Probing Nonlinear Relations in Multivariate Time Series

Milan Paluš

Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic
E-mail: `mp@uivt.cas.cz, mp@santafe.edu`

In this chapter we briefly review a method for detection and characterization of nonlinear relations in multivariate as well as in univariate time series. The method employs the technique of uni- and multivariate surrogate data and information-theoretic functionals called redundancies. The test for nonlinearity based on the redundancy – linear redundancy approach, combined with the surrogate data is described in detail in (Paluš, 1995), its multivariate version in (Paluš, 1996a). The univariate surrogate data have been introduced in (Theiler et al., 1992), and the multivariate surrogate data in (Prichard and Theiler, 1994). More details about the information-theoretic functionals can be found in (Cover and Thomas, 1991).

Consider n discrete random variables $X_1, \ldots, X_n$ with sets of values Ξ_1, ..., Ξ_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 $X_1, \ldots, X_n$:

$$R(X_1; \ldots; X_n) \; = \; \sum_{x_1 \in \Xi_1} \ldots \sum_{x_n \in \Xi_n} p(x_1, \ldots, x_n) \log \frac{p(x_1, \ldots, x_n)}{p(x_1) \ldots p(x_n)}. \qquad (1)$$

The marginal redundancy $\varrho(X_1, \ldots, X_{n-1}; X_n)$ quantifies the average amount of information about the variable X_n, contained in the $n-1$ variables X_1, ... , X_{n-1}:

$$\varrho(X_1, \ldots, X_{n-1}; X_n) \; = \; \sum_{x_1 \in \Xi_1} \ldots \sum_{x_n \in \Xi_n} p(x_1, \ldots, x_n) \log \frac{p(x_1, \ldots, x_n)}{p(x_1, \ldots, x_{n-1}) p(x_n)}. \qquad (2)$$

The relation

$$\varrho(X_1, \ldots, X_{n-1}; X_n) = R(X_1; \ldots; X_n) - R(X_1; \ldots; X_{n-1}) \qquad (3)$$

can be derived by simple manipulation. In addition to the redundancy (1) and the marginal redundancy (2), various types of redundancies can be defined, quantifying the average amounts of information between/among variables or groups of variables. Also conditional redundancies can be considered. See (Paluš, 1996a) for details.

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

$$L(X_1; \ldots; X_n) = -\frac{1}{2} \sum_{i=1}^{n} \log(\sigma_i), \tag{4}$$

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 (Morgera, 1985).

Based on (3) we define the *linear marginal redundancy* $\lambda(X_1, \ldots, X_{n-1}; X_n)$, quantifying linear dependence of X_n on $X_1, \ldots, X_{n-1}$, as

$$\lambda(X_1, \ldots, X_{n-1}; X_n) = L(X_1; \ldots; X_n) - L(X_1; \ldots; X_{n-1}). \tag{5}$$

Similarly, for any kind of general (nonlinear) redundancy its linear equivalent exists (Paluš, 1996a). The general redundancies R detect all dependences in data under study, while the linear redundancies L are sensitive only to linear structures. For detailed discussion see (Paluš, 1995).

The basic idea in the surrogate-data based nonlinearity test is to compute 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 statistic computed 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, is accepted and the data can be further analyzed and characterized 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. In the multivariate case also cross-correlations of all pairs of variables must be preserved. 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 by using different sets of random phases. In the multivariate case, the cross-correlations can be preserved by preserving the original phase differences between the variables, i.e., the phases are randomized by adding random numbers, so that for a particular frequency bin

[1] Also, preservation of histogram is usually required. A histogram transformation used for this purpose is described in (Paluš, 1995) and references within.

the same random number is added to related phases of all the variables. More details can be found in (Prichard and Theiler, 1994) and references therein.

