ISSN 0027-1349, Moscow University Physics Bulletin, 2024, Vol. 79, No. 4, pp. 415–425. © Allerton Press, Inc., 2024.
Russian Text © The Author(s), 2024, published in Vestnik Moskovskogo Universiteta, Seriya 3: Fizika, Astronomiya, 2024, Vol. 79, No. 4, 2440701.

BIOPHYSICS AND MEDICAL PHYSICS (REVIEWS)

Using Entropy in Time Series Analysis

A. M. Adelyanov[1]*, E. A. Generalov[1], Wen Zhen[1], and L. V. Yakovenko[1]

[1] *Faculty of Physics, Lomonosov Moscow State University, Moscow, 119991 Russia*

Received March 13, 2024; revised May 29, 2024; accepted May 29, 2024

Abstract—Results of experiments in the field of biophysics are often presented as time series obtained with low resolution and not always of great length. In particular, in studies of the effects of various physicochemical factors on bilayer lipid membranes, transmembrane ion currents and their fluctuations are usually measured. In this case, the mean values and variances of the currents may not differ significantly, making it difficult to determine the nature and degree of impact based on them. Therefore, the development of approaches to time series analysis has never ceased. Attempts to use the entropy of random variable distributions in such analysis have been made for a long time, but in practical work, these approaches have been difficult to implement, especially due to the requirements for the length of the series and the absence of noise. In recent decades, there have been significant changes in this area, and many new methods of time series analysis using various modifications of entropy have been proposed. In this regard, there is a need for a summary of methods based on entropy calculation, indicating their advantages and disadvantages. This is the goal of the proposed brief review of entropy-based methods for analyzing scalar time series, which can be useful in analyzing experimental data. The review considers only some of the basic approaches on which further algorithmic improvements are based. The concept of entropy sometimes causes difficulties for students, so the review can also be useful for educational purposes.

Keywords: time series, entropy, complexity

DOI: 10.3103/S0027134924700607

INTRODUCTION

Modern scientific research requires new approaches, including the analysis of the data obtained. As computational power increases, so does the need for methods to analyze various complex parameters of biological systems. This is particularly relevant for biomedical research with large sets of not very accurate data: the analysis of signal transmission through intracellular signaling cascades [1, 2], the influence of various molecules on the state of biological systems, such as organisms [3, 4], as well as differentiating between normal and pathological conditions, for example, in oncological, cardiological, and neurodegenerative diseases [5, 6].

The results of studies of systems of various natures are often presented as time series (TS)—records of the values of some observable quantity (signal) depending on time. These records can be either discrete or continuous, but in the latter case, discretization is usually used with some constant time interval. Therefore, only discrete TS will be considered further. A stationary TS can be regarded as a representation of a certain state of the system that generated it. Therefore, studying TS theoretically makes it possible to obtain information about this system. The TS itself does not reveal anything about the structure of the system, but if there is a model of the system with unknown parameters, then the analysis of the TS allows at least determining the number of independent parameters.

A scalar time series $\{x_i\}_{i=1}^{N}$ is an array of N values of the observable quantity $x(t)$, measured with a constant time step $\triangle t$ at time points $t_i = t_0 + (i-1)\triangle t$, where t_0 is the time of the first measurement, $i = 1, 2, \ldots, N$. Usually, TS analysis is associated with solving two types of problems: identifying the system that generated the signal and predicting its behaviour for some time ahead. Identification involves estimating certain parameters of the system, such as correlation dimension, entropy, complexity, etc. Prediction aims to forecast future values of the measured signal based on the available data, i.e., constructing an approximating function that can estimate the next value of the measured quantity from several previous ones. The task of prediction is not considered in this article;

*E-mail: tashxat@gmail.com

the main focus will be on methods for assessing the stochasticity and complexity of TS.

In experimental studies of nonlinear chemical, hydrodynamic, and other systems, aperiodic changes in their parameters, called deterministic chaos, are often observed. By eye, this cannot be distinguished from a purely stochastic process, so mathematical methods have been developed to classify the behaviour of such systems, which is based on constructing the corresponding phase portraits. However, the experimenter usually does not know the differential equations representing the system model, so there arises a problem of constructing a phase portrait from limited experimental data. This problem has long been solved, but it seems appropriate in this review to give a very brief description of the approaches used.

The state of a system is characterized by a single point in multidimensional (n-dimensional) phase space, which, over time, follows a certain trajectory. A phase portrait is understood as a set of such trajectories obtained for all possible initial conditions with a fixed set of control parameters. The complex task of constructing a multidimensional phase portrait can be reduced to the analysis of just one observable variable using the methods described in [7–9]. Their essence boils down to the fact that for almost any observable quantity $B(t)$ and time delay τ, an m-dimensional phase portrait constructed using vectors $\mathbf{X}_m(k) = \{B(t_k), B(t_k+\tau), \ldots, B(t_k+(m-1)\tau)\}$, where $t_k = k\triangle t$, $k = 1, 2, \ldots, \infty$ will have the same properties (e.g., the Lyapunov exponent spectrum) as one constructed from measurements of n independent variables, provided that $m \geq 2n+1$. Such a phase portrait is an embedding of the original manifold. The delay τ can be chosen almost arbitrarily: for example, its coincidence with the period in a periodic mode will lead to the degeneration of the portrait into a straight line. In practice, the value of m can be chosen to be small and then increased by one until visible changes in the portrait disappear. After constructing the phase portrait, it can be analyzed by various methods, such as constructing a Poincaré section or determining the maximum Lyapunov exponent [9], but since this review is devoted to entropy methods, other approaches will not be considered here.

Entropy methods are associated with determining the amount and quality of information obtained when analyzing TS. The methods for determining these information characteristics are considered in information theory, whose mathematical foundations were laid in the works [10, 11]. Since the amount of information needed for a complete description of the studied system was equated with the decrease in the entropy of the distribution of probable outcomes of the study, in this article, the terms entropy and information will be used interchangeably, depending on the context.

The entropy (informational) of probability distributions and the entropy of classical thermodynamics are different concepts. The main difference between them is whether the thermodynamic characteristics of the system (amount of heat, temperature, etc.) or the characteristics of the distributions of random variables related to the system are used. In the latter case, entropy is a purely mathematical concept (and does not always have a physical meaning), and its value is relative, i.e., it depends on which random variables related to the system are studied. For example, the entropy of the distribution of coin denominations does not depend on other characteristics of the coins (year of issue, defects, etc.) [12]. The relationship between informational entropy and thermodynamic entropy is consistently considered in the works [13, 14].

The essence of entropy approaches to TS analysis is to determine the entropy of samples from TS in some way and obtain information about the properties of the physical system that generated the TS. Entropy is a function of the probability distribution of a random variable and changes with transformations of distributions. Therefore, in principle, it is possible to choose such a transformation in which changes in entropy are most sensitive to the investigated characteristics of TS. Entropies corresponding to different calculation methods have received different names: approximate entropy, sample entropy, permutation entropy, dispersion entropy, etc. Some of them in application to univariate TS are briefly considered below, with the main focus on permutation entropy and its modifications, which are often used in the analysis of TS of various natures.

