Chapter 1

Entropy Analysis of Univariate Biomedical Signals: Review and Comparison of Methods

Hamed Azami

Luca Faes

Javier Escudero

Anne Humeau-Heurtier

Luiz E.V. Silva

H.A. is with Department of Neurology and Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA; hazami@mgh.harvard.edu and hmd.azami@gmail.com

L.F. is with Department of Engineering, University of Palermo, Italy, luca.faes@unipa.it

J.E. is with School of Engineering, Institute for Digital Communications, The University of Edinburgh, Edinburgh EH9 3FB, UK; javier.escudero@ed.ac.uk

A.H.-H. is with Univ Angers, LARIS - Laboratoire Angevin de Recherche en Ingénierie des Systèmes, Angers, France, anne.humeau@univ-angers.fr

L.E.V. Silva is with Department of Internal Medicine, School of Medicine of Ribeirao Preto, University of São Paulo, Ribeirão Preto, SP, Brazil, luizeduardo@usp.br

Nonlinear techniques have found an increasing interest in the dynamical analysis of various kinds of systems. Among these techniques, entropy-based metrics have emerged as practical alternatives to classical techniques due to their wide applicability in different scenarios, specially to short and noisy processes. Issued from information theory, entropy approaches are of great interest to evaluate the degree of irregularity and complexity of physical, physiological, social, and econometric systems. Based on Shannon entropy and conditional entropy (CE), various techniques have been proposed; among them, approximate entropy, sample entropy, fuzzy entropy, distribution entropy, permutation entropy, and dispersion entropy are probably the most well-known. After a presentation of the basic information-theoretic functionals, these measures are detailed, together with recent proposals inspired by nearest neighbors and parametric approaches. Moreover, the role of dimension, data length, and parameters in using these measures

is described. Their computational efficiency is also commented. Finally, the limitations and advantages of the above-mentioned entropy measures for practical use are discussed. The Matlab codes used in this Chapter are available at `https://github.com/HamedAzami/Univariate_Entropy_Methods`.

Contents

1. Introduction

Dynamical analysis is a popular and powerful method for understanding biological systems. A dynamical system is described by two elements: state and dynamics. The state of a dynamical system is determined by the values of all the variables describing the system at a specific time. Therefore, the state of a system represented by ι variables can be shown by a point in an ι-dimensional space. This space is termed the state or phase space of the system. The dynamics of the system is the set of laws or equations describing how the state of a system changes during time.[1]

Physiologists and clinicians are often confronted with the problem of distinguishing different kinds of dynamics in biomedical signals, such as heart rate tracings from infants who had an aborted sudden infant death syndrome versus control infants,[2] and electroencephalogram (EEG) signals from young versus elderly people.[3] A number of physiological time series, such as cardiovascular and brain activity recordings, show a nonlinear in addition to linear behaviour.[4–7] Moreover, several studies suggested that physiological recordings from healthy subjects have nonlinear complex relationships with ageing and disease.[8–10] Thus, there is an increasing interest in nonlinear techniques to analyse the dynamics of physiological signals.

There are a number of nonlinear dynamical analysis techniques, such as fractal dimension (FD),[11] Lyapunov exponent,[12] Lempel-Ziv complexity (LZC),[13] and entropy-based metrics.[14,15] Among them, entropy methods have emerged as a less ambitious but more practical alternative to classical techniques for the analysis of nonlinear dynamical systems due to their applicability to short and noisy processes with important stochastic components such as those describing the dynamical activity of physiological systems.[16]

Entropy approaches taken from information theory are of great interest for the evaluation of the degree of irregularity and complexity of physical, physiological, social, and econometric systems.[16] They have been successfully and widely used in many applications ranging from biomedical engineering and neuroscience to mechanical engineering,[17] financial data analysis,[18,19] climatology,[20] earth sciences,[21] cellular automata,[22,23] and others.[24,25]

Shannon entropy (ShEn), conditional entropy (CE) and information storage (IS) are the most common concepts used in the context of analysis of physiological signals.[8,14,16,26] ShEn and CE respectively show the amount of information learned and the rate of information production in a system.[8,26,27] Based on these two approaches, various techniques rooted in information theory,[14] such as approximate entropy (ApEn),[28] sample entropy (SampEn),[29] corrected conditional entropy (CorCE),[30] fuzzy entropy (FuzzyEn),[31] permutation entropy (PermEn),[32] distribution entropy (DistEn),[33] dispersion entropy (DispEn),[27] and fluctuation dispersion entropy (FDispEn)[34] have been proposed.

To show the relevance and usefulness of some of these methods, the evolution of the number of citations for ApEn,[28] SampEn,[29] PermEn,[32] FuzzyEn,[31] DistEn,[33] and DispEn[27] are reported in Fig. 1. Information was extracted from Ref[35] on September 28, 2020.

These entropy measures have been successfully employed in a wide range of applications to characterize different pathological states. For example, ApEn was used to automatically detect epileptic seizures in EEGs.[36] ApEn and SampEn were used to analyze the EEGs and magnetoencephalogram (MEGs) in patients with Alzheimer's disease (AD).[37] SampEn was used to study heart rate dynamics during episodes of mechanical ventilation and acute anoxia in rats[38] and the analysis of heart rate variability in diseases and aging.[39] CorCE was exploited to investigate the short-term cardiovascular control in a variety of pathophysiological states.[40–42] FuzzyEn-based approaches analyzed a gait maturation database to distinguish the effect of age on the intrinsic stride-to-stride dynamics and also the Fantasia database to distinguish short RR interval signals recorded from healthy young vs. elderly subjects.[43] Additionally, Fuzzy ApEn analysis was used to characterize surface electromyograms (EMGs) to evaluate local muscle fatigue.[44] DistEn was used to analyze heartbeat interval series in healthy aging and heart failure patients[33] and also to discern between emotional states of calm and negative stress (also called distress).[45] PermEn was applied in epilepsy research[46,47] and in anesthesiology[48,49]

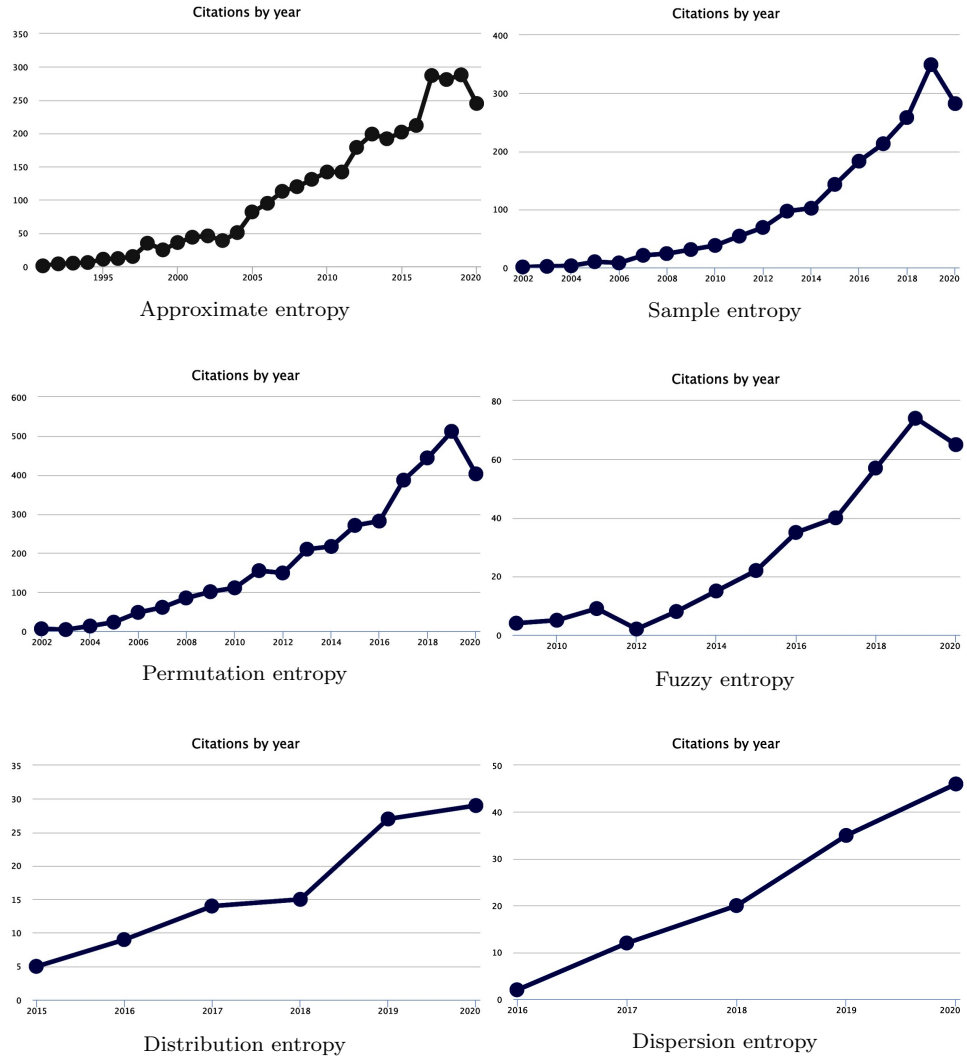


Fig. 1.: Evolution of the number of citations for ApEn,[28] SampEn,[29] PerEn,[32] FuzzyEn,[31] DistEn,[33] and DispEn[27] extracted on September 28, 2020.

using EEGs. DispEn was used to diagnose breathing and movement-related sleep disorders using electrocardiograms (ECGs) and EMGs.[50] DispEn and FDispEn were used to help clinicians in diagnosing AD and mild cognitive impairment using MEG signals.[51]

In spite of the broad relevance and applications of entropy metrics for various biological systems, a number of issues exist which often prevent a fair assessment of the performance of these approaches in addition to a correct interpretation of the

measured complexity or irregularity of the observed dynamics. First, there are various entropy methods which have not been systematically compared from different theoretical, computational and practical views. Second, the relationships between the parameters used in these entropy approaches have not been discussed as well. Third, the correlation between these entropy approaches have not studied empirically and theoretically. Here, apart from addressing the aforementioned issues, we provide a unifying framework for the definition of entropy metrics and corresponding estimation methods from time series, which serves to clarify their theoretical meanings and assess their practical significance in the assessment of the complexity of dynamic processes measured from physical systems.

The next Section of this article provides an overview of the theory used in entropy based on information theory. Section 3 details the key entropy methods used for biomedical signals. Role of dimension, data length and parameters are described in Sec. 4. The computational efficiency of the entropy approaches are evaluated in Sec. 5. The limitations and advantages are discussed in Sec. 6. Finally, the conclusions and future directions are outlined in Sec. 7.

2. Entropy measures for signal analysis: Theory

2.1. *Basic information-theoretic functionals*

This introductory section recalls the basic concepts of probability and information theory which will be used to define entropy measures for signal analysis. In a probabilistic framework,[52] a random variable is a mathematical variable whose value is subject to variations due to chance. A random variable V can be either *discrete* or *continuous*, respectively when it can take values inside a countable set of values (i.e., the alphabet $\mathcal{A}_V$), or inside a continuous set of values (i.e., the domain $\mathcal{D}_V$). If the random variable V is discrete, its probability distribution is a function mapping each possible outcome $v \in \mathcal{A}_V$ into a number $p(v) = Pr\{V = v\}$, where $Pr\{\cdot\}$ denotes the probability of an event; in the continuous case, the probability density is defined as the derivative of the cumulative distribution function $F_V(v) = Pr\{V \leq v\}$, i.e., $p(v) = \frac{dF_V(v)}{dv}, v \in \mathcal{D}_V$. These definitions extend in a straightforward way to the generic $k-$dimensional variable $\boldsymbol{W} = [W_1, \ldots, W_k]$ by defining the joint probability distribution $p(\boldsymbol{w}) = Pr\{W_1 = w_1, \ldots, W_k = w_k\}$ in the discrete case, and the joint probability density $p(\boldsymbol{w}) = \frac{dF_{\boldsymbol{W}}(\boldsymbol{w})}{d\boldsymbol{w}}$, with $F_{\boldsymbol{W}}(\boldsymbol{w}) = Pr\{\boldsymbol{W} \leq \boldsymbol{w}\}$, in the continuous case. Moreover, the conditional probability density of V given $\boldsymbol{W}$ expresses the probability of observing the value v for V given that the values $\boldsymbol{w} = [w_1, \ldots, w_k]$ have been observed for $\boldsymbol{W}$: $p(v|\boldsymbol{w}) = \frac{p(v,\boldsymbol{w})}{p(\boldsymbol{w})}$.

Probability distributions are represented in a compact way in the framework of information theory.[53] The central concept in this framework is the *Shannon information content* of a random variable V:[54] the information contained in a specific outcome v of a random variable V is the quantity $h(v) = -\log p(v)$. The information content, which is typically measured in bits (base 2) for discrete random

variables or in nats (base e) for continuous variables, is low for highly probable outcomes of the observed random variable, and very high for unlikely outcomes. Then, the *entropy* of V, quantifying the information contained in the variable intended as the average uncertainty about its outcomes, is defined as the expected value of the information content:

$$H(V) = \mathsf{E}[h(v)] = -\mathsf{E}[\log p(v)], \tag{1}$$

where $\mathsf{E}[\cdot]$ is the expectation operator. The concept of entropy defined in Eq. (1) relies on the seminal work of Shannon performed in the field of communication theory.[54] In this context, entropy quantifies the information contained in a random variable intended as the average uncertainty about its outcomes: if all observations of the variable are the same, there is no uncertainty and the entropy is zero; if, on the contrary, the variable takes different values all with the same probability of occurrence, the entropy is maximum and reflects maximum uncertainty.

