Quantification of nonlinear interdependence in complex systems dynamics: simulations and applications

N S Pyko, S A Pyko, V N Mikhailov and M I Bogachev

Radio Systems Department, St. Petersburg Electrotechnical University, 5 Professor Popov street, 197376 St. Petersburg, Russia

E-mail: rogex@yandex.com

Abstract. In our work we studied the nonlinear interdependence metric quantifying the mutual dynamics of two stochastic data series. This metric is based on the calculation of the Euclidean distances between points belonging to the trajectories of these series in the state-space. Using surrogate data as an example, the sensitivity of the metric to the autocorrelation properties of the studied data series, as well as to the amplitude and phase randomization, are investigated. We also considered the application of this metric to the analysis of backscatter signals in sea surface monitoring. We suggest that the nonlinear interdependence metric may be useful as a complementary indicator for the sea wave structure quantification and modeling.

1. Introduction

Investigation of the behavior of complex systems is commonly performed via the analysis of the data series generated by such systems. Analysis of the joint dynamics of data series generated by complex systems is a common problem in various fields of science [1]. Among other applications, considerable fraction of related works focus on the analysis of interactions in biological systems, known as the network physiology approach [2–4]. According to this approach, the human body can be considered as a network in which organs are represented by interacting nonlinear dynamical systems. Besides interconnections, interacting systems also evolve over time, resulting in a combination of cross- and auto-correlations. Network physiology focuses on revealing the relationship between network topology and physiological function of each organ, altogether characterizes the human body as a whole. Similar approaches have been applied to the analysis of climate [5, 6], infrastructural, transportation [7], as well as social and economic systems [8].

However, conventional correlation metrics are characterized by several inevitable limitations. In order to overcome these limitations, a large number of alternative metrics characterizing the mutual dynamics of jointly observed stochastic series have been proposed. In a recent work [9], we investigated the properties of five such metrics using surrogate data with different correlation properties obtained by modeling as an example, and also considered the possibility of applying them to the analysis of biological signals and climate dynamics. Our results indicate that these metrics are characterized by considerably different sensitivity and specificity with respect to the alterations in the properties of the observed data series. To the date, there is no single consensus metric that would effectively characterize mutual dynamics of various observational data series from different fields of application. The aim of this work was to study the features and possibilities of practical application for assessing the joint behavior of the series of one of the indicators proposed in [10] in the context of biomedical data analysis, denoted as the metric of nonlinear interdependence. The authors of [10] used this nonlinear
interdependence metric to identify the degree of synchronization of electroencephalogram (EEG) records from various channels for patients of the Epilepsy Center of the University of Freiburg and found it informative for an early seizure prediction. In this work, we focused on the analysis of the statistical characteristics of this nonlinear interdependence metric following similar methodology as previously reported in [9]. Besides a systematic theoretical study, we also aimed at the analysis of the possibility of the nonlinear interdependence metric application to the data series of non-biological nature.

2. Data sources
In this work, we used two types of data. To study the sensitivity of the nonlinear interdependence metric \textit{NID} to the parameters of its calculation algorithm, as well as to the type and degree of randomization between the two analyzed series, surrogate data obtained by computer simulations were used. We generated sample sequences of normally distributed random series with short-term and long-term autocorrelations. Short-term correlations were obtained using a first-order autoregressive filter $x_{i+1} = ax_i + \xi_f$, where $a = \exp(-1/\tau_c)$, $\xi_f$ are the white noise samples, $\tau_c$ is the desired correlation time. Series exhibiting long-term correlations we simulated using a filter with the absolute value of the transfer function which was specified by the expression $B(f) = \sigma^2 f^{-\gamma} (1 - \gamma)^2$, where $\gamma = 2 - 2H$, $H$ is desired value of the Hurst exponent.

Observational samples for our study included the sea surface remote sensing data represented by the reflected backscatter signal intensities from the sea surface for various types of sea waves. Consecutive observations of the sea surface with 60 m spatial displacement outside of the surf line perpendicular to the coastline have been carried out at 0.833 ms time intervals using a recently developed experimental setup [11].