An experimentalist usually deals with a multivariate time series $\{x_1(t), \ldots, x_n(t)\}$, $t = 1, \ldots, N$, which is considered as a realization of a multivariate, stationary and ergodic stochastic process $\{X_1(t), \ldots, X_n(t)\}$. Then, due to ergodicity, all redundancies can be estimated using time averages instead of ensemble averages; in particular, correlation matrices in (4) are obtained as the time averages over the series, and probability distributions, used in computation of the redundancies R, are estimated as time-averaged histograms. When the discrete variables $X_1, \ldots, X_n$ are obtained from continuous variables on a continuous probability space, then the redundancies R 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 (Paluš et al., 1993), (Paluš, 1993), (Paluš, 1995), (Weigend and Gershenfeld, 1993) and references therein). We have found that satisfactory results can be obtained by using simple box-counting method and by observing the following two rules:
a) The partition is defined by the marginal equiquantization method, i.e., the marginal histogram bins are defined not equidistantly but so that there is approximately the same number of samples in each marginal bin.
b) The relation between the number Q of quantization levels (marginal bins) and the effective[2] series length N in the computation of n-dimensional redundancy should be

$$N \geq Q^{n+1}, \tag{6}$$

otherwise results may be heavily biased.

Applying this simple recipe, the redundancy estimator should bring consistent estimates in a relative sense, not unbiased estimates of absolute values, because the absolute values of the redundancies are not important here. Subjects of interest here are differences between redundancy estimates obtained in the same numerical conditions: differences between the redundancies obtained from the scrutinized data and its surrogates, as well as behaviour of the redundancy as a function of time lags, or estimates (using the same numerical parameters) of the redundancies from data recorded in different experimental (physiological/pathological) conditions.

In the case of multivariate data the marginal equiquantization is applied to each variable separately. The marginal equiquantization effectively means a transformation of data into a uniform distribution. Pompe (This volume) also uses this kind of partitioning, however, his approach to final redundancy

[2] If a univariate series is used to construct a time-delay n-dimensional embedding, the effective series length N is $N = N_0 - (n-1)\tau$, where N_0 is the total series length, n is the embedding dimension, and τ is the time delay. In a multivariate case the effective series length is $N = N_0 - \tau$.

estimation is different.

Having chosen an appropriate type of the redundancy, say $R(X_1; \ldots; X_n)$, it is possible and useful to investigate dependences among lagged series, i.e., to evaluate redundancies of the type $R(x_1(t); x_2(t+\tau_1); \ldots; x_n(t+\tau_{n-1}))$. Due to stationarity this redundancy does not depend on the time t and is a function of the lags $\tau_1, \ldots, \tau_{n-1}$.

Like in (Paluš, 1995) we define the test statistic 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 are functions of the lags $\tau_1, \ldots, \tau_{n-1}$, and their graphs are $(n-1)$-dimensional hyperplanes. Such objects are hard to study and therefore it would be practically useful to define a few one-dimensional cuts of the hyperplanes, e.g., the cuts along the axes (i.e., setting all τ's but one to be equal to zero). When testing nonlinearity in univariate series (Paluš, 1995) also redundancies of several variables are evaluated. Those variables, however, are obtained using a one lagged variable with $\tau_i = i\tau$, with one variable τ. Evaluating the redundancies and related statistics for broad ranges of the lags can bring a problem of simultaneous statistical inference (see (Paluš, 1995), (Paluš and Novotná, 1994) and references within for details). This approach, however, can be more reliable than single-valued tests, as it was demonstrated in univariate case in (Paluš, 1995). As far as all the related considerations from (Paluš, 1995) directly apply to multivariate problems, we refer readers to (Paluš, 1995) and will not repeat them here, similarly as the discussion of the function of the linear redundancy, which is used to check the quality of the surrogate data: In some cases the surrogates can have auto-/cross-correlations different from the original data, usually due to a numerical artifact. This difference is detected by the redundancies R (or other nonlinear statistics) and can be erroneously interpreted as detection of nonlinearity in linear data. The linear redundancy-based statistic evaluates just the differences in the "linear properties", i.e., in autocorrelations in the univariate tests and in crosscorrelations in the multivariate tests. A significant result in linear redundancy-based statistic indicates a problem in the surrogates and necessity of further investigation before a conclusion about (non)linearity of data under study is made. See Refs. (Paluš, 1995), (Paluš, 1996a) for examples.