The definition of entropy extends readily to a vector variable $\boldsymbol{W}$ computing the joint information content $h(\boldsymbol{w}) = -\log p(\boldsymbol{w})$ for the generic outcome $\boldsymbol{w}$, and using the joint probability $p(\boldsymbol{w})$ in Eq. (1) to get the joint entropy $H(\boldsymbol{W})$. Then, the *conditional* information content of v given $\boldsymbol{w}$ is the quantity $h(v|\boldsymbol{w}) = -\log p(v|\boldsymbol{w}) = h(v, \boldsymbol{w}) - h(\boldsymbol{w})$, measuring the amount of information carried by the specific outcome v of the variable V when the variable $\boldsymbol{W}$ takes the value $\boldsymbol{w}$. Again, moving to expected values yields CE of V given $\boldsymbol{W}$:

$$H(V|\boldsymbol{W}) = \mathsf{E}[h(v|\boldsymbol{w})] = -\mathsf{E}[\log p(v|\boldsymbol{w})] = H(V, \boldsymbol{W}) - H(\boldsymbol{W}), \tag{2}$$

which quantifies the residual information contained in V when $\boldsymbol{W}$ is assigned, intended as the average uncertainty that remains about the outcomes of V when the outcomes of $\boldsymbol{W}$ are known.

Another important information measure characterizing the interaction between two variables is the *mutual information* (MI). The specific MI between the outcomes v and $\boldsymbol{w}$ of the two variables V and $\boldsymbol{W}$ is the information that the two outcomes contain when they are taken individually but not when they are taken together, i.e. $i(v; \boldsymbol{w}) = h(v) + h(\boldsymbol{w}) - h(v, \boldsymbol{w}) = h(v) - h(v|\boldsymbol{w})$. The expected value over all outcomes corresponds to the MI between the two variables V and $\boldsymbol{W}$:

$$I(V; \boldsymbol{W}) = \mathsf{E}[i(v; \boldsymbol{w})] = \mathsf{E}[\log \frac{p(v|\boldsymbol{w})}{p(v)}] = H(V) - H(V|\boldsymbol{W}), \tag{3}$$

measuring the amount of information shared between V and $\boldsymbol{W}$ intended as the average reduction in uncertainty about the outcomes of V obtained when the outcomes of $\boldsymbol{W}$ are known.

The definitions provided above for the main information-theoretic measures are formalized in a different way for discrete and continuous random variables, depending on how the expectation operator is implemented in the two cases. In the discrete case, given a scalar random variable V and a vector variable $\boldsymbol{W}$, the entropy of V

and the CE of V given $\boldsymbol{W}$ are obtained particularizing Eq. (1) and Eq. (2) as follows:

$$\begin{aligned} H(V) &= -\sum_{v \in \mathcal{A}_V} p(v) \log p(v), \\ H(V|\boldsymbol{W}) &= -\sum_{v \in \mathcal{A}_V} \sum_{\boldsymbol{w} \in \mathcal{A}_{\boldsymbol{w}}} p(v, \boldsymbol{w}) \log p(v|\boldsymbol{w}). \end{aligned} \tag{4}$$

When the variables are continuous, the sums extended over the alphabet are substituted by integrals over the domain, so that entropy and CE are expressed as:

$$\begin{aligned} H(V) &= -\int_{v \in \mathcal{D}_V} p(v) \log p(v) dv, \\ H(V|\boldsymbol{W}) &= -\int_{v \in \mathcal{D}_V} \int_{\boldsymbol{w} \in \mathcal{D}_{\boldsymbol{w}}} p(v, \boldsymbol{w}) \log p(v|\boldsymbol{w}) dv d\boldsymbol{w}. \end{aligned} \tag{5}$$

In either case, the MI can be computed from entropy and CE using Eq. (3).

The definitions of entropy, CE and MI provided here for generic random variables, particularized to the variables obtained sampling a random process at different time instants, are exploited in the next subsection to define the entropy measures typically used for signal analysis. Moreover, the distinction between discrete and continuous random variables is at the basis of the different approaches for entropy estimation presented in Sec. 3; for instance, the linear and nearest neighbor estimators attempt to compute the continuous probability density functions of the amplitude values of the observed time series, while the binning estimator discretizes such amplitude values and then operate through computation of discrete probability distributions.

2.2. *Shannon entropy measures for stochastic processes*

The computation of entropy measures for signal analysis is grounded in a probabilistic framework, in which estimators of the entropy measures defined in the previous subsection are applied to the signal samples in order to characterize their statistical structure. To introduce the context, let us consider a physiological system and a biomedical time series extracted from it; for instance, the system can be the brain or the heart, and the time series a sampled EEG signal acquired on the scalp or the heart rate variability measured from the ECG. The observed system is supposed to be a dynamical system (i.e., a system that assumes diverse states at different instants of time), and stochastic (i.e., a system whose current state does not depend only on its inputs and initial state but also on the outcome of a random experiment). The time series measured from the system is taken as a quantitative indicator of the system's states and, as such, is considered as a realization of the stochastic process that maps the evolution of the system over time. Then, in an information-theoretic framework, the "information" contained in the system varies with time in a way such that, in the transition from past states to a new state, the system produces new information in addition to that already carried by the past states. The information content of a dynamical system is quantified by entropy

measures computed on the associated stochastic process:[16] the entropy quantifies the information carried by the present state of the process; the conditional entropy quantifies the new information contained in the present but not in the past; and the information storage quantifies the amount of information carried by the present that can be explained by the past history of the process.

Let us denote $\mathcal{X}$ as the observed dynamical system and X as the corresponding stochastic process. A specific realization of the process is the time series $\mathbf{x} = \{x_1, \ldots, x_N\}$, where N is the time series length. Let us further denote X_n as the random variable obtained by sampling the process X at the present time n, and $\mathbf{X}_n = [X_{n\ 1}, X_{n\ 2} \ldots]$ as the vector variable describing the past of X. Separating the present state X_n from the past states $\mathbf{X}_n$ allows to account for the arrow of time and to study the dynamic interactions within the process by looking at the statistical dependencies of X_n on $\mathbf{X}_n$.[55] Indeed, the statistical structure of the process at the current time step is provided by the probability distribution $p(x_n)$, measuring the probability that X takes the value x_n at time n, while the transition from past to present is described by the conditional probability $p(x_n|\mathbf{x}_n)$, measuring the probability that X_n takes the value x_n when $\mathbf{X}_n$ assumes the value $\mathbf{x}_n = [x_{n\ 1}, x_{n\ 2}, \ldots]$. If the joint probability $p(\mathbf{x}_n)$ of the past states is assigned, the behavior of the system is completely known from a probabilistic perspective because the joint probability $p(x_n, \mathbf{x}_n) = p(x_n|\mathbf{x}_n)p(\mathbf{x}_n)$ can be computed from the transition probabilities. Moreover, we note two important properties which are commonly exploited for the estimation of probability distributions and entropy measures. The first is stationarity, which defines the time-invariance of the joint probabilities extracted from the process: $p(x_n, \mathbf{x}_n) = p(x_{n+m}, \mathbf{x}_{n+m})\ \forall\ n, m \geq 1$. The second is the Markov property, which states the absence of memory of the process for time steps larger than the Markov order m: $p(x_n|\mathbf{X}_n) = p(x_n|x_{n\ 1}, \ldots, x_{n\ m})$. In practical analysis, exploiting the Markov property allows to approximate the infinite-dimensional variable describing the past history of the process with the finite-dimensional variable describing the past m states, i.e., $\mathbf{X}_n \approx \mathbf{X}_n^m = [X_{n\ 1} \cdots X_{n\ m}]$.

The probability distributions above defined are represented in a compact way through a number entropy measures which characterize the statistical structure of the observed process.[16,55] In the following, we provide definitions of such measures computed for a stationary stochastic process. First of all, the *entropy* of the process X is defined applying Eq. (1) with $V = X_n$ to get the average information contained in the present state of X:

$$E(X) = H(X_n) = -\mathsf{E}[\log p(x_n)]. \tag{6}$$

For a stationary process, the entropy is a *static* measure, meaning that it does not take the temporal information into account since $p(x_n)$, and thus $H(X_n)$, are the same at all times. The *dynamic* information leading to define entropy rates and complexity is retrieved considering the past information contained in the system up to time $n-1$ and the total information contained the the system up to time n, obtained respectively as the joint entropy of the past variables $\mathbf{X}_n$ and the joint

entropy of the present and past variables, as given by:

$$\begin{aligned} H(\mathbf{X}_n) &= H(X_{n\ 1}, X_{n\ 2}, \ldots) = -\mathsf{E}[\log p(x_{n\ 1}, x_{n\ 2}, \ldots)] \\ H(X_n, \mathbf{X}_n) &= H(X_n, X_{n\ 1}, X_{n\ 2}, \ldots) = -\mathsf{E}[\log p(x_n, x_{n\ 1}, x_{n\ 2}, \ldots)]. \end{aligned} \tag{7}$$

These two entropies are combined as in Eq. (2) with $V = X_n$ and $\boldsymbol{W} = \mathbf{X}_n$ to quantify the *conditional entropy* (CE) of X_n given $\mathbf{X}_n$ as:

$$CE(X) = H(X_n|\mathbf{X}_n) = H(X_n, \mathbf{X}_n) - H(\mathbf{X}_n) = -\mathsf{E}[\log p(x_n|x_{n\ 1}, x_{n\ 2}, \ldots)]. \tag{8}$$

The CE measures the average uncertainty that remains about the present state of X when its past states are known, reflecting the new information that is available in the present state but cannot be inferred from the past. Entropy and CE are related to each other by the so-called *information storage*, defined as the MI computed as in Eq. (3) with $V = X_n$ and $\boldsymbol{W} = \mathbf{X}_n$:

$$IS(X) = I(X_n; \mathbf{X}_n) = \mathsf{E}[\log \frac{p(x_n|\mathbf{x}_n)}{p(x_n)}] = H(X_n) - H(X_n|\mathbf{X}_n). \tag{9}$$

The IS measures the average uncertainty about the present state of X that is resolved by the knowledge of its past states, reflecting the amount of information shared between the present and the past observations of the process.

To summarize, the entropy of a dynamical system measures the information contained in its present state. The information of the present state can then be decomposed to two parts: the new information that cannot be inferred from the past, which is measured by CE and the information that can be explained by its past, which is measured by the information storage. Consequently, entropy, CE and information storage are related to each other by the equation $IS(X) = E(X) - CE(X)$.

2.3. *Entropy rate and complexity*

Considering that a dynamical system varies its state over time, it naturally comes up the idea of quantifying how the information in the system varies with time, so that the *dynamical* properties of the system can be captured. One possibility is to calculate the average rate at which the information is produced by the system. This concept was mathematically defined by Kolmogorov[56] and Sinai,[57] leading to the so-called Komogorov-Sinai (KS) entropy:

$$KS(X) = \lim_{n\,!\,\infty} \frac{1}{n} H(X_n, \mathbf{X}_n), \tag{10}$$

where $H(X_n, \mathbf{X}_n)$ is defined in Eq. (7), representing the total information contained in the process X up to the present time (instant n). Thus, if the information of the system increases with time, the KS entropy captures the average amount of information gained at each time step n. On the contrary, if the information of the system does not change over time, in the limit $n \to \infty$, the KS entropy will be zero.

The idea of entropy rate is also intrinsically related to the concept of CE. The information created by the system at each instant of time is given by $CE(X) = H(X_n|\mathbf{X}_n)$ – Eq. (8) – and represents the information contained in the system at the present time n that cannot be explained by the past up to time n. Thus, the asymptotic value of CE ($n \to \infty$) is another way to define the average rate of information produced by the system. In fact, it can be proved[58] that, under stationary conditions, both limits exist and are equal, i.e.:

$$\lim_{n \to \infty} \frac{1}{n} H(X_n, \mathbf{X}_n) = \lim_{n \to \infty} H(X_n|\mathbf{X}_n). \tag{11}$$

For the analysis of biomedical signals, the limits of Eq. (11) cannot be satisfied, once the recorded signals are always finite in time. However, since stationarity is assumed, the CE is not expected to change over time. Therefore, in practice, the algorithms utilized to calculate CE takes into account only the available (finite) samples of the signals and can be considered as statistics (approximations) of the theoretical CE. In general, the larger the number of samples, the better the estimate provided by the algorithm. Examples of algorithms proposed to estimate the CE from signals are discussed in further sections of this chapter. They include methods to estimate probability densities based on binning, nearest neighbor techniques, or kernel functions; in particular, the use of different kernels leads to the CE estimates known as approximate entropy, sample entropy, and fuzzy entropy.

It is very important to understand the meaning of entropy rate or CE in the context of signal analysis. When most of the information produced by the system cannot be explained by its past, the process generated by the system is assumed to be highly unpredictable (high CE). In contrast, when the average information produced by the system at each time is highly explained by its past system states, the associated process is highly predictable (low CE). In other words, if the information contained in the past states of the system is sufficient to provide a good prediction of the current state, the new information produced by the transition to the current state is low and CE will also be low. On the contrary, when the past system states do not carry the necessary information to predict the current state with good accuracy, the CE will be high. Therefore, CE can be understood as a measure of unpredictability of time series: fully random processes characterize the most unpredictable situations, and thus yield maximal CE, while fully predictable processes characterize the most predictable situations, yielding CE of zero.

The information storage is also related to the predictability of signals. However, while CE is a measure of the current information not resolved by the past, IS represents the current information that is resolved by the past. As such, the IS of a fully random process is zero, as the information carried (stored) in the past values is not useful at all to predict the information of the current time. In contrast, for fully predictable processes, the information stored in the past values is useful to predict completely the current value, so that the IS takes its maximum value.

