7.2

USING LARGE MEMBER ENSEMBLES TO ISOLATE LOCAL LOW DIMENSIONALITY OF ATMOSPHERIC DYNAMICS

D. J. Patil*, Istvan Szunyogh, Brian R. Hunt,
Eugenia Kalnay, Edward Ott, and James A. Yorke
University of Maryland, College Park, MD 20742

1 INTRODUCTION

A key question in the understanding of local instabilities that lead to the breakdown in forecasts, is the complexity of their dynamics (behavior). In this paper, we demonstrate that, in spite of the atmosphere's high dimensionality, in a suitable sense the *local finite-time* atmospheric dynamics is often low dimensional, and we conjecture that these low dimensional regions have a strong relationship to dynamical instabilities.

Section 2 introduces the method of breeding (Toth and Kalnay 1993; Toth and Kalnay 1997) to produce an ensemble of forecasts from different initial conditions (see Ehrendorfer (1997) and Kalnay (2002) and references within for a review of different methods to generate the initial conditions). Section 3 describes the two types of bred vector ensembles that we utilize. One type, obtained from NCEP has also been previously discussed by Patil et al. (2001). The second ensemble type uses many more bred vectors from a replica of the NCEP model. A main aim of this work is to compare results from these two ensembles. In Section 4 we introduce a statistic, which we call the *Bred Vector dimension* (BV-dimension), which effectively determines the dimension of the subspace spanned by members of the ensemble over a geographically localized region. Using our statistic we investigate the Earth's atmospheric dynamics for the Northern Hemispheric winter. A combination of operational data and a larger membership ensemble run are used to assess the robustness and saturation of the BV-dimension. We conclude with a discussion about the potential implications of our finding for weather forecasting and data assimilation.

* *Corresponding author address*: D. J. Patil, University of Maryland, IPST, College Park, MD 20742-2431; e-mail: dpatil@ipst.umd.edu

2 BREEDING

The procedure for breeding can be outlined in the following steps: (a) add a perturbation to the base state (usually the atmospheric analysis) at a given time t_0; (b) integrate both the base state and the perturbed state forward in time for a period of $t_1 - t_0$ (for this paper $t_1 - t_0 = 24$ hours); (c) at time, t_1, subtract the base state from the perturbed state and rescale the difference, so that it has the same norm (for the NWS ensemble system, rotational kinetic energy is used) as the initial perturbation; (d) this rescaled difference is then used as a new perturbation to the base state and the process (a)-(c) are sequentially repeated. The process is illustrated in Figure 1. The iterated difference between the perturbation and the main solution is called the *bred vector*.

3 DATA

For the analyses presented in this paper, we used two different types of ensembles.

Ensemble type 1: This ensemble consists of 5 perturbed forecasts (see Toth and Kalnay (1993) and Toth and Kalnay (1997) for operational implementations). The ensemble forecasts are made available on the Internet every 24 hours by the NWS and give the forecasts at 12 hour intervals spanning 8 days. In this study we focus on the wind vector field

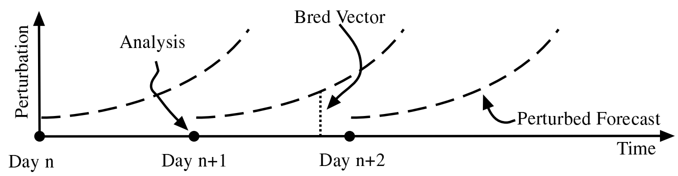

Figure 1: Schematic of the breeding cycle for ensemble forecasts

at 250mb, 500mb, and 850mb during the period of February 10, 2000 to March 31, 2000 and from December 1, 2000 to March 31, 2001. Thus our data will emphasize the northern hemispheric winter.

Ensemble type 2: In addition to the operational data, larger membership ensembles were also created using a replica NCEP Ensemble Forecasting system. Each of these ensembles consisted of 15 pairs of ensemble members and were run for the period of January 15, 2000 to February 20, 2000.

