; Du, 2002), might be the only way to possibly give a full picture of future state of the atmosphere, which is a highly nonlinear and often unstable system. The ultimate goal of ensemble forecasting is to reliably estimate the time-evolution of the probabilistic density function (PDF) of meteorological fields by quantifying uncertainties in forecasts both at the initial time and over the entire model integration (Brooks *et al.*, 1995; Du *et al.* 1997; Hamill *et al.*, 2000). However, due to model error and the imperfect IC perturbations used in any real-world Ensemble Prediction System (EPS), it is believed that this task is extremely difficult if not impossible (Lorenz, 1963; Smith, 2000; Judd and Smith, 2001, Zhang *et al.*, 2002). Currently, little is known about the impact of model error and imperfect IC perturbations on the evolution of ensemble-based PDFs within operational, state of the art numerical weather prediction (NWP) models.

In this study, using the NCEP operational Short-Range Ensemble Forecast (SREF) system (http://wwwt.emc.ncep.noaa.gov/mmb/SREF/SREF.html) (Stensrud *et al.*, 1999; Du and Tracton, 2001; Du *et al.*, 2004), the following three issues will be discussed: (1) given a near-perfect EPS, how well PDF can be predicted; (2) how can bad PDF forecast regions be identified [defined as *"unpredictable spots"* (see Section 2.2)] so that a special post-processing (such as Du *et al.*, 2000) might be applied to the forecasts over these regions before those forecasts are used by end users; and (3) what is the relative importance of model error versus imperfect IC perturbations over PDF evolution? Since there is no way to know "true" PDFs in the real atmosphere, two "perfect model" experiments were conducted to study these issues.

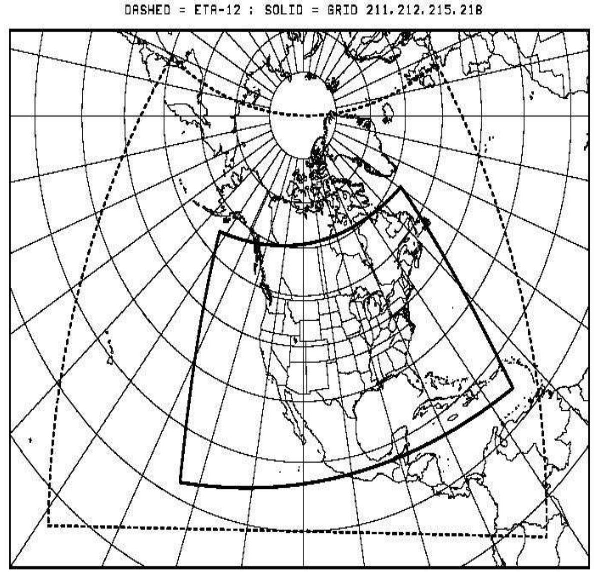

Figure 1: The model computation domain (large with dash line) and output/verification domain (small with solid line).

* *Corresponding author address:* Dr. Jun Du, Environmental Modeling Center/NCEP, 5200 Auth Road, Camp Springs, MD 20746, USA; e-mail <Jun.Du@noaa.gov>

Experiment I	09z, 8/22/03	21z, 8/22/03	09z, 8/23/03	21z, 8/23/03	09z, 8/24/03
	21z, 8/24/03	09z, 8/25/03	21z, 8/25/03	09z, 8/26/03	09z, 9/11/03
	21z, 9/11/03	09z, 9/12/03	21z, 9/12/03	09z, 9/13/03	21z, 9/13/03
	09z, 9/14/03	21z, 9/14/03	09z, 9/15/03	21z, 9/15/03	09z, 9/16/03
Experiment II	09z, 8/22/03	21z, 8/22/03	09z, 8/23/03	21z, 8/23/03	09z, 8/24/03
	21z, 8/24/03	09z, 8/25/03	21z, 8/25/03	09z, 8/26/03	

Table 1: The cases used in this study, listed by their model initiation time (UTC, mm/dd/yy). The forecast length is 63 hour with outout at every 3 hour. Model's horizontal resolution is about 48km with a large North American domain (Fig. 1). There are total 20 cases from August and September 2003.

EPS	membership	model/physics	IC/perturbations	representativeness
Eta.BMJ	5	Eta with BMJ convective scheme	EDAS/bred from Eta.BMJ	*"truth"*
RSM.SAS	5	RSM with SAS convective scheme	GDAS/bred from RSM.SAS	very *"good"* system
Eta.KF	5	Eta with KF convective scheme	EDAS/bred from Eta.KF	near *"perfect"* system

Table 2: Design of Experiment I: how well PDF can be predicted with a near perfect EPS?

