Referee #1:

The paper describes a new method to calculate bred vectors in reconstructed time series. The method is original and it is compared the usual procedure to calculate bred vectors for the Lorenz model. The paper is clear and well written and in my opinion it deserves publication as it is now.

We thank the referee for taking the time to read our manuscript and for the recommendation that the paper is ready for publication in Nonlinear Processes of Geophysics.

Referee #2:

The authors present a purely data driven method to extract from a given time sieres dynamical information about the underlying dynamical system. To this extend they combine the bred vector method with the time delay embedding method to construct the phase space of the dynamical system. Within this reconstructed phase space pairs of nearby trajectories between a control and an initially nearby trajectory are piecewise followed over a specified time interval to measure the final separation distance. At this point the bred vector idea comes into play. After a rescaling of the final separation vector a local search in phase space for trajectories close to the final control phase space point shifted by the rescaled final separation vector of the previous iteration is initiated. After a proper identification to avoid points on the control trajectory, the next piecewise tracking is initiated. The growth rates at the end of each interval are the basic information derived from the analysis. The method is applied to the Lorenz 1963 L63 model. Three set ups are compared: (1) the standard breeding using the explicit L63, the proposed model, and an intermediate one where the time delay embedding is not used in favor of the original three dim L63 structure. The methods are compared by monitoring the local growth rates along the control trajectory and using positive large values as predictors for the regime shifts between the two leaves of the L63 attractor. With respect to this metric the new method performs reasonably well and the authors conclude that the new method provides a purely data-driven way to diagnose regime shifts for dynamical systems not well or not at all described by a set of equations.

In principle the paper is worth to be published in NPGP. It contains new information e.g. the proposed method and offers (some) help in interpreting the results. However, there is no clear conclusion or message especially with respect to data requirements and/or the dimensionality of the dynamics. The authors simply state that they ensured sufficient data density in three dimensions. But it would be worth to see how the contingency table statistics degrade when the actual data density is reduced. Although it is only a result for the idealized L63 it can give hints about the performance using real data. Another nice-to-know information would be on the frequency distribution (estimated probability density) of the calculated growth rates for the three test beds. This would again give more confidence into the new method than the simple thresholding of *looking at large growth rates. So my major suggestion before publication of this paper is that the authors should provide a clear message to the reader and potential user: is it worth to apply the method to other types of (real) data because the method is generic or are the results specific to the chosen setup?*

We thank the referee for taking the time to review our manuscript and provide us with these thoughtful comments. We appreciate the suggestion that our manuscript is appropriate for publication in this journal. Below we address the points raised.

The key objective of our paper is to demonstrate that the breeding vector technique can be used to determine the stability of low-dimensional dynamical systems using the time series data of one of its variables. For this proof-of-principle, a simple and well-known Lorenz63 is used and the parameters used in the analysis are identified. It is expected that this technique may not be effective for some systems and data sets, but a comprehensive analysis is needed to address many related issues, including the case of data density. Like most early papers on a new technique, including Packard et al.(1980) that showed time delay embedding using well-known systems, the emphasis in the paper is to demonstrate the new technique. Further studies on different systems and data sets would, over time, provide answers to the practical questions and the direct message to the reader is that the breeding technique for time series data works for the chosen well-known system (Lorenz63) and thus shows promise as a new technique.

The data density required for accurate results will depend on the individual system and the availability of data – including the sampling rate, the length of the time series, the time delay and dimension required for a proper embedding, and the recurrence time of the system. One way to visualize this is to look at a recurrence plot for the embedding time and dimension selected. To form this plot, the pairwise distance between each point in the dataset is computed. If this distance places the points within a neighborhood of given radius, a symbol is plotted at the corresponding location. This method depends on having segments of trajectories that lie near the control trajectory in question. Taking the desired initial magnitude of separation between the control and perturbed trajectory as the neighborhood radius, this is represented on the recurrence plot by diagonal segments of neighboring points. For a deterministic, nonlinear system, the trajectory will return to a given neighborhood given sufficient time. One can increase the density of the points covering the attractor by observing the time series for a longer time. On the recurrence plot this can be thought of as taking successively larger square segments. For shorter the duration of the time series, resulting in a data set with lower density, there are few neighboring trajectories within the specified radius. This means that it will be difficult to identify neighboring trajectories to serve as perturbations for the control trajectory. As the duration of the time series is increased, more neighboring trajectories within the specified radius become available and the contingency statistics improve. The density we selected was sufficient such that we were able to find analogues to our perturbed initial conditions that met our target perturbation size of $\alpha = 0.1$, on average.

The application of this method to real data will require some knowledge of the characteristics of underling system and analysis of the properties of the time series.

