USE OF THE BREEDING TECHNIQUE IN THE ESTIMATION OF THE BACKGROUND COVARIANCE MATRIX FOR A QUASI-GEOSTROPHIC MODEL

M. Corazza[1,2], E. Kalnay[1], D. J. Patil[1], E. Ott[1], J. A. Yorke[1], B. R. Hunt[1], I. Szunyogh[1] and M. Cai[1]
[1] University of Maryland, College Park, 20742 MD, USA
[2] INFM – DIFI, Università di Genova, 16146 Genova, Italy

It is well known that numerical weather predictions are sensitive to small changes in the initial conditions, i.e., a rapid growth of the initial errors can lead in a relatively short time to large forecast errors. During the last decades much effort has been devoted to study and improve the methods used for the preparation of the initial conditions for numerical atmospheric models (the so-called analysis), as well as to understand the mechanisms involved in the growth of the initial errors.
The analysis is obtained as a statistical interpolation of short-range numerical forecasts (known as background) with new observations. The weight given to each of these contributions is essentially proportional to the inverse of their error covariance. It follows that a good representation of the observation and background error covariances is one of the major goals in the development of data assimilation systems (e.g. Klinker et al., 2000; Bennet et al., 1996; Houtekamer and Mitchell, 1998; Hamill and Snyder, 2000). In 3D-Variational schemes the background error covariance matrix is statistically derived from long term statistical estimations and it is maintained *constant in time* during the assimilation cycle. This implies that the large time dependence of the errors ("errors of the day") is neglected, despite its large variability (Corazza et al, 2001).

Kalnay and Toth (1994) argued that the similarity between breeding (Toth and Kalnay, 1993, 1997) and data assimilation suggests that the background errors should have a structure similar to those of bred vectors. Here we take advantage of the relationship between the bred vectors and the analysis and background errors demonstrated by Corazza et al., 2001, for a simple quasi-geostrophic model (Morss, 1999), and test whether it is possible to augment the constant forecast error covariance used in 3D-Var with "errors of the day" derived from the breeding method. We present results obtained with different methods aimed to include in the data assimilation scheme (i.e., in the representation of the background error covariance matrix) the information given by the bred vectors.

The numerical model is a quasi-geostrophic (QG) mid-latitude flow in a channel discretized by finite differences both in horizontal and vertical directions. The simulated data assimilation is performed with an algorithm similar to the operational Spectral Statistical Interpolation (SSI) at NCEP (Parrish and Derber, 1992). "Rawinsonde observations" are generated every 12 hours by randomly perturbing the true state at fixed observation locations. Bred vectors are produced using a method similar to that adopted at NCEP (Toth and Kalnay 1993,1997), rescaling the difference between the perturbed runs and the control forecast every 12 hours.

Since this is a simulation system, we can explicitly define the "true state of the atmosphere" (by integrating the model from a given initial state) and therefore study the analysis and forecast errors. A perfect model assumption is made so that our conclusions are not necessarily valid for more complex models with model errors, and similar tests have to be made with more general simulation systems and with real forecast systems.

Corazza et al. (2001) found that bred vectors in this QG simulation system are indeed closely related to the background errors, suggesting that the bred vectors can be useful in specifying the part of the background error covariance matrix that corresponds to the "errors of the day". In particular we found the following properties of the bred vectors:

- Convergence to well organized structures in the bred vectors occurs within a few (3-5) days. This indicates that it is possible to operationally use the information given by the bred vectors without waiting an infinite time for asymptotic convergence.
- Bred vectors obtained using normalizations based on the potential vorticity and on the stream function are virtually indistinguishable.
- Bred vectors obtained using the "true" atmosphere are very similar to those obtained using the "analysis" atmosphere. This is true even if we use a low density observing network, suggesting that the bred vectors are not too sensitive to the details of the flow and that the errors themselves are more likely dependent on the large scale nature of the flow.

The original data assimilation cycle (referred to as the *regular* data assimilation system) is based on the NCEP 3D-Var scheme, and is solved iteratively for the analysis state $\mathbf{x}_a$ (Morss, 1999). Given the background state (or first guess - the 12 hour forecast from the previous analysis) $\mathbf{x}_b$, and the set of observations $\mathbf{y}_o$, the equation can be written as follows:

$$(\mathbf{I}+\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})(\mathbf{x}_a-\mathbf{x}_b)=\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}_o-H(\mathbf{x}_b))$$

where **B** is the background error covariance matrix, **R** is the observation error covariance matrix, *H* is the observation operator and **H**, $\mathbf{H}^{\mathrm{T}}$ are the matrices that represent the linearized *H* and its transpose respectively. The right part of the equation is computed at the beginning of the process, and the equation is then iteratively solved for $(\mathbf{x}_a-\mathbf{x}_b)$ until the equation is satisfied with an error smaller than a given threshold.