2 Experiment I: How Well PDF Can be Predicted?

Table 1 is the list of the cases used by both experiments I and II. Table 2 describes the design of Experiment I, where Eta refers to the NCEP (National Centers for Environmental Prediction) Eta model (Black, 1994), RSM to the NCEP Regional Spectral Model (Juang and Kanamitsu, 1994), BMJ to the Betts-Miller-Janjic convective scheme (Janjic, 1994), KF to the Kain-Fritsch scheme (Kain and Fritsch, 1990, 1993), SAS to the simplified Arakawa-Schubert scheme (Arakawa and Schubert, 1974; Pan and Wu, 1995), EDAS to the Eta Data Assimilation System (Rogers *et al.*, 1996), GDAS to the NCEP Global Data Assimilation System (Parrish and Derber, 1992), and Bred to the NCEP breeding method for generating ensemble IC perturbations (Toth and Kalnay, 1997 and 1993). Since it's reasonable to assume that the difference between any current NWP model (analysis) and the real atmospheric system (state) is much larger than that between any two "good" (widely accepted) operational NWP models (analyses), the EPS of "Eta.KF" could represent a near *"perfect"* ensemble system, while the "RSM.SAS" a very *"good"* ensemble system with respect to the *"true"* ensemble system ("Eta.BMJ"). Given such ideal EPSs as described in Table 2, how well can PDF be predicted?

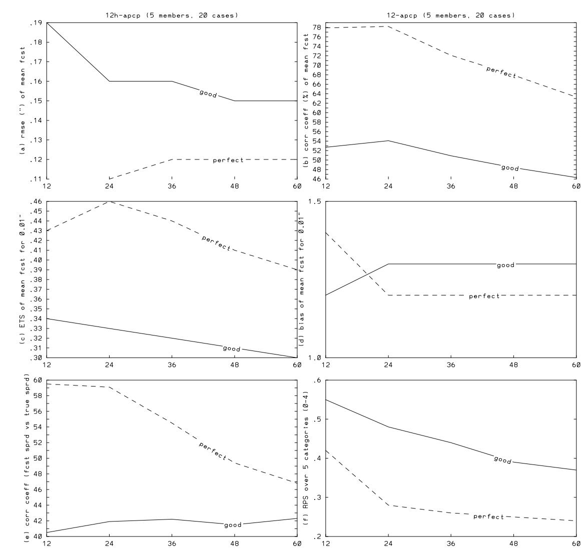

Figure 2: Scores of very *"good"* and near *"perfect"* EPS against the *"truth"*. The verification was performed over a 40km resolution, US Continental Domain (IxJ=185x129=23865 grid points, Fig. 1) averaged over the all 20 cases during August and September 2003 (Table 1). (a)-(d) for ensemble mean [root-mean-squared error, correlation coefficient, Equitable Threat Score and bias (value 1.0 represents no bias), respectively], (e) correlation coefficient between forecast spread and true spread; and (f) Ranked Probabilistic Score of probability distribution over 5 MECE (mutually exclusive, collectively exhaustive) categories (less than 0.01", 0.01"-0.25", 0.25"-0.5", 0.5"-1.0", and greater than 1.0").

2.1 General Performance

For simplicity, only the 12h-accumulated precipitation forecast (12h-apcp, so chosen because of its importance) was investigated as an illustration. The model output or verification domain is shown in Fig. 1, and has 185x129 grid points with 40km grid spacing covering the entire continental United States. Figure 2 shows the domain-averaged values of various standard scoring matrics as a measure of EPS's general performance. In particular, readers are referred to Epstein (1969) and Murphy (1971) for the original definition of Ranked Probabilistic Score (RPS, Fig. 2f), to Schaefer (1990) for the definition of Equitable Threat Score (ETS), and also to the Appendix of Du *et al.* (1997) for a brief summary of both RPS and ETS. Note that the

category	1	2	3	4	5	6
RPS value	0	(0,1)	[1,2)	[2,3)	[3,4)	4
meaning	perfect	good	useful	bad	worse	worst

Table 3: Definition of RPS score category. For a probability distribution over 5 MECE categories, the perfect RPS score is 0.0 and the worst is 4.0 (see the Appendix of Du *et al.* (1997)).

smaller the RPS value is, the better a probabilistic forecast is (zero being the perfect score), while the larger the ETS, the better the forecast. As expected, for both *"good"* and near *"perfect"* EPSs, their general performances are reasonably good in all aspects, including the ensemble mean (Fig. 2a-d), spread (Fig. 2e) and probability distribution (Fig. 2f) verifying against the *"truth"*.