The predictability of signals is a key concept in the study of the complexity of biomedical signals. CE and IS can be directly utilized to represent the system complexity, with the rationale that the more unpredictable the signal is, the more complex is the underlying system that generated this signal. However, it is worth recalling that the hallmark of complex systems is the huge number of elements and the nonlinear interdependence among them.[59] Thus, the characterization of the system's complexity through a single signal, i.e., by the study of the dynamics of a single output of the system (univariate analysis), configures an important limitation. This is similar to the attempt of studying the properties of a three-dimensional object through its projection in the bi-dimensional plan. Information will be lost. For this reason, many approaches for calculating entropy from more than one signal (multivariate analysis) have been proposed in recent years.[60–64]

It should be remarked that there is a different interpretation in which the entropy, or the degree of unpredictability of signals, cannot be directly used to represent the level of complexity. This interpretation relies on the assumption that neither a completely ordered nor a completely disordered system configure a complex systems.[65,66] For physiological signals, it is argued that the most complex scenario occurs when the system is operating with its high integrity (healthy conditions), and in such cases, the dynamics of the system (measured by the signals) are not completely predictable nor completely unpredictable, but some situation in between these two extremes. Thus, since entropy is essentially a measure of the signal's unpredictability, it could not be used directly to quantify the level of complexity.[65] On the other hand, another important feature of complex systems is the presence of structures at multiple scales, both spatial and temporal.[65] Many studies have shown that the calculation of entropy for different time scales of a signal can reasonably represent the physiological complexity of living organisms, i.e. assigning higher levels of complexity to healthy systems and lower to diseased or elderly individuals.[8,67,68]

Whether the complexity of biomedical signals can be characterized by a direct or an indirect measure of entropy remains a matter of debate and depends on the understanding of complexity. However, one thing is certain: no matter what formulation is proposed to quantify the level of complexity, entropy will certainly take part in it.

3. Methods for entropy estimation

This section introduces some of the most used entropy methods applied to biomedical signal analysis. They are all derived from the theoretical definitions presented in the previous section, representing different algorithms for the estimation of Shannon entropy, CE and IS.

3.1. *Basic definitions*

Here we provide some basic definitions and notations that appears more than once in the definition of entropy methods. Let's consider an arbitrary signal $\mathbf{x}= \{x_1, \ldots, x_N\}$, where N is the length (number of points) of the signal. Note that the terms "signal" and "time series" are nomenclatures used interchangeably in this text, representing a sequence of any quantity estimated over time. The considered signal is taken as a realization of the stochastic process X which quantifies the states visited over time by the underlying dynamical system. Assuming stationarity and ergodicity of the process, the entropy measures defined above can be estimated from the available signal through the methods presented in this section. Moreover, to estimate dynamic entropies that involve probabilities computed on the past history of X, the process itself is assumed as a Markov process, so that a finite number of time-lagged variables can be used to approximate the past; for a Markov process of order m, the past history $\mathbf{X}_n$ is covered by the m-dimensional vector $\mathbf{X}_n^m = [X_{n\ \ 1} \cdots X_{n\ \ m}]$. For the process realization $\mathbf{x}$, the past with memory m is represented by the vector $\mathbf{x}_n^m = \{x_{n\ \ 1}, \ldots, x_{n\ \ m}\}$, which is often denoted as *embedding vector*. Thanks to stationarity, this is equivalent to define the vector $\mathbf{x}_i^m$ as the sequence of values in $\mathbf{x}$ from i to $i+m-1$, i.e. $\mathbf{x}_i^m = \{x_i, \ldots, x_{i+m\ \ 1}\}$. Moreover, when there is the need to introduce a temporal spacing between samples (e.g., in the presence of oversampled signals), the embedding vector may be defined as $\mathbf{x}_i^m(L) = \{x_i, x_{i+L}, \ldots, x_{i+(m\ \ 1)L}\}$, where L is the time delay. In most of the definitions provided in the following sections, $L = 1$ is adopted, unless explicitly state. However, extensions for $L > 1$ are straightforward.

Some of the entropy measures presented in this work make use of distances to estimate probability distributions. The most used metric is the Chebyshev distance, or maximum norm. The Chebyshev distance between two vectors $\mathbf{x}_i^m$ and $\mathbf{x}_j^m$ is defined as:

$$d[\mathbf{x}_i^m, \mathbf{x}_j^m] = \max_{0 \ \ k \ \ m \ \ 1} |\mathbf{x}_{i+k}^m - \mathbf{x}_{j+k}^m|, \tag{12}$$

and represents the maximum pointwise difference between the vectors $\mathbf{x}_i^m$ and $\mathbf{x}_j^m$.

3.2. *Standard binning estimates*

The most intuitive approach for the estimation of entropy measures in signal analysis is the so-called binning estimator. This approach is based on performing uniform quantization of the observed time series and then estimating the entropy approximating probabilities with the frequency of visitation of the quantized states, or bins.

Considering a stationary stochastic process X that takes values in the continuous domain $\mathcal{D}_X = [X_{min}, X_{max}]$, quantization is an operator transforming X into a process X^q which takes values in the discrete alphabet $\mathcal{A}_X$ formed by Q symbols. Correspondingly, quantization transforms the the continuous variable X_n that samples the process X at time n into a discrete variable X_n^q with alphabet

$\mathcal{A}_X = \{1, \ldots, Q\}$. In practice, the realizations of X_n are samples of the time series $\mathbf{x} = \{x_1, \ldots, x_N\}$, which is coarse grained spreading its dynamics over Q quantization levels, or bins. The most common strategy is uniform quantization, which uses bin of equal amplitude $r = (X_{max} - X_{min})/Q$,[69] but alternatives exists which implement a variable bin size chosen to keep constant the number of signal samples falling into the bin.[70,71] The utilization of different transformations from the continuous to the discrete domain is illustrated in Sec. 3.3, where the related technique, DispEn, is described.

Quantization assigns to each sample $x_n, n = 1, \ldots, N$, the number of the bin to which it belongs, so that the quantized time series $\mathbf{x}^q = \{x_1^q, \ldots, x_N^q\}$ is a sequence of discrete values belonging to the alphabet $\mathcal{A}_X$. Then, under the assumption of stationarity, the probability of the i^{th} bin, $i = 1, \ldots, Q$, is estimated simply as the frequency of occurrence of the bin across the quantized time series, i.e. $p(i) = Pr\{x_n^q = i\} = N_i/N$, where N_i is the number of time series points that fall into the bin. The probability estimated in this way is then plugged into Eq. (4) to estimate the entropy of the present state of X according to the definition of Eq. (6):

$$H_{BIN}(X_n) = -\sum_{i=1}^{Q} p(i) \log p(i), \tag{13}$$

As quantization can be performed also for vector variables, this allows to estimate measures like the dynamic entropies and the CE defined in Eq. (7) and Eq. (8). Specifically, the quantization of the past state vector $\mathbf{x}_n^m = [x_{n\ \ 1}, \ldots, x_{n\ \ m}]$ builds a partition of the m-dimensional state space into Q^m disjoint hypercubes of size r, such that all patterns $\mathbf{x}_n^m$ obtained from the time series which fall within the same hypercube are associated with the same m-dimensional bin, and are thus indistinguishable within the tolerance r. The same operation can be performed in the $(m+1)$-dimensional space spanned by the realizations of the present and past states $[X_n, \mathbf{X}_n^m]$ (vectors $[x_n, \mathbf{x}_n^m]$). In both cases, the discrete probabilities of the quantized vector variables are estimated as the fraction of patterns falling into the hypercubes, and can be exploited to compute the dynamic entropy measures (7) as:

$$\begin{aligned} H_{BIN}(\mathbf{X}_n^m) &= -\sum_{i=1}^{Q^m} p(i) \log p(i), \\ H_{BIN}(X_n, \mathbf{X}_n^m) &= -\sum_{i=1}^{Q^{(m+1)}} p(i) \log p(i), \end{aligned} \tag{14}$$

from which the binning estimate of the CE is easily obtained according to (8):

$$CE_{BIN}(X) = H_{BIN}(X_n, \mathbf{X}_n^m) - H_{BIN}(\mathbf{X}_n^m). \tag{15}$$

Note that the sums in Eq. (14) are extended to the bins which contain at least one embedding vector.

In addition to being intuitive and simple the binning method is very fast, as the transformation of the time series values into integer numbers and the application

of a sorting procedure to the integer labels allows the efficient computation of the relative frequencies of the embedding vectors.[72] However, unfortunately the binning method provides biased estimates of the CE. The bias arises from the fact that, at increasing the embedding dimension m, the embedding vectors become more and more isolated in the state space, and this isolation results in an increasing number of vectors $\mathbf{x}_n^m$ found alone inside an hypercube of the m-dimensional quantized space (single vectors); when $\mathbf{x}_n^m$ is a single vector, the vector $[x_n, \mathbf{x}_n^m]$ is also single inside an hypercube of the $(m+1)$-dimensional space, so that the contribution brought by these two vectors to the CE is null; this results in an artificial reduction of the CE estimate that gives a false indication of signal predictability. To counteract this effect, which is exacerbated increasing the embedding dimension, a *corrected* CE (corCE) was defined by adding to the CE estimated as in (15) a corrective term that takes the percentage of single vectors in the m-dimensional space into account.[30]

The correction described above compensates the bias in the CE arising from the lack of reliability of probability estimates due to the shortness of the data sequence relative to the embedding dimension. Interestingly, it can also serve the crucial choice of how to embed the past history of the observed process: while the embedding dimension m is typically constrained to low values to allow reliable statistics,[28] or is set according to complex non-uniform embedding techniques, the finding that the corCE shows a minimum as a function of m[30] provides an objective criterion for the setting of this parameter. Besides embedding, the fundamental issue in binning estimation of entropy measures is the strategy adopted to discretize the observed time series, and the choice about the number of bins to use. Alternative strategies to uniform quantization are seen in Sec. 3.3 in the context of DispEn. As regards the number of bins, there is no "optimal" choice for it. Similarly to the case of other entropy estimators (e.g., SampEn in Sec. 3.4), the choice of the number of bins represents a trade-off between the precision of the probability estimations and the robustness to noise.[73] Several rules of thumb have been proposed, including taking the square root of the number of samples in the signal ($\sqrt{N}$), and proposals for non-Gaussian data, such as Doane's formula.[74]

Binning estimates of entropy measures, mostly the CE but also the implementation of the IS obtained using the estimates in Eq. (13) and Eq. (15) in the definition of Eq. (9), have been extensively used to characterize the complexity of physiological systems. Applications range from animal models to the study of human physiology, analyzed through biomedical time series of heart rate variability, arterial pressure, sympathetic nerve activity, respiration and others, recorded to infer the complexity of homeostatic regulation in different physiological states and pathological conditions.[30,40,42,72,75] In some studies, binning estimates of physiological complexity were compared to other approaches reviewed in this chapter, such as approximate and sample entropy, nearest neighbor estimates, and parametric methods.[40,41]

3.3. *Dispersion Entropy*

Dispersion Entropy (DispEn) is a recently proposed entropy metric based on the application of Shannon entropy to sequences of symbols derived from the levels of amplitude in the samples of $\mathbf{x}$, exploring a different coarse-graining process to that covered in Sec. 3.2. DispEn seeks to provide reliable entropy estimations for short time signals and have a fast computational time.

The DispEn algorithm resembles that of PermEn in some of its steps (see Sec. 3.9), but it considers the amplitude values in the time series after the conversion of samples of $\mathbf{x}$ to different symbols by means of the application of a mapping function.[27] DispEn was introduced in[27] and, since then, the concept has been applied in a variety of settings in biomedical signal analysis, including EEG,[76,77] MEG,[78] cardiac activity,[79,80] and heart sounds,[81] among others.

Formally, DispEn is computed as follows:

(1) The samples in $\mathbf{x}$ (seen as realisations of X_n) are mapped to c discrete classes, which can be denoted with integers ranging from 1 to c. In this step, a number of linear and, more commonly, nonlinear mapping functions can be considered. In most cases, the mean and standard deviation of $\mathbf{x}$ are computed and the samples in $\mathbf{x}$ are transformed using a sigmoid-like function. The transformed values are assigned into c bins of equal size depending on their level of amplitude after the transformation. This results in a temporal sequence of symbols $\mathbf{v}^c = \{v_1^c, \ldots, v_N^c\}$.[34]

(2) The coarse-grained sequence $\mathbf{v}^c$ is used to create patterns in embedding dimension m, number of classes c, and time delay L, in a similar way to how PermEn creates patterns of length m. That is, the *dispersion patterns* $\mathbf{V}_i^{m,c}(L)$ are formed as:

$$\mathbf{V}_i^{m,c}(L) = \{v_i^c, v_{i+L}^c, \ldots, v_{i+L(m\ \ 1)}^c\}, i = 1, 2, \ldots, N - L(m-1). \tag{16}$$

(3) For each of c^m potential dispersion patterns $\phi_{v_0 \ldots v_m}$, its relative frequency of appearance is obtained by counting the number of sequences with that pattern and dividing it by the total number of patterns extracted from the signal. If $p(\phi_{v_0 \ldots v_m})$ denotes the relative frequency of dispersion pattern $\phi_{v_0 \ldots v_m}$, we have

$$p(\phi_{v_0 \ldots v_m}) = \frac{\#\text{ of } i, \text{ such that } \mathbf{V}_i^{m,c} \text{ has type } \phi_{v_0 \ldots v_m}}{N - L(m-1)}. \tag{17}$$

(4) Finally, based on Shannon entropy definition, the DispEn value of $\mathbf{x}$ is calculated as follows:

$$DispEn = -\frac{1}{\log(c^m)} \sum_{=1}^{c^m} p(\phi_{v_0 \ldots v_m}) \cdot \log p(\phi_{v_0 \ldots v_m}), \tag{18}$$

where the factor $\frac{1}{\log(c^m)}$ simply normalises the output to be in the range $[0, 1]$.

The maximum value of 1 is achieved when all possible dispersion patterns $\phi_{v_0 \ldots v_m}$ have equal probability to appear. In contrast, when there is only one pattern across the whole signal, DispEn becomes 0, thus indicating a completely regular time series. It is also worth noting that the number of possible dispersion patterns assigned are c^m.[34] Hence, combining a large number of classes with long pattern lengths may lead to having to store extremely high numbers of distinct patterns $\phi_{v_0 \ldots v_m}$.

