Physics Letters A 175 (1993) 203–209
North-Holland

PHYSICS LETTERS A

Information theoretic test for nonlinearity in time series

Milan Paluš [1], **Vladimír Albrecht and Ivan Dvořák**

Laboratory for Applied Mathematics and Bioengineering, Prague Psychiatric Center [2], *Ústavní 91, CS-181 03 Prague 8, Czech Republic*

Received 13 November 1991; revised manuscript received 14 September 1992; accepted for publication 12 February 1993
Communicated by A.P. Fordy

A diagnostic test for identifying nonlinear dynamical relationships in time series, based on mutual information and redundancy, functionals introduced in information theory, is proposed. Its ability to distinguish (noised) multiperiodic and random time series from time series generated by chaotic dynamical systems is demonstrated. The latter are characterized by specific behaviour of marginal redundancies reflecting the increase of uncertainty in time due to positive information production rate.

1. Introduction

Algorithms for analysis of experimental time series, based on the inverse problem of nonlinear dynamical systems, can in principle serve for identification and quantification of underlying chaotic dynamics [1,2]. Analyzing experimental and usually short and noisy data, however, ordinary estimators of dimensions (see e.g. refs. [3,4]) or Lyapunov exponents (see e.g. refs. [5,6]) can be fooled e.g. by simple autocorrelation of the series under study and can consider as chaotic a process which is in fact linear and stochastic [7]. These complications evoked the necessity of developing methods testing for basic properties of chaotic systems like nonlinearity, independently of the dimensional or Lyapunov exponent algorithms used [8–11].

In this paper we propose an original method for distinguishing time series generated by continuous nonlinear and especially chaotic dynamical systems on one side, from noised multiperiodic and random signals on the other side. The method is based on evaluation of redundancies (multidimensional mutual information) of the time series and its delayed versions in two ways: as a functional of the series probability distribution densities and from the series covariance matrix. We demonstrate that the former can reflect nonlinearities in the data while the latter, the special case of the former, is sensitive to linear relationships only. Moreover, the general (nonlinear) redundancy can measure the information production rate – metric Kolmogorov–Sinai entropy of the chaotic dynamical systems [12–17].

The theoretical concept of redundancies is introduced in section 2. Remarks on algorithms for their numerical estimation can be found in appendix A. The basic ideas of the proposed test are explained in section 3. In section 4 we demonstrate the power of the proposed methodology using numerically generated data which are described in detail in appendix B.

2. Mutual information and redundancies

Let x, y be random variables with probability distribution densities $p_x(x)$ and $p_y(y)$. The entropy of the distribution of a single variable, say x, is defined as

$$H(x) = -\int p_x(x)\log[p_x(x)]\,\mathrm{d}x\,. \qquad (1)$$

[1] Present address: Santa Fe Institute, 1660 Old Pecos Trail, Santa Fe, NM 87501, USA. E-mail: mp@santafe.edu.

[2] Affiliated to the 3rd School of Medicine, Charles University, Prague.

For the joint distribution $p_{x,y}(x, y)$ of x and y the joint entropy is defined as

$$H(x,y) = -\iint p_{x,y}(x,y)\log[p_{x,y}(x,y)]\,\mathrm{d}x\,\mathrm{d}y\,. \qquad (2)$$

The average amount of information that the variable x bears about the variable y is quantified by the *mutual information* $I(x; y)$,

$$I(x;y) = H(x) + H(y) - H(x,y)\,. \qquad (3)$$

Clearly, $I(x; y) = 0$ iff $p_{x,y}(x, y) = p_x(x)p_y(y)$, i.e. iff x and y are statistically independent. For more details see e.g. refs. [18–20].

For n variables $x_1, ..., x_n$ the extension of (2) is straightforward,

$$H(x_1, ..., x_n) = -\int ... \int p(x_1, ..., x_n) \times \log[p(x_1, ..., x_n)]\,\mathrm{d}x_1 ... \mathrm{d}x_n\,. \qquad (4)$$

Then analogously as in (3) we define

$$R(x_1; ...; x_n) = H(x_1) + ... + H(x_n) - H(x_1, ..., x_n)\,. \qquad (5)$$

This difference between the sum of the individual entropies and the entropy of the n-tuple $x_1, ..., x_n$ vanishes iff there is no dependence among these variables. Quantity (5) is called the *redundancy* of $x_1, ..., x_n$.

