Neural Network World 7(4,5) (1997) 451-460 .

FROM NONLINEARITY TO PREDICTABILITY

M. Paluš* and D. Novotná†

June 5, 1997

Abstract

Detection of nonlinearity in experimental time series is usually based on rejection of a linear null hypothesis by a statistical test. Typically, the null hypothesis is a Gaussian process or a Gaussian process passed through a static nonlinear transformation or a similar simple alternative. Rejection of such a null is frequently interpreted as a detection of a deterministic nonlinear relation in data under study, which is, however, only one of possible alternatives. We show how variable variance, or seasonality in variance could lead to spurious identification of deterministic nonlinearity and discuss how to distinguish actual determinism in studied time series.

1 Introduction

Let $\{y(t)\}$ be a time-series, i.e., a series of measurements done on a system in consecutive instants of time $t = 1, 2, \ldots$. We will discuss approaches to processing and prediction for such a data.

The time series $\{y(t)\}$ can be considered as a realization of a stationary linear stochastic process $\{Y(t)\}$. Without loss of generality we can set its mean to zero. Then the linear stochastic process $Y(t)$ can be written as:

$$Y(t) = Y(0) + \sum_{i=1}^{\infty} a(i)Y(t-i) + \sum_{i=0}^{\infty} b(i)N(t-i), \tag{1}$$

where $b(0) = 1$, $\sum_{i=1}^{\infty} |a(i)| < \infty$, $\sum_{i=0}^{\infty} |b(i)| < \infty$, $\{N(t)\}$ is an independent, identically distributed (iid), normally distributed process with zero mean and finite (constant) variance. (For more details see [18].)

Alternatively, the time series $\{y(t)\}$ can be considered as a projection of a trajectory of a dynamical system, evolving in some measurable d-dimensional state space. To be more specific, let X_t denote a state vector in R^d. Then the measurements $y(t)$ are obtained as $y(t) = g(X_t)$, where $g(.)$ is a projection (measurement function), and temporal evolution of X_t may be described by a discrete-time dynamical system (a difference equation):

$$X_t = F(X_{t-1}), \tag{2}$$

with $X_0 \in R^d$ and for $t \geq 1$. The vector-valued function F is also called a map. If the map F was linear, then for $t \to \infty$, either $| X_t | \to \infty$ (the unstable case) or $| X_t | \to c$ (a constant), such that $c = F(c)$ (the stable case). These types of behaviour are obviously not interesting for time series analysts and therefore we will consider nonlinear maps F, which provide the following three possibilities: 1) periodic solution (for $t \to \infty$, $X_t \to \{c_1, ..., c_p\}$, $F(c_1) = c_2, ..., F(c_{p-1}) = c_p, F(c_p) = c_1$), 2) quasi-periodic solution (for $t \to \infty$, $X_t \to$ a sum, or some other smooth function, of a finite number of periodic functions with incommensurate

*Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic

†Institute of Atmospheric Physics, Academy of Sciences of the Czech Republic, Boční II/1401, 141 31 Prague 4, Czech Republic

periods); and 3) deterministic chaos (for $t \to \infty$, $X_t \to$ non-degenerate closed subspace of R^d, called strange attractor).

Traditionally, processing and prediction in complex experimental time series were dominated by linear methods based on spectral analysis and ARMA (auto-regressive moving average) models such as presented in Eq. (1). In eighties, new methods for nonlinear signal processing have been introduced, which have not arisen as an extension of linear analysis, but have been conceived due to an inspiration provided by theory of nonlinear dynamical systems. While the above cases 1 and 2, i.e., periodic and quasiperiodic dynamics are of a limited interest when considering complex experimental time series, the case of deterministic chaos attracts attention of researchers due to the finding that deterministic chaotic systems can possess complex, aperiodic, and irregular (seemingly random) solutions to simple mathematical formulas. New and innovative techniques for time series analysis based on the idea of deterministic chaos have provided experimentalists with new ways of understanding the implications of their data, though the limitations of these new techniques have not always been understood and necessary precautions fully appreciated.