1. SHANNON ENTROPY

In 1948, Shannon [10, 11] proposed a quantitative measure of information based on the evaluation of the probability distribution $\{p_i\}_{i=1}^N$ of possible responses (messages) to a well-defined question Q. In this context, numbering all possible responses and evaluating their probabilities based on the existing knowledge X about the possible answers plays an important role (it is impossible to estimate the amount of information upon receiving a response completely unrelated to the question). A well-defined question means that all the uncertainty of the situation should be contained only in which message will be received. The probability distribution function $H(Q/X) = H(p_1, \ldots, p_N)$, which can serve as a quantitative measure of information and which Shannon called entropy, is uniquely determined if four constraints (Shannon's postulates) are imposed on it. The amount of information in a message is defined as the difference in entropy due to

changes in knowledge about the possible responses before receiving the message (X) and after receiving it (X') : $I = H(Q/X) - H(Q/X')$. Thus, entropy is a quantitative measure of the uncertainty in knowledge about the system—it is the information that is lacking for complete knowledge. Shannon entropy forms the basis of all the methods considered below.

Shannon's work primarily pursued practical goals. The mathematically rigorous presentation and generalization of Shannon's ideas to the case of finite probabilistic schemes were first made by Khinchin [15, 16], so it is customary in information theory to use the formulations of the basic concepts and theorems of information theory in his interpretation. If a complete system of events $A_1, A_2, \ldots, A_N$ is given along with their probabilities $p_1, p_2, \ldots, p_N \left(p_i \geq 0, \sum_{i=1}^{N} p_i = 1\right)$, it is said that a finite scheme is given:

$$A = \begin{pmatrix} A_1 A_2 \ldots A_N \\ p_1 p_2 \ldots p_N \end{pmatrix}.$$

Every finite scheme is associated with the uncertainty about which event will occur since only the probabilities of their outcomes are known. A convenient measure of the uncertainty of a finite scheme is Shannon entropy defined by the equation:

$$H(p_1, \ldots, p_N) = -K \sum_{i=1}^{N} p_i \ln p_i = -\sum_{i=1}^{N} p_i \log_a p_i,$$

$$K > 0 \left(\log_a p_i = \frac{\log_b p_i}{\log_b a} = K \log_b p_i\right),$$

where in the second equality, the logarithm base a can be any, but the same in all terms (the constant K is associated with the choice of this base). Entropy is proportional to the amount of information obtained when realizing the finite scheme A, and the choice of the proportionality coefficient determines the unit of measurement of the amount of information. Thus, with $K = 1$ and $a = 2$, the amount of information is measured in bits. This probability distribution function is uniquely defined if the mentioned Shannon–Khinchin postulates are satisfied:

1. $H(p_1, \ldots, p_N)$ is continuous with respect to all arguments.

2. For a uniform distribution $p_i = \frac{1}{N} \forall i$, the function $H(p_1, \ldots, p_N) = H(N)$ monotonically increases with N and is maximal compared to other probability distributions; if an impossible event A_{N+1} $(p_{N+1} = 0)$ is added to the finite scheme, then $H(p_1, \ldots, p_{N+1}) = H(p_1, \ldots, p_N)$.

3. If two mutually independent finite schemes A and B with numbers of events N and M are given, then the set of events $A_k B_l (1 \leq k \leq N, 1 \leq l \leq M)$ is called the union of the schemes A and B and forms a new scheme AB with the event probabilities $p_{kl} = p_k p_l$, and $H(AB) = H(A) + H(B)$.

4. For dependent finite schemes A and B, the probability of event B_l in scheme B given that event A_k occurred in scheme A will be the conditional probability $q_{kl} = p(B_l/A_k)$, so $p_{kl} = p_k q_{kl}$; then $H(AB) = H(A) + \sum_{k=1}^{N} p_k \times \sum_{l=1}^{M} q_{kl} \log q_{kl} = H(A) + \sum_{k=1}^{N} p_k H_k(B) = H(A) + H_A(B)$.

In point 4, the last term on the right-hand side represents the mathematical expectation of $H(B)$ in scheme A, i.e., it is the mathematical expectation of additional information when realizing scheme B after obtaining information when realizing scheme A, so $H(B) \geq H_A(B)$ always holds. Entropy is defined only for discrete random variables. When transitioning to continuous random variables, the absolute value of entropy, defined analogously to Shannon's formula by replacing summation with integration and probabilities with probability density, turns out to be infinite. This is evident, for example, from the second Shannon–Khinchin postulate, as entropy increases indefinitely with increasing N, which occurs during the limiting transition from summation to integration. Additionally, uncertainty arises because a dimensional quantity appears under the logarithm. However, in this case, one can introduce the so-called relative or differential entropy which has properties different from Shannon entropy and will not be considered in this review.

If two initially noninteracting systems A and B with numbers of microstates N_A and N_B are combined into a single system with a number of states N_{A+B}, then the entropy of such a system, according to the third Shannon–Khinchin postulate, equals the sum of the entropies of the subsystems forming it: $S(AB) = S(A) + S(B)$, i.e., entropy is proportional to the number of elements in the system when it is large. This holds when the subsystems are statistically independent (or almost independent), in other words, when only local correlations can exist in the systems. (Spatial and temporal correlations between the microscopic state parameters arise due to subsystem interactions. If the interactions are short-range and such that the correlation time and length are very small, they will not affect the observable macroscopic parameters. This case is referred to here as local correlations.) The additivity of entropy means the same as extensiveness, provided that $N_{A+B} = N_A N_B$. For

a random TS, any permutation of its members will not lead to changes in the moments of the distribution, which have the form $\sum_i (x_i - \langle x \rangle)^n$, where $\langle x \rangle$ is the average value of the series, and n is the order of the moment, so statistics do not provide information about the organization of the TS. In analyzing the randomness of a TS, it is important not just to determine whether the series is random but to assess the degree of randomness. Randomness is associated with complexity: a simple TS has an obvious structure that allows accurately predicting its next values, whereas, with maximum randomness, future values are unpredictable, meaning a stochastic series has no structure. This is related to the possibility of "compressing" (archiving) the original information contained in the TS: for a maximally random TS, compression is impossible. To calculate entropy, it is necessary to estimate the probabilities of events p_i, which are usually determined in practice by the frequencies of various outcomes, even in a TS of not very great length. However, this provides a biased estimate of the entropy value.

1.1. Systematic Error in Entropy Estimation from a Finite Sample

In biophysical research, it is often necessary to deal with stationary point Poisson processes (all events are independent of each other). Therefore, below we consider the entropy estimation for such a case, although a similar systematic error arises for other probability distributions of random variables when estimating probabilities from frequencies [17].