Besides (5) we define the *marginal redundancy* $\mathcal{R}(x_1, ..., x_{n-1}; x_n)$ quantifying the average amount of information about the variable x_n contained in the variables $x_1, ..., x_{n-1}$,

$$\mathcal{R}(x_1, ..., x_{n-1}; x_n) = H(x_1, ..., x_{n-1}) + H(x_n) - H(x_1, ..., x_n)\,. \qquad (6)$$

The following relation between redundancy (5) and marginal redundancy (6) can be obtained by a simple manipulation,

$$\mathcal{R}(x_1, ..., x_{n-1}; x_n) = R(x_1; ...; x_n) - R(x_1; ...; x_{n-1})\,. \qquad (7)$$

Let now $x_1, ..., x_n$ be an n-dimensional random variable with normal distribution with zero mean and covariance matrix $\mathbf{C}$. In this special case redundancy (5) can be computed straightforwardly from the definition [21]

$$R(x_1; ...; x_n) = \tfrac{1}{2}\sum_{i=1}^{n}\log(c_{ii}) - \tfrac{1}{2}\sum_{i=1}^{n}\log(\lambda_i)\,, \qquad (8)$$

where c_{ii} are diagonal elements (variances) and λ_i are eigenvalues of the $n \times n$ covariance matrix $\mathbf{C}$.

Formula (8), obviously, may be associated with any positive definite covariance matrix. Thus we use formula (8) to define the *linear redundancy* $L(x_1; ...; x_n)$ of an arbitrary n-dimensional random variable $x_1, ..., x_n$, whose mutual dependences are described by the corresponding covariance matrix $\mathbf{C}$,

$$L(x_1; ...; x_n) = \tfrac{1}{2}\sum_{i=1}^{n}\log(c_{ii}) - \tfrac{1}{2}\sum_{i=1}^{n}\log(\lambda_i)\,. \qquad (9)$$

If formula (5) is evaluated using the correlation matrix instead of the covariance matrix, then particularly $c_{ii} = 1$ for every i, and we obtain

$$L(x_1; ...; x_n) = -\tfrac{1}{2}\sum_{i=1}^{n}\log(\lambda_i)\,. \qquad (10)$$

Furthermore, in analogy with (7) we can define the *marginal linear redundancy* of $x_1, ..., x_{n-1}$ and x_n as

$$\mathcal{L}(x_1, ..., x_{n-1}; x_n) = L(x_1; ...; x_n) - L(x_1; ...; x_{n-1})\,. \qquad (11)$$

3. The test

Consider the typical "chaos inverse" problem: There is an experimental one-dimensional time series $Y(t)$ and we want to assess whether it is chaotic (i.e. generated by a low-dimensional nonlinear dynamical system in the chaotic regime) or not. The first step is the construction of an n-dimensional series $x_i(t)$ using the time-delay method based on the embedding theorem of Takens [22],

$$x_i(t) = Y(t + (i-1)\tau)\,, \quad i = 1, ..., n\,, \qquad (12)$$

where τ is a time delay and n is the so-called embedding dimension [22]. Searching for structures in the data the redundancies of the type

$$R(Y(t); Y(t+\tau); ...; Y(t+(n-1)\tau)) \qquad (13)$$

are of interest. Assuming stationarity of the series the redundancy

$$R^n(\tau) = R(Y(t); Y(t+\tau); ...; Y(t+(n-1)\tau)) \quad (14)$$

is clearly independent of t.

Analogously we denote the marginal redundancy

$$\mathscr{R}^n(\tau) = \mathscr{R}(Y(t), Y(t+\tau), ..., Y(t+(n-2)\tau); Y(t+(n-1)\tau)), \quad (15)$$

the linear redundancy

$$L^n(\tau) = L(Y(t); Y(t+\tau); ...; Y(t+(n-1)\tau)), \quad (16)$$

and the marginal linear redundancy

$$\mathscr{L}^n(\tau) = \mathscr{L}(Y(t), Y(t+\tau), ..., Y(t+(n-2)\tau); Y(t+(n-1)\tau)). \quad (17)$$

Relations (7) and (11) can be rewritten as

$$\mathscr{R}^n(\tau) = R^n(\tau) - R^{n-1}(\tau) \quad (18)$$

and

$$\mathscr{L}^n(\tau) = L^n(\tau) - L^{n-1}(\tau), \quad (19)$$

respectively.

