Nonlinear Processes in Geophysics (2001) 8: 341–345

Nonlinear Processes in Geophysics
©European Geophysical Society 2001

Probing the linearity and nonlinearity in the transitions of the atmospheric circulation

A. A. Tsonis

Dept. of Mathematical Sciences, Atmospheric Sciences Group, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

Received: 5 October 2000 – Revised: 20 August 2001 – Accepted: 4 September 2001

Abstract. In this paper, we apply the principles of information theory that relate to the definition of nonlinear predictability, which is a measure that describes both the linear and nonlinear components of a system. By comparing this measure to a measure of linear predictability, one can assess whether a given system has a strong nonlinear or a strong linear component. This provides insights as to whether the system should be modeled by a nonlinear model or by a linear model. We apply these ideas to a known dynamical system and to a time series that describe the transitions in atmospheric circulation.

1 Introduction

The underlying principle behind the notion of linearity in dynamical systems is the fact that a linear combination of all fundamental solutions of a system provides the general solution of that system. Accordingly, since the general solution of linear dynamical systems is often periodic, periodic evolutions (whether produced by linear or nonlinear deterministic systems) can be adequately described by linear models. This linear character of periodic evolutions is associated with long-term predictability. A stochastic process is defined as linear when it can be represented as a linear combination of a sequence of white noise values with a zero mean and variance σ^2, which is analogous to the condition for linearity in dynamical systems. This leads to a variety of random processes, many of which realistically simulate natural processes. Nonlinearity can arise from chaotic dynamical systems that exhibit aperiodic behaviour and a rapid loss in predictability due to the sensitivity to initial conditions. It can also arise from nonlinear stochastic processes, which often are generated when white noise is applied to nonlinear dynamical systems. According to the above, a given process can exhibit linearity (due to periodicities and/or linear stochastic influences) and/or nonlinearity (due to chaotic dynamics and/or nonlinear stochastic influences).

A major concern in many scientific disciplines is whether a given process should be modeled as linear or as nonlinear. It is currently well accepted that many natural systems are nonlinear with feedbacks over many space and time scales. However, certain aspects of these systems may be less nonlinear than others and the nature of nonlinearity may not be always clear. The work here uses a newly developed method (Darbellay, 1998; Darbellay and Vajda, 1999) that provides insights into the nature of linearity and/or nonlinearity, thus allowing one to decide whether a specific process should be modeled with a linear or a nonlinear model. We apply this method to a time series from a known dynamical system (this series serves as our control example) and then to an observable from our climate system, which describes the transitions in the atmospheric circulation. While exploring nonlinearity in other geophysical records using similar approaches is credited to Palus and co-workers (Palus, 1996, 1995; Palus and Novotna, 1994; Palus et al., 1993), this work reports on a novel time series.

2 Information and predictability

The method is based on information theory. Let us consider two vectors of continuous random variables $\boldsymbol{X} = (X_1, ..., X_n)$ and $\boldsymbol{Y} = (Y_1, ..., Y_m)$ taking real values. The vector $\boldsymbol{X}$ can be regarded as the input set of variables of a system and $\boldsymbol{Y}$ as its output. Each X_i, $i = 1, n$ can be understood as a predictor variable and each Y_j, $j = 1, m$ as a predicted one. The values that X_i and Y_j are taking will be denoted as x_i and y_j. Let $p_x(\boldsymbol{x})$ and $p_y(\boldsymbol{y})$ be the probability density functions of the vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, and $p_{x,y}(\boldsymbol{x}, \boldsymbol{y})$ be their joint probability density function. The densities $p_x(\boldsymbol{x})$ and $p_y(\boldsymbol{y})$ will be referred to as the marginal densities. The random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent if and only if

$$p_{x,y}(\boldsymbol{x}, \boldsymbol{y}) = p_x(\boldsymbol{x}) p_y(\boldsymbol{y}) \qquad \text{for all } \boldsymbol{x} \in R^n \text{and} \boldsymbol{y} \in R^m.$$