Overall, the number of classes c must be chosen to balance the quantity of entropy estimates with the loss of signal information. A small value of c enables us to reduce the impact of noise on the entropy estimation. However, too low c values may result in detailed information being lost in the coarse-graining process. Thus, a trade-off between large and small c values is needed. To achieve reliable estimations, it has been suggested that the number of potential dispersion patterns, c^m should be smaller than the length of the signal, N.[27] In particular, we recommend $c^m \ll N$. Therefore, due to the exponential dependency, a rule of thumb for the selection of c and m is $c^{(m \ \ 1)} < N$.

It is worth noting that the mapping function used in the first step of the DispEn algorithm to transform $\mathbf{x}$ into $\mathbf{v}^c$ has a major impact on the results. The simplest approach would imply sorting the original time series $\mathbf{x}$ and assigned the sorted samples to classes in such a way that each class c has equal range of values. However, this approach may have difficulties dealing with signals with abnormally large values and/or spikes. The reason is that such linear mapping would tend to assign the majority of the samples in $\mathbf{x}$ to too few classes when the maximum and/or minimum values in the time series are much higher in absolute value than most other samples in the signal. It is important to note as well that the use of a linear mapping would result in the quantised series $\mathbf{v}^c$ being equivalent to $\mathbf{x}^q$ in Sec. 3.2. Hence, nonlinear functions with sigmoid shapes are recommended and typically used when dealing with real world data. This recommendation is further supported by the fact that nonlinear mappings of the dynamical range of the $\mathbf{x}$ have demonstrated better performance in the separation of different kinds of biomedical recordings.[34] For a comparison of common nonlinear mapping functions, the reader is referred to Ref.[34]

Variants of DispEn have already been proposed. Multiscale and multivariate versions of DispEn have been introduced to assess patterns across several temporal scales[82] or components of a multivariate signal,[64] respectively. Another variant sought to modify the mapping process to make it robust to outliers and/or missing data due to the fact that these artefacts can be relatively common in biomedical recordings.[83]

The algorithm of DispEn has also been modified to propose fluctuation-based DispEn (FDispEn), which disregards the absolute levels of amplitude in a time series.[34] In this variant, only the difference between adjacent elements of dispersion patterns is considered. The patterns computed in this way are called 'frequency-based dispersion patterns', which have length $m-1$ and elements ranging from $-c+1$ to $+c-1$. The rest of the algorithm is applied like in the original definition

of DispEn, with the difference that we now have $(2c-1)^m$ [1] potential frequency-based dispersion patterns.[34] Preliminary research on a number of synthetic and real-world datasets related to neurological diseases suggest that FDispEn, and its multiscale version, mFDispEn, may even be superior to multiscale SampEn and multiscale DispEn in the detection of different states in the signals.[84]

3.4. *Approximate and sample entropy*

One of the first methods proposed to calculate CE (entropy rate) from finite time series was the approximate entropy (ApEn).[28] Inspired in an approximation for the KS entropy introduced by Eckmann and Ruelle (ER),[85] Pincus proposed to fix the two parameters taken as limits in the ER formulation (m and r) and estimate ApEn as a statistics for any finite-length signals. ApEn is defined as follows.

Consider the signal $\mathbf{x}$ and all of its possible vectors $\mathbf{x}_i^m$ $(1 \leq i \leq N-m+1)$ defined in Sec. 3.1. Define $C_i^m(r)$ as the probability of finding any vector in $\mathbf{x}$ whose distance (Eq. (12)) to the template vector $\mathbf{x}_i^m$ is lower than or equal to r. In mathematical terms, C_i^m can be defined as:

$$C_i^m(r) = \frac{\#\text{ of } \mathbf{x}_j^m \text{ such that } d[\mathbf{x}_i^m, \mathbf{x}_j^m] \leq r}{N-m+1}, \ (1 \leq j \leq N-m+1). \tag{19}$$

When the distance between the template vector and another vector in $\mathbf{x}$ is lower than or equal to the tolerance factor, i.e. $d[\mathbf{x}_i^m, \mathbf{x}_j^m] \leq r$, the two vectors are considered similar (vector match). The number of matches for the template vector $\mathbf{x}_i^m$ is divided by the number of possible vectors of length m, so that $C_i^m(r)$ estimates, within the tolerance r, the probability of finding vectors similar to $\mathbf{x}_i^m$ in $\mathbf{x}$; such probability corresponds to the probability of the history $\mathbf{X}_n^m$ of the investigated process estimated at the data point $\mathbf{x}_i^m$ using the Heaviside (step) kernel with parameter r.[16] Now, define

$$\Phi^m(r) = \frac{1}{(N-m+1)} \sum_{i=1}^{N-m+1} \ln C_i^m(r), \tag{20}$$

as the average logarithmic probability of finding any match in $\mathbf{x}$, considering all possible vectors of length m. Equation (20) is a negative estimate of the dynamic entropy in the first part of Eq. (7), i.e. $\Phi^m(r) \approx -H(\mathbf{X}_n^m)$. Finally, the ApEn of $\mathbf{x}$ (length N), for specific choices of m and r, is defined as:

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r). \tag{21}$$

Note that, similarly to $\Phi^m(r)$, the term $\Phi^{m+1}(r)$ is a negative estimate of the dynamic entropy in the second part of Eq. (7), i.e. $\Phi^{m+1}(r) \approx -H(X_n, \mathbf{X}_n^m)$. This makes clear that the ApEn defined in Eq. (21) is a kernel estimate of the CE defined in Eq. (8), i.e. $ApEn(m, r, N) \approx -H(\mathbf{X}_n^m) + H(X_n, \mathbf{X}_n^m) = H(X_n|\mathbf{X}_n^m)$.

In the definition proposed by ER, CE is obtained in the limits $m \to \infty$, $r \to 0$ and $N \to \infty$. However, Pincus proposed that its estimation for specific choices of

m, r and N could be of great interest as a measure of regularity.[28] Therefore, ApEn is an estimator of CE for real (and usually noisy) signals and can be interpreted as the (logarithm) probability that vectors that are similar for m points will remain similar when an additional point is considered ($m+1$). For example, if all the similar m-length vectors remain similar when the vectors size are increased to $m+1$, it means that the signal tends to be very repetitive. If one knows the previous m values of any sequence, the $(m+1)$-th value can be fully predicted. In this case, the new information carried by the current value $(m+1)$ is zero and ApEn will also be zero. On the other hand, when none of the similar m-length vectors remain similar when the vectors size are increased to $m+1$, it means that the previous m points are not useful at all to predict the $(m+1)$-th value of the sequence. In this situation, the information carried by the current value $(m+1)$ is maximum, and so is ApEn.

Although the introduction of ApEn was a hallmark, it quickly became apparent that ApEn is biased. ApEn intentionally does not discard self-matches in the comparison of vectors to avoid the occurrence of zeros in Eq. (19), which would lead to undefined entropy. However, this is at the expense of resulting in a biased estimation of the conditional probabilities, so that ApEn assigns more similarity among patterns than is really present.[2] As a consequence, ApEn is strongly dependent on the signal size (N) and does not show relative consistence, i.e. if $ApEn(m_1, r_1)(\mathbf{x}) \leq ApEn(m_1, r_1)(\mathbf{y})$, there is no guarantee that $ApEn(m_2, r_2)(\mathbf{x}) \leq ApEn(m_2, r_2)(\mathbf{y})$.

To overcome the limitations of ApEn, Richman and Moorman introduced the sample entropy (SampEn).[29] Essentially, SampEn has the same purpose as ApEn, but the algorithm utilized to calculate SampEn does not require the inclusion of self-matches when estimating the probability of occurrence of vectors. This was possible by changing the way the conditional probabilities are estimated. In ApEn, it is a template-wise procedure, so that the logarithm of the conditional probability of *each template vector* is calculated and averaged. In SampEn, however, the conditional probability is estimated from *all template vectors*, so that the logarithm is only taken after the calculation of the overall conditional probability. This drastically reduces the chance of resulting in undefined entropy.

The algorithm of SampEn was inspired in the work from Grassberger and Procaccia[86] and can be defined as follows. Consider the same signal ($\mathbf{x}$) and all of its possible vectors $\mathbf{x}_i^m$ $(1 \leq i \leq N-m+1)$ defined previously. Let

$$U_i^m(r) = \frac{\#\text{ of } \mathbf{x}_j^m \text{ such that } d[\mathbf{x}_i^m, \mathbf{x}_j^m] \leq r}{N-m-1}, \ (1 \leq j \leq N-m, j \neq i) \tag{22}$$

and

$$U_i^{m+1}(r) = \frac{\#\text{ of } \mathbf{x}_j^{m+1} \text{ such that } d[\mathbf{x}_i^{m+1}, \mathbf{x}_j^{m+1}] \leq r}{N-m-1}, \ (1 \leq j \leq N-m, j \neq i) \tag{23}$$

be the probability of finding any vector similar to $\mathbf{x}_i^m$ and $\mathbf{x}_i^{m+1}$, respectively, in signal $\mathbf{x}$. The constraints $j \neq i$ in Eq. (22) and Eq. (23) assure that self-matches

are not accounted. Also, the range of j is the same in both equations, assuring that the space of vectors evaluated for m and $m+1$ is the same.

Now, the probability of finding any vector match for sequences of size m can be obtained by averaging the probabilities for each template vector $\mathbf{x}_i^m$, $1 \leq i \leq N-m$. The same procedure can be performed for vectors of size $m+1$, resulting in

$$U^m(r) = \frac{1}{N-m} \sum_{i=1}^{N} {}^{m} U_i^m(r) \tag{24}$$

and

$$U^{m+1}(r) = \frac{1}{N-m} \sum_{i=1}^{N} {}^{m} U_i^{m+1}(r). \tag{25}$$

Finally, the SampEn of $\mathbf{x}$ (length N), for specific choices of m and r, is defined as:

$$SampEn(m, r, N) = -\ln \frac{U^{m+1}(r)}{U^m(r)}. \tag{26}$$

Considering that the space of vectors utilized to estimate $U^{m+1}(r)$ and $U^m(r)$ is the same, the ratio $U^{m+1}(r)/U^m(r)$ in Eq. (26) can be simplified and SampEn could be defined simply as:

$$SampEn(m, r, N) = -\ln \frac{A}{B}, \tag{27}$$

where A and B are the total number of matches for all template vectors of size $m+1$ and m, respectively.

As the number of vector matches (A and B) or the vector occurrence probabilities ($U^m(r)$ and $U^{m+1}(r)$) are averaged over all template vectors, it is straightforward to see that any match, found for any template vector of size $m+1$, is sufficient to assure that SampEn will return a valid entropy value. In case of ApEn, it must be satisfied for all template vectors individually, otherwise ApEn will not be defined. Moreover, as SampEn does not account for self-matches, the bias presented by ApEn were considerably improved[29] and SampEn should be preferable over ApEn whenever possible.

The extension of SampEn to a multiscale method was proposed by Costa *et al.*, which became widely known as multiscale entropy (MSE).[8,87] MSE calculates SampEn for several scaled versions of the original signal and the variations of SampEn as a function of the scale factor has been used to represent the signal complexity in many biomedical problems.[88–91] Several variants of MSE were proposed in the following years.[92]

3.5. *Fuzzy entropy approaches*

In ApEn and SampEn algorithms, the similarity definition of vectors is based on the Heaviside function, a hard and sensitive boundary function. This leads to entropy measures that are sensitive to the parameter values and may be invalid in case of small parameter values:[93,94] the value of the entropy measure is discontinuous and it can show large variations with a slight change of the tolerance r. This is due to the two states of the Heaviside function (0 and 1). It has been reported that fuzzy entropy approaches are more accurate irregularity measures than ApEn and SampEn: they show more consistency and less dependence on data length, achieve continuity, and are more robust to noise.[93]

Given the time series $\mathbf{x} = \{x_1, \ldots, x_N\}$, the algorithm to compute the fuzzy sample entropy is the following:[93,95,96]

(1) for an embedding dimension m, construct $(N - m + 1)$ vectors $\mathbf{s}_i^m = \{x_i, x_{i+1}, ..., x_{i+m\ \ 1}\} - \overline{x}_i^m$, $1 \leq i \leq N - m + 1$, where $\overline{x}_i^m$ is the baseline and is computed as

$$\overline{x}_i^m = \frac{1}{m} \sum_{j=0}^{m\ \ 1} x_{i+j}. \tag{28}$$

Notice that vectors $\mathbf{s}_i^m$ are similar to the vectors $\mathbf{x}_i^m$ defined in Sec. 3.1, with the difference that $\mathbf{s}_i^m$ removes the vector mean baseline.

(2) for a given $\mathbf{s}_i^m$, calculate its similarity with the neighboring vector $\mathbf{s}_j^m$. This is performed through the similarity degree $D_{ij}^m(n, r)$ defined by a fuzzy function

$$D_{ij}^m(n, r) = \mu_L(d_{ij}^m, n, r), \tag{29}$$

where μ_L is the fuzzy function defined as

$$\mu_L(d_{ij}^m, n, r) = \exp\left(- \frac{(d_{ij}^m)^n}{r} \right), \tag{30}$$

and d_{ij}^m is the Chebyshev distance [Eq. (12)] between $\mathbf{s}_i^m$ and $\mathbf{s}_j^m$, i.e.:

$$d_{ij}^m = d[\mathbf{s}_i^m, \mathbf{s}_j^m], \tag{31}$$

$$d_{ij}^m = \max_{0\ \ k\ \ m\ \ 1} |x_{i+k} - \overline{x}_i^m - (x_{j+k} - \overline{x}_j^m)|. \tag{32}$$

(3) determine $B^m(n, r)$ as

$$B_i^m(n, r) = \frac{1}{N - m - 1} \sum_{j=1, j \neq i}^{N\ \ m} D_{ij}^m(n, r), \tag{33}$$

and

$$B^m(n, r) = \frac{1}{N - m} \sum_{i=1}^{N\ \ m} B_i^m(n, r). \tag{34}$$

(4) in the same way, compute

$$A_i^m(n,r) = \frac{1}{N-m-1} \sum_{j=1, j\neq i}^{N-m} D_{ij}^{m+1}(n,r), \tag{35}$$

and

$$A^m(n,r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(n,r). \tag{36}$$

(5) compute the fuzzy sample entropy as

$$\text{FuzzyEn}(m,n,r) = \lim_{N \to \infty} [\ln B^m(n,r) - \ln A^m(n,r)], \tag{37}$$

which gives, for finite datasets

$$\text{FuzzyEn}(m,n,r,N) = \ln B^m(n,r) - \ln A^m(n,r). \tag{38}$$

Similarly to ApEn and SampEn, the FuzzyEn is a kernel estimate of the CE defined in Eq. (8), computed from a realization of length N of the underlying process assumed as a Markov process of order m; the difference between FuzzyEn and ApEn/SampEn stands in the use of a smooth kernel function in place of the step function realized by the Heaviside kernel.