The linear redundancy, according to its definition (10), reflects dependence structures contained in the correlation matrix **C** of the variables under study. In the special case, considered here, when all the variables are, according to eq. (12), lagged versions of the series $Y(t)$, each element of **C** is given by the value of the autocorrelation function of the series $Y(t)$ for a particular lag. As the correlation is the measure of linear dependence, the linear redundancy characterizes linear structures in the data under study.

We propose to compare the linear redundancy $L^n(\tau)$ with the redundancy $R^n(\tau)$ (or the marginal linear redundancy $\mathscr{L}^n(\tau)$ with the marginal redundancy $\mathscr{R}^n(\tau)$) considered as the functions of the time lag τ. If their shapes are the same or very similar a linear description of the process under study should be considered sufficient. Large discrepancies suggest important nonlinearities in links among the variables, or, recalling (12), among the studied time series and its lagged versions, i.e. in the dynamics of the process under study.

Let us recall that equivalence of redundancy $R^n(\tau)$ and linear redundancy $L^n(\tau)$ can be proved only for a special type of linear processes – the processes with the multivariate Gaussian distribution. In the general case, however, a possibility that differences between redundancies and linear redundancies are not due to nonlinearity but due to a non-Gaussian distribution of the studied data, cannot be neglected. Nevertheless, after the extensive numerical study we can conjecture that the shapes of the τ-dependence of redundancy $R^n(\tau)$ ($\mathscr{R}^n(\tau)$) and linear redundancy $L^n(\tau)$ ($\mathscr{L}^n(\tau)$) are approximately the same or similar also for different kinds of linear processes. Only for nonlinear processes the difference is qualitative and it is very distinct in the case of chaotic dynamics, when the time-lag dependence of the marginal redundancy $\mathscr{R}^n(\tau)$, unlike that of the marginal linear redundancy $\mathscr{L}^n(\tau)$, reflects specifically the "production of information", the typical property of chaotic dynamical systems, which is quantified by the positive metric (Kolmogorov–Sinai) entropy [12–17].

4. Assessing the power of the test by the known data

In applications of the proposed test we compare shapes of redundancies as functions of the lag τ, not particular values of the redundancies. Estimated values of $R^n(\tau)$ and $\mathscr{R}^n(\tau)$ depend on the numerical procedure used ("quantization", see appendix A), while the shapes of their τ-traces are usually consistent for a large extent of numerical parameters used in the redundancy estimations. Therefore each figure, depicting redundancies against the time lag τ, is drawn in its individual scale. Redundancies are in bits and time lags in number of samples. The different curves in each figure correspond to redundancies of the different number n of variables (embedding dimension), n is from 2 to 5, reading from the bottom to the top.

We start with noisy torus (two-periodic) time series. (For details about all the data sets see appendix B.) Time lag τ plots of linear redundancy $L^n(\tau)$ and redundancy $R^n(\tau)$ computed from the torus series jammed by 50% of uniformly distributed noise are presented in figs. 1a and 1b, respectively. We can see that these figures are almost the same, i.e. the linear description of the data is sufficient. This holds also for marginal linear redundancy $\mathscr{L}^n(\tau)$ and marginal