The "experiment" consists of N measurements of a random variable whose values are distributed across K intervals with probabilities $\mathbf{p} = \{p_1, p_2, \ldots, p_K\}$. Then, in the ith interval, there will be n_i measurement results ($i = 1, 2, \ldots, K$), and their average value will be $\langle n_i \rangle = Np_i$. According to the properties of the Poisson distribution, the variance equals the mean value of the random variable: $\left\langle (\delta n_i)^2 \right\rangle = \langle n_i \rangle = Np_i$. Defining the frequency of the event as usual, $f_i = n_i/N$, we get $\langle f_i \rangle = p_i$, $\left\langle (\delta f_i)^2 \right\rangle = p_i/N$. Then the "naive" entropy estimate is: $\tilde{H} = -\sum_{i=1}^{K} f_i \ln f_i$. We will measure entropy in bits. The Taylor series expansion of $\tilde{H}$ is:

$$\tilde{H} = -\sum_{i=1}^{K} f_i \log_2 f_i$$

$$= -\sum_{i=1}^{K} (p_i + \delta f_i) \log_2 (p_i + \delta f_i)$$

$$\approx -\sum_{i=1}^{K} p_i \log_2 p_i - \sum_{i=1}^{K} \left(\log_2 p_i + \frac{1}{\ln 2} \right) \delta f_i - \frac{1}{2} \sum_{i=1}^{K} \left(\frac{1}{p_i \ln 2} \right) (\delta f_i)^2 + \ldots .$$

We will measure entropy in bits. The Taylor series expansion of $\tilde{H}$ is:

$$\left\langle \tilde{H} \right\rangle = H - \frac{1}{2\ln 2} \sum_{i=1}^{K} \left(\frac{\left\langle (\delta f_i)^2 \right\rangle}{p_i} \right)$$

$$= H - \frac{1}{2\ln 2} \sum_{i=1}^{K} \left(\frac{p_i}{N p_i} \right) = H - \frac{K}{(2\ln 2)\, N}.$$

After averaging, the second term on the right side is set to zero, and the third gives an estimate bias.

Thus, with the "naive" calculation, the entropy estimate is always lower than the exact value, and the systematic error is proportional to the number of possible values of the random variable and inversely proportional to the number of trials [17].

2. COMPLEXITY, KOLMOGOROV ENTROPY

There are two extreme types of systems: stochastic and deterministic. The evolution of a stochastic system is accompanied by its random transitions between microstates, making it impossible to predict which state the system will transition to. The behaviour of a stochastic system is nonreproducible. For unstable complex nonlinear dynamic systems with a phase space dimension of at least three, the appearance of a strange attractor and a transition to a regime of dynamic chaos is possible. Externally, the behaviur of a chaotic system does not differ from a stochastic one, but it is reproducible.

To describe the complexity of chaotic behaviour in a dynamic system, Kolmogorov, Solomonoff, and Sinai proposed using K-entropy, defined as follows [18].

The F-dimensional phase space of the system is divided into cells ε^F with indices $(i_1, \ldots, i_d)$. Suppose there is an attractor with trajectory $\mathbf{x}(t)$. The states of the system are measured at intervals of τ; if the joint probability that $\mathbf{x}(t = \tau) \in i_1, \ldots, \mathbf{x}(t = d\tau) \in i_d$ equals $p(i_1, \ldots, i_d)$, then:

$$K = -\lim_{\tau \to 0} \lim_{\epsilon \to 0} \lim_{d \to \infty} \frac{1}{d\tau} \times \sum_{i_1, \ldots, i_d} p(i_1, \ldots, i_d) \ln p(i_1, \ldots, i_d).$$

$K = 0$, $K \to \infty$, and $K \neq 0$ in ordered, stochastic, and chaotic systems, respectively. K-entropy is a

measure of the rate of information loss. The practical use of K-entropy is difficult due to the limiting transitions and the need for very large data volumes.

Quantitative assessment of TS complexity is associated with solving some problems of dynamic systems theory, statistics, and computational mathematics, such as assessing the complexity of turbulent flow and its evolution, choosing a simpler model describing the experiment, and assessing the complexity of an algorithm. In this regard, different definitions of complexity have arisen in various fields of research. The most commonly used representation is algorithmic complexity (according to Kolmogorov), where the complexity of TS is related to the minimum length of a binary computer program that can reproduce it, divided by the size of TS (generally, in the limit of an infinite TS). However, there is no algorithm that allows estimating the complexity of all data sets in this way, so other methods for determining complexity have been proposed.

Algorithmic complexity is closely related to the measure of the amount of information (per symbol) or the Shannon entropy density, which more intuitively corresponds to the concept of complexity. A stochastic TS is "incompressible," so it has high algorithmic complexity, but the system that generated it can be physically very simple. From a physics perspective, the complexity of the TS itself is less important than the complexity of the system that generated it. Cases where TS values are fully predictable, corresponding to low algorithmic complexity, and when they are completely unpredictable, corresponding to high algorithmic complexity, can be considered as the results of measuring characteristics of simple systems. Truly complex systems are in between these extremes, which agrees, for example, with the intuitive notion of the complexity of a biological system versus the simplicity of an ideal crystal or glass. In these extreme cases, entropy is either constant or grows linearly with the system size [19]. In both cases, corrections to the asymptotic behaviour do not increase with the increase in the data set. This allows linking the slow approximation of entropy to the asymptotic limit with the complexity of the system [19, 20].

Due to the practical difficulties of using K-entropy in the case of short TS, approximate entropy [21] and later sample entropy [22] were proposed. In the limit of infinite TS, these also coincide with K-entropy and can serve as an estimate of system complexity.

3. APPROXIMATE ENTROPY AND SAMPLE ENTROPY

Approximate (AppEn) and sample entropy (SampEn) are used as measures of TS randomness in the absence of information about the system that generated this TS [23]. This approach is based on assessing the uncertainty about possible data sets of observations and is thus related to the amount of information needed for their complete description. Uncertainty refers to what is possible but unknown. All types of entropy were introduced as measures of uncertainty about an event before it occurred.