Correspondence to: A. A. Tsonis(aatsonis@csd.uwm.edu)

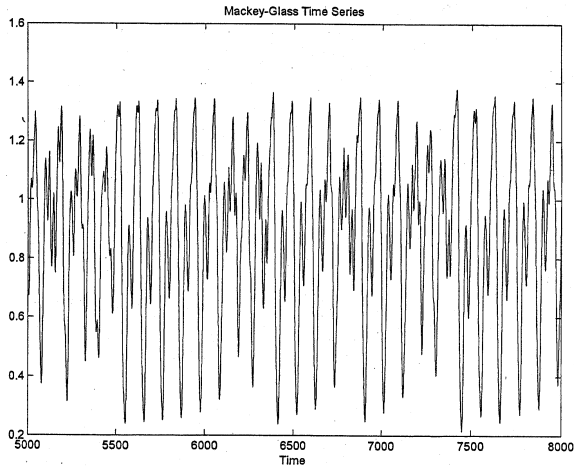


Fig. 1. A sample of the Mackey-Glass time series used here.

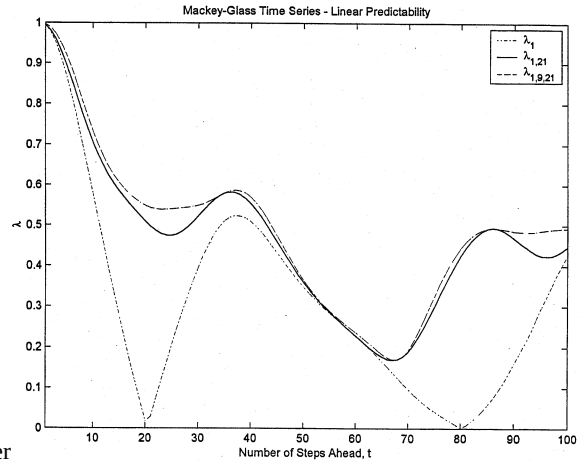


Fig. 2. Linear predictability for the Mackey-Glass time series for one, two, and three inputs. Linear predictability seems to saturate after two inputs. This indicates that the linearity of this dynamical system is due to a simple cycle (which requires an embedding dimension of two).

Given an observed system, this condition is easy to verify. What is more desirable is to have a single number, which will serve for comparing whether a pair of random vectors is independent; thus, its range can be arbitrary. We may choose zero for absolute independence and one for total dependence. One way of obtaining zero out of the above definition is to form the logarithm of the ratio of the joint probability with the product of the marginal probabilities. The expectation of the logarithm of this ratio is called the mutual information $I(\boldsymbol{X}, \boldsymbol{Y})$

$$I(\boldsymbol{X}, \boldsymbol{Y}) = \int_{R^d} p_{x,y}(\boldsymbol{x}, \boldsymbol{y}) \ln \frac{p_{x,y}(\boldsymbol{x}, \boldsymbol{y})}{p_x(\boldsymbol{x}) p_y(\boldsymbol{y})} d\boldsymbol{x} d\boldsymbol{y},$$

where $d = n + m$ is the dimension of the space spanned by the vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ if they are linearly independent. This integral, $0 \leq I \leq \infty$, plays a central role in information theory. To normalize it between zero and one, we can use the transformation

$$\rho = \sqrt{1 - e^{-2I}}.$$

The number ρ captures both the *linear* and *nonlinear* dependence between $\boldsymbol{X}$ and $\boldsymbol{Y}$ and it can be interpreted as the predictability of $\boldsymbol{Y}$ by $\boldsymbol{X}$. This measure of predictability is based on the probability distributions underlying the data and does not depend on the particular model used to predict $\boldsymbol{Y}$ from $\boldsymbol{X}$.