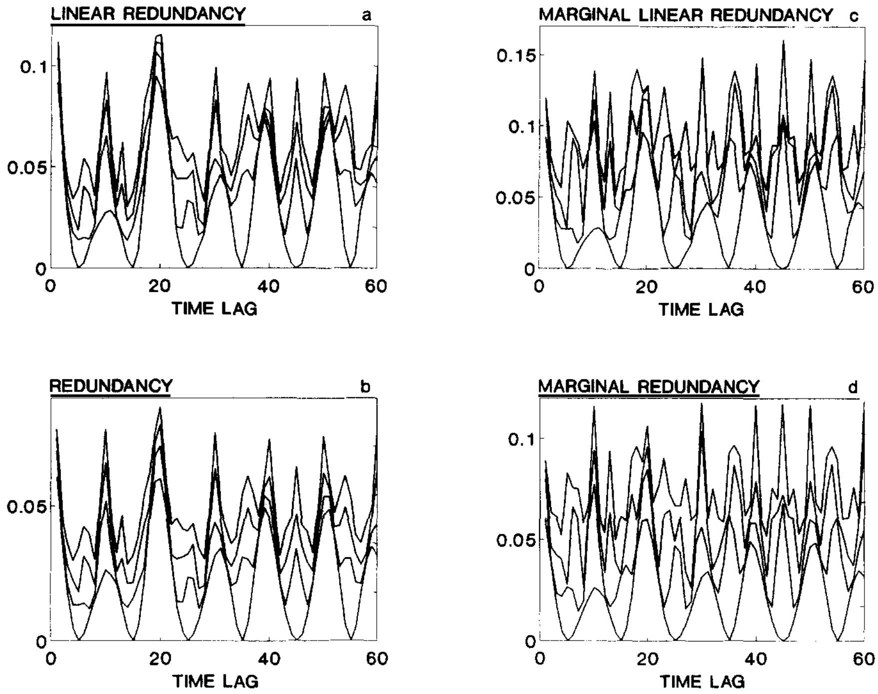

Fig. 1. (a) $L^n(\tau)$, (b) $R^n(\tau)$, (c) $\mathscr{L}^n(\tau)$ and (d) $\mathscr{R}^n(\tau)$ computed from noisy torus series. The embedding dimensions n are 2, 3, 4 and 5 reading from the bottom to the top of each figure. (The values of L^n and R^n are divided by $n-1$.)

redundancy $\mathscr{R}^n(\tau)$, illustrated in figs. 1c and 1d, respectively.

On the contrary, there are differences of the qualitative level between $\mathscr{L}^n(\tau)$ and $\mathscr{R}^n(\tau)$ computed from chaotic time series (fig. 2). The main feature of the (marginal) redundancies of the chaotic data (figs. 2b and 2d) is a long time lag decreasing trend (i.e. the decrease on the time scale greater than a period of the data oscillations reflected in redundancies). Recalling the definition of the marginal redundancy (6) we can consider the value of the quantity (const $-\mathscr{R}^n(\tau)$) as a measure of uncertainty in prediction of x_n knowing $x_1, ..., x_{n-1}$. The behaviour of $\mathscr{R}^n(\tau)$ in figs. 2b and 2d reflects the fact that this uncertainty increases with time in chaotic systems. It is the consequence of the positive information production rate – the positive metric (Kolmogorov–Sinai) entropy of chaotic systems [13–17]. The character of the dependence of the marginal redundancy $\mathscr{R}^n(\tau)$ on τ tends for embedding dimensions n greater than the system dimension and for a certain extent of τ [12] to a linearly decreasing function of the type $A-K\tau$. The coefficient K can serve as an estimate of the metric entropy of the system under study as proposed by Fraser [16,17] and Shaw [23], who introduced applications of the information theory to the nonlinear dynamics.

The linear (marginal) redundancy is not able to detect the above properties of the nonlinear chaotic systems. The results presented for the Rössler (fig. 2a) and Lorenz (fig. 2c) systems demonstrate two types of possible "misinterpretation" of the chaotic data by the linear methods: The Rössler series gives results similar to a multiperiodic signal, i.e. there is a constant nonzero level of the linear marginal redundancy or the constant finite level of uncertainty as estimated by the linear method. The linear marginal redundancy is not able to detect the production of information in this system. On the other hand, the linear marginal redundancy (like the linear redundancy) for the Lorenz system decreases quickly in an exponential or power-law way resulting in the constant close-to-zero level of the redundancy or the