Applying the techniques of nonlinear time series analysis outlined in the referenced articles will allow one to assess an appropriate embedding and whether such an approach is suitable for the data in question. Further, with the size limitations of a brief communication, we felt it best to restrict our analysis to the Lorenz attractor. Applications of this and other similar techniques to more complicated real systems are in progress and be forthcoming in future publications.

Referee #3:

The authors of "Breeding vectors in the phase space reconstructed from time series data" presented an interesting approach to detect the behavior of breeding vectors using only a one-dimensional time series. The authors use the well known embedding technique to observe the growth rate and the spatial structure of perturbations starting from a fixed distance.

The paper is interesting but I have two main issues:

a) in a seminal work of Lyapunov exponents determination starting from one dimensional time series Brown, Bryant and Abarbanel, PRA 43, 2787 (1991) "Computing the Lyapunov spectrum of a dynamical system from an observed time series" the authors discuss in detail the importance to use two dimensions in computing the exponents. This is probably related to the choice of the integer l discussed at page 1306. Please discuss the relation between the choice of this paper and that given by Brown et al. namely the first minimum in the mutual information. I think that the authors should discuss the effect of changing l in their findings.

b) The example given on the standard Lorenz model is, in my opinion, not sufficient. The authors should test their technique on more complicated models like, for example, the Lorenz 96 model where the system dimension is larger than 3 and the embedding technique becomes more difficult to be applied. After the authors address the two points I raised the paper can be considered for publication.

We thank the referee for taking the time to review our manuscript and provide us with these thoughtful comments. We appreciate the recommendation that our paper be considered for publication and address the issues raised below.

To address the first point, the exclusion of points along the control trajectory is simply to ensure that the growth of separation between the control and perturbed trajectories is due to the nonlinear dynamics of the system rather than displacements along the control trajectory, as Brown et al. discuss in the referenced paper. We based our estimation of the number of points to exclude on the estimate for the time delay for the x(t) data, which we obtained by the same method as Brown et al., i.e. the first minimum of the average mutual information function. This ensures that the neighboring points used as perturbed trajectories are nearby due to the structure of the attractor and not just correlated in time.

As long as sufficiently many points are excluded that the nearest neighbor to the perturbation tends to lie along a different orbit from the control trajectory, the numerical value of l does not have much effect on the results.

Brown et al (also the review: Abarbanel et al. RMP 1993) address the issue of spurious exponents due the inaccuracies or local features in the reconstructed space. The analysis in our paper is on the largest Lyapunov exponent and the existence or nature of spurious exponents would not affect the conclusions of the paper.

The main result of this paper is that the application of breeding to a reconstructed phase space without the use of a dynamical model is a viable data driven method. To achieve this in the brief communication we have presented our results on the Lorenz63 model. It is our hope that this paper will stimulate similar studies that explore the extension of this technique to other models and systems. We have applied similar techniques to those outlined here to real data and expect to present those in future publications. As a brief communication, it would not be possible to add additional results without altering the format of the article.

Referee #4:

Interactive comment on "Brief Communication: Breeding vectors in the phase space reconstructed from time series data" by Lynch et al.

The authors proposed a new approach, i.e. the nearest-neighbor breeding, to model and predict sudden transitions in systems represented by time series data. Furthermore, they used the Lorenz-63 model to examine the validity of this method. The results show that the dynamical properties of the standard and nearest-neighbor breeding are similar. This validates the ability of this new approach to predict regime change in a dynamical system using the time series data of one variable. Thus, this has important implications.
However, I think that the presentation needs to be improved. Here a list of points and questions should be addressed:

We thank the referee for taking the time to review our manuscript and provide these thoughtful comments. We address each point below.

1. Page 1304, line 15: The authors mentioned "the systems known to exhibit sudden regime changes in their data". In fact, I am especially interested in these systems. In addition to magnetospheric substorms and geospace storms, are there any other systems known to exhibit sudden regime changes? As for the well-known phenomena such as the haze, rainstorm and thunderbolt, could the nearest-neighbor breeding be used to model and predict them?

Our focus in the paper is on the demonstration of the breeding technique and the Lorenz63 model is used. Systems in nature that exhibit sudden changes, e. g., magnetospheric substorms (see Vassiliadis et al. GRL 1991 for computation of the largest Lyapunov exponent for substorms), are currently under study and we expect that the technique would be applicable to many natural systems that exhibit regime changes – e.g. transitions from a quiet (near equilibrium) state to a disturbed or active state. The application of the techniques outlined in this article requires the system to have low dimensional underlying dynamics such that phase space trajectories lie onto an attractor. While the application to other systems is outside the scope of this article, we hope that others will follow and look into applying this method to these systems. The applicability of the technique the phenomena such as rainstorm etc. are of interest but requires detailed studies.