3. Analysis methods
\textit{Calculation of the nonlinear interdependence metric NID.} The nonlinear interdependence metric proposed in [10] is based on the analysis of a nonlinear dynamical system trajectories in the phase space producing the attractors represented by the studied series $x(t)$ and $y(t)$. A complete study of the attractor is possible only with the knowledge of the nonlinear differential equations that determine the dynamical system. The degree of the equation determines the dimension of the phase space in which the attractor is represented, i.e., the embedding dimension $d$. In practice, these differential equations are typically unknown, and only a single observation of each data series is available for the analysis. Therefore, to construct an attractor in $d$-dimensional space, redistribution of the samples of the series $x(t)$ and $y(t)$ is carried out according to the equations

\begin{align*}
  x(t) &= \{x(t - (d - 1)\tau), \ldots, x(t - \tau), x(t)\}, \\
  y(t) &= \{y(t - (d - 1)\tau), \ldots, y(t - \tau), y(t)\}.
\end{align*}

Thus, at each single time step, $d$-dimensional vector samples are formed from each series. As a result, each of the two analyzed data series is finally represented by a set of $N d$-dimensional points, i.e., by a trajectory in a $d$-dimensional phase space.

Next, the first point on the attractor corresponding to the first series is selected, and the Euclidean distances from this point to all other $d$-dimensional points are considered. Then $K$ nearest neighboring points on the attractor are selected for this point, and the average distance from a given point to $K$ of its nearest neighbors $R_x(t)$ is calculated.

The first point is also selected on the second attractor. The number of its nearest (in terms of Euclidean distance) neighboring points is calculated. After that, the average distance from the selected point on the first attractor to $K$ neighboring points $R_y(t)$ is determined. Averaging the ratio of the obtained distances over all points of the first series attractor, non-symmetric statistic $R(x/y)$ is obtained. Next the above steps are repeated for the second series. As a result, the symmetric \textit{NID} coefficient quantifying the degree of nonlinear interactions between the two series is calculated as
\[ NID = (R(x/y) + R(y/x)) / 2. \]

**Analysis of the NID metric statistical characteristics.** The statistical characteristics of the nonlinear interdependence metric NID were studied using surrogate data with short-term and long-term correlation. In order to introduce controlled randomization between the series, the following actions were performed. For every surrogate series \( x(t) \), we first obtained a copy. We implemented both phase randomization as well as additive and multiplicative amplitude randomization with either normal or uniform distribution. Phase randomization was introduced following the approach described in detail in [12]. With respect to the series \( x(t) \) with autocorrelation properties specified as described above, represented by the sequence of samples \( x(t_i), i = 1, 2, ..., N \) the spectral density was determined using the discrete Fourier transform

\[
X(\omega) = F(x(t)) = \sum_{i=1}^{N} x(t_i) \exp(-j\omega t_i) = A(\omega) \exp(\Phi(\omega)),
\]

where \( A(\omega) \) is the amplitude spectrum, and \( \Phi(\omega) \) is the phase spectrum. Phase randomization was carried out by the addition of an independent random variable uniformly distributed in the range \([-\varphi, \varphi]\) to the phase value at each frequency. The second series \( y(t) \) was obtained using the inverse Fourier transform of the spectral density \( X(\omega) \). It exhibits the same amplitude spectrum as the initial series \( x(t) \), while the phase spectrum deviates from the initial one according to the desired randomization level. To obtain dependence on the randomization level, next similar procedures have been repeated for different \( \varphi \) values from nearly zero to \([-\pi, \pi]\].

The additive amplitude randomization was carried out by adding to the samples of the initial series \( x(t) \) the independent random values with either uniform distribution in the interval of \([-A, A]\) or with normal distribution with zero mean value. For a relevant comparison, the standard deviation of the uniform noise and the normally distributed noise were set equal. Therefore, the standard deviation of the normally distributed noise was equal to \( 2A/\sqrt{12} \).