It's also clear that with improving model physics and IC perturbations (from *"good"* to near *"perfect"* EPS), the overall performance of an EPS generally improves too: from the solid curves to the dashed curves in Fig. 2. This result certainly encourages scientists to further improve model and ensemble IC perturbation schemes in ensemble forecasting research and operations.

2.2 Local Performance

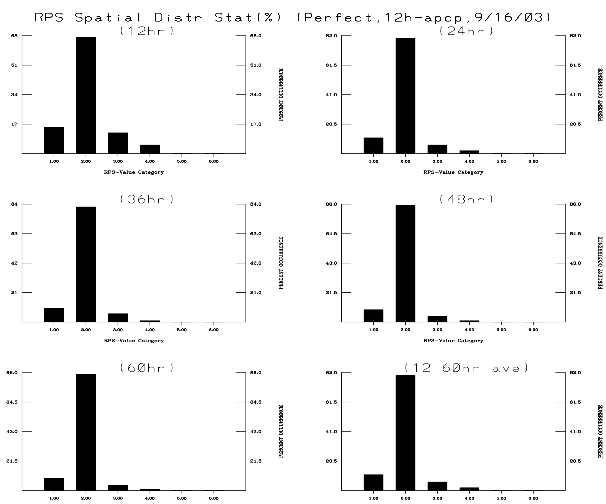

Figure 3: Percentage distribution of the number of grid points over the six RPS score categories for 12h, 24h, 36h, 48h and 60h probability forecasts as well as 12-60h average. It is from the near *"perfect"* EPS for 09z, Sept. 16, 2003 case.

Although the domain-wise performance is reasonably good, what is the spatial variation of the performance? By examing the values of Ranked Probabilistic Score (RPS) at each grid point over the entire model output domain (Fig. 1), the spatial variation in the quality of the probability forecast can be revealed. Figures 3 and 4 show the percentage distribution of the numbers of grid points over the six

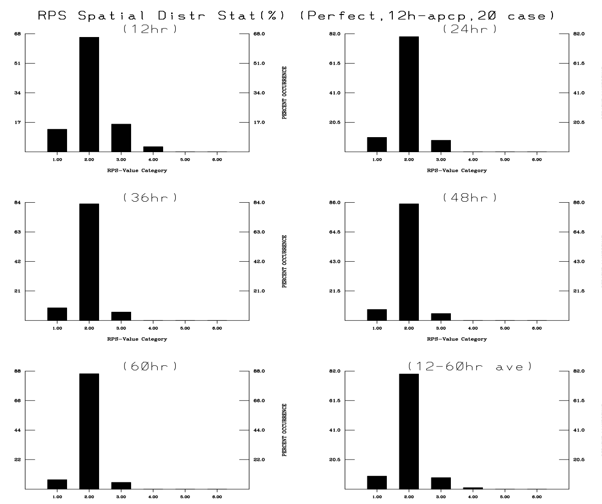

Figure 4: Same as Fig. 3 but averaged over the 20 cases.

RPS score categories (Table 3) for the near *"perfect"* EPS for a single case (Fig. 3) and the 20-case average (Fig. 4), respectively. Although the majority of grid points (about 80 %) have good probability forecasts (categories 1 and 2), about 4 % of grid points have extremely bad probability forecasts (categories 4 and 5) averaged over the 20 cases during August and September 2003. For the very *"good"* EPS (Figs. 5-6), the percentage of the numbers of grid points having very bad probability forecasts (categories 4 and 5) increased to about 5 %, with a small number of grid points even entering category 6 (a completely opposite distribution). Further study shows that over these bad-PDF regions, not only the probability forecast is bad, but ensemble spread and mean forecasts perform very poorly too (Fig. 7). These bad PDF spots can be defined as **"unpredictable spots"** since even a near-perfect ensemble prediction system cannot predict weather well over those spots by any means (mean state, uncertainty, and probability distribution).