2. Page 1305, lines 11-13: This sentence should be reformulated. It is not clear.

Original sentence: Having defined the reconstructed phase space by the time-delayed embedding, the new approach to breeding is in essence a matter of selecting the perturbed trajectories that capture the unstable directions along the control.

We will modify the sentence to read: Having defined the reconstructed phase space by the time-delayed embedding, the new approach to breeding is in essence a matter of selecting perturbed trajectories that diverge from the control trajectory along the unstable directions.

3. Page 1306, line 2: The authors said "in order to avoid selecting nearest neighbors that are on the control trajectory". Please explain the reasons.

If nearest neighbors along the control trajectory are used to initiate the perturbed trajectory, the growth will be along the orbit and will not represent the nonlinear growth of diverging trajectories. By excluding a segment of the control trajectory near the point at the start of the breeding cycle, we take as our perturbations neighboring trajectories that may tend to diverge exponentially from the control trajectory via the nonlinear dynamics of the system.

4. Page 1306, lines 1-4: Are there 2l+1 points to be excluded?

We excluded the l+1 points centered around the point on the control trajectory.

5. Page 1306, lines 7-8: "the density of the trajectory points must be high enough". That is to say, the temporal resolution of the time series Dt should be sufficiently small. Is this right?

It is important that the trajectory cover the attractor and that it do so densely enough that suitable analogues for perturbations can be found. The temporal resolution will certainly affect the density of points and the properties of the embedding and breeding, but the more important feature of the time series that contributes to adequate density is the duration of the time series. Given sufficient time, systems like those suitable for this type of analysis will revisit neighborhoods of phase space arbitrarily frequently and within an arbitrarily small neighborhood. The radius of the neighborhood, or the size of the perturbation, will be limited in practice by the duration of the time series.

6. Page 1307, line 1: Is here "$m = 3$ and $\tau = 7$" determined by the methods described in section 2 (Page 1305, lines 2-9)?

Yes, we determined the dimension and time delay by looking at estimates of the correlation dimension and mutual information function respectively.

7. Page 1307, line 17: The authors used the breeding window size $n = 8$ with $\Delta t = 0.01$ and perturbation size $\alpha = 0.10$ in all experiments. Then, what about the sensitivity of the results in this paper on n and α?

We chose these parameters to make a direct comparison of our work to the results presented by Evans et al. [Table coming]

8. Page 1307, line 21: The authors said "excluding $l = 6$ adjacent points". However, I think there should be $2l+1=13$ points to be excluded. Am I right?

We excluded l=6 points, three on either side of the control point.

9. Page 1307, line 25: The authors said "The left column of Fig. 1 shows the growth rates along the respective controls in the three experiments", but not mentioned the specific points. Are the points in the left column of Fig. 1 the control trajectory points? That is to say, do these points correspond to the time series data?

The points in the left column are the control trajectory points for which bred vector growth rates were computed. They correspond exactly to the time series data. We will modify the text to clarify this point.

10. Page 1308, line 18: The threshold value "1" seems to be unreasonable. According to Fig. 1(f), there are many red stars corresponding to the absolute value of x1 that is greater than 1. If ignoring all these stars, some information about the regime change may not be noted and used.

The threshold value of 1 is for the minimum (maximum) value attained by the x variable for orbits in the positive (negative) regime. It is true that sometimes this method will miss a regime change or predict a regime change when none occurs as is indicated by the false alarm rate.

11. Page 1309, lines 10-11: The first reason should be reformulated. I do not understand what you said.

Original sentence: First, unlike the size of a particular variable, breeding can be tested in any dynamical model.

We will modify the sentence to read: First, unlike threshold values of a particular variable, bred vector growth can be tested in any dynamical model.

12. Page 1309, lines 17-19: For the time series data of variable x when t is smaller than 10, the longer duration of the high growth rate does not indicate the next longer-lasting regime (Fig. 1d). Please clarify this phenomenon.

If the system is continuous, there should be segments along the trajectory that have a high growth rate for bred vectors. Since we have discrete data, the distribution of points for which bred vectors are computed will not always capture the entire segment. We will increase the size of the figure so that it is easier to see that there are often several red stars, or high growth rate bred vectors, preceding a long duration regime change that are very close to one another.

Modifications to the text:

Page 1305, lines 11-13: Once the phase space has been reconstructed by the time-delay embedding, the new approach to breeding is in essence of matter of selecting perturbed trajectories that tend to grow along the unstable directions with respect to the control trajectory.

Page 1307, line 26: The left column of Fig. 12 shows the points along the respective control trajectories for which bred vector growth rates were computed in each of the three experiments.

