6.4 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

1. INTRODUCTION

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.

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).

There are several methods that try to account for the errors of the day in the forecast error covariance, including 4D-Var, Kalman Filtering, and the method of representers (e.g. Klinker et al., 2000; Bennet et al., 1996; Houtekamer and Mitchell, 1998; Hamill and Snyder, 2000). Unfortunately, these methods are computationally very expensive, and can only be implemented with substantial shortcuts, such as the use of a reduced rank background error covariance matrix in Kalman Filtering and lower model resolution in 4D-Var. It follows that it is important to develop new, low-cost methods aimed to improve the estimation of the background error covariance matrix.

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. They also proposed a simple method to include the information of the bred vectors in the data assimilation cycle.

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. Figure 1 shows a typical example of the largest values of the midlevel background error in potential vorticity (solid lines) against two arbitrarily chosen bred vectors (dotted and dashed lines respectively). The patterns are in good agreement in some areas, and agree less well in other areas. In particular, the agreement is higher in those areas where the background error is larger. In general, these results are valid also for the vertical structure of the fields and at different observational densities [only 16 observations were used in Figure 1,

[*] *Corresponding author address:* Eugenia Kalnay, Meteorology Department, University of Maryland, College Park, MD 20742-2425, USA;
email: ekalnay@atmos.umd.edu

with similar results obtained with 32 and 64 “rawinsondes” observation locations].

We found (Corazza et al., 2001) 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.

Following Patil et al. (2001) we also computed a measure of the effective dimensionality of the space spanned by the bred vectors called Bred Vector Dimension (BV-Dimension). For every point a surrounding domain of 25 grid points (5×5) is selected. We consider the 25 dimensional vector composed of the values of potential vorticity over the grid points of the domain, which we refer to as a *local vector*. The BV-Dimension is a measure of the *degree of linear independence of the k local vectors* (k being the number of bred vectors) and is defined as:

$$\Psi(\sigma_1,\sigma_2,...,\sigma_k)=\left(\ \sum_{i=1}^{k}\sigma_i\right)^2/\sum_{i=1}^{k}\sigma_i^2\ ,$$

where σ_i, $i=1,...,k$ are the singular values of the $k\times k$ covariance matrix of the 25×k matrix formed by the local vectors. Note that $\Psi\le k$.

We found that for k=10 the local dimension of the subspace is always smaller than 6.5 and that in areas where the background errors are large it is usually between 2 and 4. These results did not change when a larger number of bred vectors was used, indicating that 10-15 vectors are enough (at least for a quasi-geostrophic system) to give a complete representation of the space spanned by the bred vectors.

We also defined the local angle between the forecast error and the subspace of bred vectors. We found that in most of the area the angle is confined to less than 10° (cosine larger than 0.985), suggesting that the background error is mostly confined to the subspace of the bred vectors. As in the case of the BV-Dimension, the areas with an angle larger than 10° generally had small background errors.

Similar results were obtained computing the local pattern correlation over the same 25 points between each bred vector and the background error and then identifying the maximum correlation at each point. We obtained high correlations in most of the domain of the model, which indicate that at least one bred vector has a structure similar to that of the error, in particular where the background error is large.

These results suggest that the bred vectors can indeed be useful in specifying the part of the background error covariance matrix that corresponds to the “errors of the day”. In the following section we discuss some computationally inexpensive methods to take advantage of this information.

BRED VECTORS IN DATA ASSIMILATION

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 x_a (Morss, 1999). Given the background state (or first guess - the 12 hour forecast from the previous analysis) x_b, and the set of observations y_o, the equation can be written as follows:

$$\left(\mathbf{I}+\mathbf{B}\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{H}\right)\left(x_a-x_b\right)=\mathbf{B}\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\left(y_o-H\left(x_b\right)\right)$$