In attempts to explain complicated dynamics, deterministic chaos has been usually considered as an opposite alternative to random processes. Recent results indicate, however, that low-dimensional chaos may be rather a rare than ubiquitous phenomenon, especially when open, real-world systems are considered [15]; or, the strict separation between deterministic-chaotic and stochastic dynamics may be impossible [5]. Also, we may argue in accord with Tong [21] that in reality observations rarely evolve according to system (2), simply because stochastic noise is ubiquitous, which may arise as a result of one or more of the following sources: (i) a model used is invariably inexact; (ii) there are always unexpected external random disturbances; (iii) measurements are often inexact. It is thus more realistic to replace the above states by random variables and the dynamics by a Markovian model such as

$$X_t = F(X_{t-1}, e_t), \tag{3}$$

where $t \in Z_+$, $F : R^{2d} \to R^d, \{e_t\}$ is a sequence of independent and identically distributed d-dimensional random vectors and e_t is independent of X_s, $0 \le s < t$. We shall call $\{e_t\}$ the *dynamic noise.* Following Tong [21], we shall refer to equation (2) as the *skeleton* of model (3), considering $F(X) = F(X, 0)$. For convenience, it is frequently assumed that the dynamic noise is additive so that equation (3) reduces to the *model with additive noise*

$$X_t = F(X_{t-1}) + e_t. \tag{4}$$

Exploring nonlinear models (4), a number of nonlinear prediction methods have been developed (see e.g. [22] and references therein). Such methods can principally improve reliability of time series prediction in comparison to linear (ARMA) models (Eq. 1), provided the evolution of a process underlying the experimental time series indeed possesses a nonlinear deterministic skeleton such as (2). Otherwise, training nonlinear predictors on essentially linear stochastic data can be considered as a waste of human and computer resources without any improvement over an ARMA model. Therefore objective methods for detecting nonlinear determinism are desirable.

2 A test for nonlinearity based on redundancies and surrogate data

Consider n discrete random variables $X_1, \ldots, X_n$ with sets of values $\Xi_1, \ldots, \Xi_n$, respectively. The probability distribution for an individual X_i is $p(x_i) = \Pr\{X_i = x_i\}$, $x_i \in \Xi_i$. We denote the probability distribution function by $p(x_i)$, rather than $p_{X_i}(x_i)$, for convenience. Analogously, the joint distribution for the n variables $X_1, \ldots, X_n$ is $p(x_1, \ldots, x_n)$. The redundancy $R(X_1; \ldots; X_n)$, in the case of two variables also known as mutual information $I(X_1; X_2)$, quantifies average amount of common information, contained in the n variables $X_1, \ldots, X_n$:

$$R(X_1; \ldots; X_n) = \sum_{x_1 \in \Xi_1} \cdots \sum_{x_n \in \Xi_n} p(x_1, \ldots, x_n) \log \frac{p(x_1, \ldots, x_n)}{p(x_1) \ldots p(x_n)}. \quad (5)$$

When the discrete variables $X_1, \ldots, X_n$ are obtained from continuous variables on a continuous probability space, then the redundancies depend on a partition ξ chosen to discretize the space. Various strategies have been proposed to define an optimal partition for estimating redundancies of continuous variables (see [12, 22] and references therein). Here we use the "marginal equiquantization" method described in detail in [10, 12].

Now, let the n variables $X_1, \ldots, X_n$ have zero means, unit variances and correlation matrix $\mathbf{C}$. Then, we define the *linear redundancy* $L(X_1; \ldots; X_n)$ of $X_1, X_2, \ldots, X_n$ as

$$L(X_1; \ldots; X_n) = -\frac{1}{2} \sum_{i=1}^{n} \log(\sigma_i), \quad (6)$$

where σ_i are the eigenvalues of the $n \times n$ correlation matrix $\mathbf{C}$.

If $X_1, \ldots, X_n$ have an n-dimensional Gaussian distribution, then $L(X_1; \ldots; X_n)$ and $R(X_1; \ldots; X_n)$ are theoretically equivalent.

In practical applications one deals with a time series $\{y(t)\}$, considered as a realization of a stochastic process $\{Y(t)\}$, which is stationary and ergodic. Then, due to ergodicity, all the subsequent information-theoretic functionals are estimated using time averages instead of ensemble averages, and the variables X_i are substituted as

$$X_i = y(t + (i-1)\tau). \quad (7)$$

Due to stationarity the redundancies

$$R^n(\tau) \equiv R(y(t); y(t+\tau); \ldots; y(t+(n-1)\tau)), \quad (8)$$

$$L^n(\tau) \equiv L(y(t); y(t+\tau); \ldots; y(t+(n-1)\tau)), \quad (9)$$

are functions of n and τ, independent of t.