For the choice of the fuzzy function, μ_L, several functions can be chosen, each with drawbacks and advantages, as described in Ref.[43] The fuzzy function should have the following properties: (1) being continuous so that the similarity does not change abruptly; (2) being convex so that self-similarity is the maximum.[93]

Using this fuzzy approach, and the nonlinear Sigmoid as the fuzzy function (μ_L), Xie *et al.* reported that FuzzyEn outperforms the standard SampEn measure in terms of relative consistency, freedom of parameter selection, robustness to noise and independence on the data length.[95]

Later on, Liu *et al.* proposed the fuzzy measure entropy (FuzzyMEn).[97] FuzzyMEn employs both the fuzzy local and the fuzzy global measure entropies to reflect the local and global characteristics of the time series. FuzzyMEn therefore reflects the entire complexity in the time series (global and local similarity degree).[97] This is not the case with the standard fuzzy entropy measure that only focuses on the local waveform characteristics of the signals by removing the local baselines without considering any global signal characteristics. The algorithm for FuzzyMEn, for given values m, n_L, n_G, r_L, r_G is

(1) compute $\mathbf{s}_i^m$ as mentioned above, but also vectors $\mathbf{g}_i^m = \{x_i, x_{i+1}, ..., x_{i+m-1}\} - \overline{x}_{mean}$, $1 \leq i \leq N-m+1$, where $\overline{x}_{mean}$ is the mean value of the time series $\mathbf{x}$.

(2) calculate the similarity as defined in Eqs. (29) to (32), but compute also

$$DG_{ij}^{m}(n_G, r_G) = \mu_G(dG_{ij}^{m}, n_G, r_G), \tag{39}$$

where μ_G is the fuzzy function

$$\mu_G(dG_{ij}^{m}, n_G, r_G) = \exp\left(-\frac{(dG_{ij}^{m})^{n_G}}{r_G}\right), \tag{40}$$

and dG_{ij}^{m} is the Chebyshev distance between $\mathbf{g}_i^m$ and $\mathbf{g}_j^m$, i.e.:

$$dG_{ij}^{m} = d[\mathbf{g}_i^m, \mathbf{g}_j^m], \tag{41}$$

$$dG_{ij}^{m} = \max_{0 \;\; k \;\; m \;\; 1} |x_{i+k} - \overline{x}_{mean} - (x_{j+k} - \overline{x}_{mean})|. \tag{42}$$

(3) determine $B^m(n_L, r_L)$ (Eq. (34)) and $BG^m(n_G, r_G)$, using for $BG^m(n_G, r_G)$:

$$BG_i^m(n_G, r_G) = \frac{1}{N-m-1} \sum_{j=1, j \neq i}^{N \;\; m} DG_{ij}^{m}(n_G, r_G), \tag{43}$$

$$BG^m(n_G, r_G) = \frac{1}{N-m} \sum_{i=1}^{N \;\; m} BG_i^m(n_G, r_G). \tag{44}$$

(4) in the same way, compute $A^m(n_L, r_L)$ [Eq. (36)] and $AG^m(n_G, r_G)$, using for $AG^m(n_G, r_G)$:

$$AG_i^m(n_G, r_G) = \frac{1}{N-m-1} \sum_{j=1, j \neq i}^{N \;\; m} DG_{ij}^{m+1}(n_G, r_G), \tag{45}$$

and

$$AG^m(n_G, r_G) = \frac{1}{N-m} \sum_{i=1}^{N \;\; m} AG_i^m(n_G, r_G). \tag{46}$$

(5) compute FuzzyEn [Eq. (38)] and FuzzyGEn using for FuzzyGEn:

$$\text{FuzzyGEn}(m, n_G, r_G) = \lim_{N \; ! \; 1} [\ln BG^m(n_G, r_G) - \ln AG^m(n_G, r_G)], \tag{47}$$

which gives, for finite datasets

$$\text{FuzzyGEn}(m, n_G, r_G, N) = \ln BG^m(n_G, r_G) - \ln AG^m(n_G, r_G). \tag{48}$$

(6) finally compute FuzzyMEn as:

$$\text{FuzzyMEn}(m, n_G, r_G, n_L, r_L, N) = \text{FuzzyEn}(m, n_L, r_L, N) + \text{FuzzyGEn}(m, n_G, r_G, N). \tag{49}$$

Liu *et al.* reported that FuzzyMEn has better discrimination ability than the above-mentioned fuzzy entropy approach.[97] More recently, a refined fuzzy entropy measure was also developed.[98] It consists in substituting a piecewise fuzzy membership function for the Gaussian function in conventional fuzzy entropy measure. Ji *et al.* reported that this refined fuzzy entropy measure outperforms conventional SampEn and FuzzyEn in terms of stability and robustness against additive noise. This was particularly true when the data set was relatively short.[98]

For some of the fuzzy functions, more parameters have to be set (e.g., n in the family of exponential function $\exp(-(d_{ij}/r)^n)$).[97] This can be an additional drawback compared to more standard entropy measures.

More recently, other fuzzy entropy-based measures have been proposed, as described in Refs.[99–102] Recent modified multiscale approaches have also been published.[103–106]

3.6. *Nearest neighbors metrics*

The class of estimators of entropy measures which makes use of nearest neighbor metrics exploits the intuitive notion that the local probability density around a given data point is inversely related to the distance between the point and its nearest neighbors: the larger the distance between an observation $\boldsymbol{w}$ of a variable $\boldsymbol{W}$ and its $k-$nearest neighbor, the lower the local density $p(\boldsymbol{w})$. This concept has been formalized first in the seminal work of Kozachenko and Leonenko (KL),[107] who studied the statistics of the distances between neighboring points in multidimensional spaces to provide an estimator of the differential entropy of a vector random variable defined as in Eq. (5). The difference with the kernel-based estimators seen in the previous subsections (i.e., approximate and sample entropy, fuzzy entropy) is in the fact that, instead of fixing the neighborhood size for the reference point according to a given threshold distance, the KL nearest neighbor estimator fixes the number of neighbors of the reference point and quantifies the neighborhood size by computing the distance between the reference point and its k^{th} nearest neighbor. Accordingly, the KL estimator can be thought of as "variable-width" kernel-estimator, as the variability of distances with the local density of points allows to adjust dynamically the "resolution" of the estimates. Crucially, there are consistency proofs[108,109] for nearest neighbor entropy estimators, meaning that these methods converge to the true values in the limit of infinite data size.

The KL estimator illustrates that to estimate an information measure it is not necessary to explicitly estimate the probability distribution from the available data and then plug such distribution in the functional expressing the measure. In fact, given a d-dimensional continuous random variable $\boldsymbol{W}$, the KL estimator computes the differential entropy $H(\boldsymbol{W}) = -\mathsf{E}[\log p(\boldsymbol{w})]$ as the sample average of estimates of the information content, $h(\boldsymbol{w}) = -\log p(\boldsymbol{w})$, obtained at the data points $\boldsymbol{w}_n, n =$

$1, \ldots, N^0$ (where N^0 is the number of observations available for $\boldsymbol{W}$):

$$\hat{H}(\boldsymbol{W}) = \frac{1}{N^0} \sum_{n=1}^{N'} \hat{h}(\boldsymbol{w}_n). \tag{50}$$

Then, denoting as $\epsilon_{n,k}$ twice the distance between the reference point $\boldsymbol{w}_n$ and its k^{th} nearest neighbor and as p_n the probability mass of the ϵ-ball surrounding $\boldsymbol{w}_n$, under the assumption that the unknown probability density $p(\boldsymbol{w})$ is constant within the ϵ-ball we have $p_n = \frac{k}{N'-1} = c_{d,L}\epsilon_{n,k}^d p(\boldsymbol{w})$, where $c_{d,L}$ is the volume of the d-dimensional unit ball under a given norm L ($c_{d,L} = 1$ for the maximum norm, i.e., taking the maximum distance of the scalar components). Exploiting this relation, the estimator in Eq. (50) becomes:

$$\hat{H}(\boldsymbol{W}) = -\frac{1}{N^0} \sum_{n=1}^{N'} \log \frac{k}{(N^0-1)c_{d,L}\epsilon_{n,k}^d}. \tag{51}$$

Finally, adding the bias-correction term $\log k - \psi(k)$, where $\psi(x) = \frac{\log(\Gamma(x))}{dx}$ is the digamma function and $\Gamma(x)$ the gamma function, we obtain the KL entropy estimator:[107]

$$\hat{H}_{KL}(\boldsymbol{W}) = -\psi(k) + \log(N^0-1) + \log c_{d,L} + \frac{d}{N^0} \sum_{n=1}^{N'} \log \epsilon_{n,k}. \tag{52}$$

The KL estimator makes use of only one free parameter, i.e., the number of nearest neighbors k, which allows for a certain control of the bias versus variance balance (larger k reduce statistical errors at the expense of a higher bias) even for finite samples. Such estimator possesses several properties, including its asymptotic absence of bias and consistency,[107–109] high data-efficiency[110] and the possibility to exploit the many algorithms for neighbor search to perform the estimation,[111] that make it particularly attractive for practical applications. In the computation of information measures for signal analysis, the estimator in Eq. (52) is applied in a straightforward way to compute, from a time series $\mathbf{x} = \{x_1, \ldots, x_N\}$, estimates of the static entropy of the underlying process X as defined in Eq. (6):

$$\hat{E}(X) = \hat{H}_{KL}(X_n) = -\psi(k) + \log(N-1) + \log c_{1,L} + \frac{1}{N} \sum_{n=1}^{N} \log \epsilon_{n,k}, \tag{53}$$

or, after approximating the past history of the process with m time-lagged variables ($\mathbf{X}_n \approx \mathbf{X}_n^m = [X_{n\ 1} \cdots X_{n\ m}]$), estimates of the dynamic entropy of the present and past of the process as defined in Eq. (7):

$$\hat{H}_{KL}(X_n, \mathbf{X}_n^m) = -\psi(k) + \log(N-m-1) + \log c_{m+1,L} + \frac{m+1}{N-m} \sum_{n=1}^{N} \sum^{m} \log \epsilon_{n,k}; \tag{54}$$

note that, for a fixed number of neighbors k and embedding dimension m, the only quantity that needs to be computed to estimate these entropies is the distance from

the n^{th} data point (x_n in Eq. (53) and $[x_n x_{n\ 1} \cdots x_{n\ m}]$ in Eq. (54)) to its k^{th} nearest neighbor (denoted as $\epsilon_{n,k}$ in both cases).

Then, for information measures defined as the sum of two or more entropy terms such as the CE and the IS in Eq. (8) and Eq. (9), a naive estimator for such functionals would consist of summing the individual differential entropy estimators. However, this may not be adequate in practical applications, since the variables involved in the computation of CE or mutual information terms are of different dimension (respectively, 1, m and $m+1$), and the bias of the KL entropy estimator varies with the dimension. The practical issue is that the naive application of the same neighbor search procedure in all spaces would result in different distance lengths when approximating the probability density in different dimensions, and this would introduce different estimation biases that cannot be compensated by taking the entropy differences. A solution to this problem was proposed by Kraskov, Stogbauer and Grassberger,[112] who provided an approach to adapt the KL estimator to mutual information estimation. The resulting estimator, denoted as KSG estimator, came from the insight that Eq. (52) holds for any number of neighbors k, and therefore can be iterated varying k and keeping fixed the distance $\epsilon_{n,k}$. Thus, to estimate a MI measure, or more generally any information measure that can be decomposed in terms of differential entropies, a neighbor search can be performed in the highest-dimensional space to find the distance between any reference point and its k^{th} neighbor, and then this distance can be used in a lower dimensional space as the range inside which the neighbors are counted. This allows to use the same range of distances in the estimation of the different entropy terms, thus achieving bias compensation. In our context, to estimate the CE and the IS defined in Eq. (8) and Eq. (9), we compute first the joint entropy $\hat{H}(X_n, \mathbf{X}_n^m)$ as in Eq. (54), where $\epsilon_{n,k}$ is twice the distance from $(x_n, \mathbf{x}_n^m)$ to its k^{th} nearest neighbor, and then, given the distances $\epsilon_{n,k}$, the entropies in the lower-dimensional spaces are estimated through a range search:

$$\begin{aligned}\hat{H}(\mathbf{X}_n^m) &= \log(N-m-1) + \log c_{m+1,L} + \frac{1}{N-m}\sum_{n=1}^{N}{}^{m}(m\log\epsilon_{n,k} - \psi(N_{\mathbf{x}_n^m})),\\ \hat{H}(X_n) &= \log(N-m-1) + \log c_{m+1,L} + \frac{1}{N-m}\sum_{n=1}^{N}{}^{m}(\log\epsilon_{n,k} - \psi(N_{x_n})),\end{aligned} \tag{55}$$

where N_{x_n} and $N_{\mathbf{x}_n^m}$ are the number of points whose distance from x_n and $\mathbf{x}_n^m$, respectively, is smaller than $\epsilon_{n,k}/2$. Then, the KSG estimates of CE and IS are obtained applying Eq. (54) and Eq. (55) to the definitions $CE(X) = H(X_n, \mathbf{X}_n^m) - H(\mathbf{X}_n^m)$ and $IS(X) = -H(X_n, \mathbf{X}_n^m) + H(\mathbf{X}_n^m) + H(X_n)$, yielding:

$$CE_{KSG}(X) = -\psi(k) + \frac{1}{N-m}\sum_{n=1}^{N}{}^{m}(\log\epsilon_{n,k} + \psi(N_{\mathbf{x}_n^m})), \tag{56}$$

$$IS_{KSG}(X) = \psi(k) + \log(N-m-1) + \log c_{m+1,L} - \frac{1}{N-m}\sum_{n=1}^{N}{}^{m}(\psi(N_{x_n}) + \psi(N_{\mathbf{x}_n^m})). \tag{57}$$

Note that, in both the estimation of CE and IS, many of the terms in each entropy estimate cancel out when are summed, so that each entropy is only implicitly estimated.