AppEn is calculated by the following algorithm [21, 23]. For the series $\{x_i\}_{i=1}^{N}$, m-dimensional vectors are constructed in the state space (embedding dimension vectors of size m):

$$\mathbf{X}(i) = \{x(i), x(i+\tau), \ldots, x(i+(m-1)\tau)\},$$
$$1 \le i \le N - m\tau.$$

The similarity of vectors $\mathbf{X}(i)$ and $\mathbf{X}(j)$ is determined by the Chebyshev distance:

$$d[\mathbf{X}(i), \mathbf{X}(j)] = d_{i,j} = \max_{0\le k\le(m-1)} (|\mathbf{X}(i)_k - \mathbf{X}(j)_k|) = \max_{0\le k\le(m-1)} (|x(i+k) - x(j+k)|).$$

The proportion of vectors $\mathbf{X}(j)$ at a distance no greater than r from $\mathbf{X}(i)$ is $C_i^{(m)}(r) = \frac{N_i^{(m)}(r)}{N-m\tau}$, where $N_i^{(m)}(r)$ is the number of indices j for which $d_{i,j} \le r$, $1 \le j \le N - m\tau$. Then the average value of the logarithm of this proportion of vectors is calculated as $\Phi^{(m)}(r) = \frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau} \ln C_i^{(m)}(r)$. After this, the dimension is increased by one and $\Phi^{(m+1)}(r)$ is calculated. Approximate entropy is defined as follows: $\mathrm{AppEn}(m, r, \tau) = \Phi^{(m)}(r) - \Phi^{(m+1)}(r)$.

Sample entropy is calculated similarly, but comparisons of vectors with themselves are excluded [22, 23], which causes a bias towards greater similarity in vectors when calculating AppEn. The proportion of vectors $\mathbf{X}_m(j)$ a distance no greater than r from $\mathbf{X}_m(i)$ is $A_i^{(m)}(r) = \frac{N_i^{(m)}(r)}{N-m\tau-1}$, $d_{i,j} \le r$, $1 \le j \le N - m\tau$, $j \ne i$. The average value over $1 \le i \le (N - m\tau)$ equals $\Psi^{(m)}(r) = \frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau} A_i^{(m)}(r)$. Then the dimension is increased by one and $\Psi^{(m+1)}(r)$ is calculated: $m \to (m+1) \to \Psi^{(m+1)}(r)$, after which sample entropy is defined by the formula: $\mathrm{SampEn} = -\ln \frac{\Psi^{(m+1)}(r)}{\Psi^{(m)}(r)}$.

AppEn and SampEn have been used to analyze electroencephalograms (EEG) and magnetoencephalograms in Alzheimer's disease, and SampEn has been applied to the analysis of heart rate variability in rats under different conditions. Both types of entropy become undefined or lead to incorrect results when analyzing short TS and are critically

dependent on the choice of the parameter r, making them inconvenient for practical calculations. The computational complexity of AppEn and SampEn grows quadratically with the length of the TS.

4. PERMUTATION ENTROPY AND ITS VARIANTS

Permutation entropy (PE) is defined as the entropy of the distribution of ordered segments of a time series (TS), to which a specific permutation symbol corresponds. However, the same frequency distribution of permutations can be obtained for different absolute amplitude values of the TS, as the algorithm only considers their relative changes. This leads to the loss of some original information contained in the TS. Numerous modifications of the original algorithm have been proposed to account for the information contained in amplitude changes. However, the assessment of the complexity of the system characterized by the given TS is not necessarily related to the absolute amplitude values.

Permutation entropy (PE) allows for a quick quantitative assessment of the uncertainty regarding the temporal structure of a TS obtained from measurements of an observable quantity characterizing the studied system. The proposed algorithm does not consider the absolute amplitude values of the observed quantities but only the sequence of their relative values. Consequently, some information contained in the TS is lost. Nevertheless, PE often allows determining whether the system behaves chaotically, stochastically, or regularly. With increasing TS length, PE for any chaotic system tends towards Kolmogorov–Sinai entropy [24, 25].

The essence of the algorithm is as follows [26]. The scalar TS $\{x_i\}_{i=1}^N$ is transformed into a sequence of vectors (embedding dimension w)

$$\mathbf{X}_w(i) = \left(x_i, x_{i+\tau}, \ldots, x_{i+(w-1)\tau}\right), \quad i = 1, 2, \ldots, N-(w-1)\tau. \tag{1}$$

Typically, $\tau = 1$ and $3 \le w \le 7$ are used, and the condition $N \gg w!$ should be satisfied. Then the components of each vector are sorted: $x_{i+(j_1-1)\tau} \le x_{i+(j_2-1)\tau} \le \cdots \le x_{i+(j_w-1)\tau}$, where $j_k (k = 1, \ldots, w)$ are the indices of the vector components after sorting (rank). If the components are equal, their order is preserved: $x_{i+(j_{k1}-1)\tau} < x_{i+(j_{k2}-1)\tau}$, if $j_{k1} < j_{k2}$. Each vector is assigned a "symbol" or "word": $X(i) \to A(i) = (j_1, \ldots, j_{k1}, j_{k2}, \ldots, j_w)$, with $w!$ such words possible from w letters. Next, the probability distribution of the symbols is evaluated: $\{p_1, p_2, \ldots, p_k\}$, $k \le w!$ and permutation entropy is calculated using Shannon's formula (e.g., in bits): $H_{PE}(w) = -\sum_{i=1}^k p_i \log_2 p_i$. The maximum entropy value $H_{PE}(w) = \log_2(w!)$ corresponds to a uniform distribution: $p_i = 1/w!$ and $k = w!$. Normalization is often used: $h_{PE} = H_{PE}(w)/\log_2(w!)$, where $0 \le h_{PE} \le 1$. For regular and chaotic series, $\lim_{w\to\infty} h_{PE} = 0$. The computational complexity of PE grows linearly with the number of TS members.

In the original algorithm, the random variable distribution was assumed to be continuous, so equal values of vector components should rarely occur and practically should not affect the result. The authors suggested adding a small random perturbation to eliminate the problem of equal values, but this is rarely used in practice as it is suitable for continuous distributions or high-resolution measurements. It was shown that in cases of discrete distributions, low resolution, or physiological TS studies such as electrocardiograms (ECG) or electroencephalograms (EEG), the original algorithm can lead to correlations that result in false conclusions [27]. Moreover, even for a TS obtained from a completely predictable dynamic system, PE does not become zero [28]. Consequently, various modifications and improvements to the original algorithm have been proposed.

4.1. Modified Permutation Entropy (mPE)

In [29], it was proposed to assign the same rank to equal components of vectors. For example, the two vectors $\mathbf{X}_5(i) = (0.2, 0.5, 0.1, 0.2, 0.7)$ and $\mathbf{X}_5(j) = (0.2, 0.5, 0.1, 0.24, 0.7)$ in the PE algorithm would correspond to the same symbol $A(i) = A(j) = (3, 1, 4, 2, 5)$. The modified algorithm: if $x_{i+(j_{k1}-1)\tau} = x_{i+(j_{k2}-1)\tau}$, and $j_{k1} < j_{k2}$, then both values are assigned j_{k1} : $A'(i) = (j_1, \ldots, j_{k1}, j_{k1}, \ldots, j_w)$, resulting in symbols: $A'(i) = (3, 1, 1, 2, 5)$ and $A'(j) = (3, 1, 4, 2, 5)$. The probability distribution of the symbols is determined using a frequency estimate: $\{p'_1, p'_2, \ldots, p'_k\}$, $k \le k_w(w)$, where $k_w(w)$ is the upper bound of k values. The authors used a recursive method to compute it. The values obtained for some frequently used w values are shown in the table. The mPE value is calculated similarly to PE: $H'_{PE}(w) = -\sum_{i1}^k p'_i \ln p'_i$; normalization: $h'_P = H'_{PE}(w)/\ln k_w(w)$, $0 \le h'_{PE} \le 1$.