The reason for choosing the above transformation is that ρ reduces to some well-known measure of linear dependence when the joint probability density of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is Gaussian. For a d-dimensional Gaussian random vector $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y})$ the mutual information takes the form

$$I(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{2} \ln \frac{\det \Sigma_{xx} \det \Sigma_{yy}}{\det \Sigma},$$

where Σ, Σ_{xx}, and Σ_{yy} are the dxd, nxn, and mxm variance-covariance matrices of $\mathbf{Z}$, $\mathbf{X}$, and $\mathbf{Y}$, respectively. It can be shown that for $\mathbf{Z}$, the mutual information depends only on the coefficients of linear correlation (Darbellay, 1998). In fact, for $n = m = 1$, i.e. $(\boldsymbol{X}, \boldsymbol{Y}) = (X, Y)$, the above equation becomes

$$I(X, Y) = -1/2 \ln \left(1 - r^2(X, Y)\right),$$

where r is the correlation coefficient between X and Y. Given the definition of ρ, this indicates that if (X, Y) is Gaussian, then $\rho(X, Y) = |r(X, Y)|$. It follows that, in general, the linear predictability of $\boldsymbol{Y}$ by $\boldsymbol{X}$ is given by

$$\lambda(\boldsymbol{X}, \boldsymbol{Y}) = \sqrt{1 - \frac{\det \Sigma}{\det \Sigma_{xx} \det \Sigma_{yy}}}.$$

3 Applications

Despite the fact that ρ is an excellent generalization of λ, their difference cannot be equated to the nonlinear part of the predictability. However, their difference does signal the inadequacy of a linear model on the grounds that linear correlations capture only linear relationships. As such a difference between ρ and λ indicates that nonlinear models best describe the process in question. Note that the above formulations can be extended to time series, $x(t)$, by considering inputs as past values of the time series, and outputs as future values of the time series. For example, Fig. 1 shows a time series obtained from the Mackey-Glass model. This model is a nonlinear dynamical system described by the following time-delay differential equation

$$\frac{dx(t)}{dt} = \frac{a\, x(t - \tau)}{1 + x^c(t - \tau)} - b\, x(t),$$

where $a = 0.2$, $b = 0.1$, and $c = 10$ are constants and $\tau = 30$ is the delay parameter. For these values, the system has a chaotic attractor of dimension 3.6 (Wolf et al., 1985). Thus, to fully examine the dynamics, one needs a four-dimensional

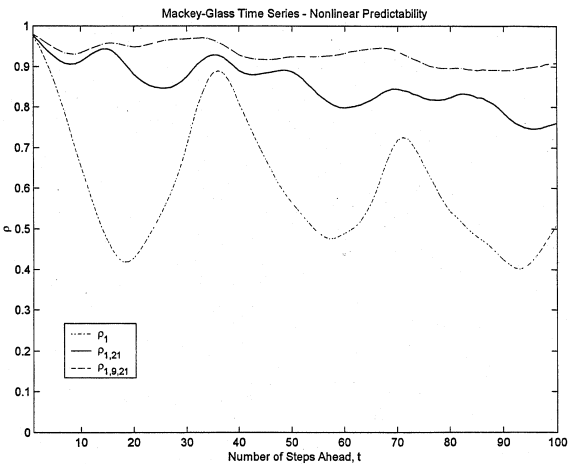
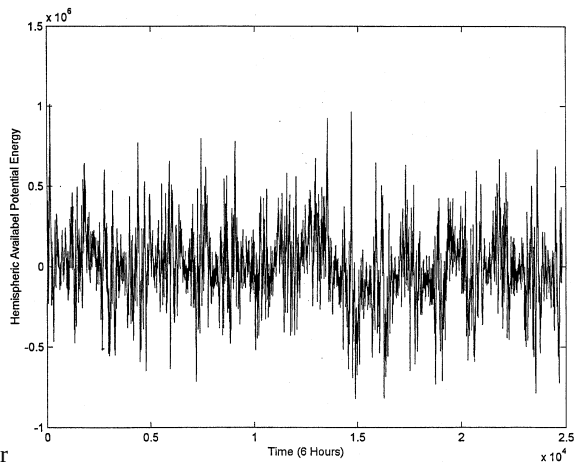