4 THE BRED VECTOR DIMENSION

In this section we outline how the bred vectors can be analyzed to obtain a useful measure of the dimensionality of the space in which instabilities in the forecasts are likely to grow the fastest. We begin with considering regions at a fixed pressure level of 5 by 5 grid points (about 1100 km squared) by choosing a grid point as the center of the region as well as an array of 24 neighbors so that the array best covers the 1100 km square (at high latitudes some points are skipped in longitude to keep an approximate uniform distance). Given N fields at each point (e.g., temperature, wind speeds, etc.), the values at these 25 points, for each bred vector, are ordered to form a $25N$ dimensional column vector which we call a *local bred vector.*

The issue we want to address is the linear independence of the k local bred vectors. That is, we want to determine the effective dimensionality of the subspace spanned by the local bred vectors. To do this we use empirical orthogonal functions (EOF) [also known as principal component analysis] (Scheick 1997). The underlying concept is to find the lowest dimensional subspace that, in a least squares sense, optimally represents the majority of the data.

The k local bred vectors form the columns of a $25N \times k$ matrix, B. The $k \times k$ covariance matrix of B is $C = B^T B$, where B^T is the transpose of B. Since the covariance matrix is nonnegative definite and symmetric, its k eigenvalues λ_i are nonnegative ($\lambda_i \geq 0$), and its eigenvectors, after multiplying by B and normalizing, form an orthonormal set of vectors v_i which span the column space of B. We order the eigenvalues by $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$. The singular values of B are $\sigma_i = \sqrt{\lambda_i}$. The eigenvalues λ_i are a measure of the extent to which the k column vectors making up B point in the direction v_i, and each $\sigma_i^2 = \lambda_i$ is said to represent the amount of variance in the set of the k unit vectors that is accounted for by v_i. By identifying how many of the vectors of v_i represent most of the variance of B, we can identify an effective dimension spanned by the k local bred vectors. For example if two out of five singular values are zero, then the subspace spanned by the k local bred vectors is three dimensional. However, if some of the singular values are nonzero but small, the issue becomes more difficult. One option is to use thresholding (as is often done when trying to isolate the dominant modes in EOF), but there is difficulty in determining the "best" value of the threshold. Instead we opt to define the following statistic on the singular values which we call the *Bred Vector dimension* (BV-dimension):

$$\psi(\sigma_1, \sigma_2, \ldots, \sigma_k) = \frac{(\sum_{i=1}^k \sigma_i)^2}{\sum_{i=1}^k \sigma_i^2}. \qquad (1)$$

As examples of the statistic, we describe several cases with $k = 5$ vectors of unit length. If the k local bred vectors comprising B were all the same, then the singular values would be $\sqrt{5}, 0, 0, 0, 0$. This would yield a statistic of $\psi(\sqrt{5}, 0, 0, 0, 0) = 1$. If the k local bred vectors were equally distributed between two orthogonal unit vectors v_1 and v_2, in the sense that each one accounts for half the variance, then the singular values would be $\sqrt{5/2}, \sqrt{5/2}, 0, 0, 0$, and our statistic would yield $\psi(\sqrt{5/2}, \sqrt{5/2}, 0, 0, 0) = 2$. If the local bred vectors again lie in the two dimensional subspace spanned by unit vectors v_1 and v_2, but the two are not equally represented, then this could give $2 \geq \psi \geq 1$. For example, if 4 of 5 unit local bred vectors were pointing in the same direction while the other pointed in an orthogonal direction, then the singular values would be $2, 1, 0, 0, 0$ and $\psi(2, 1, 0, 0, 0) = 1.8$. While the dimension of the space spanned by the local bred vectors is 2, our statistic gives an intermediate value reflecting the degree of dominance of one direction over the other. In general our statistic returns a real value between 1 and k. Note that while small perturbations due to noise or numerical error will typically cause the dimensionality of the space spanned by the k local bred vectors to be k, the effective dimension ψ may be substantially lower and is insensitive to small changes in the σ_i due to noise or numerical error.

5 REGIONS OF LOW BV-DIMENSION

5.1 ENSEMBLE TYPE 1

We now consider the two horizontal (zonal and meridional) wind vector components ($N = 2$) of the ($k = 5$) bred vectors from the NWS ensemble system. An example is shown in Figure 2. The BV-dimension was calculated at each spatial point on the grid and contours are shown for values less than three. A large region of relatively low dimensionality (BV-dimension less than 3) is evident over western North America. This indicates that in this region, the local bred vectors effectively span a space of substantially lower dimension than that of the full space.