2.3 How To Identify "Unpredictable Spots"?

What we see here is that given a near-perfect EPS, it can produce a reasonably good ensemble forecast (PDF, spread and mean) over the majority of model grid points, but it's almost certain that it will, on the other hand, also generate extremely bad, misleading forecasts (in all aspects of

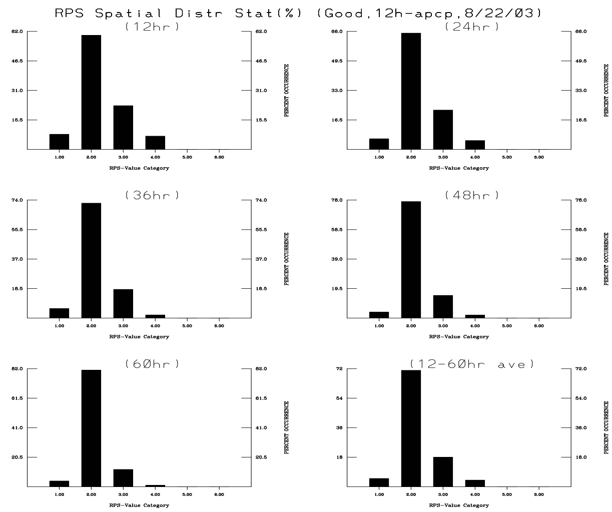

Figure 5: Same as Fig. 3 but from the very *"good"* EPS for 21z, Aug. 22, 2003 case.

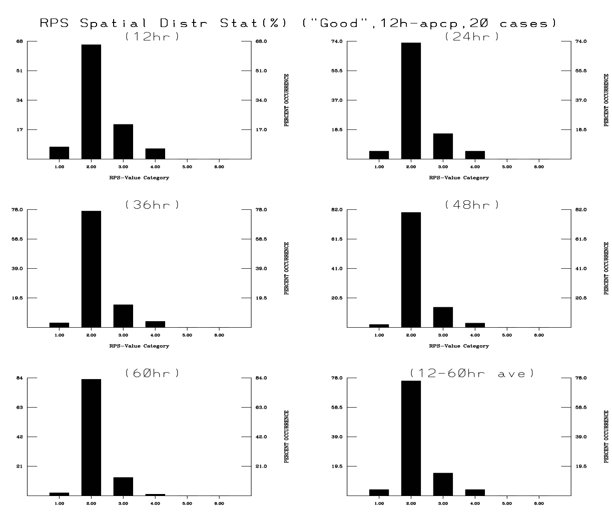

Figure 6: Same as Fig. 5 but averaged over the 20 cases.

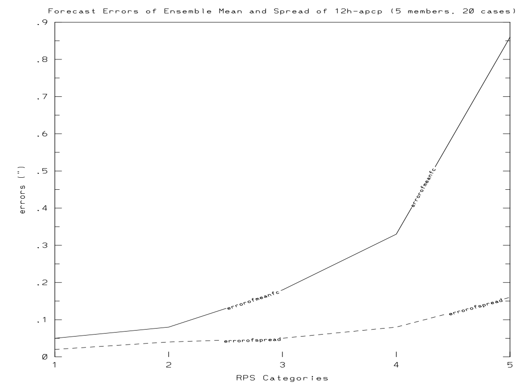

Figure 7: Increase of forecast errors with the decrease of probability forecast accuracy (from category 1 to 5). Solid curve is the absolute difference of ensemble mean forecasts between the *"good"* EPS and the *"true"* EPS, while the dash curve the absolute difference of ensemble spread between the same two EPSs, both averaged over regions with a same RPS category. Result is the average over all 20 cases.

mean, spread and probability) at some locations (the unpredictable spots). Those "unpredictable spots" are likely to be flow-dependent as well as model dependent. Their location varies from time to time and case to case. However, almost all current statistical post-processing methods (for bias, spread, PDF calibrations) for an EPS are based on the general performance (statistically) of its *past* forecasts; therefore, they are unlikely to help in correcting a *future* forecast over those "unpredictable spots" where calibration is really needed the most (while over other regions the forecasts are already reasonably good and don't need much correction!). Therefore, *a case-dependent or location-, time- and flow-dependent dynamical post-processing method is strongly desired.* The author of this paper is now researching several approaches into this issue. The Hybrid Ensembling method is one of these approaches (Du, 2004).