The authors [29] analyzed RR interval series in the ECGs of healthy young and elderly people and heart failure patients from open databases (MIT-BIH Fantasia database; BIDMC congestive heart failure database, part of the PhysioNet database, https://physionet.org) and showed that PE does not distinguish ECGs in these groups, whereas mPE reliably differentiates between them.

Table 1. Upper bound of k values for mPE calculations

w	3	4	5	6	7
k_w	13	73	501	4051	37 633

The drawbacks of mPE include a very large k_w (w) value for $w \geq 4$, which imposes a restriction on the minimum record length (see Table 1). Furthermore, mPE does not have a maximum value for Gaussian white noise, as it should. Like PE, mPE is highly sensitive to the noise level in the analyzed signal. Therefore, an "improved" PE (IPE) was proposed [30].

4.2. Improved PE

The IPE calculation algorithm compensates for the above shortcomings to some extent and also takes signal amplitudes into account. After constructing vectors (1), a uniform quantization (UQ) of the first column $\mathbf{X}(:,1)$ of the matrix formed by the row vectors is performed:

$$UQ(\mu) = \begin{cases} 0, & x_{\min} \leq x < x_{\min} + \triangle, \\ 1, & x_{\min} + \triangle \leq x < x_{\min} + 2\triangle, \\ \vdots & \\ L-1, & x_{\min} + (L-1)\triangle \leq x \leq x_{\max}, \end{cases}$$

where $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the analyzed TS, μ are the input data, $\triangle = (x_{\max} - x_{\min})/L$, L is the quantization level. The UQ value is an integer in the range from 0 to $L-1$ and is the "symbol" of $\mathbf{X}(:,1)$. Denote it: $S(:,1)$. For the kth column $\mathbf{X}(:,k)$, $2 \leq k \leq w$, the value $S(:,k)$ is calculated by the formula:

$$S(j,k) = S(j,1) + \left[\frac{X(j,k) - X(j,1)}{\triangle}\right],$$
$$1 \leq j \leq N - (w-1)\tau, \quad 2 \leq k \leq w.$$

The rows of the resulting matrix are considered as "words" $A(l), 1 \leq l \leq L^w$. Then the probability distribution of these words $\{p_l\}$ and the normalized IPE value are calculated:

$$H_{\text{IPE}}(w,\tau,L) = \frac{-\sum_{l=1}^{h} p_l \ln p_l}{\ln L^w},$$

where $h \leq L^w$, and $\ln L^w$ is the maximum value of H_{IPE}, which is achieved only for a uniform distribution.

The main differences between IPE and PE are as follows. First, it accounts for information about amplitudes and their fluctuations. Second, equal values are assigned the same symbols. Third, IPE is more resistant to noise. Finally, the number of possible "words" in IPE is L^w rather than $w!$.

The authors analyzed the influence of parameters on IPE estimates and recommend using $w = 4$, $\tau = 1$, $L = 4$ for most practical situations. Analysis of synthetic and natural TS showed that IPE is significantly more sensitive than PE and mPE.

4.3. Ensemble PE, EPE

An important parameter in computing PE is the dimensionality w_j of the vectors $\mathbf{X}_{w_j}(i)$. Usually, the condition $(w_{\max}+1)! \leq N$ must be met, i.e., $2 \leq w_j \leq w_{\max}$. In [31], to reduce the dependence of PE on the choice of w_j, it was proposed to use the value of PE averaged over all permissible values of w_j:

$$h_{\text{PE}}(\mathbf{X}_{w_j}, w_j) = H_{\text{PE}}(\mathbf{X}_{w_j}, w_j)/\log_2(w_j!);$$
$$h_{\text{EPE}} = \frac{\sum_{j=2}^{w_{\max}} h_{PE}(\mathbf{X}_{w_j}, w_j)}{w_{\max} - 1}.$$

The authors demonstrated through various examples that this approach reduces the dependence of calculation results on noise and generally better distinguishes between TS.

4.4. Ensemble Improved Permutation Entropy (EIPE)

In [32], in contrast to the original IPE calculation algorithm, the authors propose first normalizing the original TS values $\{x_i\}_{i=1}^N$ using the normal distribution function. The new TS members will then have the following values:

$$y_i = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_i} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt, \qquad (2)$$

where μ and σ^2 are the mean value and variance of the original TS. Then, the window width w and the time step τ are selected, and a "phase space" is constructed: $\Upsilon(j,:) = [y_j, y_{j+\tau}, \ldots, y_{j+(w-1)\tau}]$, where $\Upsilon(j,:)$ is the jth row of the matrix, and $j = 1, 2, \ldots, N - (w-1)\tau$.

We denote $y_{\min}$ and $y_{\max}$ as the minimum and maximum values of the TS members. Then, the number of quantization levels L is to be chosen, the quantization step is to be determined $\triangle = (y_{\max} - y_{\min})/L$, and a uniform probability function (UPF) is to be constructed:

$$\mathrm{UPF}\,(u) = \begin{cases} 0 & \text{for} \quad y_{\min} \le u < y_{\min} + \triangle, \\ 1 & \text{for} \quad y_{\min} + \triangle \le u < y_{\min} + 2\triangle, \\ \vdots & \\ L-1 & \text{for} \quad y_{\max} - \triangle < u \le y_{\max}. \end{cases}$$

Thus, UPF transforms u into an integer symbol from 0 to $L-1$. Applying this transformation to the elements of the first column $\Upsilon\,(:,1)$ yields the column of symbols $S\,(:,1)$. For the kth column $(2 \le k \le w)$, the transformation is used:

$$S\,(j,k) = S\,(j,1) + \lfloor [\Upsilon\,(j,k) - \Upsilon\,(j,1)]/\triangle \rfloor\,,$$

where $1 \le j \le [N-(w-1)/\tau]$, and $\lfloor \cdot \rfloor$ denotes the floor function, which rounds down to the nearest integer. The resulting matrix S has each row considered as a symbol (or word). To calculate IPE, the probabilities of these symbols p_i must be determined. Since each symbol contains w elements that can take L different values, the total number of symbols is L^w. IPE is calculated using Shannon's formula and normalized to $\ln L^w$:

$$H_{\mathrm{IPE}}\,(w,L,\tau) = -\sum_{i=1}^{L^w} p_i \ln p_i / \ln L^w.$$

Many properties of IPE strongly depend on the value of L. To reduce this dependence to some extent, the authors propose averaging over several values of L:

$$H_{\mathrm{EIPE}}\,(w,\tau) = \frac{1}{L_{\max} - L_{\min}} \sum_{1=L_{\min}}^{L_{\max}} H_{\mathrm{IPE}}\,(w,i,\tau),$$

where $L_{\min}$ and $L_{\max}$ are the chosen minimum and maximum values of L. Small L values increase noise resistance but lose some information contained in the TS.