The regions of low BV-dimension can be shown to be statistically significant and not due to random fluctuations, by applying the BV-dimension to surrogate data (Theiler et al. 1992) that consists of bred vectors chosen from different days which are substantially far apart (removing temporal correlations). When this is done no regions with low BV-dimension are observed (Patil et al. 2001).

On average, we find that more than 20% of the Earth's surface is covered by regions of low BV-dimension (i.e., less than three). Table 1 lists the percentage of the Earth's surface that is covered on average by different BV-dimensions. On individual days we find that regions of low BV-dimension cover as much as 46.74% at 250mb, 30.76% at 500mb, and 32.06% at 850mb. The maximum percentage of area covered on any given day by the lowest values of the BV-dimension (less than two) are 6.40% at 250mb, 4.72% at 500mb, and 4.23% at 850mb. These large regions of low BV-dimension develop and persist on the order of less than a week and move eastward at about one third of the wind speed at 500mb. Figure 2 depicts the development of a region of low BV-dimension over North America from a forecast initiated on March 5, 2000. Note that the regions of low BV-dimension depend on the background flow as demonstrated by the good agreement between the regions of low BV-dimension at different forecast times (Figure 2 top panels) and the corresponding analysis times (Figure 2 bottom panels).

Table 1: **Average Percentage of Surface Covered by BV-dimensions**

Local Dimension	250mb	500mb	850mb
1-2	1.14%	0.69%	0.69%
2-3	29.77%	22.20%	23.67%
3-4	60.79%	63.74%	63.37%
4-5	7.7%	12.77%	11.68%

5.2 ENSEMBLE TYPE 2

It is important to understand the behavior of the BV-dimension as the number of members in the ensemble is substantially increased. For this we use ensemble integrations consisting of 15 members. Figure 3 shows the maximum and minimum BV-dimension at 500 mb over the entire globe as the number of members in the ensemble is increased. There does not appear to be any saturation of the maximum BV-dimension. However, the minimum BV-dimension grows very slowly from approximately BV-dimension 2 with 5 bred vectors to BV dimension 3 when 15 bred vectors are used. The lack of saturation of the maximum BV-dimension is consistent with the bred vectors pointing in random directions when there is no growth of errors (e.g., members of the ensemble show good agreement and the bred vectors are simply random perturbations of the background flow). On the other hand, the very slow growth of the minimum BV-dimension gives an indication of the robustness of