A major question is how to identify those "unpredictable spots" *beforehand*, and is it even possible to do so? Since those "unpredictable spots" might be likely associated with highly unpredictable regions where the ensemble spread should be large, the correlation between RPS score and ensemble spread was calculated over the entire model output domain and shown in Fig. 8. If PDF forecast accuracy is truly related to flow predictability, a positive correlation between RPS and ensemble spread should be observed. Unfortunately, Fig. 8 shows that "unpredictable spots" are not closely correlated to ensemble spread and only slightly positively correlated in general. It is suspected that this correlation might improve but only by focusing on those areas with extreme spread (either very large or near zero) rather than on the entire domain. This weak correlation implies that the task of locating those "unpredictable spots" is not an easy one, if not impossible, and that ensemble spread alone might not be a good indicator in this regard.

One might argue that as long as a majority of model grid points are well predicted, who cares about the small percentage number of grid points that have bad forecasts? First of all, one might expect that those "unpredictable spots" are possibly related to weather of interest (to be investigated) since extreme or high impact events are often associated with unstable and complex flow situations. Secondly, as a model's resolution becomes higher and higher and the end-users' requirement more and more sophisticated and detailed, such as mesoscale convection modes, cloud structures, city air pollution and nuclear material release in an urban area, these few locations could be very important to many users under certain circumstances.

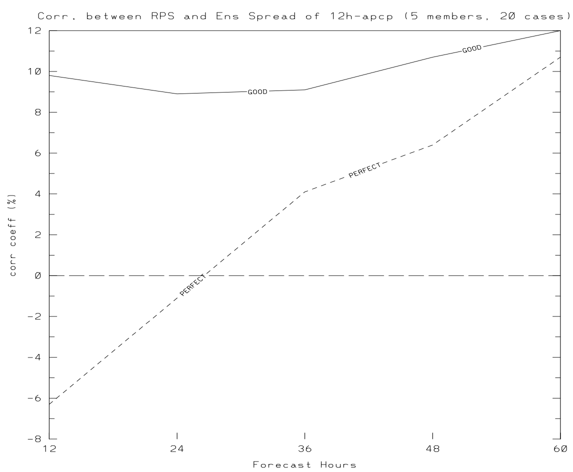

Figure 8: Spatial correlation between RPS score and ensemble spread for near *"perfect"* (dash) and *"good"* (solid) EPSs, averaged over the 20 cases.

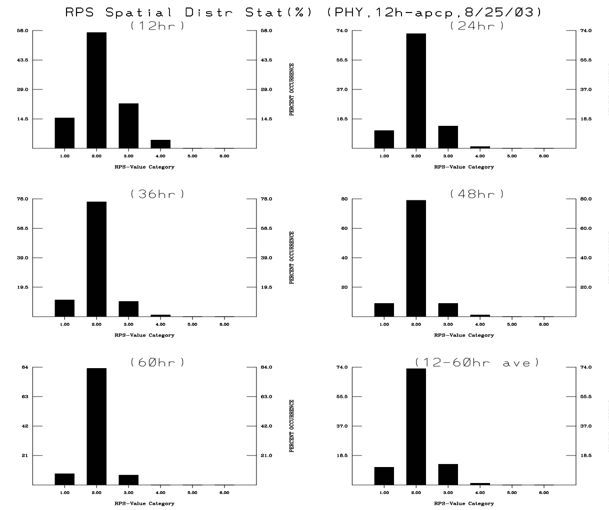

Figure 9: Same as Fig. 3 but for "perfect IC perturbations but slight model error" scenario ("Eta.KF" in Table 4) for 21z, Aug. 25, 2003 case.

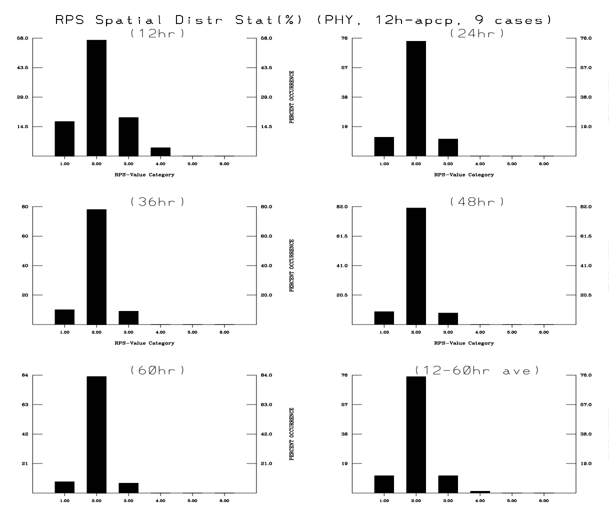

Figure 10: Same as Fig. 9 but averaged over the 9 cases.