Comparing the results of applying different types of entropy (PE, weighted PE, dispersion entropy, which will be considered in Subsection 4.3, and EIPE) to the analysis of synthesized noise series (white, pink, and brown) with $20 \le N \le 700$, $w = 4$, $\tau = 1$, and $2 \le L \le 8$ showed that white noise has the maximum entropy in all cases, and brown noise has the minimum entropy. However, EIPE reliably distinguishes the type of noise even for minimal TS lengths, whereas other types of entropy cannot.

Analysis of the logistic map $x_{n+1} = r x_n\,(1-x_n)$ with r changing from 3.5 to 3.99 in steps of 0.001 and with $N = 10\,000$ showed that starting from $r = 3.57$, there is a positive correlation of EIPE with increasing chaos of the series. For values corresponding to periodic modes (particularly near 3.84), there is a sharp decrease in EIPE. This algorithm also showed better results in analyzing natural signals such as ECG, EEG, vibrations of healthy and faulty bearings, and marine vessels.

4.5. Dispersion Entropy (DE)

Among other types of entropy, dispersion entropy [33] has good discrimination ability for TS types, noise resistance, and the ability to account for amplitudes. Its calculation consists of the following steps.

The values of the elements of the series $\{x_i\}_{i=1}^N$ are mapped to the interval [0, 1], usually using the normal cumulative distribution function (2), resulting in the sequence $\{y_i\}_{i=1}^N$. The members y_i of the new series are distributed among c classes and replaced with integers from the interval [1, c]. The authors recommend using the rounding function: $r_c(j) = cy_j + 0.5$, $z_c(j) = \mathrm{round}(r_c(j))$, $j = 1, 2, \ldots, N$, and choosing the number of classes from the interval $[3, 9]$ such that $c^w < N$. For the embedding dimension w and time delay τ, the embedding vectors $\mathbf{Z}_{w,c}(i) = \{z_c(i), z_c(i+\tau), \ldots, z_c(i+(w-1)\tau)\}$, $i = 1, 2 \ldots, N-(w-1)\tau$, are constructed. Each vector $\mathbf{Z}_{w,c}(i)$ is assigned a "dispersion pattern" $u_{v_0 v_1 \ldots v_{w-1}}$, where $v_0 = z_c(i)$, $v_1 = z_c(i+\tau), \ldots,$ $v_{w-1} = z_c(i+(w-1)\tau)$. The number of such patterns is c^w. Then, the probability of each pattern is estimated:

$$p\left(u_{v_0 v_1 \ldots v_{w-1}}\right) = \frac{Number\left\{t \,|t \le N-(w-1)\,\tau\,, \text{when } \mathbf{Z}_{w,c}\,(i) \text{ has the pattern } u_{v_0 v_1 \ldots v_{w-1}}\right\}}{N-(w-1)\,\tau},$$

where $Number\{\cdot\}$ is the count function.

To clarify the sequence of calculation steps, consider a specific example. Let $\{x_i\}_{i=1}^N = (9, 8, 1, 12, 5, -3, 1.5, 8.01, 2.99)$. Mapping to the interval $[0, 1]$ gives $\{y_i\}_{i=1}^N = (0, 82, 0.75, 0.21, 0.94, 0.52, 0.05, 0.24, 0.75\,0.35)$. Rounding with $c = 3$ gives $\{z_i\}_{i=1}^N = (3, 3, 1, 3, 2, 1, 1, 3, 2)$. The embedding

vectors for $\tau = 1$ and $w = 2$ are: $\mathbf{Z}_{2,3}(1) = (3,3)$, $\mathbf{Z}_{2,3}(2) = (3,1)$, $\mathbf{Z}_{2,3}(3) = (1,3), \ldots, \mathbf{Z}_{2,3}(8) = (3,2)$ corresponding to the dispersion patterns $u_{33}, u_{31}, u_{13}, \ldots, u_{32}$, with the number of possible patterns being $3^2 = 9$. Therefore, the pattern probabilities will be: $p(u_{11}) = \frac{1}{8}$, $p(u_{12}) = 0$, $p(u_{13}) = \frac{1}{4}$, $p(u_{21}) = \frac{1}{8}, \ldots, p(u_{33}) = \frac{1}{8}$.

Dispersion entropy is calculated using Shannon's formula:

$$\mathrm{DE}(\mathbf{X}, w, c, \tau) = -\sum p(u_{v_0 v_1 \ldots v_{w-1}}) \ln p(u_{v_0 v_1 \ldots v_{w-1}}),$$

and is usually normalized to the maximum possible DE value (when all patterns are equally probable):

$$\mathrm{NDE}(\mathbf{X}, w, c, \tau) = \frac{\mathrm{DispEn}(\mathbf{X}, w, c, \tau)}{\ln c^w}.$$

In the given example, DE ≈ 1.844.

For real TS, assigning series members to classes using the rounding function is not always unambiguous. Therefore, in [34], it was proposed to replace the rounding function with a "fuzzy" function, allowing a TS member to belong to two classes simultaneously, and each embedding can correspond to no more than 2^w patterns. This does not increase the computational complexity of the algorithm (as for DE, it grows linearly with TS length) but allows using shorter TS.

Another modification of DE, coded DE (CDE), was proposed in [35]. The coding operation was previously used to improve the properties of PE [36], but other PE modifications gave better results. In the application to DE, it involves additional quadratic partitioning of embeddings $\mathbf{Z}_{w,c}(i)$ and also eliminates the drawbacks of using the rounding function. This is done as follows.

Each embedding $\mathbf{Z}_{w,c}(i)$ with the same dispersion pattern is associated with an embedding $\mathbf{X}_w(i)$ of the original series, and the elements of the embedding $\mathbf{Z}_{w,c}(i)$ are replaced with the mean values of the corresponding elements of $\mathbf{X}_w(i)$. Then the average pattern is formed:

$$\langle u_{v_1 \ldots v_w} \rangle = \left\{ \left\langle z^1_{v_1 \ldots v_w} \right\rangle, \left\langle z^2_{v_1 \ldots v_w} \right\rangle, \ldots, \left\langle z^i_{v_1 \ldots v_w} \right\rangle \right\},$$

