WILLIAM & MARY
CHARTERED 1693

W&M ScholarWorks

2016

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

Erin Lynch

Daniel Kaufman
Virginia Institute of Marine Science

A. Surjalal Sharma

Eugenia Kalnay

Kayo Ide

Recommended Citation

Nonlinear Processes
in Geophysics

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

Erin Lynch[1], Daniel Kaufman[2], A. Surjalal Sharma[3], Eugenia Kalnay[1,4,5], and Kayo 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: Erin 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 Advanced Composition Explorer (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.

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 control. 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, respectively. 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-neighbor 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-D 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, strong 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 timescale 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 as shown in Fig. 1c.

The regime transitions are analyzed by performing and comparing the following three breeding experiments. Ex-
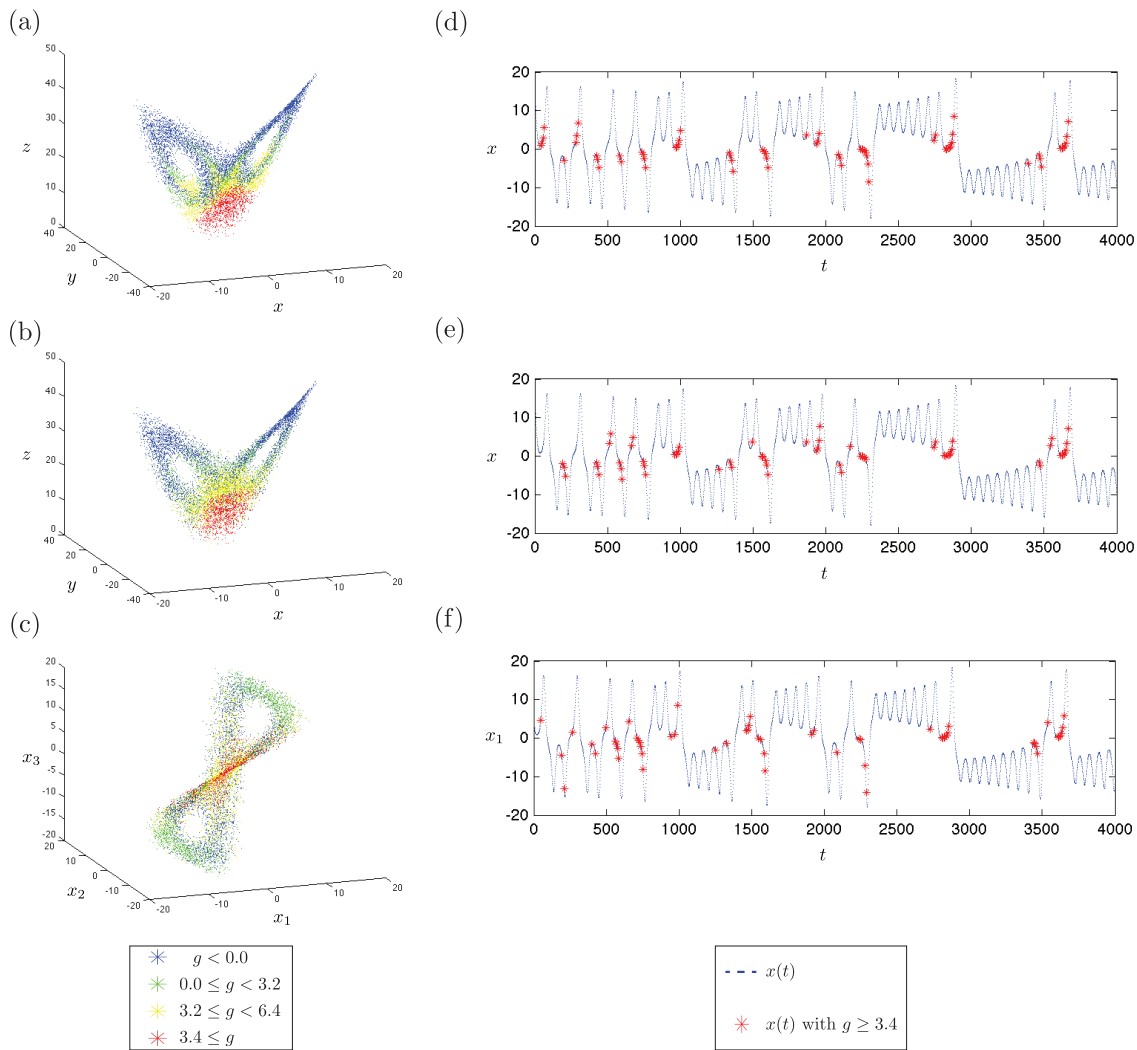
Figure 1. The first column of figures depicts the growth rates of bred vectors in the Lorenz system using three different methods: **(a)** standard breeding in the phase space (x, y, z); **(b)** nearest-neighbor breeding in the phase space (x, y, z); and **(c)** nearest-neighbor breeding in the reconstructed 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).

periment (a) is the standard breeding in the 3-D 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., $\hat{x}_i = \{x(t_i), y(t_i), z(t_i)\}$. 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-delay 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 (tar-

geted) 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 $l = 3$ 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 along the respective controls in the three experiments. The magni-

tudes 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 ($0 \leq g_i < 3.2$) in green, medium-growth points ($3.2 \leq g_i < 6.4$) in yellow, and high-growth points ($g_i \geq 6.4$) 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, in the original (Fig. 1b) and 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 capabilities 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, 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 increase 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

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

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 a predictor of fast growth in the dynamical system are two-fold. First, unlike the size of a particular variable, breeding can be tested in any dynamical model. 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. The bred vectors, by their ability to capture nonlinearity in instability growth, can characterize instability in dynamical systems in a more general manner than Lyapunov vectors and provide a way to obtain regime change in cases where the latter are computed from the time series data (Vassiliadis et al., 1991).

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).

Edited by: S. Vannitsem
Reviewed by: four anonymous referees

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. Meteor. 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, UK, 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 dimension is low, its predictability, high, Rev. Geophys., 33(S1), 645–650, doi:10.1029/95RG00495, 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 aspects of magnetopheric dynamics in local-linear filter, J. Geophys. Res., 107, 1369, doi:10.1029/2001JA009160, 2002.

Valdivia, J. A., Sharma, A. S., and Papadopoulos, K.: Prediction of magnetic storms by nonlinear models, Geophys. Res. Lett., 23, 2899–2902, 1996.

Vassiliadis, D., Sharma, A. S., and Papadopoulos, K.: Lyapunov exponent of magnetospheric activity from AL time series, Geophys. Res. Lett., 18, 1643–1646, 1991.

Vassiliadis, D., Klimas, A. J., Baker, D. N., and Roberts, D. A.: A description of the solar wind-magnetosphere 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.