In the study presented in this volume (Hoyer et al., this volume) only 2-variable mutual information $I(X;Y)$ was applied: the univariate version $I(X(t); X(t+\tau))$ when dynamical properties and nonlinearity of individual series (variables) were studied, and the bivariate version $I(X(t); Y(t+\tau))$ when dynamical relations between two variables were investigated. The mutual information $I(X;Y)[\mathrm{o}]$ from the scrutinized data and the mean mutual information $I(X;Y)[\mathrm{s}]$ from the surrogates, as well as the test statistics, defined above, were plotted as functions of lag τ. Significant differences found between $I(X;Y)$[o] and $I(X;Y)$[s] were used to infer nonlinearity in dynamics of a variable (in univariate case), or in a relation between two variables (in bivariate case). The values of $I(X;Y)$[o] indicate a "coherence" or predictability of a variable, i.e., the dependence between $x(t)$ and $x(t+\tau)$ (in univariate case), or a strength of the link between two variables (in bivariate case), both as a function of the lag τ. As noted above, the absolute value of $I(X;Y)$ is not of interest here, but the relative differences between $I(X;Y)$ obtained in different physiological states, as well as behaviour of $I(X;Y)$ as a function of the time lag were evaluated.

In order to help a reader to understand the presented methodology, we present here several examples of processing numerically generated data.

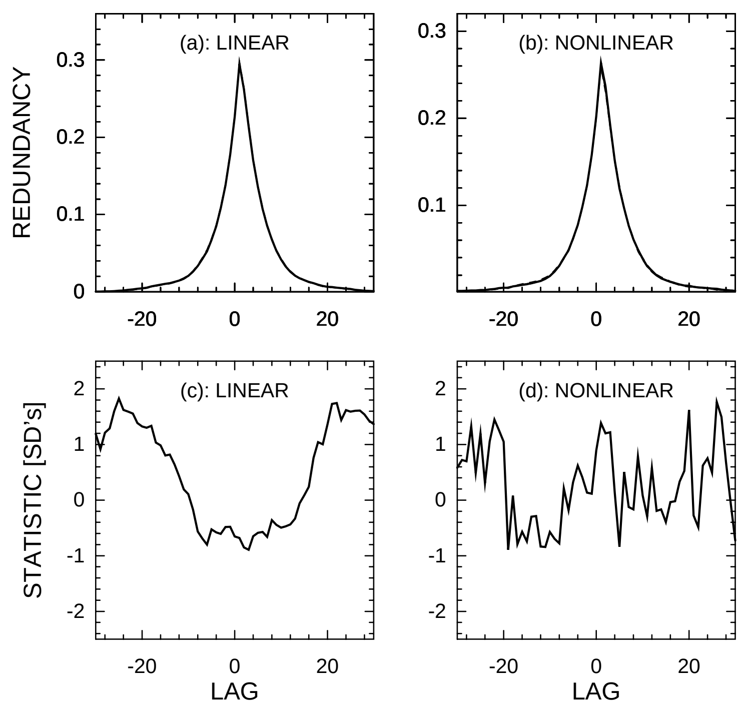

Fig. 1. *a): Linear redundancy $L(x(t); y(t+\tau))$, b): nonlinear (general) redundancy $R(x(t); y(t+\tau))$, for a bivariate linear autoregressive process and (coinciding curves) for related isospectral surrogates (mean of a set of 30 realization of the surrogates); c): linear (L-based), and d): nonlinear (R-based) statistics; as functions of the time lag τ.*

Consider a bivariate series $\{x(t), y(t)\}$, generated by the linear AR model:

$$x(t) = 0.9x(t-1) + \sigma_1(t),$$

$$y(t) = 0.3x(t-1) + 0.3y(t-1) + \sigma_2(t),$$

where $\sigma_1(t)$ and $\sigma_2(t)$ are Gaussian deviates with zero means and unit variances. The results – the linear redundancy $L(x(t); y(t+\tau))$, the redundancy $R(x(t); y(t+\tau))$, the linear (linear redundancy L-based) statistic and the nonlinear (redundancy R-based) statistic as functions of the time lag τ are presented in Fig. 1. The redundancies for the data and for the surrogates (Figs. 1a,b) coincide. Both the linear and nonlinear statistics (Fig. 1c and 1d, respectively) are confined between the values -2 and 2 SD's, i.e., the data are not significantly different from the surrogates. The linear stochastic hypothesis is accepted in agreement with the origin of the series.