where $\left\langle z^i_{v_1 \ldots v_w} \right\rangle$ is the mean value of the embedding element, and $\langle u_{v_1 \ldots v_w} \rangle$ is the overall pattern. For example, for the considered example, the original embeddings are $\mathbf{X}_2(1) = (9,\ 8)$, $\mathbf{X}_2(2) = (8,\ 1)$, $\ldots, \mathbf{X}_2(8) = (8.01,\ 2.99)$. The pattern u_{13} corresponds to the embeddings $\mathbf{Z}_{2,3}(2)$ and $\mathbf{Z}_{2,3}(7)$, which correspond to the embeddings $\mathbf{X}_2(2) = (8,\ 1)$ and $\mathbf{X}_2(7) = (1.5,\ 8.01)$. Then the mean values are $\left\langle z^1_{12} \right\rangle = 4.75$, $\left\langle z^2_{12} \right\rangle = 4.005$, so the average pattern $\langle u_{13} \rangle = \{4.75, 4.005\}$. In [35], it was proposed to compare the "unrounded" values $r_c(i)$ with the mean value and perform partitioning using the following criterion:

$$v(i) = \begin{cases} 2, & r_c(i) > \left\langle z^i_{v_1 \ldots v_w} \right\rangle, \\ 1, & r_c(i) = \left\langle z^i_{v_1 \ldots v_w} \right\rangle, \\ 0, & r_c(i) < \left\langle z^i_{v_1 \ldots v_w} \right\rangle. \end{cases}$$

However, in practice, the comparison was made with the values of the original series embeddings, i.e., $\langle u_{13} \rangle$ was compared element-wise with $\mathbf{X}_2(2) = (8,\ 1)$ and $\mathbf{X}_2(7) = (1.5,\ 8.01)$ and the criterion was applied. In our case, we get $v(2) = (2,\ 0)$, $v(7) = (0,\ 2)$.

The result is a vector $\mathbf{v}_{w,c}(i) = \{v(i), v(i+\tau), \ldots, v(i+(w-1)\tau)\}$. Combining both partitions (into classes and the last one), we get a "combined" vector $\{\mathbf{Z}_{w,c}(i), \mathbf{v}_{w,c}(i)\}$, corresponding to the pattern $u_{v_1 v_2, \ldots, v_{2w}}$, with the number of possible patterns increasing to $c^w \cdot 3^w$ because the criterion has three possibilities.

In practice, $v(i) = 1$ values are very rare, so such cases can be disregarded without loss of accuracy, leading to a simplified second partition criterion:

$$v(i) = \begin{cases} 1, & r_c(i) > \left\langle z^i_{v_1 \ldots v_w} \right\rangle, \\ 0, & r_c(i) \le \left\langle z^i_{v_1 \ldots v_w} \right\rangle, \end{cases}$$

with the number of possible patterns reducing to $c^w \cdot 2^w$. The simplified CDE (SCDE) is calculated using the usual scheme: pattern probabilities are estimated:

$$p(u_{v_1 v_2 \ldots v_{2w}}) = \frac{Number\{u_{v_1 v_2 \ldots v_{2w}}\}}{N - (w-1)\tau},$$

then Shannon's formula is used:

$$\mathrm{SCDE} = -\sum_{u=1}^{c^w \cdot 2^w} p(u_{v_1 v_2 \ldots v_{2w}}) \ln p(u_{v_1 v_2 \ldots v_{2w}})$$

and normalization is performed if necessary:

$$\mathrm{NSCDE} = \frac{\mathrm{SCDE}}{\ln(c^w 2^w)}.$$

The calculation results for CDE and SCDE for synthesized and natural signals (EEG, bearing vibrations, ship noises) exceed other types of entropy (PE, CPE, DE) in recognition accuracy by at least 10%, but SCDE requires approximately three times less computation time than CDE.

CONCLUSIONS

This brief overview coveres only the basic information on the application possibilities of entropy methods in the analysis of time series (TS), which may be of interest to experimenters and "big-data" analysts. The primary focus is on permutation entropy and its modifications, as these types of entropy are widely used in various applications. Many other methods developed based on the approaches discussed (such as bubble, fuzzy, multivariate, multiscale, and other types of entropy) are not included in the review, as well as methods using, for example, Tsallis nonextensive entropy, although these may be more effective in some cases (see, for example, [37–53]).

Overall, permutation entropy often yields good results in the classification of TS, especially for sufficiently long series. In the case of very short TS or when it is necessary to detect the short-term impact of some factors on the system, it is better to use other types of entropy discussed above. To reduce computation time, small values of parameters w and L are usually used, but sometimes it makes sense to perform additional analysis of the dependence of the calculation results on these parameters.

Software for all the described methods is available in open access on the internet (see, for example, [54]).

FUNDING

This work was supported by ongoing institutional funding. No additional grants to carry out or direct this particular research were obtained.

CONFLICT OF INTEREST

The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. E. A. Generalov, N. T. Levashova, A. E. Sidorova, P. M. Chumakov, and L. V. Yakovenko, Biophysics **62**, 717 (2017). https://doi.org/10.1134/s0006350917050086
2. S. G. Gaidin, S. A. Maiorov, D. P. Laryushkin, V. P. Zinchenko, and A. M. Kosenkov, J. Neurochem. **164**, 583 (2023). https://doi.org/10.1111/jnc.15729
3. N. G. Kulchenko and E. V. Yatsenko, Eksp. Klin. Urologiya, No. 3, 158 (2019). https://doi.org/10.29188/2222-8543-2019-11-3-158-163
4. E. A. Generalov, E. Yu. Simonenko, N. G. Kulchenko, and L. V. Yakovenko, Biomed. Khim. **68**, 403 (2022). https://doi.org/10.18097/pbmc20226806403
5. E. Generalov, T. Clarke, L. Iddamalgoda, V. S. Sundararajan, P. Suravajhala, and A. Goltsov, in *Companion and Complementary Diagnostics: From Biomarker Discovery to Clinical Implementation*, Ed. by J. T. Jørgensen (Academic, 2019), p. 27. https://doi.org/10.1016/B978-0-12-813539-6.00003-1
6. K. A. Kritskaya, E. I. Fedotova, and A. V. Berezhnov, Biomedicines **12**, 282 (2024). https://doi.org/10.3390/biomedicines12020282
7. H. Whitney, Ann. Math. **37**, 645 (1936). https://doi.org/10.2307/1968482
8. N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980). https://doi.org/10.1103/PhysRevLett.45.712
9. J.-C. Roux, R. H. Simoyi, and H. L. Swinney, Phys. D: Nonlinear Phenom. **8**, 257 (1983). https://doi.org/10.1016/0167-2789(83)90323-8
10. C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948). https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
11. C. E. Shannon, Bell Syst. Tech. J. **27**, 623 (1948). https://doi.org/10.1002/j.1538-7305.1948.tb00917
12. D. S. Chernavskii and M. L. Khazin, Bull. Lebedev Phys. Inst., No. 10, 33 (2001).
13. E. T. Jaynes, Phys. Rev. **106**, 620 (1957). https://doi.org/10.1103/PhysRev.106.620
14. E. T. Jaynes, Phys. Rev. **108**, 171 (1957). https://doi.org/10.1103/PhysRev.108.171
15. A. Ya. Khinchin, Usp. Mat. Nauk **8** (3), 3 (1953). https://www.mathnet.ru/eng/rm8203.
16. A. Ya. Khinchin, Usp. Mat. Nauk **11** (1), 17 (1956). https://www.mathnet.ru/eng/rm7742.
17. W. S. Bialek, *Biophysics: Searching for Principles* (Princeton Univ. Press, Princeton, N.J., 2012).
18. P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983). https://doi.org/10.1103/PhysRevA.28.2591
19. W. Bialek, I. Nemenman, and N. Tishby, Neural Comput. **13**, 2409 (2001). https://doi.org/10.1162/089976601753195969
20. P. Grassberger, Int. J. Theor. Phys. **25**, 907 (1986). https://doi.org/10.1007/bf00668821
21. S. M. Pincus, Proc. Natl. Acad. Sci. U. S. A. **88** (6), 2297 (1991). https://doi.org/10.1073/pnas.88.6.2297
22. J. S. Richman and J. R. Moorman, Am. J. Physiol.-Heart Circ. Physiol. **278**, H2039 (2000). https://doi.org/10.1152/ajpheart.2000.278.6.H2039
23. A. Delgado-Bonal and A. Marshak, Entropy **21**, 541 (2019). https://doi.org/10.3390/e21060541
24. O. A. Rosso, H. A. Larrondo, M. T. Martin, A. Plastino, and M. A. Fuentes, Phys. Rev. Lett. **99**, 154102 (2007). https://doi.org/10.1103/PhysRevLett.99.154102