The multiplicative amplitude randomization, at first sight, should have been formed by multiplying the samples of the series copy by random samples \( n(t_i) \)

\[
y(t_i) = x(t_i) \cdot n(t_i),
\]

where \( n(t_i) \) is a sequence of independent random uniformly or normally distributed values. However, with this approach, in the absence of randomization the second series would be zero, while it would have to coincide with the first series. Therefore, when simulating a series with given amplitude multiplicative randomization, the following expression was used

\[
y(t) = x(t) \cdot (1 + n(t)).
\]

In the latter scenario, at zero randomization level the two series \( x(t) \) and \( y(t) \) are identical. With increasing the intensity of the noise level \( A \), the degree of randomization between the series also increases. For each type of the noise and each value of \( \varphi \) in the case of phase randomization or \( A \) in the case of amplitude randomization, 1000 independent random generator runs have been performed. In each modeling procedure, the noise intensity increased until the values of the estimated coefficients characterizing the mutual dynamics of the series stabilized. After that, the coefficients of the mutual behavior of the series were determined and their statistical characteristics, such as medians and quartiles, have been evaluated.

**4. Results**

The method for the nonlinear interdependence metric NID estimation described above has two free parameters, including the embedding dimension \( d \), and the number of nearest neighboring attractor points \( K \), that can significantly affect the resulting estimates. Figure 1 illustrates the influence of the algorithm parameters on the coefficient of the nonlinear interaction NID estimates. As an example, the results obtained during phase randomization between test series with short-term autocorrelations characterized by 3s correlation time. In this figure, the horizontal axis represents the phase
randomization levels $\varphi$ (in radians), while the vertical axis indicates the corresponding $NID$ metric values. The medians of the $NID$ coefficients in the graphs are depicted by different markers and connected by full lines. The error bars indicate the interquartile ranges.

**Figure 1.** (a) Influence of the number of nearest neighboring points $K$ on the $NID$ metric estimates, (b) Influence of the embedding dimension $d$ on the $NID$ metric estimates.

Figure 1 indicates that the sensitivity of the nonlinear $NID$ metric to the phase randomization between the series increases with decreasing the embedding dimension $d$, as well as with decreasing the number $K$ of nearest neighboring points of the attractor. Similar results were obtained for series with other correlation properties and other types of randomization between test series.

Our results also indicate that, while maintaining the sensitivity of the method, a substantial increase in the embedding dimension $d$ and in the number of nearest neighboring points $K$ seems impractical. If large data arrays are available for the analysis, in order to increase the accuracy of the obtained estimates of the nonlinear interaction metric, it is reasonable to calculate $NIS$ in (non-overlapping) gliding windows of certain size with subsequent averaging of the estimates. Figure 2 shows the statistical characteristics of the $NID$ metric for various sizes of the gliding window. Figure 2 illustrates the behavior of the analyzed metric for the embedding dimension $d = 5$ and for the number of nearest neighboring points $K = 10$ while the right panels show this behavior for $K = 20$. Like in figure 1, the obtained results correspond to the case of phase randomization between test series with a short-term correlation and a correlation time of 3 seconds.

Figure 2 indicates that the sensitivity of the studied metric of nonlinear interdependence to the level of randomization between the series increases with increasing the size of the gliding window. At the same time, when the window size increases from 500 to 1000 samples, the rate of the $NID$ metric decay increases significantly, while further increase of the window size to 1500 and 2000 samples, respectively, results in smaller decay of the $NID$ coefficient. Based on the obtained dependencies, when studying the statistical characteristics of the $NID$ metric on the surrogate data, the algorithm parameters were fixed at $d = 4$, $K = 10$, and the size of the analysis window was set at 2000 samples.
Figure 2. Statistical characteristics of the NID coefficient for various sizes of the gliding window.

Figure 3 illustrates the behavior of the statistical characteristics of the NID metric depending on the phase randomization level. The graphs in the left panel of figure 3 correspond to the surrogate data with short-term correlations characterized by various correlation times. The graphs in the right panel of figure 3 were obtained for the surrogate data with long-term correlation characterized by different Hurst exponents.

Figure 3. Statistical characteristics of the NID coefficient depending on the phase randomization level.

Figure 3 indicates that the obtained estimates of the NID metric are characterized by small variations even at maximum randomization level. Moreover, the sensitivity of the metric to the randomization level increases with the increase of the correlation time for short-term correlated data, and with the increase of the Hurst exponent for long-term correlated data, respectively. It can be also noted that the sensitivity
of the metric to the phase randomization for short-term correlated data series is higher than for the long-term correlated series.

