Data Assimilation via Local Ensemble Kalman Filtering

E. Kalnay, M. Corazza[1], S.-C. Yang (Department of Meteorology), B. Hunt, E. Kostelich, E. Ott, D. J. Patil, I. Szunyogh, J. Yorke, A. Zimin

Institute for Physical Science and Technology
University of Maryland, College Park 20742

Abstract

"Errors of the day" due to instabilities of the background flow, dominate analysis and forecast errors. Bred vectors are the difference between two nonlinear model integrations, periodically rescaled to avoid nonlinear saturation of the instabilities of interest. We test the conjecture by Kalnay and Toth (1994) that bred vectors represent the same instabilities that generate the "errors of the day" using a simulated data assimilation system based on a quasi-geostrophic model. We first show that bred vectors obtained by perturbing the analysis (not the truth) have shapes similar to the dominant "errors of the day", and that 10 bred vectors are enough to explain 97% of the forecast variance. Surrogate bred vectors, obtained by using 10 bred vectors corresponding to 10 randomly chosen days are able to locally explain 87% of the variance. This should be representative of the level attainable by the constant background error covariance used in 3D-Var data assimilation. We argue that the use of local bred vectors, rather than global Lyapunov vectors, reduces substantially the number of vectors required to represent well the background errors.

A Local Ensemble Kalman Filtering formulation has been developed by Ott et al (2002), in which bred vectors are computed globally, and a local Kalman Filter analysis is used to rescale them within the subspace of the locally dominant bred vectors. Preliminary results of the application of this method to the QG system, including the addition of random perturbations to the bred vectors, are very encouraging.

1.   Introduction

Forecast errors can originate from errors in the initial conditions that, due to the chaotic nature of the atmosphere, grow with time, or from model deficiencies (which we do not consider in this paper). Because the error growth is not uniform, but is associated with instabilities of the background flow, mid-latitude forecast errors tend to be dominated by relatively large "errors of the day" intermittent in space and in time. Errors of the day tend to have the same geographical location and similar shapes for forecasts of different lengths verifying at the same time. This suggests that they also dominate analysis errors since they originate from errors in the forecasts used as a background field within the analysis cycle.

Kalnay and Toth (1994) conjectured that bred vectors, the periodically rescaled difference between two nonlinear model integrations, represent the same instabilities that generate the errors of the day and can therefore be used to estimate the shape of forecast errors. The intermittent presence of these instabilities in the background flow provides a simple explanation of the observation of Patil et al (2001) that ensembles of bred vectors show the existence of regions of low dimensionality with well defined 3-dimensional structure, and which evolve in time with life cycles of the order of 3-7 days.

In section 2 we review several properties of bred vectors in a quasi-geostrophic data assimilation simulation system (Morss et al, 2001), show that bred vectors represent very well forecast errors (Corazza et al, 2002), and suggest why fewer bred vectors are needed when they are used locally rather than globally. In section 3

---

[1] INFM-DIFI, University of Genoa

we briefly discuss a potentially accurate and efficient approach called "Local Ensemble Kalman Filtering" and developed by Ott et al (2002) that takes advantage of this property, and present preliminary results.

2. Bred vectors in a QG simulated data assimilation system

In this section we present results using breeding in the quasi-geostrophic data assimilation system developed by Morss et al (2001), dealing first with the relationship between bred vectors and forecast errors from Corazza et al (2002). Preliminary results of the application of a Local Ensemble Kalman Filtering (LEKF) data assimilation using the method of Ott et al (2002) are presented in the following section.

The control data assimilation is performed using 3D-Var based on Parrish and Derber (1992). "Truth" in this simulation is defined as a very long model run, from which observations are obtained at randomly located "rawinsondes stations" with random observational errors. As in Morss et al (2001), the QG channel model is used for both the truth and the data assimilation (a "perfect model" assumption). Breeding is performed using only the analysis (without knowledge of the truth), as it would be done in an operational setting (Toth and Kalnay, 1993).