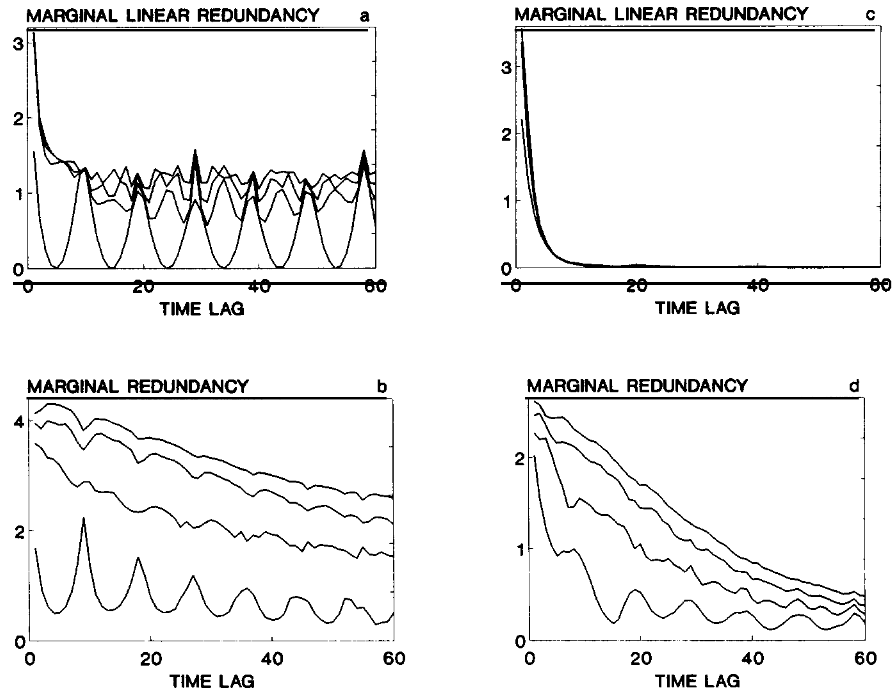

Fig. 2. (a) $\mathscr{L}^n(\tau)$ and (b) $\mathscr{R}^n(\tau)$ computed from the time series generated by the chaotic Rössler system; (c) $\mathscr{L}^n(\tau)$ and (d) $\mathscr{R}^n(\tau)$ computed from the time series generated by the chaotic Lorenz system. The embedding dimensions n are 2, 3, 4 and 5 reading from the bottom to the top of each figure.

constant high (noise-like) level of uncertainty (measured by the linear method). In this case the linear (marginal) redundancy is not able to detect any relationships in the data from a certain (low) value of τ.

This difference can be explained by various levels of "chaoticity" of the above two systems – the Rössler system is "weakly" chaotic and the Lorenz system "strongly" chaotic. To be more specific, the metric entropy or the positive Lyapunov exponent of the Lorenz system is more than ten times greater than that of the Rössler system (see e.g. refs. [5,6]).

Application of our test to the linear stochastic processes with the same spectra as the above series (the so-called surrogate data, see refs. [8,9]) brought interesting results; marginal redundancies $\mathscr{R}^n(\tau)$ and marginal linear redundancies $\mathscr{L}^n(\tau)$ are the same and they both are practically the same as the marginal linear redundancies of the original data. (Clearly, this holds also for redundancies $L^n(\tau)$ and $R^n(\tau)$.) This means the best for nonlinearity for these processes gave negative results consistent with the way of generating these data as linear and stochastic processes. The more interesting fact is that the notion of linearity as considered here (see section 3) coincides with randomness, i.e. using redundancies (like using spectra) one cannot distinguish (noised) linear oscillation from (filtered) noise. Nonlinear and chaotic oscillations, however, are clearly detectable.

"Coloured noises" is the term used for the random processes which can be generated by backward Fourier transform from a spectrum of the type $1/f^{\alpha}$ and uniformly distributed random phases. Osborne and Provenzale [7] showed that such processes can exhibit a finite correlation dimension. The value of the dimension depends on the spectrum decay coefficient α [7,24]. This coefficient also determines the shape of the process probability distribution: the larger α, the smaller the dimension, but also the larger the difference of the probability distribution from the Gaussian one [7]. It means that some differences between $\mathscr{R}^n(\tau)$ and $\mathscr{L}^n(\tau)$ of the coloured noises can occur for larger α, but in any case these differences are not qualitative and, moreover, the behaviour of $\mathscr{R}^n(\tau)$ for n greater than the estimated dimension is not like those of chaotic systems. Therefore the coloured noises are easily discernible from low-dimensional chaotic processes by our test, while only by the estimation of the dimensionality they are not.