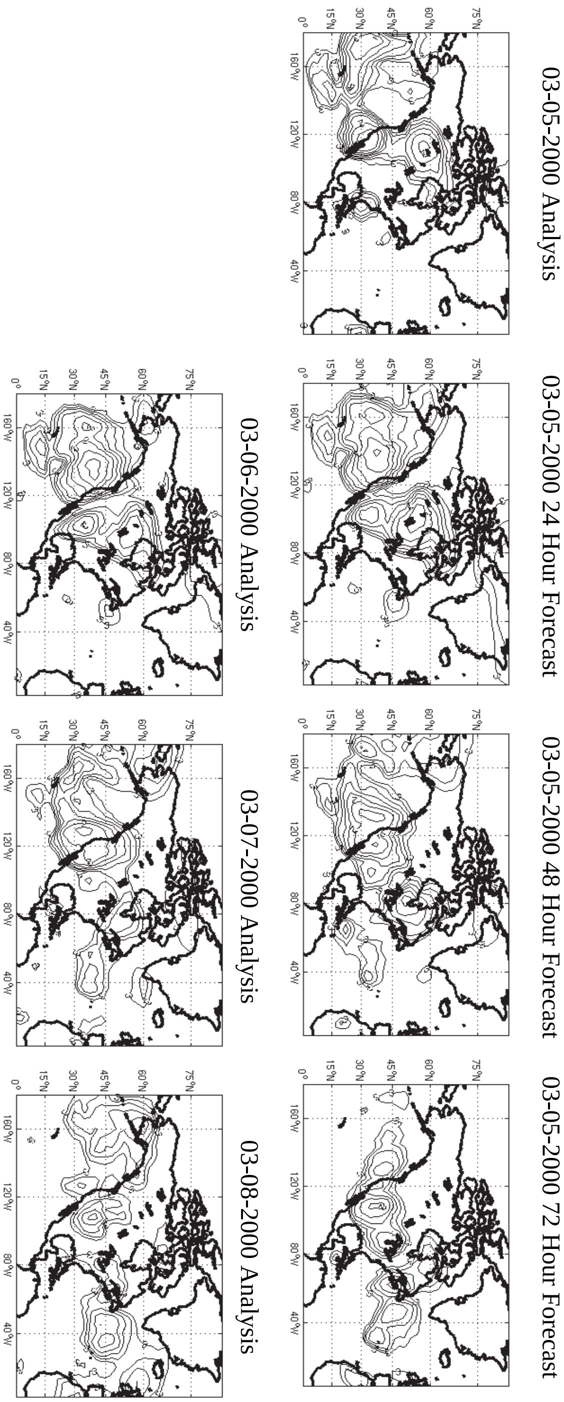

Figure 2: Development and propagation of a region of BV-dimension less than 3 at 500 mb. The upper panel series depicts the development of a region of BV-dimension less than 3 from a forecast initiated on March 5, 2000. The lower panel series shows regions of BV-dimension less than 3 analysis cycle corresponding to the forecast times in the upper panels.

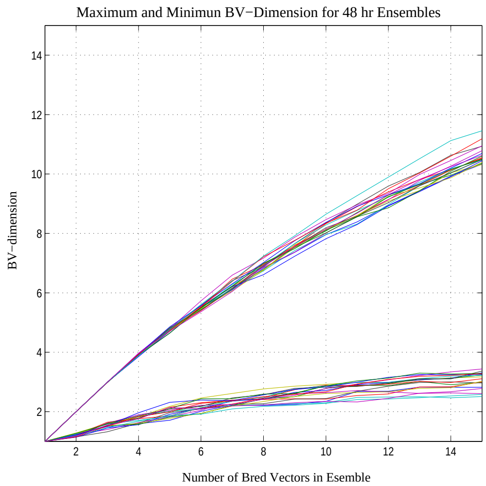

Figure 3: Maximum and minimum BV-dimension for the globe at 500 mb as a function of the number of members in the ensemble. Each curve represents a different 48 hour forecast.

lower values of BV-dimension. Further evidence of for the spatial robustness of regions of low BV-dimension (less than 3.5) can be seen in Figure 4; which shows regions of low BV-dimension as the number of members in the ensemble are increased.

6 CONCLUSION

The main result of this paper is a means of identifying local low dimensional behavior (the BV-dimension) in the atmosphere. We have provided evidence that these low dimensional, dynamical instabilities (regions of low BV-dimension) cover a significant portion of the globe, that they are intrinsic to the dynamics of the atmospheric system, that they typically last for several days, and that they are robust when the number of members of the ensemble is increased.

The analysis presented in this paper was conducted using the wind fields. We have conducted similar analysis using other fields such as temperature, geopotential height, and relative vorticity. These different fields all yield consistent results for the BV-dimension. This indicates that the local low dimensional behavior is robust to the choice of the atmospheric field as well as to the size of the membership of the ensembles.

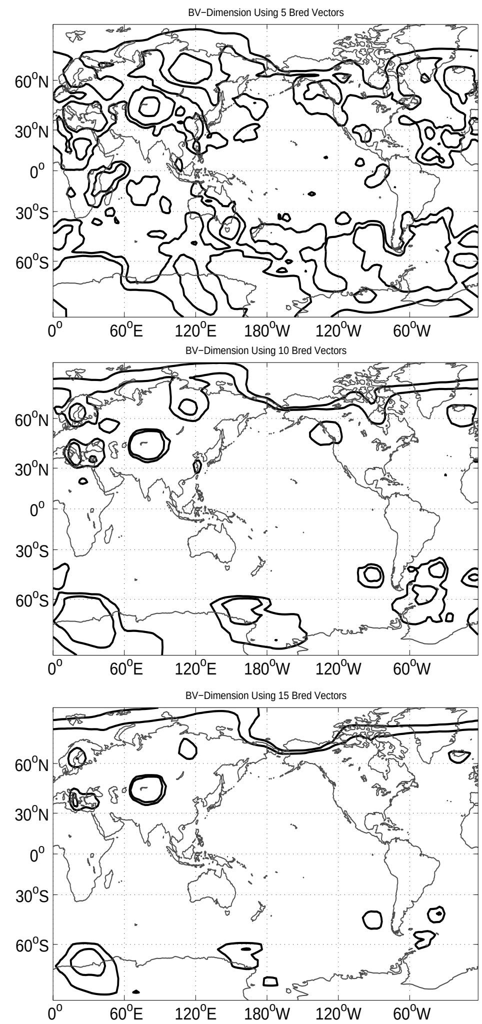

Figure 4: Contours of BV-dimension less than 3.5 at 500 mb computed with ensembles of five, ten and fifteen members for February 22, 2000 at 00Z.