3 Experiment II: Relative Importance Between Model Error and Imperfect IC Uncertainty Over PDF Evolution

Table 4 shows the design of Experiment II. Nine cases from August 2003 (Table 1) were investigated. Results are shown in Figs. 9-10 (a "model error only" scenario) and Figs. 11-12 (an "imperfect IC perturbations only" scenario). Figures 9-10 tell us that as long as a model has error (always the case in the real world), it is almost certain that there are some spots (about 2 % of grid points in caterories 4 and 5 for the average of the 9 cases) which cannot be predicted well even given perfect IC perturbations when initiating an EPS. Figures 11-12 imply that only when given a perfect model plus very realistic IC perturbations ("Eta.BMJx"), will good probability forecasts at almost all model grid points become a possibility, though

EPS	membership	model/physics	IC/perturbations	representativeness
Eta.BMJ	5	Eta with BMJ scheme	EDAS/bred from Eta.BMJ	truth
Eta.BMJx	5	Eta with BMJ scheme	EDAS/bred from Eta.KF	slight diff IC pert/no model error
Eta.KF	5	Eta with KF scheme	EDAS/bred from Eta.BMJ	same IC pert/slight model error

Table 4: Design of Experiment II: what is the relative importance between model error and imperfect IC perturbations over PDF evolution?

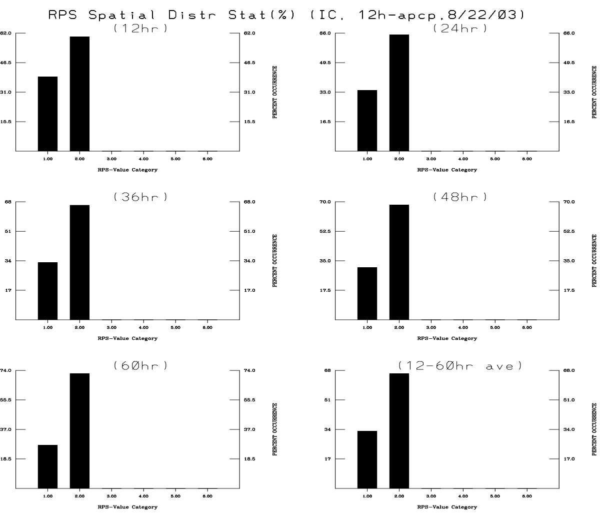

Figure 11: Same as Fig. 3 but for "perfect model but slightly imperfect IC pertuebations" scenario ("Eta.BMJx" in Table 4) for 21z, Aug. 22, 2003 case.

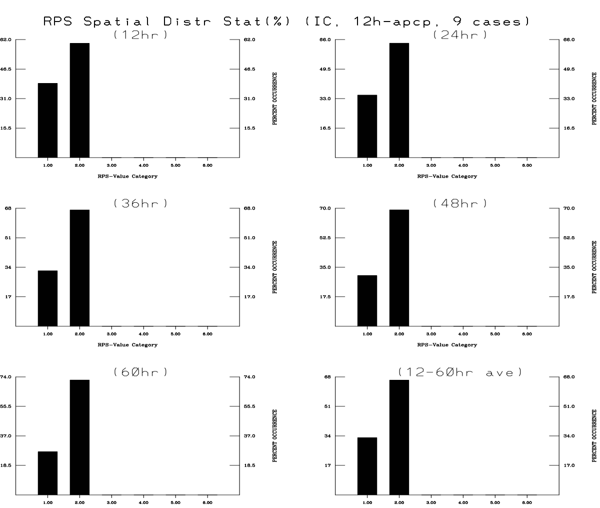

Figure 12: Same as Fig. 11 but averaged over the 9 cases.