The peculiarities of the nearest neighbor technique described above, i.e., its adaptive resolution obtained through changing the distance scale[110] and the bias compensation implemented through distance projection,[112] allow the method to work with short data sets even when working in relatively high-dimensional spaces as those naturally considered after the embedding of raw signals. These properties have made the approach eligible especially for the computation of entropy measures for multivariate time series where the issue of dimensionality is more serious.[113–115] Nevertheless, the properties of course hold also in the univariate case, and indeed the approach has gained popularity also when implemented for the computation of entropy measures for individual biomedical time series, as documented by the recent applications in neuroscience[116] and cardiovascular control.[16,41,117,118]

3.7. *Parametric methods*

The parametric methods to estimate information-theoretic quantities make the assumption that the probability distributions involved into the computation of the desired entropy measures belong to a certain family, and start by first inferring the parameters of the family that best fit the observed data samples. Afterwards, the estimated parameters are plugged into the functional that relates the probability distribution to the entropy measure, thus providing an estimate of the latter. For some families of continuous probability densities, the parametric approach provides an analytical insight on the dependence of the desired information-theoretic functional on the parameters. The most common case, which will be treated in this section, is the multivariate Gaussian distribution, for which entropies can be computed analytically from parameters related to the covariance structure of the underlying variables.[119] This case is of particular relevance thanks to the fact that many real-world data tend to the Gaussian distribution, and also because of the close link existing between linear parametric models and Gaussian distributions.[119,120] Extensions to other distributions are given in the work of Hlavackova *et al.*,[121] who provided the derivation of differential entropy for several parametric probability functions including the generalized normal, log-normal and Weinman densities.

The linear Gaussian estimator starts from the well-known joint probability density of the generic d-dimensional variable $\boldsymbol{W}$ belonging to the Gaussian family:

$$p(\boldsymbol{w}) = \frac{1}{\sqrt{2\pi^d|\boldsymbol{\Sigma_W}|}} e^{\ \frac{1}{2}(\boldsymbol{w}\ \ \boldsymbol{\mu_W})\boldsymbol{\Sigma_W^{-1}}(\boldsymbol{w}\ \ \boldsymbol{\mu_W})^T}, \tag{58}$$

where $\boldsymbol{\mu_W} = \mathsf{E}[\boldsymbol{W}]$ and $\boldsymbol{\Sigma_W} = \mathsf{E}[(\boldsymbol{W}-\boldsymbol{\mu_W})^T(\boldsymbol{W}-\boldsymbol{\mu_W})]$ are the mean vector and the covariance matrix of $\boldsymbol{W}$. Then, the joint entropy of $\boldsymbol{W}$ can be computed by

using Eq. (58) in Eq. (1):[53]

$$H(\boldsymbol{W}) = \frac{1}{2}\ln((2\pi e)^d|\boldsymbol{\Sigma}_{\boldsymbol{W}}|); \tag{59}$$

when a scalar variable V is considered, Eq. (59) reduces to $H(V) = \frac{1}{2}\ln(2\pi e\sigma_V^2)$, where σ_V^2 is the variance of V, which shows how the entropy of a Gaussian variable is a function of its variance only. Moreover, the use of Eq. (59) in the two terms appearing in the right hand side of Eq. (2) allows to express the CE of V given $\boldsymbol{W}$ as:

$$H(V|\boldsymbol{W}) = \frac{1}{2}\ln(2\pi e\sigma_{V\jmath\boldsymbol{W}}^2), \tag{60}$$

where $\sigma_{V\jmath\boldsymbol{W}}^2 = \frac{\jmath\boldsymbol{\Sigma}_{V\boldsymbol{W}}\jmath}{\jmath\boldsymbol{\Sigma}_{\boldsymbol{W}}\jmath}$ is the so-called partial variance of V given $\boldsymbol{W}$ ($\boldsymbol{\Sigma}_{V\boldsymbol{W}}$ is the covariance matrix of the joint variable $[V, \boldsymbol{W}]$). Additionally, the reported formulations can be exploited to draw a parallel between information measures computed for Gaussian variables and linear prediction. In fact, considering a multivariate linear regression of V on $\boldsymbol{W}$ represented as $V = \boldsymbol{A}\boldsymbol{W} + B + U$, where $\boldsymbol{A}$ and B are the d-dimensional vector of the regression coefficients and the scalar intercept, and U is a zero-mean random variable modeling the prediction error, it can be shown that the partial variance of V given $\boldsymbol{W}$ is equivalent to the prediction error variance, i.e. $\sigma_{V\jmath\boldsymbol{W}}^2 = \sigma_U^2$.[120]

The above derivations can be exploited in a straightforward way to estimate entropy measures in univariate signal analysis. Assuming that the observed stochastic process X is a Gaussian Markov process of order m, the roles of the generic variables V and $\boldsymbol{W}$ are taken by the present state of the process, X_n, and by its past states truncated to lag m, $\mathbf{X}_n^m = [X_{n\ 1} \dots X_{n\ m}]$. Then, assuming stationarity of the process, sample estimates of the process variance, $\hat{\sigma}_X^2 = \hat{\sigma}_{X_n}^2$, and of the covariance matrix of present and past states, $\hat{\boldsymbol{\Sigma}}_{X_n\mathbf{X}_n^m}$, can be easily obtained from the time series $\mathbf{x} = \{x_1, \dots, x_N\}$ measured as a realization of X. Moreover, the linear regression model $X_n = \boldsymbol{A}\mathbf{X}_n^m + B + U_n$ can be identified via ordinary least squares methods to obtain estimates of the regression parameters and of the prediction error variance $\hat{\sigma}_U^2$.[122] From these parameter estimates, plug-in estimates of the static entropy of the process, of the dynamic entropy of present and past states, of the CE of the present given the past, and of the IS measured as MI between present and past states are finally obtained as:

$$E_G(X) = \hat{H}(X_n) = \frac{1}{2}\ln(2\pi e\hat{\sigma}_X^2), \tag{61}$$

$$\hat{H}(X_n, \mathbf{X}_n^m) = \frac{1}{2}\ln((2\pi e)^{m+1}|\hat{\boldsymbol{\Sigma}}_{X_n\mathbf{X}_n^m}|), \tag{62}$$

$$CE_G(X) = \hat{H}(X_n|\mathbf{X}_n^m) = \frac{1}{2}\ln(2\pi e\hat{\sigma}_U^2), \tag{63}$$

$$IS_G(X) = \hat{I}(X_n; \mathbf{X}_n^m) = \frac{1}{2}\ln\frac{\hat{\sigma}_X^2}{\hat{\sigma}_U^2}; \tag{64}$$

in the above expressions, the subscript $_G$ stands for the use of the linear Gaussian approximation to compute the relevant entropy measures. In particular, Eq. (63) provides the linear estimate of the complexity of a time series (LinCE).

The main advantage of the methods leading to the computation of entropy measures as in Eqs. (61-64) is the high data-efficiency guaranteed by the parametric representation, which allows indeed accurate estimation of the information measures based on the parameter estimates. As regards the linear parametric description formulated here, the existence of well-established parameter estimation methods for the identification of time series regression models[123] assures accurate and efficient computation of the parameters and consequently of the entropy measures. Note that, while the formulation presented here holds exactly only for Gaussian processes for which the linear representation captures the whole the entropy variations in the analyzed system, extensions to non-linear representations are possible exploiting non-Gaussian parametric distributions.[121] On the other hand, the success of parametric estimators depends on the correctness of the assumptions made about the family of probability densities described by the estimated parameters: if the empirical distribution of the observed data deviates strongly from the assumed distribution, heavy misinterpretations may be expected in the inference of system properties from the estimated entropy measures. Nevertheless, the assumption of joint Gaussian distribution for the analyzed processes is reasonably met in several applicative contexts also in the analysis of biomedical signals. This is documented by the fact that the corresponding parametric representation leading to the measures presented here has been widely used to characterize the patterns of complexity and regularity of biomedical time series, particularly in the context of cardiovascular and cardiorespiratory dynamics.[16,41,118,124,125] Moreover, CE measures derived from linear models have been proposed as a model-based alternative of SampEn as the basis for multiscale entropy analysis, allowing the computation of complexity across multiple time scales in a data-efficient way.[126–128]

3.8. *Distribution entropy*

Distribution entropy (DistEn) has been proposed to improve the robustness of complexity assessment for short-term time series.[33] In the DistEn algorithm, the distribution of vector-to-vector distances in the state-space is used as an interpretation of the spatial structures, and is supposed to increase with complexity.

The algorithm to compute DistEn of $\mathbf{x}$ is the following[33]

(1) construct the vectors $\mathbf{x}_i^m$ as defined in Sec. 3.1.
(2) define the distance matrix $\mathbf{D}^m = \{d_{i,j}^m\}$ among all vectors $\mathbf{x}_i^m$ and $\mathbf{x}_j^m$ as the Chebyshev distance between $\mathbf{x}_i^m$ and $\mathbf{x}_j^m$. Note that $\mathbf{D}^m$ is symmetrical.
(3) estimate the empirical probability density function (ePDF) of $\mathbf{D}^m$ with the histogram (binning) approach. As described in Sec. 3.2, the binning procedure involves the quantization of $\mathbf{D}^m$ values into Q levels, yielding the quantized

sequence $\mathbf{D}_q^m$, $q = 1, \ldots, Q$. Then, the probability of occurrence of the i^{th} bin is estimated by the frequency of occurrence of each bin, i.e. $p(i) = Pr\{\mathbf{D}_q^m = i\}$. To reduce the bias, elements with $i = j$ are excluded when estimating ePDF.

(4) finally, compute DistEn expressed in bits with the formulation provided in Eq. (13):

$$\overline{\text{DistEn}}(m, Q) = -\sum_{i=1}^{Q} p(i) \log_2 p(i), \tag{65}$$

where $p(i) \log_2 p(i) = 0$ when $p(i) = 0$. $\overline{\text{DistEn}}$ varies between 0 and $\log_2(Q)$ (one peak and flat ePDF, respectively). A normalized DistEn value can be computed as:

$$\text{DistEn}(m) = -\frac{1}{\log_2(Q)} \sum_{i=1}^{Q} p(i) \log_2 p(i). \tag{66}$$

In the latter case, DistEn varies between 0 and 1. Here again, $p(i) \log_2 p(i) = 0$ when $p(i) = 0$.

When applied to synthetic signals, DistEn has shown to have lower sensitivity on parameters (more consistency than SampEn and than a fuzzy approach) and more stability for short signals.[33] Moreover, it has been reported that the stability, consistency, and classification performance is not much influenced by changes in m values.[129] Even for a data length of 50 samples, DistEn has proved to give acceptable results.[33] Furthermore, DistEn does not depend on the variance of the signal and does not change after an amplitude-rescaling procedure.[33] On short-term RR interval time series, it has been shown that DistEn overperforms SampEn and a fuzzy approach.[33,130,131] However, it should be noted that DistEn is difficult to compare straightly to ApEn/SampEn and to the other CE measures reviewed in this chapter, because it explores a different concept of complexity: while CE measures quantify an entropy rate intended as the information contained in the present state of the system but not in its past m states, DistEn quantifies the entropy of the embedding vectors intended as the information contained in the dissimilarities among observations of m system states.

As mentioned by Li *et al.*,[33] the influence of sampling frequency should now be studied thoroughly. As $\mathbf{D}^m$ is symmetrical, the estimation of ePDF can be performed by using only the upper or lower triangular matrix. This property can be used to reduce the time of calculation. However, computation time has to be analyzed more deeply and compared with other complexity measures. The distribution entropy measure to assess the complexity of complex-valued signal has also been proposed.[132] Some authors have suggested the elimination of vector distances apart more than 10 samples in $\mathbf{D}^m$. This procedure drastically reduces the computational cost of the calculation, with acceptable lost of precision in the estimations of the

entropy value.[133] Recently, by using the concepts of DistEn and FuzzyEn, modified versions of DistEn have been proposed for the detection of epileptic seizures[134,135] or for rolling bearing.[136] Furthermore, the permuted DistEn has recently been proposed to detect the changes of irregularity caused by the fluctuation of the complex time series.[137] Moreover, a joint DistEn has also been published.[138] Based on DistEn, Wang and Shang recently proposed the cumulative residual DistEn to study complexity of time series and the cumulative residual DistEn model extended to the fractional order.[139] The multiscale Rényi DistEn has also been proposed and applied to study financial time series.[140] Other investigations in this way deserve future attention.