Page 1308, lines 18-19: We note that we slightly modified the empirical rule for the experiments using nearest neighbor breeding, by excluding red stars when they occurred on orbits with extrema of x1 whose absolute value was greater than 1 as the trajectory approaches x1 = 0, since they led to false alarms in the prediction of regime change.

Page 1309, lines 10-11: First, unlike threshold values of a particular variable, bred vector growth can be tested in any dynamical model. For many systems, there will not be a correspondence between the numerical value of a particular variable and regime change.

Figure 1 has been updated.

Manuscript prepared for Nonlin. Processes Geophys. Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 23 February 2016

Brief Communication: Breeding vectors in the phase space reconstructed from time series data

E. Lynch[1], D. Kaufman[2], A. S. Sharma[3], E. Kalnay[1,4,5], and K. Ide[1,4,5,6]

[1]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, MD 20742, USA
[2]Virginia Institute of Marine Science, College of William & Mary, Gloucester Point, VA 23062, USA
[3]Department of Astronomy, University of Maryland, College Park, College Park, MD 20742, USA
[4]Institute for Physical Science and Technology, University of Maryland, College Park, College Park, MD 20742, USA
[5]Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, MD 20742, USA
[6]Center for Scientific Computation and Mathematical Modeling, University of Maryland, College Park, College Park, MD 20742, USA

Correspondence to: E. Lynch (elynch@umd.edu)

Abstract

Bred vectors characterize the nonlinear instability of dynamical systems and so far have been computed only for systems with known evolution equations. In this article, bred vectors are computed from a single time series data using time-delay embedding, with a new technique, nearest-neighbor breeding. Since the dynamical properties of the standard and nearest-neighbor breeding are shown to be similar, this provides a new and novel way to model and predict sudden transitions in systems represented by time series data alone.

1 Introduction

Prediction of sudden regime changes in the evolution of dynamical systems is a challenging problem. For systems with known dynamical models, such as the Earth's atmosphere, simulated trajectories under judiciously chosen, finite-size perturbations can provide useful information regarding regime changes by detecting fast-growing instabilities along the model representation of the system evolution, called the "control." Breeding is a technique to generate an ensemble of such perturbations, developed for operational ensemble forecasting of the numerical weather prediction (Toth and Kalnay, 1993, 1997; Kalnay, 2003), and the resulting perturbations are called "bred vectors." Evans et al. (2004) demonstrated that the growth rate of the bred vectors could be used as a means of predicting the regime changes in the chaotic Lorenz (1963) system (Lorenz, 1963). They found that the appearance of high growth rate typically indicated that a regime change would occur upon completion of the current orbit, and that the longer the duration of the high growth rate, the longer the next regime would last.

Models of most natural systems like the Earth's atmosphere are described by a very large number of dynamical variables and thus are high dimensional. However, the variables or degrees of freedom are nonlinearly coupled, and consequently in dissipative systems the dimensionality of the phase space is significantly reduced. This is the basis for the time-delay embedding method in the reconstruction of phase space (Packard et al., 1980; Takens, 1981). The ability of this method to yield the dynamics inherent in observational data, independent of modeling assumptions, has stimulated the modeling of dynamics using time series data, and has led to the development of forecasting tools. For example, reconstruction of the dynamics of the Earth's magnetosphere using time series data has led to low dimensional models and forecasts of space weather (Sharma et al., 1993; Sharma, 1995). The data-derived models of magnetospheric substorms and geospace storms (Vassiliadis et al., 1995; Ukhorskiy et al., 2002; Valdivia et al., 1996) now provide near real-time forecasts using the solar wind data monitored by the ACE spacecraft at the first Lagrange point (L1).

This paper presents a novel extension of the original breeding technique to the phase space reconstructed from time series data using the time-delay embedding method. Because dynamic instabilities are intrinsically low dimensional (Patil et al., 2001), such an extension is an appealing approach for the systems known to exhibit sudden regime changes in their data. The predictive capabilities of this new breeding technique are tested using data taken from the chaotic Lorenz system (Lorenz, 1963).