2.1. Dependence of the bred vectors on the norm

Figure 1 shows an example of bred vectors obtained by rescaling with a potential vorticity squared (left) and a streamfunction squared norm (right), superimposed over the background errors corresponding to the same analysis time. This figure shows that a) the errors of the day have strong similarity with the bred vectors, and b) that the characteristics of the bred vectors are not sensitive to the choice of norm. We have found that these results hold true with both dense and sparse "observation networks", indicating that the instabilities that dominate the errors of the day are more sensitive to the large scale characteristics of the background flow, and less to the analysis errors.
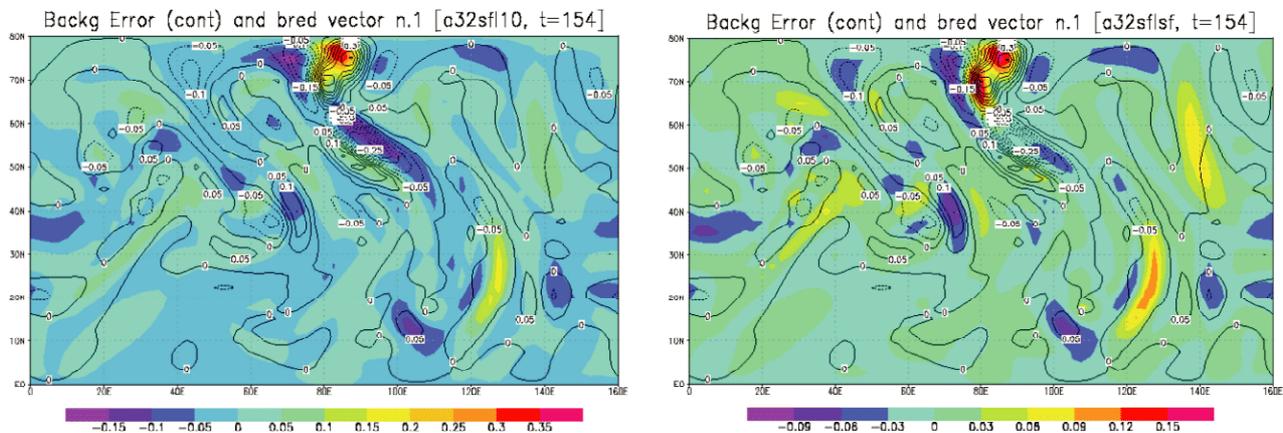

Figure 1 Example of background error field (contours) superimposed with bred vectors corresponding to the same analysis time, derived using the enstrophy norm (left) and the streamfunction squared norm (right).

2.2. Explained variance of the forecast error by the bred vectors

The example in Figure 1 is typical, and we have found that with 10 bred vectors it is possible to explain 97% of the variance of the "errors of the day" (Fig. 2a). Figure 2a also shows that if surrogate bred vectors are used (corresponding to randomly chosen times), about 87% of the background error variance is explained. This may be representative of the level attainable by the constant background error covariance used in 3D-Var data assimilation, since the "NMC method" used to create a constant background error covariance is based on an ensemble of differences between 2-day forecasts minus 1-day forecasts verifying at the same time, and these differences should be dominated by bred vector structures. Figure 3 is a simple schematic

example that illustrates why using bred vectors locally to determine the shapes of the unstable subspace (even though they are computed globally) is more efficient than using either global bred vectors, or Lyapunov vectors that require global orthogonalization (Kalnay et al, 2002).
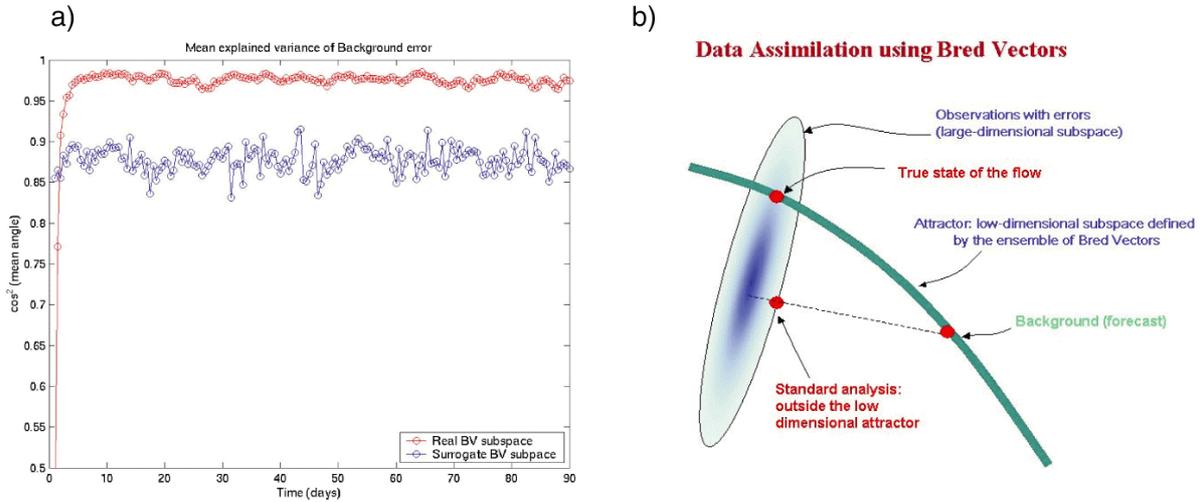
a)   b)