Fig. 3. Nonlinear predictability for the Mackey-Glass time series for the same inputs used in Fig. 2. Nonlinear predictability increases as the number of inputs increases. This occurs because the nonlinearity of this system needs a four-dimensional embedding and as a result, ρ increases as the number of inputs increases.



center

Fig. 4. The average hemispheric available potential energy (APE) as a function of time from 1 January 1979 to 31 December 1995. The time interval between successive values is 6 hours.

embedding. Figure 2 shows the linear predictability λ of $\boldsymbol{Y} = Y = x(t)$ as a function of time steps ahead ($t = 1, 100$) using: (1) one input, $\boldsymbol{X} = X_1 = x(-1)$ (i.e. only the value 1 step ago is used as a predictor of the future values; this is denoted as λ_1), (2) two inputs, $\boldsymbol{X} = (X_1, X_2)$ with $X_1 = x(-1)$ and $X_2 = x(-21)$ (i.e. the values 1 time step ago and 21 time steps ago are used to predict the future values; this is denoted as $\lambda_{1,21}$), and (3) three inputs, $\boldsymbol{X} = (X_1, X_2, X_3)$ with $X_1 = x(-1)$, $X_2 = x(-21)$, and $X_3 = x(-9)$ (i.e. the values 1 time step ago, 21 time steps ago, and 9 time steps ago are used to predict the future values; this is denoted as $\lambda_{1,9,21}$).

Figure 3 shows the nonlinear predictability ρ for the same conditions. For the above calculations, we use 100 000 values. The algorithm for the estimation of the nonlinear predictability is described in Darbellay (1998) and Darbellay and Vajda (1999). The general idea behind this algorithm is the partitioning of R^d into finite, non-intersecting subsets (cells) of R^d whose sum is the whole R^d. First, the entire R^d is partitioned into 2^d cells by dividing each of its d edges into two equiprobable intervals. Then, an estimation of the various probability distributions is achieved by finding the frequencies, i.e. by dividing the points in a cell by the total number of points. Subsequently, each cell is partitioned into 2^d subcells and the procedure is repeated. A cell is not subdivided any further if local independence on that cell has been achieved. Local independence is achieved when the following condition is satisfied

$$\frac{P_{\boldsymbol{X},\boldsymbol{Y}}(C_{kl})}{P_{\boldsymbol{X}}(C_{kl})\ P_{\boldsymbol{Y}}(C_{kl})} = \frac{P_{\boldsymbol{X},\boldsymbol{Y}}(C_k)}{P_{\boldsymbol{X}}(C_k)\ P_{\boldsymbol{Y}}(C_k)},$$

where P_X, P_Y are the probability distributions of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively, $P_{X,Y}$ is the joint probability distribution of $(\boldsymbol{X}, \boldsymbol{Y})$, C_k is the cell in question and C_{kl} is a subpartition of that cell (for more details, see Darbellay and Vajda, 1999). Finally, the mutual information is obtained by summing over all independent cells:

$$I(\boldsymbol{X}, \boldsymbol{Y}) = \sum_k P_{\boldsymbol{X},\boldsymbol{Y}}(C_k) \ln \frac{P_{\boldsymbol{X},\boldsymbol{Y}}(C_k)}{P_{\boldsymbol{X}}(C_k)\ P_{\boldsymbol{Y}}(C_k)}.$$

Going back to Figs. 2 and 3, we observe (1) a clear difference between the two graphs as expected and (2) the nonlinear predictability increases as the number of inputs increases, whereas the linear predictability remains virtually unchanged after two inputs. This indicates that the linearity of this dynamical system is due to a simple cycle (which requires an embedding dimension of two). The nonlinearity, however, needs a four-dimensional embedding and as a result, nonlinear predictability increases as the number of inputs increases. Note that the long-term decrease in ρ and λ is a manifestation of the loss in predictability associated with chaotic systems.