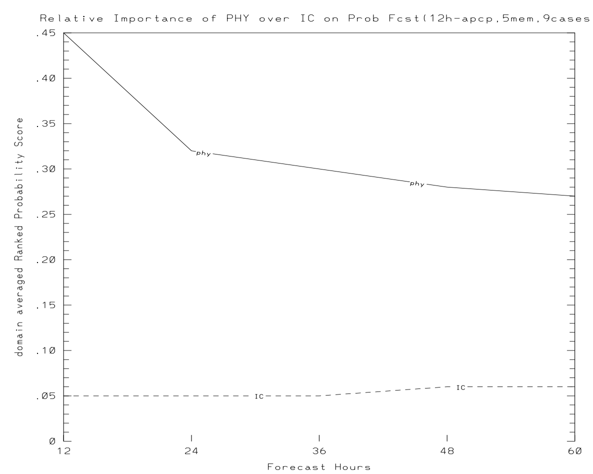

Figure 13: The domain averaged value of RPS scores for "model error only" (solid line) and "imperfect IC perturbations only" (dash line) scenarioes. The result is averaged over the 9 cases from August of 2003.

they are still not perfect forecasts (the majority of them are in category 2 and a small number of points (about 1 %) are still in category 3 which is not a good forecast).

Figure 13 is the domain averaged value of RPS scores for the nine cases for the scenarios "model error only" (corresponding to Fig. 10) and "imperfect IC perturbation only" (corresponding to Fig. 12). The result shows that the correctness of the model physics might play a more important role than that of IC perturbations in obtaining a better PDF forecast based on these limited cases. However, it is fully recognized that this result is not yet conclusive but needs further research. It is suspected that model error might depict "the big picture", such as what can and cannot be resolved (which is critical to the existence of a PDF for a particular event) and model climate shift (which affects the central location of a PDF), while IC uncertainties might give "the small picture", such as location, amounts, timing and detail structures. If so, the correctness of IC perturbations could be much more important than that of model physics in a practical sense (not in measuring scores) and from the users' point of view, since "big pictures" are relatively easier to systematically correct in post processing, such as in bias correction, but "small pictures" are basically unpredictable.

Experiment II vividly illustrates how tough it is to correctly predict PDF based on an EPS in operational environment where the model always has errors and the IC perturbation scheme is far from perfect.

4 Summary

Although a "good" ensemble system can produce good mean, spread and probability forecasts at the majority of model grid points, it's almost certain that it also can generate extremely bad and misleading forecasts in some locations, which are defined as "unpredictable spots". As long as the model used is imperfect, which is always the case in the reality, the number of "unpredictable spots" will never diminish even if the IC perturbations used in an EPS are perfect.

Since those areas have bad forecasts, "unpredictable spots" are, needless to say, the areas where calibration is needed the most. Given that the location of "unpredictable spots" varies from time to time and case to case, flow-dependent dynamical post-processing methods (rather than traditional statistically based approaches) are, therefore, needed to calibrate an ensemble prediction system. Some research is currently underway by the author of this paper (Du 2004) and at other locations such as the University of Washington (Grimit, personal communication).

Identifying the locations of "unpredictable spots" is important before doing any post-processing calibration of a forecast. Unfortunately, our result shows that it's difficult, if not impossible, to predict their locations because those spots are not well correlated with ensemble spread (predictability) in general. Further research is needed to study if "unpredictable spots" are related to particular weather systems or flow patterns to help in identifying them. It's also reasonable to expect that the forecaster's human expertise might play an important role in locating the "unpredictable spots".

The results from Experiment II imply that the correctness of model physics might be more important than that of IC perturbations in making a correct PDF forecast from an EPS. However, this is not a conclusive result, and needs further research. Answer could be different from different angles such as measuring by numerical scores or by its utility. In Experiment II, we see that only if given a perfect model (impossible in reality) and very realistic IC perturbations (difficult to achieve), is an ensemble system able to produce good (but still not perfect) forecasts in terms of ensemble mean, spread and probability distribution over nearly the entire model domain.

In short, the task of correctly predicting probability distribution or PDF using ensembles is extremely challenging if not impossible.

Acknowledgement Thanks to Prof. Fuqing Zhang of Texas A&M University and Dr. Eric Grimit of University of Washington for their scientific comments and to Mary Hart of NCEP for English language check.

References

Arakawa, A., and W. H. Shubert, 1974: The interaction of a cumulus cloud ensemble with the large scale environment, Part I. *J. Atmos. Sci.*, **31**, 674-701.

Black, T., 1994: The new NMC mesoscale Eta model: Description and forecast examples. *Wea. Forecasting*, **9**, 265-278.

Brooks, H. E., M. S. Tracton, D. J. Stensrud, G. DiMego, and Z. Toth, 1995: Short-range ensemble forecasting: Report from a workshop, 25-27 July 1994. *Bull. Amer. Meteor. Soc.*, **76**, 1617-1624.