The surrogate-data based nonlinearity tests [19, 12] consist of computing a *nonlinear* statistic from data under study and from an ensemble of realizations of a linear stochastic process, which mimics "linear properties" of the studied data. If the computed statistic for the original data is significantly different from the values obtained for the surrogate set, one can infer that the data were not generated by a linear process; otherwise the null hypothesis, that a linear model fully explains the data, cannot be rejected and the data can be analyzed, characterized and predictions can be obtained by using well-developed linear methods. For the purpose of such test the surrogate data must preserve the spectrum[1] and consequently, the autocorrelation function of the series under study. An isospectral linear stochastic process to a series can be constructed by computing the Fourier transform (FT) of the series, keeping unchanged the magnitudes of the Fourier coefficients, but randomizing their phases and computing the inverse FT into the time domain. Different realizations of the process are obtained using different sets of the random phases.

To evaluate the test, the test statistic is defined as the difference between the redundancy obtained for the original data and the mean redundancy of a set of surrogates, in the number of standard deviations (SD's) of the latter. Thus both the redundancies and redundancy-based statistics can be evaluated as functions of the lag τ and the embedding dimension n.

The redundancy $R^n(\tau)$, based on probability distributions, measures general dependences among the series $\{y(t)\}$ and its lagged versions, whereas the linear redundancy $L^n(\tau)$, based on correlations, reflects

[1] Also, preservation of histogram is usually required. A histogram transformation used for this purpose is described in [12] and references within.

only their linear relations. Comparing the plots of $R^n(\tau)$ and $L^n(\tau)$ can provide an informal test for important nonlinearities in the studied data [8]. This approach, in [12] referred to as *qualitative testing*, or, *qualitative comparison*, can bring some additional information to the results of the quantitative (surrogate data based) test. Moreover, one can not always construct good surrogate data. That is, despite theoretical expectations, in numerical practice linear properties of the surrogates may differ from those of the data under study. Changes in linear properties are reflected in nonlinear statistics as well, and thus may result in spurious detection of nonlinearity in linear data [12]. Therefore, we also evaluate the statistic based on the linear redundancy $L^n(\tau)$, which specifically reflects changes in linear properties. Then, only those significant differences in the nonlinear statistic can reliably count for nonlinearity, which are not detected in the linear statistic [12].

3 The null hypothesis and its negations

Consider that the above described test yielded a significant result, i.e., the null hypothesis was reliably rejected. The null hypothesis was equivalent[2] to a linear stochastic process such as that described by the ARMA model (1). It is very common in nonlinear dynamics literature to consider the rejection of the null (1) as an evidence for a process such as (4) with a nonlinear skeleton (2). This is, however, only one of possible negations of (1). A number of different processes should be considered, which possess a linear deterministic skeleton[3], i.e., a linear AR part – the first sum in (1), or no deterministic skeleton at all (MA processes), however, their innovations $\{N(t)\}$ do not fulfill the conditions given above. Generally, one or more of the following properties could reject the null (1):

1. The innovations $\{N(t)\}$ are not Gaussian.
2. The innovations $\{N(t)\}$ are not an iid process, where iid means that the innovations should be not only uncorrelated, but generally independent.
3. The variance of $\{N(t)\}$ is not constant.

4 An example of atmospheric pressure series

A series of daily recordings of atmospheric surface pressure {P(t)} ($t = 1 - 65{,}536$ days; i.e., 180 years) was described and analyzed in [11]. Here we analyze differences from long-term daily averages. This transformation of data (almost entirely) removes the oscillations with the period of one year (seasonality in mean). We ask the question about possible long-term nonlinear dependence and perform the linear redundancy – redundancy surrogate data test, described above. The results are presented in Fig. 1. For lags larger than a few days there is only a weak linear dependence, as measured by the linear redundancy (Fig. 1a) and reflected in the surrogates, but the (nonlinear) redundancy detects a clear dependence as an oscillatory structure with a yearly periodicity. This difference is highly significant (10 – 30 SD's, Fig. 1d), while no significance in the linear statistic (Fig. 1c) confirm the quality of the surrogates, which reflect correctly "the linear properties" of the data (in the sense of the model (1)). Can this result be understood as an evidence for the model (4) with a nonlinear periodic skeleton F, which could provide predictability of the atmospheric pressure for several years in advance?

The seasonality in mean present in this data (Fig. 2, upper panel) was mostly removed by considering differences from the long term averages.[4] The problem is that also the variance of this data is not constant, but clearly seasonal (Fig. 2, lower panel, standard deviation (SD) is the square root of variance). This

[2]Principal equivalence and technical differences between the surrogate data constructed by using the Fourier transform and the ARMA modeling are discussed in [19, 20].