2 Phase space reconstruction

To extend the breeding technique to a system represented by a time series $x(t)$ at discrete times $t = t_1, t_2, \ldots, t_N$, we first give a brief introduction of the time-delay embedding and set the notation used in this study. The state of the system at t_i in the reconstructed phase space is defined by an m component vector:

$$\boldsymbol{x}_i = \{x_1(t_i), x_2(t_i), \ldots, x_m(t_i)\}, \tag{1}$$

where $x_k(t_i) = x(t_i - \tau(k-1)\Delta t)$ is the time-delay coordinate for $k = 1, \ldots, m$, τ is the delay time, and Δt is the temporal resolution of the time series. The vectors $\boldsymbol{x}_i$ specify the dynamical behavior inherent in the data in the form of a trajectory. The time-delay embedding (Sauer et al., 1991; Abarbanel et al., 1993; Kantz and Schreiber, 1997) requires only two parameters, viz. the dimension m and time delay τ. Among the several techniques available for obtaining these parameters from the data, the convergence of the correlation among the trajectories (Sauer et al., 1991; Abarbanel et al., 1993; Kantz and Schreiber, 1997) and the minimum in the mean-square prediction error (Chen and Sharma, 2006) yield the optimum values of m. The time delay τ also depends on the correlations, and the minimum of the mutual information function yields reliable values (Sauer et al., 1991; Abarbanel et al., 1993; Kantz and Schreiber, 1997; Chen and Sharma, 2006).

3 Bred vectors in reconstructed phase space

Having defined the reconstructed phase space by the time-delayed embedding, the new approach to breeding is in essence a matter of selecting the perturbed trajectories that capture the unstable directions along the controldiverge from the control trajectory along the unstable directions. Over a breeding cycle with window size n, the control starting from $\boldsymbol{x}_i$ evolves to $\boldsymbol{x}_{i+n}$ and its neighbor (perturbed) $\boldsymbol{x}_j$ to $\boldsymbol{x}_{j+n}$. The corresponding growth rate of the bred vector is given by

$$g_i = \frac{1}{n\Delta t} \ln\left(||\boldsymbol{\delta x}_i^{\mathrm{f}}||/||\boldsymbol{\delta x}_i^0||\right) \tag{2}$$

where $\boldsymbol{\delta x}_i^0 = \boldsymbol{x}_j - \boldsymbol{x}_i$ and $\boldsymbol{\delta x}_i^{\mathrm{f}} = \boldsymbol{x}_{j+n} - \boldsymbol{x}_{i+n}$ are the initial and final perturbations of the breeding cycle. To select the perturbed trajectory of the next breeding cycle around the control starting from $\boldsymbol{x}_{i+n}$, we follow the spirit of the standard breeding in which the initial perturbation is given by $\boldsymbol{\delta x}_{i+n}^0 = \alpha \boldsymbol{\delta x}_i^{\mathrm{f}}/||\boldsymbol{\delta x}_i^{\mathrm{f}}||$, i.e., the final perturbation of the previous cycle is rescaled and bred as the new perturbation. Here the perturbation size α is constant for all breeding cycles. In the reconstructed phase space, however, the trajectory is defined by discrete points, and the rescaled position $\boldsymbol{x}_{i+n} + \alpha \boldsymbol{\delta x}_i^{\mathrm{f}}/||\boldsymbol{\delta x}_i^{\mathrm{f}}||$ may not be a trajectory point. We thus search and select the nearest point $\boldsymbol{x}_{j^*}$ and refer to the distance between these two points as the displacement distance. In order to avoid selecting nearest neighbors that are on the control trajectory, points $\boldsymbol{x}_{i+j\pm l}$ immediately adjacent to $\boldsymbol{x}_{i+n}$ for small l are excluded from the search for $\boldsymbol{x}_{j^*}$ so as to ensure that $\boldsymbol{\delta x}_{i+n}^0 = \boldsymbol{x}_{j^*} - \boldsymbol{x}_{i+n}$ captures the instability around the control. We call this technique the "nearest-neighbor breeding". Like the standard breeding, it involves two parameters, viz., the window size n and the target perturbation size α. For successful applications of the nearest-nerighbor search, the density of the trajectory points must be high enough that, on average, the displacement distance in the nearest-neighbor search is small with respect to target initial perturbation size α and the correlation between $\boldsymbol{\delta x}_i^{\mathrm{f}}$ and $\boldsymbol{\delta x}_{i+n}^0$ is nearly 1.

4 Results and discussion

To test whether the nearest-neighbor breeding shares with the standard breeding the ability to predict regime changes in the reconstructed phase space, we use the 3-dimensional Lorenz (1963) system (Lorenz, 1963) and generate a time series data set. Along with its simplicity, this system possesses dynamical properties desirable for this study, namely, high nonlinearity that manifests in a low dimensional attractor and chaotic transitions between two regimes (Fig. 1). The model equations are given by:

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= \sigma\,(y-x) \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= -xz+rx-y \\
\frac{\mathrm{d}z}{\mathrm{d}t} &= xy-bz
\end{aligned} \tag{3}$$

with the commonly used parameter values $r=28$, $b=8/3$, and $\sigma=10$. Forward integration of the model is performed using a fourth order Runge–Kutta with time step 0.01. To reconstruct the phase space from a single time series $x(t_i)$ with the temporal resolution $\Delta t=0.01$, we use the embedding dimension $m=3$ and time delay $\tau=7$, which correspond to the time-scale of the mutual information function (Sauer et al., 1991; Abarbanel et al., 1993; Kantz and Schreiber, 1997; Chen and Sharma, 2006). The reconstructed system is a discrete set of trajectory points that exhibits the dynamical features of the attractor (Sauer et al., 1991; Abarbanel et al., 1993; Kantz and Schreiber, 1997) as shown in Fig. 1c.