Next, we apply the above method to atmospheric data, namely mean hemispheric available potential energy (APE), which can loosely be defined as the portion of the atmosphere's potential energy that can be converted into kinetic energy. APE is a very important variable since its variability determines transitions in the atmospheric circulation. Details on the calculation of APE can be found in Wintels and Gyakum (2000). Figure 4 shows the time series which extends from 1 January 1979 to 31 December 1995 at a sampling rate of 6 hours (24 836 points). Figure 5 shows the linear predictability for 1, 2, and 3 inputs ($x(-1)$, $x(-7)$, and $x(-22)$) and Fig. 6 shows the nonlinear predictability for the same inputs. Here again we observe the differences between linear and nonlinear predictability, indicating that APE exhibits a nonlinear component. Note that APE also exhibits a linear component. This component is associated with the diurnal variability (a periodic component) and it manifests itself as an oscillation in both functions (remember that ρ contains information about both linear and nonlinear structures in the data). As can be seen from Fig. 7 (which shows the difference between ρ and λ), the linear component competes

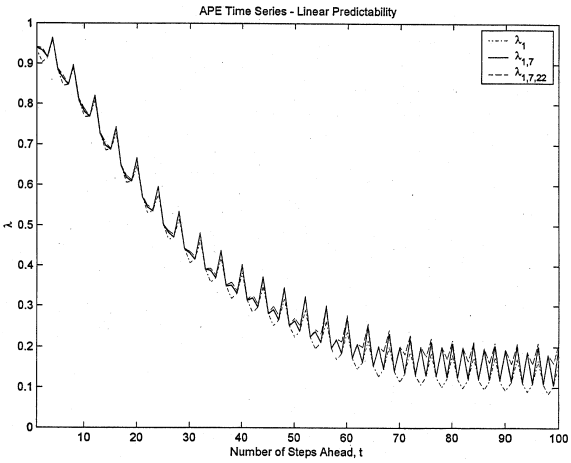


Fig. 5. Linear predictability of APE for one, two, and three inputs. Here again linear predictability saturates after two inputs, indicating a cycle, which corresponds to diurnal variability.

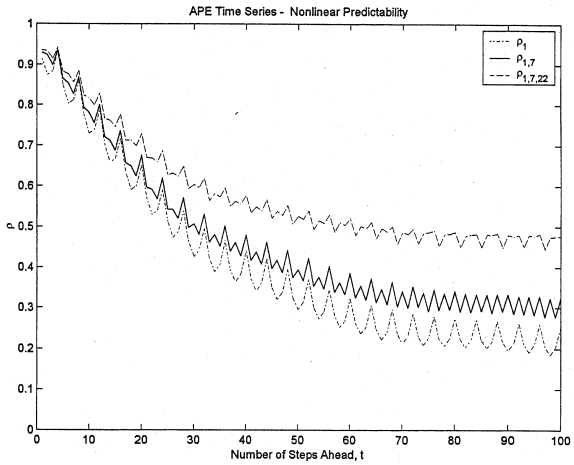


Fig. 6. Nonlinear predictability of APE for the same inputs used in Fig. 5. Here again we observe that nonlinear predictability increases as the number of inputs increases, and that the nonlinear component in APE is stronger than the linear one.

with the nonlinear component over short time scales where $\rho - \lambda \approx 0$, whereas the nonlinear component becomes dominant at longer time scales. The above results indicate that the variability superimposed on the periodic component is nonlinear and as such, it should be modeled with a nonlinear model.

The question we need to ask next is: what is the source of nonlinearity in APE? Is it dynamical or stochastic? In the case of the Mackey-Glass system, we dealt with a long time series and we knew a priori that the data come from a nonlinear dynamical system. Thus, the differences between ρ and λ were expected and easy to interpret. However, when the method is applied to an observed and possibly short time series from a system whose mathematical formulation and physics may not be accurately known, one has to be careful in interpreting the differences between ρ and λ. For example, as the number of inputs and outputs increases, the estimation of the frequencies is done in high dimensional spaces and as such, it requires large amounts of data. Thus, if the dynamical nonlinearity in the data is high dimensional and the sample size is small, then the dynamical component will not be clearly delineated and will not be statistically significant. It follows that the nonlinearity in the data is nonexistent or if the sample size is not adequate, then $\rho \approx \lambda$.