Du, J., 2004: Hybrid Ensemble Prediction System: a New Ensembling Approach. *Symposium on the 50th Anniversary of Operational Numerical Weather Prediction*, University of Maryland, College Park, Maryland, June 14-17, 2004, Amer. Meteor. Soc., CD-ROM (paper p4.2, 5pp). [also available on the NCEP SREF web.]

Du J., 2002: Present Situation and Prospects of Ensemble Numerical Prediction. *Journal of Applied Meteorological Science*, Vol. 13, No. 1, 16-28. Meteorological Press, Beijing, China.

Du, J., J. McQueen, G. DiMego, T. Black, H, Juang, E. Rogers, B. Ferrier, B. Zhou, Z. Toth and M. S, Tracton, 2004: The NOAA/NWS/NCEP short-range ensemble forecast (SREF) system: evaluation of an initial condition vs multi-model physics ensemble approach. Preprints (CD), *16th Conference on Numerical Weather Prediction*, Seattle, Washington, Amer. Meteor. Soc.. [also available on the NCEP SREF web]

Du, J., S. L. Mullen, and F. Sanders, 2000: Removal of distortion error from an ensemble forecast. *Mon. Wea. Rev.*, **128**, 3347-3351.

Du, J., S. L. Mullen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, **125**, 2427-2459.

Du, J., and M. S. Tracton, 2001: Implementation of a real-time short-range ensemble forecasting system at NCEP: an update. Preprints, 9th Conference on Mesoscale Processes, Ft. Lauderdale, Florida, Amer. Meteor. Soc., 355-356. [also available on the NCEP SREF web]

Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.*, **8**, 985-987.

Hamill, T. M., S. L. Mullen, C. Snyder, Z. Toth, and D. P. Baumhefner, 2000: Ensemble forecasting in the short to medium range: report from a workshop. *Bull. Amer. Meteorol. Soc.*, **81**, 2653-2664.

Janjic, Z. I., 1994: The step-mountain eta coordinate model: further developments of the convection, viscous sublayer and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927-945.

Juang, H-M, and M. Kanamitsu, 1994: The NMC nested regional spectral model. *Mon. Wea. Rev.*, **122**, 3-26.

Judd, K., and L. A. Smith, 2001: Indistinguishable states I: Perfect model scenario. *Physics D.*, **151**, 125-136.

Kain, J., and J. Fritsch, 1993: Convective parameterization for mesoscale models: the Kain-Fritsch scheme. *Representation of Cumulus in Numerical Models.* Meteorological Monographs, Emanuel and Raymond Eds., **24**, 165-170.

Kain, J., and J. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784-2802.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409-418.

Lorenz, E. N, 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130-141.

Mullen, S. L., J. Du, and F. Sanders, 1999: The dependence of ensemble dispersion on analysis-forecast systems: implication to short-range ensemble forecasting of precipitation. *Mon. Wea. Rew.*, **127**, 1674-1686.

Murphy, J. M., 1971: A note on the ranked probability score. *J. Appl. Meteor.*, **10**, 155-156.

Palmer, T.N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parameterization in weather and climate prediction models. *Q. J. Meteorol. Soc.*, **45**, 198-207.

Palmer, T.N., R. Mureau, and F. Molteni, 1990: The Monte Carlo forecast. *Weather*, **45**, 198-207.

Pan, H.-L., and W.-S. Wu, 1995: Implementating a mass flux convection parameterization package for the NMC medium-range forecast model. *NMC Office Note 409*, 40pp. [Available from NCEP/NOAA.]

Parrish, D., and J. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Mon. Wea. Rev.*: **120**, 17471763.

Rogers, E., T. Black, D. Deaven, G. DiMego, Q. Zhao, M. Baldwin, N. W. Junker, and Y. Lin, 1996: Changes to the operational "early" eta analysis/forecast system at the National Centers for Environmental Prediction. *Wea. Forecasting*, **11**, 391-413.

Shaefer, J., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570-575.

Smith, L. A., 2000: Disentangling uncertainty and error: On the predictability of nonlinear systems. *Nonlinear Dynamics and Statistics*, A. I. Mees, Ed., Birkhauser, 31-64.

Stensrud, D., H. Brooks, J. Du, M. S. Tracton, and E. Rogers, 1999: Using Ensembles for Short-Range Forecasting, *Mon. Wea. Rev.*, **127**, 433-446.

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

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

Zhang, F., C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the 'surprise' snowstorm of 24-25 January 2000. *Mon. Wea. Rev.*, **130**, 1617-1632.