For a sequence of independent random variables, usually called white noise, any redundancy should be zero (in practical computations it is close to zero) for any $\tau > 0$. This condition is fulfilled automatically by the proposed test.

Remark. One should ask how the proposed test behaves when applied to other types of processes, e.g. time series generated by Hamiltonian or near Hamiltonian systems, linear or nonlinear autoregressive processes, noisy chaotic data or nonstationary data. These questions are important and will be discussed in a more extended paper together with examples of physical and medical experimental data [25].

5. Conclusion

A test for detecting nonlinear dynamical relationships in time series based on comparing two types of redundancies has been proposed. On several typical examples it has been demonstrated that this technique is able to discern chaotic from random or noised multiperiodic time series.

The results presented are encouraging, however, they must not be overestimated by absolutizing the method proposed. We recommend to use it as a part of a battery of methods and algorithms for testing the nonlinear dynamics and deterministic chaos in time series together with order techniques for estimation of dimensions, entropies and/or Lyapunov exponents.

Acknowledgement

The authors are grateful to A. Fraser for inspiring discussions and valuable advices.

This study was performed as a part of the projects Information and entropy properties of the spontaneous brain electrical activity, supported within the framework of the IBM Academic Initiative in the CSFR, and Application of EEG complexity measures for quantification of changes of the brain state, supported by the Ministry of Health of the Czech Republic.

Appendix A

The linear redundancies were computed according to eq. (10). Eigenvalues of each correlation matrix were obtained by its decomposition using SVDCMP routine described in ref. [26, p. 52]. The results presented here were computed using 5120 samples of the total series length, but almost the same results can be obtained even from several hundreds of samples of the effective series length. (A part of the series is "spent" in the embedding reconstruction by the time delay method. For instance, with the maximum lag 100 samples and dimension 5 the effective series length is the total length minus 400.)

The algorithm for computing the redundancy, proposed by Fraser [16,17] or Fraser and Swinney [18], is a rather complicated one. The simple box-counting method we have found sufficient. The only "special prescriptions", based on our extensive numerical experience, concern the way of the data quantization:

(a) The type of quantization: We propose to use the marginal equiquantization method, i.e. the boxes for box-counting are defined not equidistantly but so that there is approximately the same number of samples in each marginal box.

(b) The number of quantization levels (marginal boxes): We have found that the requirement for the effective series length N using Q quantization levels in the computation of the n-dimensional redundancy is $N \geqslant Q^{n+1}$, otherwise the results can be heavily biased. Usually better results are obtained with $Q < N^{1/(n+1)}$ than with $Q > N^{1/(n+1)}$. Redundancies computed with $Q < N^{1/(n+1)}$ can be underestimated in the values, but graphs of R^n versus τ are, even for $Q = 4$–6, similar to those obtained from long time series and $Q = N^{1/(n+1)}$. In the case of $Q > N^{1/(n+1)}$ the redundancies can be overestimated and the results could be distorted in a qualitative way, e.g. the dependence of redundancies on τ is lost or τ-trends in marginal redundancies of chaotic data lead to an absurd result of negative metric entropy (see ref. [12]).

The results of redundancies presented here were obtained from 1 024 000 samples of the total series length. These extensive computations were performed in a study of estimations of the metric entropy and are not necessary for realization of the proposed test – a much shorter time series is sufficient when only qualitative results are of interest.

Appendix B

Two-periodic noisy data were generated according to the following formula,

$$Y(t)=[R_1+R_2\sin(\omega_2 t+\phi)]\sin(\omega_1 t)+\xi,$$

where $R_1:R_2=5:4$, $\omega_1:\omega_2=10:9$, $\phi=1.3\pi$, ξ are random numbers uniformly distributed between $-\Xi$ and Ξ. The term "50% of noise" means that $R_1:R_2:\Xi=5:4:9$.