[3]Obviously, for a linear function F, the model (4) is a special case of (1).

[4]The seasonality in mean can be entirely removed by filtration in spectral domain, as done in [11].

property is "nonlinear" in a sense that the surrogate data and the model (1) possess a constant variance and cannot reproduce the seasonality in variance. After rescaling the data in order to obtain a constant variance, the effect of the false long-term nonlinear dependence is lost (Fig. 3). In the rescaled data there is no long-term dependence except of weak linear link due to not entirely removed seasonality in mean.

5 Discussion and conclusion

The influence of variable variance on the redundancy – surrogate data nonlinearity test, introduced in Sec. 2 and in detail in [12], has been demonstrated in the previous section. The effect of non-Gaussian innovations $\{N(t)\}$ was discussed in [12], and a possible influence of non-iid $\{N(t)\}$ (i.e., innovations containing (nonlinear) temporal structures) is understandable. It is important to note that similar effects of "defective" innovations in a process under study would effect not only this particular test for nonlinearity, but all tests which use some type of FT/ARMA surrogates, and also any method which contain the process (1) at least implicitly in its construction, e.g., so-called qualitative comparison of linear redundancy and redundancy (Sec. 2, Ref. [12]), or comparison of linear correlations with so-called generalized correlations, based on mutual information [4]. Also, all entropy-related statistics, that is, not only the above information-theoretic functionals, but also, for instance, statistics based on correlation integral [17], are extremely sensitive to variable variance and/or to (non)Gaussianity of data/innovations. Therefore one must very carefully assess results of nonlinearity tests in order to avoid confusing this kind of effects with actual functional dependence in the data under study. Therefore assessing (in)variability of variance should be a part of basic statistical description of the data before any testing. The other above discussed effects are hard to detect without a priori knowledge of the underlying model/mechanism. A certain help could be provided by using the statistics which test directly properties of functional dependence – e.g., the presence of a vector field of a hypothetical underlying dynamical system [6, 7], and continuity and differentiability of a hypothetical functional dependence [16]. Unfortunately, even these "direct deterministic" statistics are relativized by the ubiquity of noise and the necessity to set up critical values or confidence intervals by some kind of bootstrap or surrogate data technique. For instance, the behaviour of the Kaplan-Glass direct test for determinism [6, 7] has been found practically equivalent to that of the above redundancy test [9]. An interesting graphical tool for assessing nonlinear dependence in time series can be provided by using so-called residual-delay plots, proposed by Casdagli [3].

In conclusion we state that there is a long way from nonlinearity to predictability, i.e., from formal detection of nonlinearity in a test to evidence for nonlinear functional dependence and its utilization by nonlinear prediction methods. It is necessary to combine several tests for nonlinearity and determinism with different properties to avoid spurious detection of functional dependence. Another way is developing special types of surrogate data, tailored to specific properties of data under study and expected dependencies.

Acknowledgements

The authors were supported by the Grant Agency of the Czech Republic (grant No. 205/97/0921).

References

[1] T.W. Anderson, *The Statistical Analysis of Time Series* (Wiley, New York, 1971).

[2] Box G.E.P., G.M. Jenkins (1976) *Time series analysis, forecasting and control.* Holden-Day, San Francisco.

[3] Casdagli M.C., Iasemidis L.D., Sackellares J.C., Roper S.N., Gilmore R.L. and Savit R.S., Characterizing nonlinearity in invasive EEG recordings from temporal lobe epilepsy. *Physica D* **99(2-3)** (1996) 381-399.

[4] Darbellay G., A non-linear correlation measure and its application to financial time series. *Neural Network World* **4** (1995) 401-405.

[5] Ellner S., P. Turchin (1995) Chaos in noisy world: new methods and evidence from time series analysis. *The American Naturalist* **145(3)** 343-375.

[6] D.T. Kaplan and L. Glass, Direct test for determinism in a time series, *Phys. Rev. Lett.* **68 (4)** (1992) 427-430.

[7] D.T. Kaplan and L. Glass, Coarse-grained embeddings of time series: random walks, Gaussian random process, and deterministic chaos, *Physica D* **64** (1993) 431-454.

[8] M. Paluš, V. Albrecht and I. Dvořák, Information-theoretic test for nonlinearity in time series, *Phys. Lett. A* **175** (1993) 203-209.