If however, there is a difference between ρ and λ, then the nonlinear part in ρ may be due to (1) chaotic (deterministic) dynamics, or (2) nonlinear noise (recall the discussion in the first paragraph of the Introduction), or (3) both.

The following procedure, which has become a common practice in testing for the nonlinear structure in the data, can be considered to address these issues. In general, when we test for dynamical properties (such as dimensions, Lyapunov exponents, etc.), one sets the null hypothesis so that the observed data come from some appropriate stochastic surrogate process. Depending on the application, this appropriate surrogate process may be a linear or a nonlinear stochastic (Gaussian) process, with the same autocorrelation structure as the observed data. The former process is generated by taking the Fourier transform of the data, randomizing the phases and then taking the inverse Fourier transform. The latter involves similar steps, but now the amplitudes are adjusted to produce a nonlinear transformation of a linear Gaussian process (Theiler et al., 1992; Schreiber and Schmitz, 1996). In our case, since we are investigating nonlinear components, it is more appropriate that the null hypothesis involves nonlinear surrogates. Note that since the surrogates preserve the autocorrelation structure of the data, it is more or less guaranteed that $\lambda^{\text{data}} = \lambda^{\text{surr}}$. However, due to the randomization of the phases and the amplitude adjustment, $\rho^{\text{data}} \neq \rho^{\text{surr}}$. This means that if $\rho - \lambda$ from the observed data is significantly different from that of the surrogates, then the observed nonlinearity is most likely due to nonlinear dynamics and not due to this particular nonlinear noise. Accordingly, we proceed with the following approach: for a given set of inputs, we calculate the difference $\rho - \lambda$ from the observed time series. Then, using the amplitude-adjusted Fourier transform method, we generate many nonlinear surrogate time series of the same length as the observed data, and for each one and for the same set of inputs, we find $\rho^s - \lambda^s$. Finally, we compare the surrogates and the observed data. Figure 8 shows $\rho_{1,7,22} - \lambda_{1,7,22}$ for the APE data. It also shows $\rho^s_{1,7,22} - \lambda^s_{1,7,22}$ for several surrogate time series. The curves for the surrogate data are "bunched" together at about $\rho - \lambda = 0.08$. The results in Fig. 8 clearly show that the surrogates deviate significantly from the observed data. From a set of 1000 surrogate time series, we find that for any time step ahead that is greater than 24 (4 days), the null hypothesis is rejected at a significance level greater than 95% (the level increases with t). This means that the nonlinearity in APE is not due to the particular nonlinear stochastic process used in the null hypothesis. The null hypothesis is also rejected at shorter time

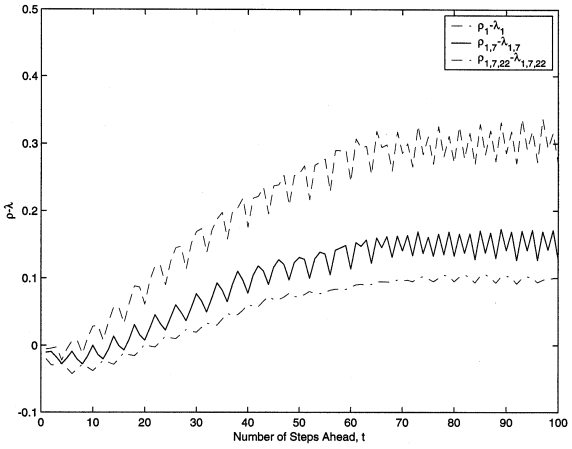


Fig. 7. This plot shows the difference between ρ and λ and indicates that the linear component competes with the nonlinear component over short time scales (where $\rho - \lambda \approx 0$), whereas the nonlinear component becomes dominant at longer time scales. Note that since ρ is the sum of the linear and nonlinear components, $\rho - \lambda$ cannot be negative. The small negative values for small time steps is probably due to numerical artifacts.