At any given time t_0, there is inevitably a discrepancy $\vec{\Delta}(t_0)$ between the true atmospheric state and its representation in the computer model (analysis). Now consider a later time $t_1 > t_0$, and suppose that in a region of interest there is a low BV-dimension at time t_1. This suggests (ignoring model error here) that any local discrepancy $\vec{\Delta}(t_1)$ between the true state and its representation in the computer model (forecast error) lies predominantly in the "unstable subspace", the space spanned by the few vectors that contribute most strongly to the the low BV-dimension. We conjecture that in many cases this information can yield a substantial improvement in forecasting. In particular, the implication is that the data assimilation algorithm should correct the computer model state by moving it closer to the observations along the direction of the unstable subspace since that is where the true state most likely lies (Kalnay and Toth 1994). Current data assimilation techniques (e.g., that used by the NWS) do not take this into account.

Our results indicate that the BV-dimension identifies instabilities that have grown due to some dominant local dynamical stretching. If this is true, then BV-dimension would indicate regions in which increased observations may lead to the greatest improvement in reducing errors in the forecast (preliminary results for a quasi-geostrophic model are discussed in Corazza et al. (2001)). Since these regions are quite large, a targeted approach may be suitable (see Bishop and Toth (1999), Bishop et al. (2001), Szunyogh et al. (2001) and references within) in which it may be possible to find the location where (in both space and time) observations are most needed to reduce the forecast errors.

ACKNOWLEDGMENTS

The authors would thank Dr. Ming Cai, Dr. Zoltan Toth, Aleksey Zimin, Matteo Corazza, and Alfredo Nava-Tudela for insightful discussions. This project was supported by the W.M. Keck Foundation.

REFERENCES

Bishop, C., B. Etherton, and S. Majumdar (2001). Adaptive sampling with the ensemble transform kalman filter part 1: Theoretical aspects. *Monthly Weather Review 126*, 420–436.

Bishop, C. and Z. Toth (1999). Ensemble transformation and adaptive observations. *J. Atmos. Sci. 56*, 2536–2552.

Corazza, M., E. Kalnay, D. J. Patil, E. Ott, J. A. Yorke, I. Szunyogh, and M. Cai (2001). Use of the breeding technique in the estimation of the background covariance matrix for a quasi-geostrophic model. (See this issue).

Ehrendorfer, M. (1997). Predicting the uncertainty of numerical weather forecasts. *Meteorol Zeitschrift 6*, 147–183.

Kalnay, E. (2002). *Atmospheric Modeling, Data Assimilation, and Predictability*, Chapter Chapter 6. Cambridge, U.K: Cambridge University Press.

Kalnay, E. and Z. Toth (July 18-22, 1994). Removing growing errors in the analysis. In *Preprints; Tenth Conference on Numerical Weather Prediction*, pp. 212–215. Amer. Meteor. Soc.

Patil, D., B. Hunt, J. Yorke, E. Kalnay, and E. Ott (2001). Local low dimensionality of atmospheric dynamics. *Phys. Rev. Let. 86*, 5878–5881.

Scheick, J. (1997). *Linear Algebra with Applications.* New York: McGraw-Hill.

Szunyogh, I., Z. Toth, S. Majumdar, and A. Persson (2001). On the propagation of the effect of targeted observation: The 2000 winter storm reconnaissance program. *Submitted to Monthly Weather Review.*

Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. Farmer (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D 58*, 77–94.

Toth, Z. and E. Kalnay (1993). Ensemble forecasting at NMC - the generation of perturbations. *Bulletin of the American Meteorological Society 74*, 2317–2330.

Toth, Z. and E. Kalnay (1997). Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review 125*, 3297–3319.