[9] Paluš, M.: Testing for Nonlinearity in the EEG, In: Jansen, B.H., Brandt, M.E. (eds.): Nonlinear Dynamical Analysis of the EEG, World Scientific, Singapore 1993, pp. 100-115.

[10] M. Paluš, Identifying and quantifying chaos by using information-theoretic functionals, in: *Time Series Prediction: Forecasting the Future and Understanding the Past,* A.S. Weigend and N.A. Gershenfeld, (eds.), Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XV (Addison–Wesley, Reading, Mass., 1993) pp. 387-413.

[11] M. Paluš and D. Novotná, Testing for nonlinearity in weather records, *Phys.Lett. A* **193** (1994) 67-74.

[12] Paluš M. (1995) Testing for nonlinearity using redundancies: Quantitative and qualitative aspects. *Physica D* **80** 186-205.

[13] Paluš M., L. Pecen, D. Pivka (1995) Estimating predictability: Redundancy and surrogate data method. *Neural Network World* **4** 537-552.

[14] Paluš M. (1996) Detecting nonlinearity in multivariate time series. *Phys. Lett. A*, **213(3,4)** 138.

[15] Paluš M. (1996) Nonlinearity in Normal Human EEG: Cycles, temporal asymmetry, nonstationarity and randomness, not chaos. *Biological Cybernetics* **75 (5)** (1996) 389-396.

[16] Pecora, L.M., Carroll, T.L. and Heagy, J.F. (1996) Statistics for Continuity and Differentiability: An Application to Attractor Reconstruction from Time Series. In C.D. Cutler and D.T. Kaplan (Eds.), *Nonlinear Dynamics and Time Series: Building a Bridge Between the Natural and Statistical Sciences*, Fields Institute Communications, Vol. 11, American Mathematical Society, Providence, Rhode Island, pp. 49-62.

[17] D. Prichard and J. Theiler, Generalized redundancies for time series analysis, *Physica D* **84** (1995) 476-493.

[18] M. Priestley, *Spectral Analysis and Time Series* (Academic Press, New York, 1981).

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

[20] J. Theiler and D. Prichard, Constrained-realization Monte-Carlo method for hypothesis testing, *Physica D* **94(4)** (1996) 221.

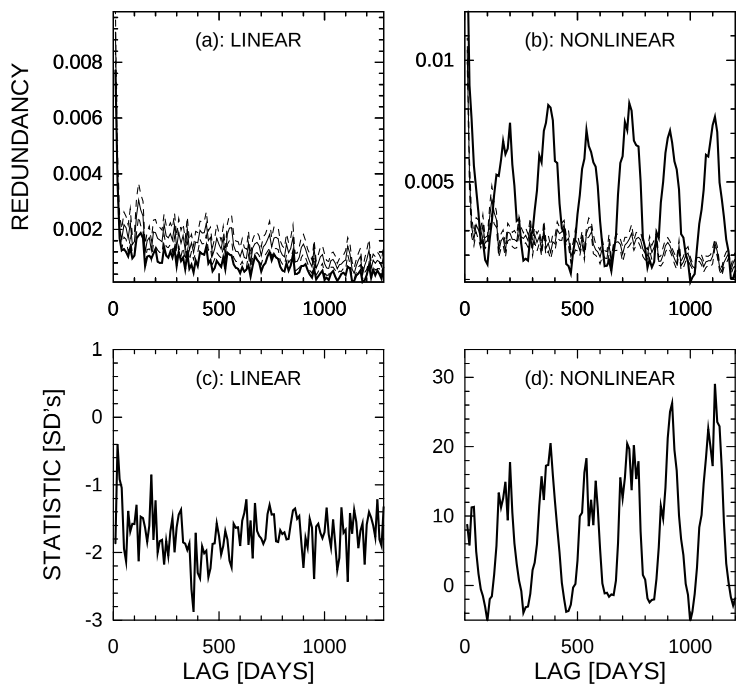

Figure 1: a*): Linear redundancy* $L(y(t); y(t+\tau))$ *(solid line),* b*): nonlinear (general) redundancy* $R(y(t); y(t+\tau))$ *(solid line), for a series of differences from the long term averages of the surface atmospheric pressure (Prague-Klementinum station) and for its FT surrogates (thin solid and dashed lines present mean and mean±SD, respectively, of a set of 30 surrogate realizations); c): linear (L-based), and d): nonlinear (R-based) statistics; as functions of the time lag* τ*, measured in days.*

[21] H. Tong, A Personal Overview of Nonlinear Time Series Analysis from a Chaos Perspective, *Scandinavian Journal of Statistics,* **22 (4)** (1995) 399-445.