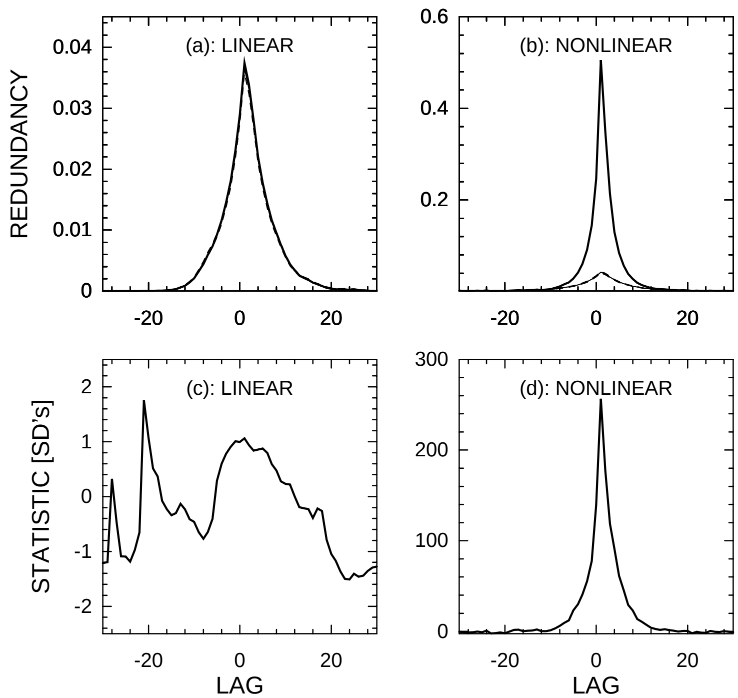

Fig. 2. *a): Linear redundancy $L(x(t); y(t+\tau))$, b): nonlinear (general) redundancy $R(x(t); y(t+\tau))$, for a bivariate autoregressive process with a quadratic link (thick lines) and for related isospectral surrogates (mean of a set of 30 realization of the surrogates – thin full lines, mean±SD – thin dashed lines); c): linear (L-based), and d): nonlinear (R-based) statistics; as functions of the time lag τ.*

Now, consider a bivariate series $\{x(t), y(t)\}$, generated by the following model:

$$x(t) = 0.9x(t-1) + \sigma_1(t),$$
$$y(t) = 0.3x(t-1)^2 + 0.3y(t-1) + \sigma_2(t),$$

where $\sigma_1(t)$ and $\sigma_2(t)$ are again Gaussian deviates with zero means and unit variances. The difference from the previous bivariate linear AR model is that here the equation for the variable y does not contain the variable x in a linear, but in a quadratic term. The results, presented in Fig. 2, are now different from those in Fig. 1.

The linear redundancy $L(x(t); y(t+\tau))$ from the data coincides with $L(x(t); y(t+\tau))$ from the surrogates (Fig. 2a), also there are no significant differences in the linear statistic (Fig. 2c). This implies that the surrogates are of a good quality and preserve the linear properties of the data. The (nonlinear) redundancy $R(x(t); y(t+\tau))$ for the data, however, is clearly different from $R(x(t); y(t+\tau))$ of the surrogates (Fig. 2b) and the nonlinear statistic (Fig. 2d) reaches values over 200 SD's. This result means a reliable detection of nonlinearity in the link between the variables x and y, in the agreement with the generating model, which contains the quadratic link between x and y.