The regime transitions are analyzed by performing and comparing the following three breeding experiments. Experiment (a) is the standard breeding in the 3-dimensional model phase space using Eq. (3) as in Evans et al. (2004). Experiment (b) is the nearest-neighbor breeding applied to a discrete time series data in the model phase space, i.e., $\hat{\boldsymbol{x}}_i = \{x(t_i), y(t_i), z(t_i)\}$. This experiment reveals whether the nearest-neighbor breeding, without any knowledge about the model equations, gives comparable results to the standard breeding in the original model phase space. Finally, Experiment (c) is our new technique, i.e., the nearest-neighbor breeding in the phase space reconstructed by the time-delayed embedding of a *single* time series as in Eq. (1).

In all experiments, the growth rate g_i is computed using the breeding window size $n = 8$ with $\Delta t = 0.01$ and (targeted) perturbation size $\alpha = 0.10$ over 10 000 total breeding cycles after an initial spin-up to make sure that the trajectory has reached the attractor. To ensure sufficient data density for the nearest-neighbor search in Experiments (b) and (c), the respective data sets are constructed from 80 000 data points in the original phase space. By excluding $l = 6$ adjacent points on the control trajectory from the nearest-neighbor search, the average displacement distance and vector correlation between the standard and the nearest neighbor breeding are 0.17 and 0.97 in Experiment (b); they are 0.12 and 0.98 in Experiment (c).

The left column of Fig. 1 shows the ~~growth rates~~ the points along the respective ~~controls in~~ control trajectories for which bred vector growth rates were computed in each of the three experiments. The magnitudes of the growth rate are represented by colors using the same empirical thresholds as in Evans et al. (2004): negative growth points ($g_i < 0$) in blue, low growth points ($0 \le g_i < 3.2$) in green, medium growth points ($3.2 \le g_i < 6.4$) in yellow, and high growth points ($g_i \ge 6.4$) in red. As shown in Evans et al. (2004) for the standard breeding, all experiments show high growth at points concentrated in the regime transition region, while the regions with different growth rates are well separated. The nearest-neighbor breeding, both in the original (Fig. 1b) and in the reconstructed (Fig. 1c) phase spaces, successfully captures the features found in the standard breeding (Fig. 1a), although the separation between the different growth rates is less sharp. The right column of Fig. 1

shows the time series of the first phase space coordinate (x) for the first 500 breeding cycles for each of the three experiments. Note that, by construction, the first coordinate x in phase space coincides with the first coordinate x_1 in the embedded space.

As pointed out by Evans et al. (2004) and apparent in the right column of Fig. 1, high bred vector growth rate, marked in red, is a very good predictor of regime change in the standard breeding in the Lorenz model. Thus we test the predictive capacities of the bred vectors for the regime changes in terms of the binary (YES-NO) forecasts based on the rule suggested by Evans et al. (2004): the presence of a red star in the previous orbit renders the regime change (YES), while the absence means the continuation of the current regime (NO). We note that we slightly modified the empirical rule for the experiments using nearest neighbor breeding, by excluding red stars when they occurred on orbits with extrema of x_1 whose absolute value was greater than 1 as the trajectory approaches $x_1 = 0$, since they led to false alarms in the prediction of regime change.

Table 1 is the contingency table (Wilks, 1995) of the forecast/event pairs for the three experiments, where individual forecasts (FCST) are made by the rule and the observed events (OBS) are based on the actual occurrence or non-occurrence of the transition. Corresponding accuracy measures for these binary forecasts (Wilks, 1995) are shown in Tables 2 using the Hit Rate (HR), Threat Score (TS), and False Alarm Rate (FAR). It is apparent that the three experiments succeed in predicting with similar accuracy the change of regime. The HRs and TSs for the three methods are close, varying from 82 to 87, and 72 to 76 %, respectively. The FARs are about 6 % for the standard breeding, but it increases to 11 and 13 % when nearest-neighbor breeding is used in the original model phase space and in the reconstructed phase space, respectively.