$\mathbf{B}$ is the background error covariance matrix, $\mathbf{R}$ is the observation error covariance matrix, H is the observation operator and $\mathbf{H}$, $\mathbf{H}^{\mathbf{T}}$ 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, then the equation is iteratively solved for $\left(x_a-x_b\right)$ 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 $\mathbf{B}$ with the ensemble average of the outer product of the bred vectors. We can build a new background covariance matrix as $\sum_{i=1}^{k}\mathbf{b}_i\mathbf{b}_i^{\mathbf{T}}/k$ where $\mathbf{b}_i$ is the i^{th} bred vector defined over the entire domain. The substitution of $\mathbf{B}$ with the new matrix can be done at a negligible computational cost; moreover, this implementation allows to apply $\sum_{i=1}^{k}\mathbf{b}_i\mathbf{b}_i^{\mathbf{T}}/k$ and $\mathbf{B}$ simultaneously (Hamill and Snyder, 2000) so that the data assimilation scheme can be generalized to:

$$\left(\mathbf{I}+\left(\alpha\ \ \frac{c}{k}\sum_{i=1}^{k}\mathbf{b}_i\mathbf{b}_i^{\mathbf{T}}+(1-\alpha)\mathbf{B}\right)\ \mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{H}\right)\ \left(x_a-x_b\right)=$$
$$\left(\alpha\ \ \frac{c}{k}\sum_{i=1}^{k}\mathbf{b}_i\mathbf{b}_i^{\mathbf{T}}+(1-\alpha)\mathbf{B}\right)\ \mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\left(y_o-H\left(x_b\right)\right)$$

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 $\mathbf{B}$ were tuned to optimize the regular 3D-Var.

In Figure 2 we show the squared analysis and forecast (up to 72 hours) errors in mid-level potential vorticity for the optimized regular system (diamonds) and for the simulation obtained using 10 bred vectors with α=0.4 (squares). The results are obtained averaging the errors over the horizontal domain every 12 hours for a one-year simulation. The use of the bred vectors allows decreasing the squared error by a factor between 15 and 20% (around 8-10% in the error). Fig. 2 shows that 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.

We want to understand whether the bred vectors are able to represent the evolution of the fast growing errors of the day associated with the short term forecast starting from the previous analysis. The errors of the analysis are due to forecast (background) as well as to observational errors, which constantly introduce random errors. However, the "classical" formulation of the bred vectors takes into account only the errors due to the forecast. For this reason, we modified the method to generate the bred vectors including "reseeding" with random "observational" errors. After every renormalization step (12 hours), we generate random errors Δy at the observation locations simulating the impact of the observational errors. Then we apply the transpose of the observation operator to obtain $\Delta x = \mathbf{H}^{\mathbf{T}} \Delta y$, and add Δx to the rescaled bred vectors.

The third line (bottom - triangles) of Figure 2 shows the results obtained for this system using the same data assimilation scheme as before. The squared error in midlevel potential vorticity is now 40% smaller than that of the regular system (22% in the error). The relative improvement decreases slightly within the forecast, but it is still around 30% after 72 hours. More work has to be done to better understand why adding random errors to the bred vectors leads to such a large improvement. It may be due in part to the fact that adding random noise does not allow bred vectors to collapse into too few directions, and therefore to span a larger subspace that better represents the fast growing modes. The fact that after 72 hours the improvement is still far better than in the standard bred vector simulation indicates that the role of the random errors evolved for 12 hours upon the bred vectors is not limited to a mere correction of mostly decaying analysis errors uniformly distributed in phase space.

Figure 3 shows the relative improvement (with respect to the regular 3D-Var data assimilation system) in the squared error in potential vorticity at midlevel, for the simulation obtained using bred vectors reseeded with random errors, and for different values of α. There is a remarkable impact of the background error covariance matrix derived from the bred vectors even for very small values of α, indicating that the new matrix is able to focus the corrections in the background state along the most unstable directions even when it is not the dominant part of $\mathbf{B}$. It is interesting to note that for large values of α, when the role of the statistically derived $\mathbf{B}$ 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 to be neglected when using global methods to include the bred vectors in the data assimilation cycle. This was also observed using local methods (not shown).

4. SUMMARY AND DISCUSSION

-Variational data assimilation scheme for a quasi-geostrophic channel model (Morss, 1999) to study the structure of the background error and its relationship to the corresponding bred vectors. The results of Corazza et al. (2001) show that the bred vectors have spatial and temporal characteristics similar to those of the background "errors of the day". They pointed out some properties of the bred vectors, including fast convergence to well organized structures, independence on the choice of the renormalization norm (potential vorticity and stream function) and similarity between bred vectors generated from the "truth" and from the analysis. We also showed that the subspace spanned by the bred vectors is able to locally describe most of the structure of the background error.