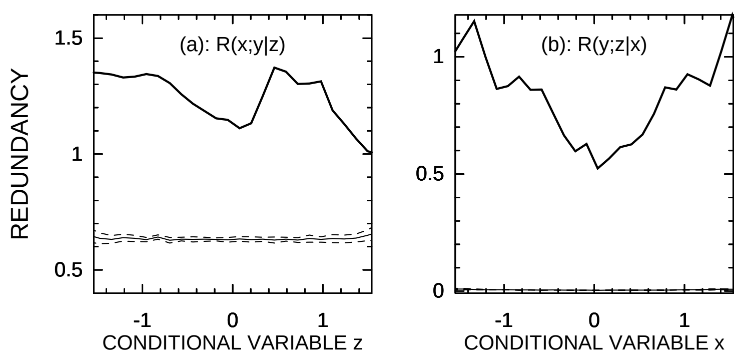

Fig. 3. a) *The conditional redundancy $R(x; y|z)$ as a function of variable z for the Lorenz data (thick line) and their trivariate surrogates (mean of 30 realizations of the surrogates – thin full line, mean±SD – thin dashed lines).* b) *The conditional redundancy $R(y; z|x)$ as a function of variable x for the Lorenz data (thick line) and their trivariate surrogates (mean of 30 realizations of the surrogates – thin full line, mean±SD – thin dashed lines).*

The three-variable series $\{x(t), y(t), z(t)\}$, obtained by integrating the chaotic Lorenz system (Lorenz, 1963):

$$(dx/dt, dy/dt, dz/dt) = (10(y-x), 28x - y - xz, xy - 8z/3), \qquad (7)$$

was analyzed in (Paluš, 1996a). Here we present results from applying conditional redundancies $R(x;y|z)$ and $R(y;z|x)$ (Fig. 3) in which we evaluate the "strength" of the link between two variables (with zero lag) conditionally given the third variable is equal to a defined value. That is, we evaluate how the "strength" of the link between two variables is changed when changing the value of the third variable. The conditional redundancy $R(x;y|z)$ as a function of the variable z is displayed in Fig. 3a. Influence of the variable z on the link between x and y is complicated, $R(x;y|z)$ has a local minimum in $z=0$, however, it is globally decreasing with increasing z. $R(x;y|z)$ obtained from the surrogates is smaller than $R(x;y|z)$ from the Lorenz data, but it is still positive, indicating the fact that there is also a linear link between x and y. The dynamics of the Lorenz system, however, is not translated into the surrogates, therefore $R(x;y|z)$ for the surrogates does not depend on z. The influence of the variable x on the relation between y and z is unambiguous: The variables y and z are related only through quadratic members xy and xz (Eq. 7), so that $R(y;z|x)$ as a function of x (Fig. 3b) has its minimum at $x=0$ and is approximately symmetric around zero, while $R(y;z|x)$ for the surrogates is equal to zero for any value of x, because there is no linear link between y and z.

The introduced method, which combines the redundancy – linear redundancy approach with the surrogate data technique, is not only a reliable method for detection of nonlinearity in univariate and multivariate time series, but can also bring further information about specific relations between/among variables under study and about changes in these relations in dynamics of underlying processes.

The author was supported by the Grant Agency of the Czech Republic (grant No. 201/94/1327) and by the Academy of Sciences of the Czech Republic (grant No. 230404).

References

Cover T.M., Thomas J.A. (1991): *Elements of Information Theory* (J. Wiley & Sons, New York)

Hoyer D. et al., this volume.

Lorenz E.N. (1963): Deterministic nonperiodic flow. *J. Atmos. Sci.* **20** 130-141

Morgera S.D. (1985): Information theoretic covariance complexity and its relation to pattern recognition. *IEEE Trans. on Systems, Man, and Cybernetics* **SMC-15 No. 5** 608-619

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

Paluš M. (1993): Identifying and quantifying chaos by using information-theoretic functionals. In: Weigend A.S., Gershenfeld N.A. (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.) 387-413

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

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

Paluš M. (1996a): Detecting nonlinearity in multivariate time series. *Phys. Lett. A,* in press.
Electronic preprint: `ftp://ftp.santafe.edu/pub/mp/multi/mu00pr.ps`.

Pompe B.: Ranking and entropy estimation in nonlinear time series analysis, this volume.

Prichard D., Theiler J. (1994): Generating surrogate data for time series with several simultaneously measured variables. *Phys. Rev. Lett.* **73** 951-954

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

Weigend A.S., Gershenfeld N.A. (eds.) (1993): *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.)