Figures 4 and 5 indicate how the statistical characteristics of the NID coefficient alternate with additive amplitude randomization between test series, respectively, with either uniform or normal distribution. Like in figure 3, left panels correspond to surrogate data with short-term correlations, while right panels correspond to long-term correlations.

**Figure 4.** Statistical characteristics of the NID coefficient affected by additive uniform noise.

**Figure 5.** Statistical characteristics of the NID coefficient affected by additive Gaussian noise.

Similarly, figures 6 and 7 indicate similar results for multiplicative amplitude randomization.
To summarize, the considered nonlinear interdependence metric exhibits pronounced sensitivity to all considered randomization scenarios between the analyzed series. With increasing randomization level, the NID coefficient decreases nonlinearly from one to a certain value. The variability of the obtained estimates is extremely small; for amplitude noise, it is less than for phase randomization.

The coefficient NID showed greatest sensitivity to additive Gaussian noise. The sensitivity of the NID metric to multiplicative noise increases with increased persistence of the data, indicated either by its correlation time or by its Hurst exponent, respectively. In the case of additive noise, the sensitivity of the method increases with the increase of the persistence of the data series, indicated either by the Hurst exponent for long-term correlated, or by the correlation time for the short-term correlated data series.

For observational data analysis, selection of adequate NID metric calculation algorithm parameters has to be reconsidered, based on the available sample sizes and the range of values obtained. The choice of the algorithm parameters is largely guided by the range of the resulting estimates to avoid saturation near the limiting values 0 or 1.
In this study, we focused on the sea surface backscatter signal intensities obtained during remote sensing in a coastal area. The detailed data description and conventional analysis are presented in [11]. Briefly, the conventional approach is based on the characterization of the sea surface conditions based on the two-dimensional spectrum of the radio waves backscatter intensity obtained as

\[ S(\omega, \theta) = S(\omega)Q(\omega, \theta), \]

where \( S(\omega, \theta) \) is the one-dimensional wave spectrum, indicating the frequency distribution of the sea wave energy, \( Q(\omega, \theta) \) is the angular (azimuth) energy distribution, \( \theta \) is the azimuth angle, and \( \omega \) is the angular frequency of the sea waves. Surface waves induced by the wind are usually described by the so-called fully developed wave spectrum [13].

Figure 8 illustrates the results of the nonlinear interdependence metric \( NID \) applied to the analysis of the mutual dynamics of 1024 consecutive backscatter sea surface signals.

**Figure 8.** \( NID \) coefficient matrices for all backscatter signal pairs for beginning (a), developing (b) and fully developed (c) sea waves.

Figure 8 shows the matrices containing the nonlinear coefficients \( NID \) for all pairs of consecutive signals for the initial (a), developing (b), and fully developed (c) sea waves. The main diagonal of the matrices, by definition, is filled by ones.

Of note, figure 8 also indicates the appearance of characteristic clusters during wave formation, which are most pronounced for the fully developed sea waves.
5. Conclusion
Potential application of the considered nonlinear interdependence metric to the observational data series characterizing various natural phenomena and/or technological processes in order to reveal complementary information on their mutual dynamics could be beneficial for a more detailed overview of the phenomenon or process under study. The nonlinear interdependence metric considered in our work, as shown by both computer simulations and test data analysis, is characterized by reasonable sensitivity, being capable of reflecting changes in both auto- and cross-correlation properties as well as reaction to either phase- or amplitude randomization levels. We believe that using additional characteristics of the mutual dynamics obtained from remote measurements of the sea surface could provide researchers with additional information useful in describing and modeling of the sea wave dynamics.

Finally, we believe that the studied NID metric could be applicable for the additional quantification of the nonlinear mutual dynamics in other complex systems exhibiting similar correlation patterns such as biomolecular structures [14, 15], river runoff [16, 17] or telecommunication traffic [18–20].

Acknowledgments
We would like to acknowledge the financial support of this work by the Russian Science Foundation (research project no. 16-19-00172).