3.9. *Original and modified permutation entropy*

Permutation entropy (PermEn) is based on the comparison of neighboring values[32] within patterns in a time series. PermEn has the advantage of being very simple and is very fast in terms of computation time. For the time series $\mathbf{x} = \{x_1, \ldots, x_N\}$ the algorithm to compute PermEn is the following:[32]

(1) compute the vectors $\mathbf{x}_i^m(L)$ defined in Sec. 3.1.
(2) the elements of these vectors are associated with numbers from 1 to m and then arranged in an increasing order. There are $m!$ possible patterns π (also called permutations) for an m-tuple vector.
(3) let $f(\pi)$ be the frequency of the permutation π and let $p(\pi)$ be its relative frequency. The permutation entropy is computed as

$$PermEn(m, L) = -\sum_{i=1}^{m!} p(\pi) \ln p(\pi). \tag{67}$$

According to this algorithm, PermEn is an estimate of the dynamic entropy defined in Eq. (7) when the dynamics are represented using ordinal patterns encompassing m system states; an extension to measuring the CE as the difference between the entropies of ordinal patterns of length $m+1$ and m has been proposed recently.[41,141] In either case, it is important to stress that the permutation patterns are different from the patterns created with the other approaches presented in this chapter. Instead of defining patterns according to the magnitude of each sample, permutation patterns account only for the rank of samples within the patterns. Thus, the magnitude of changes is not taken into account, only the ordering of the values. While using ranks simplifies the dynamics favoring entropy computation for short data sequences and helps in dealing with nonstationarity effects, this is also the drawback of PermEn: no information besides the order structure is used when extracting the ordinal patterns for the time series.[142]

For practical purposes, Bandt and Pompe suggested to work with $3 \leq m \leq 7$.[32] Myers and Khasawneh proposed a way to automatically choose the parameters m

and L.[143] Wang *et al.* proposed a parameter optimization strategy for multiscale PermEn (selection of the embedding dimension and time delay).[144]

Variants for PermEn have been introduced. Some of them rely on the measure itself (as detailed below) while others have proposed to modify the multiscale step (see Refs.[145–154]). Some measures have been applied to biomedical time series, while others have not yet but could lead to interesting results for such data. We mention, below, the main variants of PermEn that have been proposed by now.

Fadlallah *et al.* proposed the weighted-permutation entropy to overcome the limitation of permutation that disregards the information contained in the amplitude values.[155] The method assigns weights for each extracted vector when computing the relative frequencies associated with every motif. However, this method is only dependent on the variance (the variance or energy of each neighbor vector is used to compute the weights), and its importance in the metric is always kept constant. Other authors proposed the use of weights to compute the weighted multiscale permutation entropy.[156,157]

In 2016, Azami and Escudero proposed the amplitude-aware permutation entropy.[158] This measure is sensitive to the changes in the amplitude of the signal, in addition to its frequency. Moreover, this new measure deals with samples that have equal amplitude values. Other studies focused on the drawback of possible ambiguities due to equal values in time series for PermEn.[142,159–163]

For the biomedical field, Bandt proposed – in 2017 – a new version of PermEn that he interpreted as distance to white noise.[164] In the latter work, distance to white noise was used as a parameter to measure depth of sleep.[164] This new permutation entropy version is supported by a statistical model which allows to compute significance, and shows how the parameters can be optimized.[164] For the biomedical field too (to identify typical ECG RR interval series, among others), some authors proposed the permutation ratio entropy.[165,166]

Zhang *et al.* proposed permutation entropy based on Hill's diversity number to study financial time series.[167] For the financial markets, the fractional order generalization of information entropy has been used and led to the weighted fractional permutation entropy and fractional sample entropy.[168] Weighted permutation entropy based on different symbolic approaches was also proposed to study financial time series.[169]

In 2018, Chen *et al.* modified the permutation entropy and weighted permutation entropy definitions replacing the Shannon entropy by the Rényi entropy.[170] In the same way, Cánovas *et al.* studied several entropy functions in the permutation entropy algorithm to assess what entropy measure is the best to reveal structural changes in time series.[171] Bai *et al.* proposed a combination of permutation and Lempel-Ziv complexity to analyze electroencephalogram data.[172] Liu *et al.* proposed the multifractal weighted permutation entropy analysis using the estimation of Rényi entropy of the system.[173]

In 2018 also, Tao *et al.* proposed the construction of a visualization scheme for

permutation entropy from a non-uniform time series embedding.[174] This scheme matches permutation entropy with the topological characteristics of the signal under study that are the optimal embedding dimension and set of time delays. Tylová *et al.* proposed the use of a hash table as a special data structure to calculate permutation entropy for window lengths up to 30.[175]

In 2019, Chen *et al.* proposed the improved permutation entropy that, as above, takes into account the amplitude information of the time series and assigns the same symbol to equalities.[176] The authors mention that, compared with permutation entropy, weighted permutation entropy, and amplitude-aware permutation entropy, their improved permutation entropy shows better performance when processing signals with numerous repeated values and can better characterize signals with diverse predictability.[176] Moreover, it gives higher recognition rate for classifying ships under noisy conditions.[176]

Li and Shang proposed, in 2019 also, the multiscale Tsallis permutation entropy to better assess the hidden temporal correlations in signals.[177]

During the same year, Watt and Politi revisited the definition of permutation entropy by introducing a second window-length to allow controlling the partition size.[178] This second window implicitly defines the resolution of the underlying partition. The authors mention that this modification in the definition of permutation entropy increases the flexibility of ordinal-pattern analysis of generic time-series.[178]

Yan *et al.* proposed the network permutation entropy where the connections between different vectors are introduced and the network of the time series is built using the Bandt-Pompe patterns and their weights.[179] Gao *et al.* proposed the multiscale permutation transfer entropy to analyze the coupling properties between the motor and sensory areas during a three-level hand grip task.[180]

In 2020, in the framework of the permutation analysis, Zhao *et al.* proposed the permutation transition entropy to measure the dynamical complexity of signals.[181] Zhang *et al.* proposed the fuzzy permutation entropy which combines multiscale entropy, permutation entropy, and fuzzy entropy.[182]

4. Role of dimension, data length and parameters

Dependencies between some nonlinear metrics are expected due to their shared foundations. For example, it is reasonable to assume that the results of ApEn, SampEn, and FuzzyEn, when applied to the same dataset, will be correlated . Likewise, the results of PermEn and its variants should be consistent when applied to the same signals.

4.1. *Dependencies and relationships*

In particular, some methods share parameters between them. Consider the embedding dimension (m): the length of the pattern (vector) which indicates how many samples are included in each vector. Most entropy estimators include this notion

of embedding dimension into their formulation. Longer m allows more detailed reconstruction of the dynamic process. However, a large value of m may also be unfavourable because of the need for a very large number of sample points and very high computation time. However, the role of m in the estimators varies depending on whether they are based on the concept of CE or not, and this is reflected in the requirements when setting up values for this parameter. For ApEn, SampEn, and FuzzyEn, it is recommended to have at least $10^m - 20^m$ sample points in the analysed signal.[29] For PermEn, DispEn, and FDispEn, it is recommended to have at least $(m+1)!$,[32] $\ln(c^m)$,[27] and $\ln((2c-1)^{m\ \ 1})$[34] sample points, respectively. The latter set of techniques see a pattern as essentially an arrangement and characterized by the order of the elements of which it is made, rather than by the intrinsic nature of these elements.[183,184] Therefore, it is highly recommended to have at least two elements in a pattern and so we suggest to set $m > 1$ for entropy approaches that count patterns (i.e., PermEn, DispEn, and FDispEn). We also note that a number of so-called "non-uniform embedding" methods have been proposed in the recent past, which employ strategies for the selection of the optimal time-lagged components (i.e., selection of the lag l of variables $X_{n\ \ l}, 1 \leq l \leq M, M > m$, that contribute most to X_n) in order to cover sufficiently the past history of the observed process while keeping small the number m of components effectively selected (e.g., see Ref.[69] and references therein).

The parameter r used in ApEn and SampEn represents a trade-off between the quality of logarithmic likelihood estimates and losing too much detail in the signal. When r is too small, poor conditional probability estimates are achieved. Furthermore, to avoid the effect of noise on data, larger r is recommended. In contrast, for a large r value, too much detailed data information is lost. Therefore, a trade-off between large and small r values is needed.[29,31] Lake *et al.* proposed an approach to optimally select r.[185] However, as it is needed to calculate SampEn for a range of r and pick the value that optimizes an efficiency metric, this may be time-consuming.[186] To alleviate this problem, a method based on the heuristic stochastic model was proposed to automatically determine r.[186] However, this approach still considers a number of r values leading to the computational burden. In the literature, it is common to set the threshold r as a constant (usually between 0.1 to 0.3) multiplied by the standard deviation (SD) of the original signal.[8,29,31] This strategy makes SampEn a scale-invariant measure.[29,185]

Methods derived from SampEn sought to remove the dependency on r. This was, at least partially, the motivation behind DistEn.[33] Likewise, FuzzyEn tries to avoid the hard threshold imposed by the use of a Heaviside function to determine the presence of a match. This is, instead, replaced by a fuzzy membership function (MFs). Potential types of MFs include triangular, trapezoidal, Z-shaped, bell-shaped, Gaussian, constant-Gaussian, and exponential functions.[43] Still, the shape and parameters of the MF still relates to the trade-off represented by r for SampEn and ApEn. A defuzzification approach using a surrogate parameter called 'center of

gravity' (C_r) has recently been proposed to re-enable a fair and direct comparison between different types of entropy estimators in this family. It was suggested setting $C_r = 0.1$ for all the MFs.[43]

The parameter c in DispEn and FDispEn, similarly to the parameter Q in CorCE, balances the quantity of entropy estimates with the loss of signal information. To avoid the impact of noise on signals, a small c is recommended. Nonetheless, for a small value of c, too much detailed data information is lost, leading to poor entropy estimates. In contrast, a large value of c increases the computation time considerably. Thus, a trade-off between large and small c values is needed. The number of classes c in DispEn and FDispEn, as well as the number of quantization levels Q in corCE, is inversely related to the threshold value r used in the ApEn and SampEn algorithms.[34]

According to the previous entropy-based approaches, the time delay (L) is usually set to 1. Nevertheless, if the sampling frequency is considerably larger than the highest frequency component of a time series (i.e., the signal is highly oversampled), the first minimum or zero crossing of the autocorrelation function or MI can be used for the selection of an appropriate time delay.[187] As an alternative, a time series may be downsampled before calculating entropy approaches to adjust its highest frequency component to its Nyquist frequency ($fs/2$).[188]

4.2. *Experimental validation*

In order to explore the dependencies between the results of some entropy estimators empirically, here we consider three different kinds of commonly used benchmarks in biomedical signal analysis: 1) Bonn epilepsy EEG dataset; 2) Fantasia blood pressure dataset; and 3) BIDMC congestive heart failure RR interval data. We then apply exemplars of entropy estimators to the signals in those datasets.

The first dataset selected for this experiment is part 'Bonn epilepsy dataset' curated by R.G. Andrzejak.[4,189] This dataset includes five separate sets (denoted 'A' to 'E') with 100 single-channel EEG signals each. The length of each signal is 23.6s and they were sampled at 173.61Hz, leading to a total length of 4097 samples. The signals were extracted from continues multi-channel EEG acquisitions following a visual inspection for artefacts. The signals were hardware-filtered between 0.5Hz and 85Hz, and the stationarity of the segments was verified.[4] Out of the five sets of signals, we used in our analysis dataset 'A', which corresponds to scalp EEG signals recordings from five healthy subjects with eyes open. We focus on this set as it contains activity typically representative of broadband signals with a dominant rhythm (α band in this case) and without the dominant nonlinear discharges that characterise ictal signals. For further information about these datasets, please, see.[4,189]

The second dataset is Fantasia (publicly-available at[190,191]). The dataset includes 10 young (21-34 years old) and 10 old (68-85 years old) rigorously-screened healthy individuals who underwent about 120 minutes of continuous supine resting

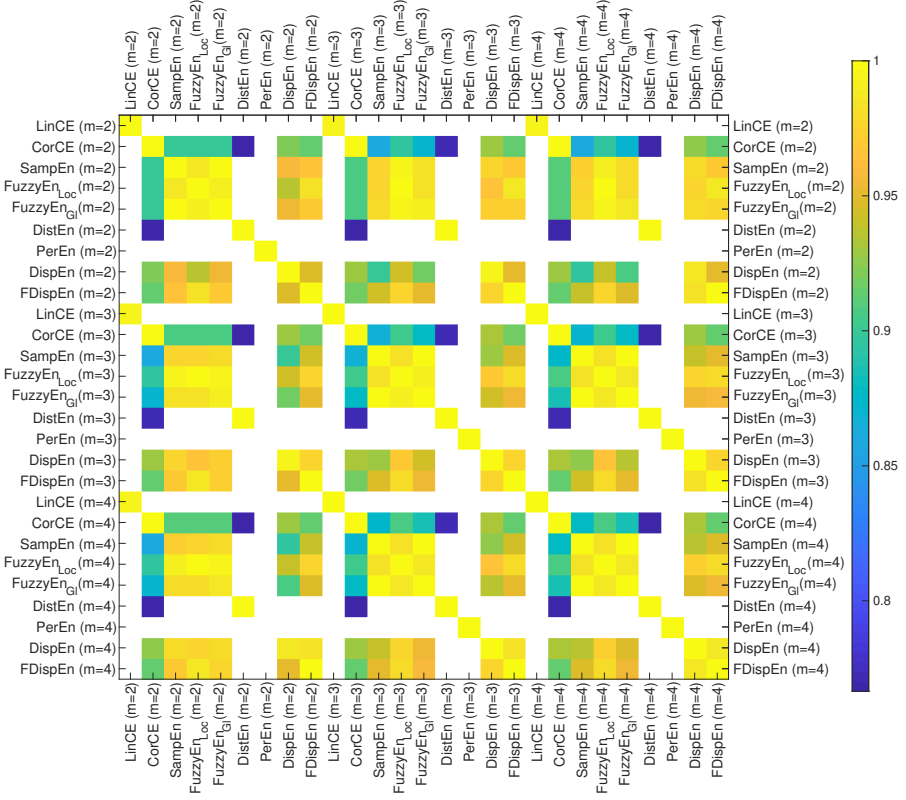
while uncalibrated non-invasive blood pressure signals were recorded. Each group consisted of 5 women and 5 men.[192] All 20 individuals remained in an inactive state in sinus rhythm when watching the movie Fantasia (Disney, 1940) to help to maintain wakefulness. For each subject, the time series were digitized at 250 Hz. Detailed information can be found in.[192]

The third dataset is BIDMC congestive heart failure (publicly-available at[190,191]). This database includes long-term ECG recordings from 15 individuals (11 men, aged 22 to 71, and 4 women, aged 54 to 63) with severe congestive heart failure (NYHA class 3–4). The recordings are about 20 hours in duration for each person (we used the first 1 hour data for each subject) and sampled at 250 samples per second. We analyzed RR interval time series extracted from the ECG data.

