Uncertainty in numerical weather forecasting via ensembles of initializations

Yulia Gel,
University of Washington.

Joint work with Adrian Raftery and Tilmann Gneiting,
University of Washington

May 26, 2002

Due to the fact that the atmosphere is a chaotic system, any small error in the initial conditions may lead to increasing errors in the forecast, which eventually leads to a partial or total loss of predictive information. However, it may be possible to obtain some information on the predictability by analyzing a set of ensembles of initializations (initial conditions) which are within a certain neighborhood of the true estimate of the atmosphere. Ensemble forecasting is useful for improving average forecasting accuracy, estimating the range and distribution of possible outcomes, and even suggesting some extra locations of the observations. There is an extensive literature on the error analysis in numerical weather forecasting with application to the global synoptic models (Sivillo et al(1997), Berliner(2001)). So ensemble forecasting is becoming an increasingly popular technique in numerical weather prediction. Thus, the problem of optimal initial perturbations is an open and keen question.

Mesoscale models such as the MM5 need three main types of input data: a) the background field; b) lateral boundary conditions; c) settings of physical parameters. For numerical weather prediction the first two types are the main source of uncertainty. Hence, it will be our main target for analysis.

The main framework for producing ensemble of forecasts (and initializations) and integrating multisource information is that of Bayesian melding (Poole et al (2000), Fuentes et al). To calculate uncertainty, defined as a posterior predictive distribution, the following sources of information are available:

- current initial conditions from a global synoptic model on a grid, $\tilde{X}$.
- current forecast from MM5 on a grid, $\tilde{Y}$.
- historic information on $\tilde{X}$ on a grid.
- historic information about direct observations at stations, irregularly spaced, $\hat{X}$.

The problem of assessing uncertainty in the forecast may then be formulated as folllows:

Let X be the unknown "true" observations. To describe uncertainty, let X be modeled as a random spatio-temporal process. The task is to forecast a function of X, $M(X)$, where M is some mesoscale model, e.g. MM5. Thus, given a random vector X with joint probability function $q(X)$, find the implied probability function $\pi(Y)$ of the random vector $Y = M(X)$.

For now the main focus will be on developing a prior distribution of the observations X given the background field

$\tilde{X}$, which leads to the idea of simulating the initializations. An alternative is to calculate the posterior predictive distribution of the "true" forecast Y given the MM5 forecasts $\tilde{Y}$. For the implementation of the suggested methods the spatial correlation of the data is investigated and taken into account. This talk will overview these two approaches for simulations of ensembles of initializations and ensembles of forecasts.

Acknowledgments.

This research was partially supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the US Office of Naval Research under Grant N00014-01-10745.

References

[1] Berliner, L. Mark. (2001). Monte Carlo based ensemble forecasting. *Statistics and Computing* **10**, 269–275.

[2] Fuentes, M. and Raftery, A.E. (2001). Model validation and spatial interpolation by combining observations with outputs from numerical models via Bayesian melding . *Technical Report, Department of Statistics, University of Washington*, no. 403.

[3] Poole, D.J. and Raftery, A.E. (2000). Inference for deterministic simulation models: The Bayesian melding approach. To appear in *Journal of American Statistical Association*.

[4] Sivillo, J.K. and Ahlquist, J.E. (1997). An ensemble forecasting primer. *American Meteorological Society* **12**, 809–818.