References
[1] Cohen R and Havlin S 2010 Complex networks: structure, robustness and function (New York: Cambridge university press)
[2] Gallos L K, Song C, Havlin S and Makse H A 2007 Scaling theory of transport in complex biological networks Proceedings of the National Academy of Sciences 104 (19) 7746–7751
[3] Bartsch R, Kantelhardt J W, Penzel T and Havlin S 2007 Experimental evidence for phase synchronization transitions in the human cardiorespiratory system Physical Review Letters 98 (5) 054102
[4] Bashan A, Bartsch R P, Kantelhardt J W, Havlin S and Ivanov P C 2012 Network physiology reveals relations between network topology and physiological function Nature communications 3 p 702
[5] Tsonis A A and Roebber P J 2004 The architecture of the climate network Physica A: Statistical Mechanics and its Applications 333 497–504
[6] Donges J F, Zou Y, Marwan N and Kurths J 2009 The backbone of the climate network EPL (Europhysics Letters) 87 (4) 48007
[7] Li G, Reis S D S, Moreira A A, Havlin S, Stanley H E and Andrade J S 2010 Towards design principles for optimal transport networks Physical review letters 104 (1) 018701
[8] Havlin S, Kenett D Y, Ben-Jacob E, Bunde A, Cohen R, Hermann H and Portugali J 2012 Challenges in network science: Applications to infrastructures, climate, social systems and economics The European Physical Journal Special Topics 214 (1) 273–293
[9] Pyko N S, Pyko S A, Markelov O A, Karimov A I, Butusov D N, Zolotukhin Y V, Uljanitskiy Y D and Bogachev M I 2018 Assessment of cooperativity in complex systems with non-periodical dynamics: Comparison of five mutual information metrics Physica A: Statistical Mechanics and its Applications 503 1054–1072
[10] Mirowski P, Madhavan D, Le Cun Y and Kuzniecky R 2009 Classification of patterns of EEG synchronization for seizure prediction Clinical Neurophysiology 120 1927–1940
[11] Kulikova D, Gorbunov I, Mikhailov V, Veremyev V and Bogachev M 2020 Experimental Remote Measurements of the Sea Surface with a Modified SHF Doppler Radar 2020 Signal Processing Workshop Proceedings
[12] Prichard D and Theiler J 1994 Generating Surrogate Data for Time Series with Several Simultaneously Measured Variables Physical Review Letters 73 (7) 951–954
[13] Lemaire D, Sobieski P and Guissard A 1999 Full-range sea surface spectrum in nonfully
developed state for scattering calculations *IEEE Trans. on Geosci. and Remote Sens.* 37 (2) 1038–1051

[14] Bogachev M I, Kayumov A R, Markelov O A and Bunde A 2016 Statistical prediction of protein structural, localization and functional properties by the analysis of its fragment mass distributions after proteolytic cleavage *Scientific Reports* 6 22286

[15] Bogachev M I, Markelov O A, Kayumov A R and Bunde A 2017 Superstatistical model of bacterial DNA architecture *Scientific reports* 7 43034

[16] Bogachev M I and Bunde A 2012 Universality in the precipitation and river runoff *EPL (Europhysics Letters)* 97 (4) 48011

[17] Bunde A, Bogachev M I and Lennartz S 2012 Precipitation and river flow: Long-term memory and predictability of extreme events *Extreme Events and Natural Hazards: The Complexity Perspective* 196 139–152

[18] Tamazian A, Nguyen V D, Markelov O A and Bogachev M I 2016 Universal model for collective access patterns in the Internet traffic dynamics: A superstatistical approach *EPL (Europhysics Letters)* 115 (1) 10008

[19] Markelov O, Duc V N and Bogachev M 2017 Statistical modeling of the Internet traffic dynamics: To which extent do we need long-term correlations? *Physica A: Statistical Mechanics and its Applications* 485 48–60

[20] Nguyen V D, Markelov O A, Serdyuk A D, Vasenev A N and Bogachev M I 2018 Universal rank-size statistics in network traffic: Modeling collective access patterns by Zipf's law with long-term correlations *EPL (Europhysics Letters)* 123 (5) 50001