[22] A.S. Weigend and N.A. Gershenfeld, (eds.), *Time Series Prediction: Forecasting the Future and Understanding the Past,* Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XV (Addison–Wesley, Reading, Mass., 1993)

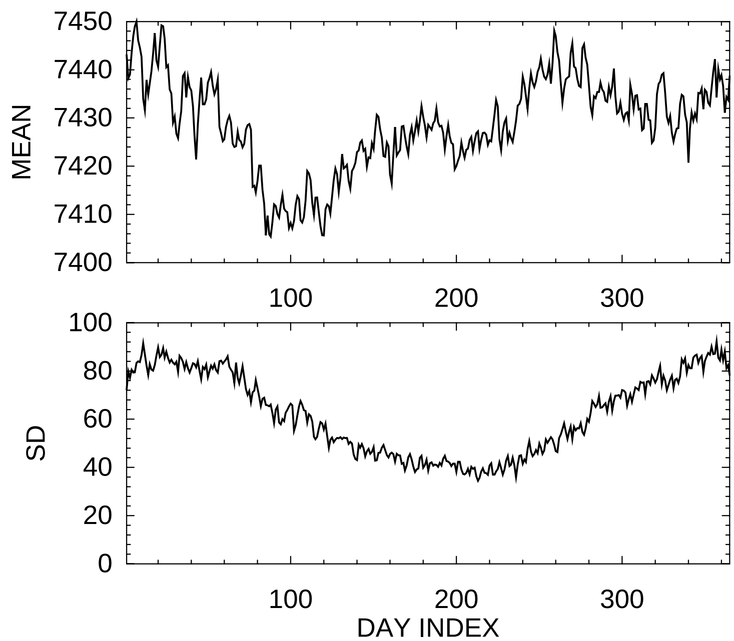

Figure 2: *Seasonality in mean (upper panel) and in variance (lower panel) of the surface atmospheric pressure series – long term means (upper panel) and standard deviations (square root of variance, lower panel) for each day in a year. The days are consecutively numbered, January 1 has the index 1, January 2 has the index 2, ..., February 1 has the index 32, etc.*

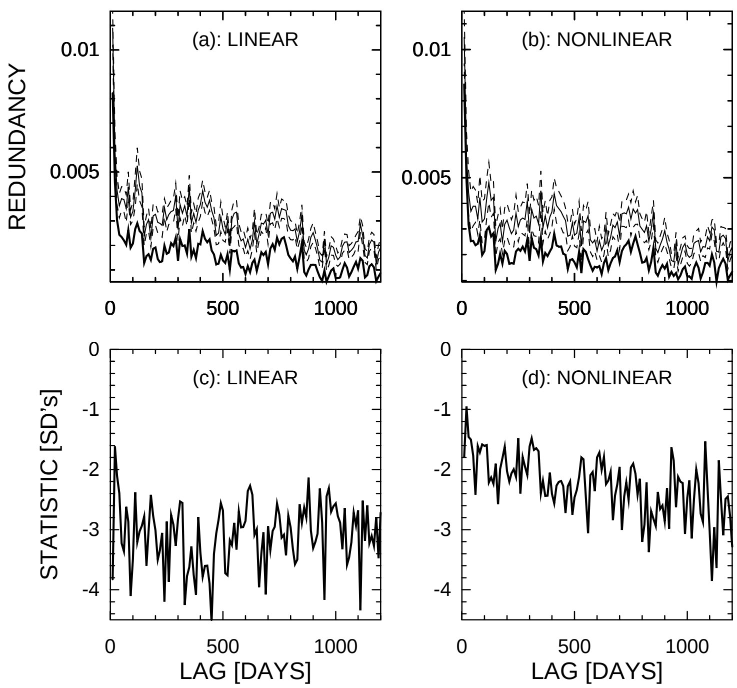

Figure 3: a*): Linear redundancy* $L(y(t); y(t+\tau))$ *(solid line), b): nonlinear (general) redundancy* $R(y(t); y(t+\tau))$ *(solid line), for a series of differences from the long term averages of the surface atmospheric pressure, rescaled in order to have a constant variance, and for its FT surrogates (thin solid and dashed lines present mean and mean±SD, respectively, of a set of 30 surrogate realizations); c): linear (L-based), and d): nonlinear (R-based) statistics; as functions of the time lag* τ*, measured in days.*