Starting from these considerations we described a global approach to include the bred vectors in the data assimilation scheme and we showed that this method is able to reduce the squared error in the analysis and forecasts up to a factor of 40%. Moreover, we considered a modified version of the bred vectors aimed to take into account the 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".

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 optimize 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). If we assume for simplicity that the forecast error covariance is given locally by a single bred vector, $\mathbf{B} = \mathbf{b}\mathbf{b}^{\mathbf{T}}$, that the observational error covariance is $\mathbf{R} = \rho^2 \mathbf{I}$, and perform a preliminary analysis, then the Optimal Interpolation solution

$$\delta x = (\mathbf{BH^T})(\mathbf{R}+\mathbf{HBH^T})^{-1} \quad (y_o - H(x_b))$$

becomes simply

$$\delta x = x_a - x_b = \frac{\mathbf{bb^T P^T H^T}(y_o - H(x_b))}{\mathbf{b^T b} + \rho^2},$$

where $\mathbf{P^T}$ is the projector operator from the global domain to the local domain. This system has a very low computational cost. A similar formula is valid for a set of locally orthogonalized bred vectors. As long as the preliminary analysis remains confined to the unstable subspace of the bred vectors, the computational cost remains low.

REFERENCES

Generalized inversion of a global NWP model., *Meteor. Atmos. Phys.*, **60**, 165-178.

Corazza, M., E. Kalnay, D. J. Patil, R. Morss, M. cai, 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.

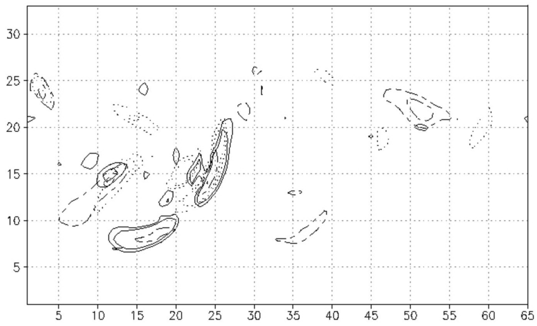

Fig. 1. Background error (solid line) and two randomly chosen bred vectors (dotted and dashed lines respectively) after 36 days from their initialization. The fields are normalized by their root mean square and contour lines are plotted for the values 2.3 and 3.

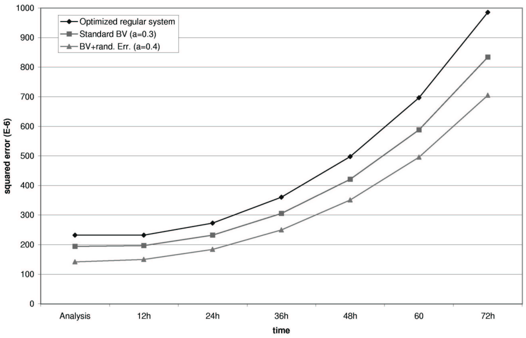

Fig. 2. Squared errors in midlevel potential vorticity for the regular system, the standard bred vector system (α=0.3) and the bred vectors + random noise system (α=0.4).

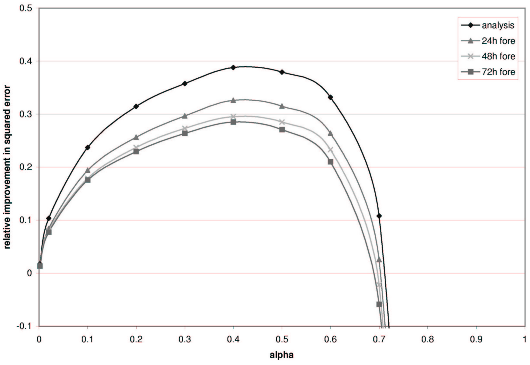

Fig. 3. Relative improvement with respect to the regular data assimilation system in the analysis and forecast squared errors in midlevel potential vorticity varying α from 0 to 1, for the simulation with bred vectors obtained adding random noise after every normalization step.