Figure 2 a) mean explained variance of the background errors using 10 bred vectors of the day (red) and10 surrogates (bred vectors corresponding to randomly chosen times, in blue). b) Schematic suggesting why the subspace of bred vectors should be used to correct the background errors along the attractor of the system, rather than with a statistical approximation obtained with a constant background error covariance.
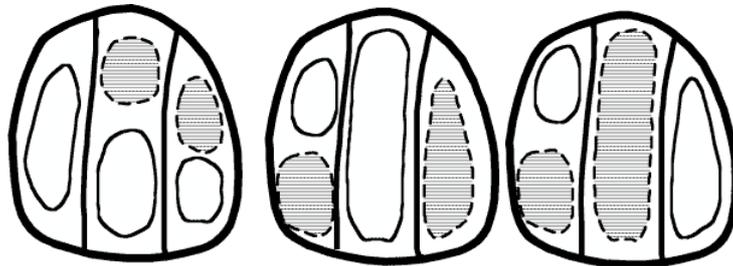

Figure 3 Schematic showing why the use of local bred vectors is more efficient than global bred vectors. The total domain in the example has three independent regions of instability, each of which can have an unstable mode with either wave number 1 or wave number 2, and 3 bred vectors are shown. Negative perturbations are shaded. From a local point of view, the first two bred vectors are sufficient to represent all possible instabilities, whereas from a global point of view, the second and third bred vectors are independent, and a larger number of BV are required to represent all independent modes.

3.   Application to data assimilation

The results presented in Figure 2a suggest that the use of local bred vectors (instead of global vectors, see Kalnay et al, 2002), provides the potential for computationally efficient data assimilation including the errors of the day. Ott et al (2002, posted at arXiv.org/abs/physics/0203058) have formulated a Local Ensemble Kalman Filtering (LEKF) method. They take advantage of the local low-dimensionality found by Patil et al (2001), suggesting that the analysis should also lie within this subspace (Fig. 2b), so that operations are done on relatively low dimensional matrices.

In the LEKF, the data assimilation is done locally allowing for massively parallel computations, and the resulting local Kalman Filter analysis error covariances are used to rescale the bred vectors, creating initial global states for forecasting to the next analysis time. Figure 4 compares background errors and the

corresponding analysis increments obtained with observations at a given analysis time. It is apparent that the corrections in the 3D-Var method, based on a constant error covariance, tend to be isotropic. In the LEKF, on the other hand, the same observations lead to corrections much closer to the background errors, because they are based on bred vectors that know about the errors of the day.
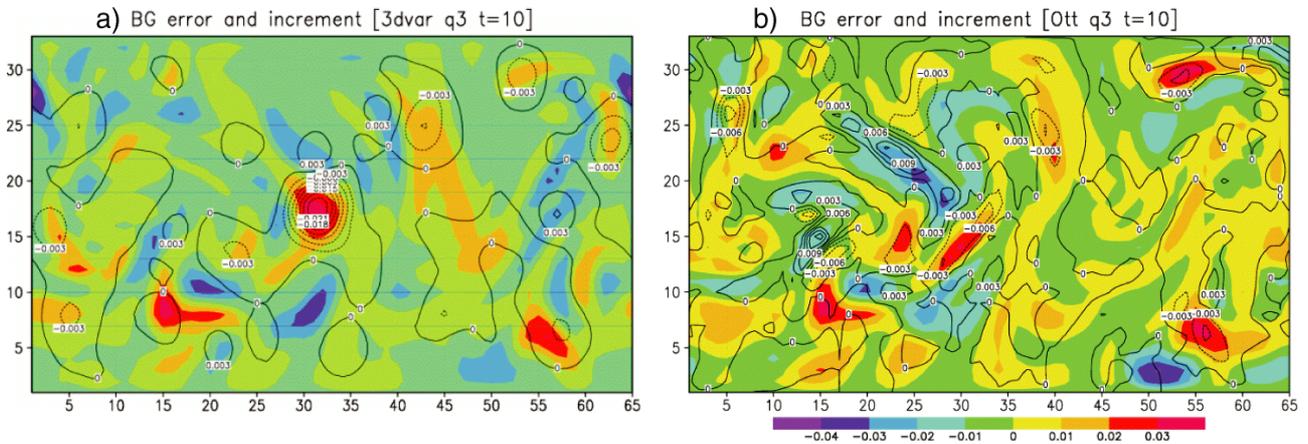

Figure 4 Example of background errors (colors) and analysis increments (contours) using a) 3D-Var and b) Local Ensemble Kalman Filtering (right).