25. K. Keller, T. Mangold, I. Stolz, and J. Werner, Entropy **19**, 134 (2017). https://doi.org/10.3390/e19030134
26. Ch. Bandt and B. Pompe, Phys. Rev. Lett. **88**, 174102 (2002). https://doi.org/10.1103/PhysRevLett.88.174102
27. L. Zunino, F. Olivares, F. Scholkmann, and O. A. Rosso, Phys. Lett. A **381**, 1883 (2017). https://doi.org/10.1016/j.physleta.2017.03.052
28. J. S. A. Eyebe Fouda and W. Koepf, Commun. Nonlinear Sci. Numer. Simul. **27**, 216 (2015). https://doi.org/10.1016/j.cnsns.2015.03.008
29. Ch. Bian, Ch. Qin, Q. D. Y. Ma, and Q. Shen, Phys. Rev. E **85**, 21906 (2012). https://doi.org/10.1103/physreve.85.021906
30. Zh. Chen, Ya. Li, H. Liang, and J. Yu, Complexity **2019**, 1403829 (2019). https://doi.org/10.1155/2019/1403829
31. H. Azami, S. Sanei, and T. K. Rajji, Knowledge-Based Syst. **256**, 109876 (2022). https://doi.org/10.1016/j.knosys.2022.109876
32. Zh. Chen, X. Ma, J. Fu, and Ya. Li, Entropy **25**, 1175 (2023). https://doi.org/10.3390/e25081175
33. M. Rostaghi and H. Azami, IEEE Signal Process. Lett. **23**, 610 (2016). https://doi.org/10.1109/lsp.2016.2542881
34. M. Rostaghi, M. M. Khatibi, M. R. Ashory, and H. Azami, IEEE Trans. Fuzzy Syst. **30**, 3785 (2022). https://doi.org/10.1109/tfuzz.2021.3128957
35. Yu. Li, B. Geng, and B. Tang, Nonlinear Dyn. **111**, 9327 (2023). https://doi.org/10.1007/s11071-023-08339-4
36. H. Kang, X. Zhang, and G. Zhang, Entropy **22**, 187 (2020). https://doi.org/10.3390/e22020187
37. W. Chen, Zh. Wang, H. Xie, and W. Yu, IEEE Trans. Neural Syst. Rehab. Eng. **15**, 266 (2007). https://doi.org/10.1109/TNSRE.2007.897025
38. M. Staniek and K. Lehnertz, Phys. Rev. Lett. **100**, 158101 (2008). https://doi.org/10.1103/PhysRevLett.100.158101
39. R. Bose and S. Chouhan, Phys. Rev. E **83**, 051918 (2011). https://doi.org/10.1103/PhysRevE.83.051918
40. R. Hanel, S. Thurner, and M. Gell-Mann, Proc. Natl. Acad. Sci. U. S. A. **108**, 6390 (2011). https://doi.org/10.1073/pnas.1103539108
41. M. Lobier, F. Siebenhühner, S. Palva, and J. M. Palva, NeuroImage **85**, 853 (2014). https://doi.org/10.1016/j.neuroimage.2013.08.056
42. P. Li, K. Li, Ch. Liu, D. Zheng, Z.-M. Li, and Ch. Liu, IEEE Trans. Biomed. Eng. **63**, 2231 (2016). https://doi.org/10.1109/TBME.2016.2515543
43. P. Tempesta, Proc. R. Soc. A: Math., Phys. Eng. Sci. **472**, 20160143 (2016). https://doi.org/10.1098/rspa.2016.0143
44. Ch. Hsu, S.-Y. Wei, H.-P. Huang, L. Hsu, S. Chi, and C.-K. Peng, Entropy **19**, 550 (2017). https://doi.org/10.3390/e19100550
45. G. Manis, M. Aktaruzzaman, and R. Sassi, IEEE Trans. Biomed. Eng. **64**, 2711 (2017). https://doi.org/10.1109/TBME.2017.2664105
46. Li, Gao, and Wang, Sensors **19**, 5203 (2019). https://doi.org/10.3390/s19235203
47. H. Azami, L. E. V. da Silva, A. C. M. Omoto, and A. Humeau-Heurtier, Signal Process.: Image Commun. **75**, 178 (2019). https://doi.org/10.1016/j.image.2019.04.013
48. G. Qin and P. Shang, Int. J. Bifurcation Chaos **31**, 2150128 (2021). https://doi.org/10.1142/S0218127421501285
49. B. Deka and D. Deka, Chaos, Solitons Fractals **158**, 112101 (2022). https://doi.org/10.1016/j.chaos.2022.112101
50. X. Jiang, Y. Yi, and J. Wu, Front. Phys. **11**, 1163767 (2023). https://doi.org/10.3389/fphy.2023.1163767
51. G. Ji, Front. Phys. **11**, 1146493 (2023). https://doi.org/10.3389/fphy.2023.1146493
52. C. Tsallis, J. Stat. Phys. **52**, 479 (1988). https://doi.org/10.1007/BF01016429
53. Ch. Li and P. Shang, Phys. A: Stat. Mech. Its Appl. **523**, 10 (2019). https://doi.org/10.1016/j.physa.2019.01.031
54. M. W. Flood and B. Grimm, PLoS One **16**, e0259448 (2021). https://doi.org/10.1371/journal.pone.0259448