We note that, in addition to large bred vector growth rate, two other methods have been also proposed to predict regime changes in the Lorenz three variable system. In his original paper (Lorenz, 1963), Lorenz pointed out that regime changes were associated with large values of the variable z. Yadav et al. (2005) showed that large absolute magnitudes of the x variable are also a good predictor. We have implemented the method used in Yadav et al. (2005) for the time series $x(t)$ and got equally good results. However, the main objective of this paper is to determine whether bred vectors can predict stability from a single time series data, i. e., in the reconstructed phase space. The reasons for the choice of bred vectors as predictor of fast growth in the dynamical system are two-fold. First, unlike ~~the size~~ threshold values of a particular variable, ~~breeding~~ bred vector growth can be tested in any dynamical model. For many systems, there will not be a correspondence between the numerical value of a particular variable and regime change. Second, bred vector perturbations and their growth have a clear physical meaning in that they detect instabilities (Hoffman et al., 2009) and are akin to the leading local Lyapunov vector and their corresponding growth (Norwood et al., 2013). Thus while predictions based on threshold values of a single variable work well for the Lorenz model, bred vector growth rate may be suitable for making predictions in a broad range of dynamical systems.

For the binary forecasts, Evans et al. (2004) noted that the next regime tended to be longer-lasting when the duration of the high growth rate is longer using the standard breeding (Fig. 1d). This tendency is also found using the nearest neighbor breeding in the reconstructed phase space in Experiment (c). However, in Fig. 1f, the cases of high growth rate may be separated by slower growth rates (e.g., at time $t \sim 23$) due to the time delay involved in the construction of the embedded time series.

5 Conclusions

The ability to predict regime change in a dynamical system using the time series data of just one of its many variables, demonstrated in this paper, has important implications. For most systems in nature and in laboratory, the time series observations of only a limited number of physical variables, often a single variable, are available. In many cases even the actual number of variables is not known. This paper presents and demonstrates that the nearest neighbor breeding enables the prediction of regime change in systems for which regime change follows the appearance of instabilities, thus extending the predictive capability beyond the cases whose time evolution equations are known. Further, when regime change is associated with large changes in the dynamical states, this technique can lead to the prediction of large or extreme events in the cases where nonlinear dynamical predictions are made using time series data, e.g., in the Earth's magnetosphere and space weather (Chen and Sharma, 2006).

Acknowledgements. This work is supported by JPL grant 1485739, CICS grants NOAA-CISC-EKEK_12 and CICS-EKEK_14, NSF/IIP grant 1338634, and ONR grant N00014-10-1-0557. This paper is in part Contribution No. 3465 of the Virginia Institute of Marine Science, College of William & Mary.

References

Abarbanel, H. D. I., Brown, R., Sidorowich, J. J., and Tsimring, L. S.: The analysis of observed chaotic data in physical systems, Rev. Mod. Phys., 65, 1331–1392, 1993.

Chen, J. and Sharma, A. S.: Modeling and prediction of the magnetospheric dynamics during intense geospace storms, J. Geophys. Res., 111, A04209, doi:10.1029/2005JA011359, 2006.

Evans, E., Bhatti, N., Kinney, J., Pann, L., Pena, M., Yang, S. C., Kalnay, E., and Hansen, J.: RISE undergraduates find that regime changes in Lorenz's Model are predictable, B. Am. Meteorol. Soc., 85, 521–524, 2004.

Hoffman, M. J., Kalnay, E., Carton, J. A., and Yang, S.-C.: Use of breeding to detect and explain instabilities in the global ocean, Geophys. Res. Lett., 36, L12608, doi:10.1029/2009GL037729, 2009.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, Cambridge, UK, 2003.

Kantz, H. and Schreiber, T.: Nonlinear Time Series Analysis, Cambridge University Press, Cambridge, 1997.

Lorenz, E. N.: Deterministic Nonperiodic Flow, J. Atmos. Sci., 20, 130–141, 1963.

Norwood, A., Kalnay, E., Ide, K., Yang, S.-C., and Wolfe, C.: Lyapunov, singular and bred vectors in a multi-scale system: an empirical exploration of vectors related to instabilities, J. Phys. A-Math. Theor., 46, 254021, doi:10.1088/1751-8113/46/25/254021, 2013.

Packard, N., Crutchfield, J., Farmer, D., and Shaw, R.: Geometry from Time Series, Phys. Rev. Lett, 45, 712–716, 1980.

Patil, D. J., Hunt, B., Kalnay, E., Yorke, J. A., and Ott, E.: Using bred vectors to establish local low dimensionality of atmospheric dynamics, Phys. Rev. Lett., 86, 5878–5881, 2001.