The easiest way to introduce the bred vectors in this equation is to globally substitute **B** with the ensemble average of the outer product of the bred vectors. We can build a new background covariance matrix as $\sum_{i=1}^{k}\mathbf{b}_i\mathbf{b}_i^{\mathrm{T}}/k$ where $\mathbf{b}_i$ is the i^{th}

bred vector defined over the entire domain. The substitution of **B** with the new matrix can be done at a negligible computational cost. Moreover, this implementation allows to apply $\sum_{i=1}^{k} \mathbf{b}_i \mathbf{b}_i^{\mathrm{T}} / k$ and **B** simultaneously (Hamill and Snyder, 2000) so that the data assimilation scheme can be generalized to:

$$\left(\mathbf{I}+\left(\alpha\frac{c}{k}\sum_{i=1}^{k}\mathbf{b}_i\mathbf{b}_i^{\mathrm{T}}+(1-\alpha)\mathbf{B}\right)\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)(\mathbf{x}_a-\mathbf{x}_b)=\left(\alpha\frac{c}{k}\sum_{i=1}^{k}\mathbf{b}_i\mathbf{b}_i^{\mathrm{T}}+(1-\alpha)\mathbf{B}\right)\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}_o-H(\mathbf{x}_b)\right)$$

where α is a number between 0 (for the regular system) and 1 (background error covariance matrix based fully on bred vectors) and *c* is a normalization factor kept constant in the results presented here. It should be noted that the covariances in **B** were tuned to optimize the regular 3D-Var.

The use of the bred vectors allows decreasing the squared error of the analysis (averaged over the horizontal domain) by a factor between 15 and 20% (around 8-10% in the error) for α equal to 0.4. The percentage improvement continues throughout the 72 hour forecast, suggesting that the correction to the analysis due to the bred vectors affects, at least in part, the growing errors. This method is able to reduce the squared error in the analysis and forecasts up to a factor of 40% considering a modified version of the bred vectors aimed to take into account random observational errors introduced in the analysis step. Simulations using these vectors show a remarkable improvement of the performance of the assimilation cycle with respect to the one based on the standard bred vectors without random "reseeding". It is interesting to note that for large values of α, when the role of the statistically derived **B** is small, the augmented system is not able to maintain the error small. This indicates that the space spanned by the bred vectors is not large enough to represent all the error directions, and that the contribution of the regular part of the assimilation scheme cannot be neglected when using global methods to include bred vectors in the data assimilation cycle. This was also observed using local methods (not shown).

Motivated by the studies performed by Patil et al. (2001) we are presently testing new methods to locally use the bred vectors in the data assimilation system. The use of local methods is desirable in order to optimise the information given by two or more bred vectors, which may be, for example, positively correlated in one area and negatively correlated in another area far away. The background error correlations should vanish beyond a limited horizontal extent, whereas our use of global bred vectors implies correlations over the global domain. An example of local use of the bred vectors can be derived as a generalization of the method proposed by Kalnay and Toth (1994).

REFERENCES

Bennett, A. F., B. S. Chua, and L. M. Leslie, 1996: Generalized inversion of a global NWP model., *Meteor. Atmos. Phys.*, **60**, 165-178.

Corazza, M., E. Kalnay, D. J. Patil, R. Morss, I. Szunyogh, B. R. Hunt, E. Ott, and M. Cai, 2001: Use of the breeding technique to estimate the structure of the analysis "errors of the day". Submitted to Nonlinear Processes in Geophysics.

Hamill, T. M., and C. Snyder, 2000: A Hybrid Ensemble Kalman Filter – 3D Variational Analysis Scheme., *Mon. Wea. Rev.*, **128**, 2905-2919.

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

Kalnay, E., and Z. Toth, 1994: Removing growing errors in the analysis cycle., *Tenth Conference on Numerical Weather Prediction – Amer. Meteor. Soc.*, pp. 212-215.

Klinker, E., F. Rabier, G. Kelly, and J.-F. Mahfouf, 2000: The ECMWF operational implementation of four dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration., *Quart. J. Roy. Meteor. Soc.*, **126**, 1191.

Morss, R. E., 1999: *Adaptative observations: Idealized sampling strategies for improving numerical weather prediction.*, Ph. D. thesis, Massachusetts Institute of Technology, 225 pp.

Parrish, D. F., and J. D. Derber, 1992: The National Meteorological Center spectral statistical interpolation analysis system., *Mon. Wea. Rev.*, **120**, 1747-1763.

Patil, D. J. S., B.R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott, 2001: Local Low Dimensionality of Atmospheric Dynamics, *Phys. Rev. Lett.*, **86**, 5878.

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

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP: the breeding method., *Mon. Wea. Rev.*, **125**, 3297- 3318.