In Figure 5a we present the analysis errors (compared with the truth) of the 3D-Var data assimilation, the LEKF method without variance inflation, and LEKF with the addition of small random perturbations (as if they were analysis errors) to the bred vectors, which increase slightly the size of the bred vectors and the dimension of the subspace that they span. The latter method produces remarkably good results, similar to those obtained by inflating the background error variance by a factor of 1.1 (as discussed, e.g., in Whitaker and Hamill, 2002, not shown). Figure 5b indicates that the advantages of LEKF are maintained during the forecasts. Experiments performed with the Lorenz 40-variable model have also yielded similar promising results.
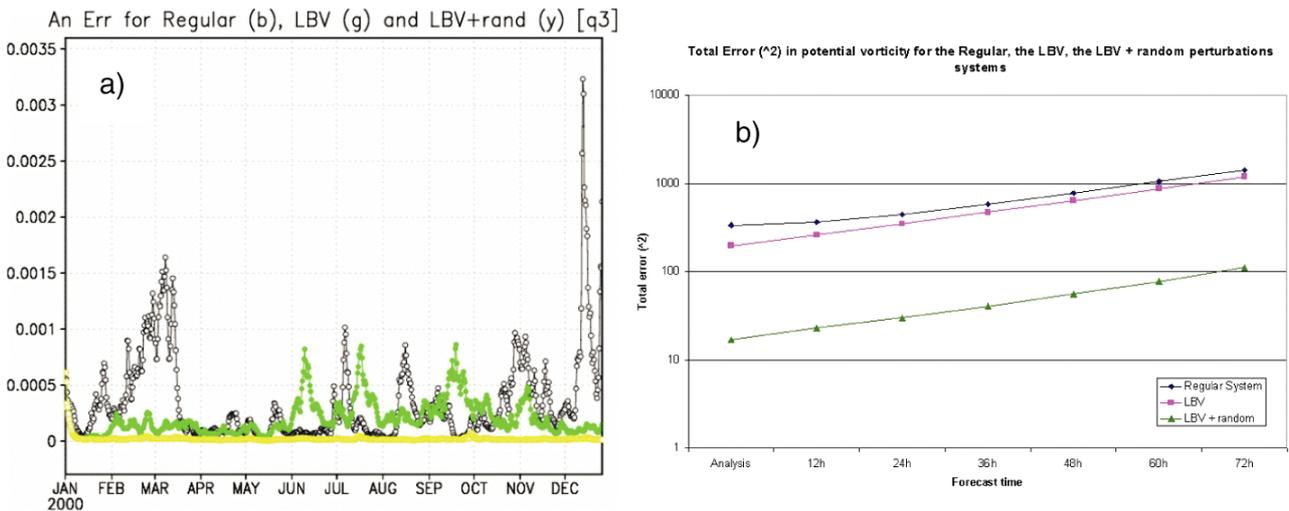

Figure 5 a) One year of domain averaged analysis errors using regular 3D-Var (black), the Local Ensemble Kalman Filtering (green) and LEKF with the addition of random errors (yellow). b) Average forecast errors computed from the 3 analyses in a).

4.   Final comments

Although we presented "identical twin" experiments with simple models, we believe that similar encouraging results could be obtained with more realistic models. We are developing a version of the LEKF with the NCEP global data assimilation system. The question of model errors may be addressable with the use of multi-systems, including a poor person ensemble of operational systems.

5.   References:

Corazza, M., E. Kalnay, D. J. Patil, I. Szunyogh, B.R. Hunt, E. Ott and J. A. Yorke, 2002:Use of the breeding technique to estimate the structure of the analysis "errors of the day". *Nonlinear Processes in Geophysics* (under review).

Kalnay, E., and Z. Toth, 1994: Removing growing errors in the analysis. Preprints, 10th *AMS Conference on Numerical Weather Prediction*, Portland, OR, 212-215.

Kalnay, E., 2002: *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press (in press)

Kalnay, E., M. Corazza and M. Cai, 2002: Are bred vectors the same as Lyapunov vectors? *AMS Symposium on observations, data assimilation and probabilistic prediction*, Amer. Meteor. Soc., pp 173-177.

Morss, R. E., K. A. Emanuel and C. Snyder, 2001: Idealized adaptive observation strategies for improving numerical weather prediction. *J. Atmos. Sci.*, 58, 210-234.

Ott, E., B. R. Hunt, I. Szunyogh, M. Corazza, E. Kalnay, D. J. Patil, J. A. Yorke, A. V. Zimin, E. J. Kostelich, 2002: Exploiting local low dimensionality of the atmospheric dynamics for efficient ensemble Kalman Filtering. *Posted at arXiv.org/abs/physics/0203058*

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., B. R. Hunt, E. Kalnay, J. A. Yorke and E. Ott, 2001: Local low dimensionality of atmospheric dynamics. Phys. Rev. Lett., **86**, 5878-5881.

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

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **127**, 3297-3318.

Whitaker, J. and T. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.,* **130**, 1913-1924.