Sauer, T., Yorke, J. A., and Casdagli, M.: Embedology, J. Stat. Phys., 65, 579–616, 1991.
Sharma, A. S.: Assessing the magnetosphere's nonlinear behavior: its dimenstion is low, its predictability, high (US National Report to IUGG, 1991–1994), Rev. Geophys. Suppl., 33, 645–650, 1995.
Sharma, A. S., Vassiliadis, D., and Papadopoulos, K.: Reconstruction of low dimensional magnetospheric dynamics by singular spectrum analysis, Geophys. Res. Lett., 20, 335–338, 1993.
Takens, F.: Detecting strange attractors in turbulence, in: Dynamical Systems and Turbulence, Warwick 1980, Lecture Notes in Mathematics 898, Springer, Berlin, 1981.
Toth, Z. and Kalnay, E.: Ensemble forcasting at NMC: the generation of perturbations, B. Am. Meteorol. Soc., 74, 2317–2330, 1993.
Toth, Z. and Kalnay, E.: Ensemble forecasting at NCEP and the breeding method, Mon. Weather Rev., 125, 3297–3318, 1997.
Ukhorskiy, A. Y., Sitnov, M. I., Sharma, A. S., and Papadopoulos, K.: Global and multi-scale properties of magnetopheric dynamics in local-linear filter, J. Geophys. Res. Lett., 107, 1369, doi:10.1029/2001JA009160, 2002.
Valdivia, J. A., Sharma, A. S., and Papadopoulos, K.: Prediction of magneti stroms by nonlinear models, Geophys. Res. Lett., 23, 2899–2902, 1996.
Vassiliadis, D., Klimas, A. J., Baker, D. N., and Roberts, D. A.: A description of the solar wind-magnetospher coupling based on nonlinear filters, J. Geophys. Res., 100, 3495–3512, doi:10.1029/94JA02725, 1995.
Wilks, D.: Statistical Methods in the Atmospheric Sciences: An Introduction, International Geophysical Series, vol. 59, Academic Press, Oxford, UK, 1995.
Yadav, R. S., Dwivedi, S., and Mittal, A. K.: Prediction rules for regime changes and length in a new regime for the Lorenz model, J. Atmos. Sci., 62, 2316–2321, 2005.

Table 1. Contingency tables based on the rule that regime change will occur in the orbit following the appearance of high growth rate bred vectors using three different methods. In (b) and (c) using the nearest-neighbor breeding, high growth rate points in orbits with absolute values of extrema above 1 are excluded. OBS and FCST stand for observed and forecast, respectively; (a)–(c) are the same as in Fig. 1.

			OBS		
			Yes	No	Total
		Yes	374	38	412
(a)	FCST	No	80	573	653
		Total	454	611	1065
		Yes	396	67	463
(b)	FCST	No	58	544	602
		Total	454	611	1065
		Yes	383	77	460
(c)	FCST	No	71	534	605
		Total	454	611	1065

Table 2. Measures of forecast accuracy in terms of the Hit Rate (HR), Threat Score (TS), and False Alarm Rate (FAR); (a)–(c) are the same as in Fig. 1. The final row shows the values when the threshold of $x(t_i)$ rule is used.

	HR (%)	TS (%)	FAR (%)
(a)	82.4	76.0	5.2
(b)	87.2	76.0	11.0
(c)	84.4	72.1	12.6

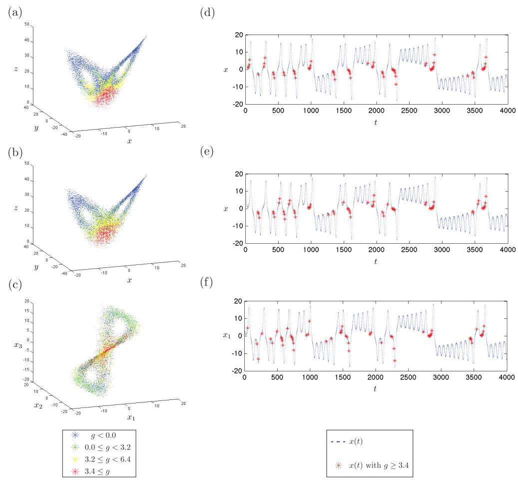

Figure 1. The first column of figures depicts the growth rates of bred vectors in the Lorenz system using three different methods: **(a)** standard breeding in the phase space (x,y,z); **(b)** nearest-neighbor breeding in the phase space (x,y,z); and **(c)** nearest-neighbor breeding in the reconstructed phase space (x_1,x_2,x_3). The colored points correspond to negative (blue), low (green), medium (yellow) and high (red) growth. The second column of figures depicts the first coordinate of phase space as a function of time, with red stars indicating the points with high growth rate ($g_i \geq$ 6.4) bred vectors for the three different methods: **(d)** standard breeding in the phase space (x,y,z); **(e)** nearest-neighbor breeding in the phase space (x,y,z); and **(f)** nearest-neighbor breeding in the reconstructed phase space (x_1,x_2,x_3).