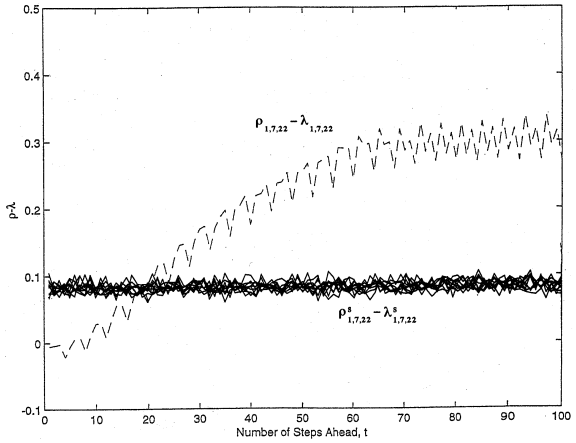


Fig. 8. This figure shows the difference between nonlinear and linear predictability for the APE data and for several surrogate data. The difference for the surrogate data is fluctuating at about $\rho - \lambda \approx 0.08$, whereas for APE, it shows a steady increase with the time step ahead. This result indicates that APE has for $t > 24$ a strong nonlinear component akin to nonlinear deterministic dynamics, whereas for $t < 16$, it exhibits a significant linear component (see text for details).

steps ($t < 3$ days) but this time it is rejected because of the presence of an appreciable linear component. The data for the three inputs were shown in this step because they delineate most of the nonlinearity. These results support the view that the observed nonlinearity in APE is a dynamical signal, superimposed on the linear component that corresponds to the diurnal variability.

4 Conclusions

We have applied a method that is very effective in delineating both the linear and nonlinear components in a time series, that describes the transitions in the atmospheric circulation. This is important because it makes it possible to infer information about the nature of the linearity and/or nonlinearity in this observable and to decide whether or not it should be treated as linear (deterministic or stochastic) nonlinear (dynamical or stochastic), or a combination of both. In the growing area of nonlinear time series analysis the method presented is an additional and complimentary tool to other already available tools (such as dimension estimates, Lyapunov exponents, nonlinear prediction). As with all other nonlinear methods (Tsonis, 1992; Tsonis et al., 1994) this method is not devoid of problems. Therefore, proper analysis and statistical tests must always be considered.

Acknowledgements. I would like to thank two anonymous reviewers for their comments. This work was party supported by NSF grant ATM-9727329.

References

Darbellay, G. A.: Predictability: An information-theoretic perspective, In: Signal Analysis and Prediction, (Eds) Prochazka, A., et al., Birkhauser, Boston, 249–262, 1998.

Darbellay, G. A. and Vajda, I.: Estimation of the information by adaptive partitioning of the observation space, IEEE Trans. Inform. Theory, 45, 1315–1321, 1999.

Palus, M.: Testing for nonlinearity using redundancies-Quantitative and qualitative aspects, Physica D., 80, 186–205, 1995.

Palus, M.: Detecting nonlinearity in multivariate time series, Phys. Lett., A, 213, 138–147, 1996.

Palus, M. and Novotna, D.: Testing for nonlinearity in weather records, Phys. Lett., A, 193, 57–74, 1994.

Palus, M., Albrecht, V., and Dvorak, I.: Information theoretic test for nonlinearity in time series, Phys. Lett.. A. 175, 203–209, 1993.

Schreiber, T. and Schmitz, A.: Improved surrogate data for nonlinearity tests, Phys. Rev., E, 77, 635–638, 1996.

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

Tsonis, A. A.: Chaos: From theory to applications, Plenum, New York, 1992.

Tsonis, A. A., Triantafyllou, G. N., and Elsner, J. B.: Searching for determinism in observed data: a review of the issues involved, Nonlinear Processes in Geophysics, 1, 12–25, 1994.

Wintels, W. and Gyakum, J.R.: Synoptic climatology of Northern hemisphere available potential energy collapses, Tellus, B52, 347–364, 2000.

Wolf, A., Swift, J. B., Swinney, H. L. and Vastano, J. A.: Determining Lyapunov exponents from a time series, Physica. D., 16, 285–317, 1985.