Chaotic data were generated by numerical integration based on the Bulirsch–Stoer method [26, p. 563] of the Rössler system [27],

$$\left(\frac{dx}{dt},\frac{dy}{dt},\frac{dz}{dt}\right)=(-z-y, x+0.15y, 0.2+z(x-10)),$$

with initial values (11.120979, 17.496796, 51.023544), integration step 0.314 and accuracy 0.0001; and the Lorenz system [28]

$$\left(\frac{dx}{dt},\frac{dy}{dt},\frac{dz}{dt}\right)=(10(y-x), 28x-y-xz, xy-8z/3),$$

with initial values (15.34, 13.68, 37.91), integration step 0.04 and accuracy 0.0001. The component x was used in both cases.

Coloured noises were generated by backward fast Fourier transform [26, p. 397] of Fourier coefficients obtained from a power spectrum of type $1/f^{\alpha}$ and random phases uniformly distributed between 0 and 2π. Random processes with the same spectra as deterministic series were generated by forward FFT of the related series followed by the randomization of the phases and backward FFT to the time domain as in the case of the coloured noises.

References

[1] G. Mayer-Kress, ed., Dimensions and entropies in chaotic systems (Springer, Berlin, 1986).
[2] N.B. Abraham, A.M. Albano, A. Passamante and P.E. Rapp, eds., Measures of complexity and chaos (Plenum, New York, 1989).
[3] P. Grassberger and I. Procaccia, Phys. Rev. Lett. 50 (1983) 346.
[4] P. Grassberger and I. Procaccia, Physica D 9 (1983) 189.
[5] A. Wolf, J.B. Swift, H.L. Swinney and J.A. Vastano, Physica D 16 (1985) 285.
[6] X. Zeng, R. Eykholt and R.A. Pielke, Phys. Rev. Lett. 66 (1991) 3229.
[7] A.R. Osborne and A. Provenzale, Physica D 35 (1989) 357.
[8] J. Theiler, Phys. Lett. A 155 (1991) 480.
[9] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian and J.D. Farmer, Physica D 58 (1992) 77.
[10] D.T. Kaplan and L. Glass, Phys. Rev. Lett. 68 (1992) 427.
[11] G. Sugihara and R.M. May, Nature 344 (1990) 734.
[12] M. Paluš, Kolmogorov entropy from time series using information functionals, submitted to Physica D.
[13] A.N. Kolmogorov, Dokl. Akad. Nauk SSSR 124 (1959) 754.
[14] Ya.G. Sinai, Dokl. Akad. Nauk SSSR 124 (1959) 768.
[15] H.G. Schuster, Deterministic chaos: An introduction (Physik Verlag, Weinheim, 1984).
[16] A.M. Fraser, IEEE Trans. Inf. Theory IT-35 (1989) 245.
[17] A.M. Fraser, in: Dimensions and entropies in chaotic systems, ed. G. Mayer-Kress (Springer, Berlin, 1986) p. 82.
[18] A.M. Fraser and H.L. Swinney, Phys. Rev. A 33 (1986) 1134.
[19] R.G. Gallager, Information theory and reliable communication (Wiley, New York, 1968).
[20] S. Kullback, Information theory and statistics (Wiley, New York, 1959).
[21] S.D. Morgera, IEEE Trans. Syst. Man Cybern. SMC-15 (1985) 608.
[22] F. Takens, in: Lecture notes in mathematics, Vol. 898. Dynamical systems and turbulence, eds. D.A. Rand and D.S. Young (Springer, Berlin, 1981) p. 365.
[23] R.S. Shaw, The dripping faucet as a model chaotic system (Aerial, Santa Cruz, 1985).
[24] T. Higuchi, Physica D 46 (1990) 254.
[25] M. Paluš, in: Predicting the future and understanding the past: a comparison of approaches, Santa Fe Institute Studies in the Sciences of Complexity, eds. A.S. Weigend and N.A. Gershenfeld (Addison-Wesley, Reading, MA, 1993), to be published.
[26] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, Numerical recipes: the art of scientific computing (Cambridge Univ. Press, Cambridge, MA, 1986).
[27] O.E. Rössler, Phys. Lett. A 57 (1976) 397.
[28] E.N. Lorenz, J. Atmos. Sci. 20 (1963) 130.