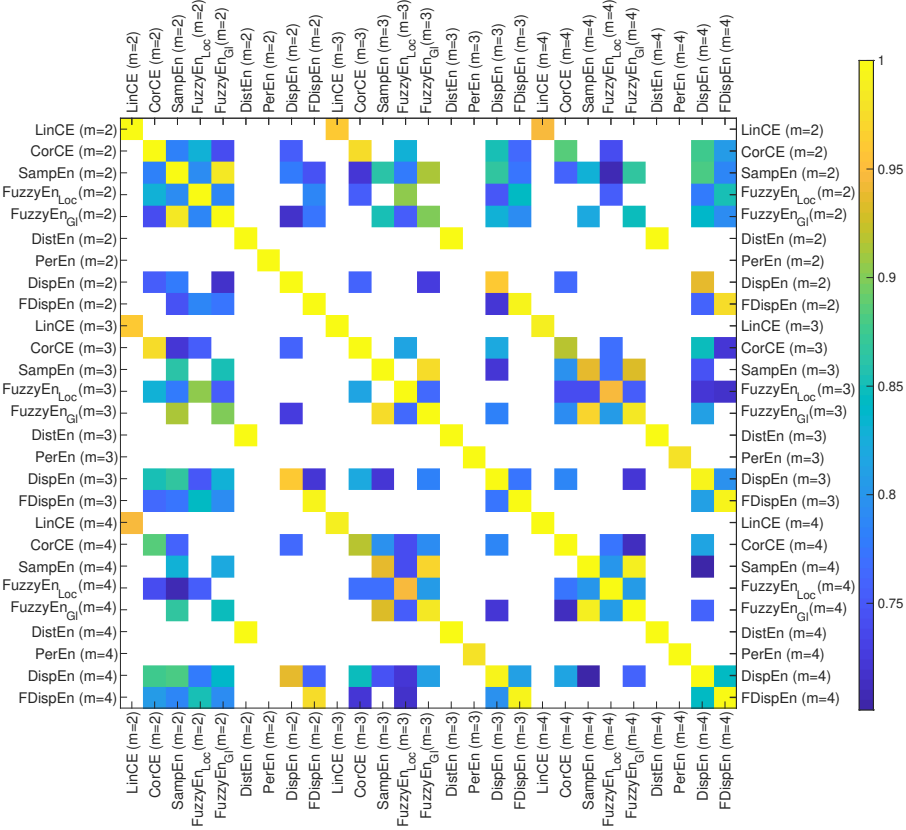
For each dataset, linear conditional entropy (LinCE), corrected CE (CorCE), SampEn, $\text{FuzzyEn}_{\text{Loc}}$ (FuzzyEn based on local characteristics[43]), $\text{FuzzyEn}_{\text{Gl}}$ (FuzzyEn based on global characteristics[43]), DistEn, PermEn, DispEn, and FDispEn values for $m = 2, 3, 4$ in addition to KL entropy with $k = 1$ were calculated. To measure the strength of the relationship between the values obtained from different entropy methods, the correlation coefficients of these entropy values were obtained (see Fig. 2). Only the coefficients greater than 0.7 (strong positive relationship) are shown. The threshold r for SampEn was 0.2 times the SD of the signal.[29] The number of quantization levels for CorCE was $Q = 6$.[30] C_r for FuzzyEn with the Gaussian membership function was 0.1253 times the standard deviation of the signal.[43] The number of classes c for DispEn and FDispEn was set 6.[34] The histogram's bin size was 512 for DistEn.[33] The time delay was equal to 1 for all the methods.

Overall, the results for the three different kinds of biomedical data (i.e., EEGs, RR interval data, and blood pressure) are consistent. For each entropy estimator other than PermEn, the entropy values obtained with different values of m are strongly correlated. Across techniques, the results also show that PermEn, and DistEn values are not strongly correlated with the other entropy methods, agreeing with the fact that the theoretical foundations of these methods are relatively distinct to the others, as described in Sec. 3. The values based on CorCE, SampEn, $\text{FuzzyEn}_{\text{Loc}}$, and $\text{FuzzyEn}_{\text{Gl}}$ are strongly correlated, in agreement with the fact that their algorithms are based on the same principles. Similarly, DispEn and FDispEn values are strongly correlated for different m values. Interestingly, the DispEn and FDispEn values are correlated with those for SampEn, $\text{FuzzyEn}_{\text{Loc}}$, and $\text{FuzzyEn}_{\text{Gl}}$.

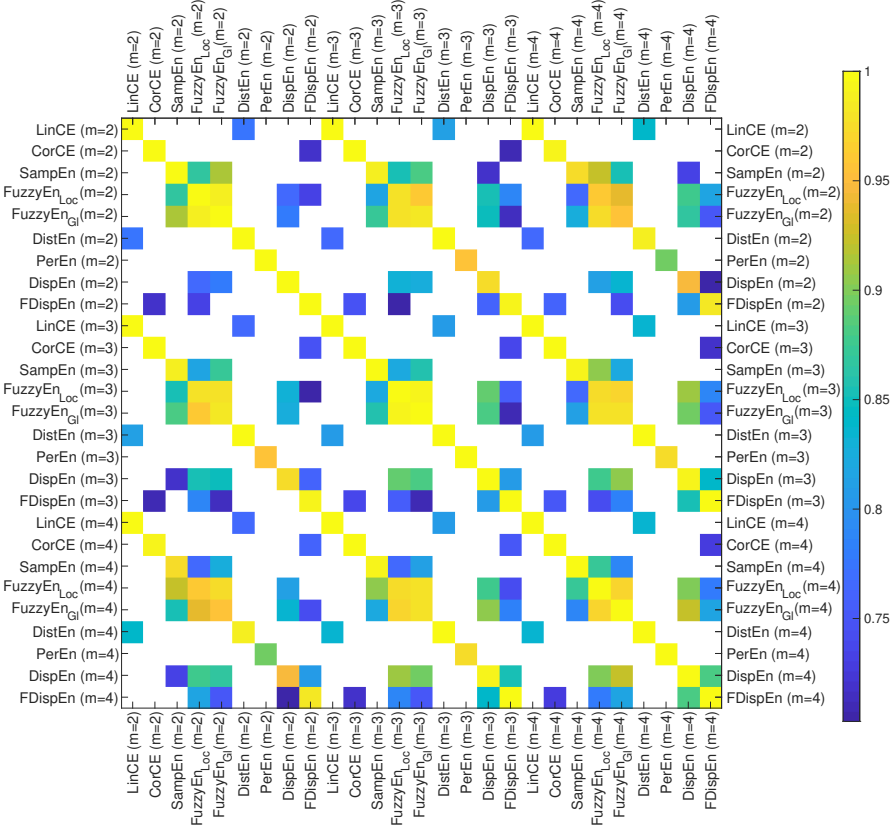
Occasionally, previous research had also studied the dependencies, and correlations or complementarity, in the outputs of nonlinear signal descriptors applied to the same dataset. An example of highly correlated results is the work by Abásolo *et al.* that showed a high level of correlation between ApEn and the rate of decrease of the automutual information when applied to EEG activity.[193] Porta *et al.* compared linear estimates of the CE with a range of nonlinear CE estimates (including binning estimates, ApEn and SampEn, and nearest neighbor estimates) in recordings



(a) Bonn epilepsy EEG data.



(b) Fantasia blood pressure data.



(c) BIDMC congestive heart failure RR interval data.

Fig. 2.: Correlation coefficients obtained from LinCE, CorCE, SampEn, $\text{FuzzyEn}_{\text{Loc}}$, $\text{FuzzyEn}_{\text{Gl}}$, DistrEn, PermEn, DispEn, and FDispEn with $m = 2, 3, 4$ for (a) Bonn epilepsy EEG data, (b) Fantasia blood pressure data, and (c) BIDMC congestive heart failure RR interval data. Only the coefficients greater than 0.7 are shown.

of heart rate variability measured during postural stress, reporting high correlation between the linear model-based approach and the nonlinear model-free methods.[41] García *et al.* reviewed the performance of diverse nonlinear estimators when applied to electroencephalographic recordings for emotion recognition.[194] Their work showed that categories of nonlinear analyses based on different principles may lead to slightly different results when applied to the same problems.[194] For this reason, it is essential to bear in mind the theoretical foundations of the range of nonlinear analysis available when selecting what analysis tools will be selected for a particular problem. Overall, divergences in the results computed with different nonlinear analysis techniques (if chosen appropriately) should not considered as detrimental to the work but, instead, as potentially complementary in providing alternative view-

points of the same phenomenon. Likewise, wisely chosen linear parameters could complement the characterisation of signals achieved with nonlinear methods.[195]

5. Computational efficiency

It is important to consider the computational cost and efficiency of the entropy estimators when used to analyse large datasets and/or long signals. The computational time of some of the techniques covered in this chapter varies widely. Entropy measures such as those covered in Sec. 3.2 quantise the time series and they tend to be fast. The parametric methods described in Sec. 3.7 rely on estimating linear regressions in the time series, something that results in high efficiency as well. In contrast, the algorithms of other entropy metrics such as ApEn, SampEn, FuzzyEn, DistEn, PermEn, DispEn, and FDispEn require the comparison of patterns along the signal. Depending on their approach and implementation, this can lead to very substantial differences in their computational time.

Metrics derived from ApEn need to scan the time series for patterns of length m and $m+1$ and then either to identify when a match is found (ApEn and SampEn)[29] or to record the distances between them (FuzzyEn and DistEn).[33,43] This results in implementations that depend quadratically on the length of the signal: their computational time is $O(N^2)$. Some other parameters, such as the length of the patterns (m) or the precise definition of the fuzzy function in FuzzyEn (n)[43] play a smaller role in the computational complexity of the algorithms, although in certain computing platforms – such as Matlab – the efficiency of the exponential functions may depend on whether the exponent is interger or not.

In contrast, methods such as PermEn, DispEn, and FDispEn simply evaluate the frequency of patterns of symbols derived from m-tuples of samples taken along the signal.[27,32] This means that the time series has to be scanned only once and, as a result, the computational time depends linearly on the number of samples: $O(N)$.

In order to show empirically the dependency of the computational time on the length of the signals, Table 1 displays the computational time of LinCE, CorCE, SampEn, FuzzyEn, DistEn, PermEn, DispEn, and FDispEn when applied to white Gaussian noise sequences of varying length. All the simulations in this article have been carried out using a PC with Intel (R) Xeon (R) CPU, E5420, 2.5 GHz and 8-GB RAM by MATLAB R2019a. The embedding dimension values change from 2 to 4 for all the methods.

The results demonstrate the different dependency of the entropy estimators on the data length. Methods with an $O(N^2)$ computational cost (i.e., SampEn and FuzzyEn) show steeper increases in time with data length. In addition, the results in Table 1 also allow us to explore the effects of varying values of pattern length (m). Overall, the computation times of LinCE, SampEn, FuzzyEn, and DistEn with various values of m are similar, while the effect of m is stronger for CorCE, PermEn, DispEn, and FDispEn. The results also confirm that SampEn is faster than FuzzyEn due to its simpler use of a Heaviside function to determine matches

Table 1.: Computation time of LinCE, CorCE, SampEn, FuzzyEn, DistEn, PermEn, DispEn, and FDispEn with $m = 2, 3, 4$ for white Gaussian noise with different lengths (300, 1000, 3,000, 10,000, and 30,000 sample points).

Number of samples →	300	1,000	3,000	10,000	30,000
LinCE ($m = 2$)	3.7278e-05 s	2.9181e-05 s	2.9029e-05 s	2.8987e-05 s	2.8714e-05 s
LinCE ($m = 3$)	3.1635e-05 s	2.878e-05 s	2.9121e-05 s	2.8755e-05 s	2.9218e-05 s
LinCE ($m = 4$)	3.0366e-05 s	2.9708e-05 s	2.8874e-05 s	2.8442e-05 s	2.8578e-05 s
CorCE ($m = 2$)	0.0033 s	0.0028 s	0.0038 s	0.0102 s	0.0196 s
CorCE ($m = 3$)	0.0128 s	0.0164 s	0.0216 s	0.0447 s	0.1023 s
CorCE ($m = 4$)	0.0192 s	0.0511 s	0.1113 s	0.2932 s	0.6412 s
SampEn ($m = 2$)	0.0004 s	0.0033 s	0.0308 s	0.3495 s	2.9524 s
SampEn ($m = 3$)	0.0004 s	0.0033 s	0.0311 s	0.3527 s	2.9937 s
SampEn ($m = 4$)	0.0004 s	0.0033 s	0.0301 s	0.3525 s	2.9656 s
FuzzyEn ($m = 2$)	0.0006 s	0.0031 s	0.0378 s	0.4595 s	3.8613 s
FuzzyEn ($m = 3$)	0.0006 s	0.0032 s	0.0378 s	0.4742 s	3.7652 s
FuzzyEn ($m = 4$)	0.0006 s	0.0032 s	0.0385 s	0.4667 s	3.9627 s
DistEn ($m = 2$)	0.0011 s	0.0074 s	0.0674 s	0.7556 s	6.7645 s
DistEn ($m = 3$)	0.0010 s	0.0074 s	0.0680 s	0.7388 s	6.7904 s
DistEn ($m = 4$)	0.0011 s	0.0075 s	0.0673 s	0.7592 s	6.8346 s
PermEn ($m = 2$)	0.0007 s	0.0018 s	0.0053 s	0.0167 s	0.0519 s
PermEn ($m = 3$)	0.0010 s	0.0029 s	0.0082 s	0.0273 s	0.0831 s
PermEn ($m = 4$)	0.0027 s	0.0079 s	0.0225 s	0.0740 s	0.2281 s
DispEn ($m = 2$)	0.0002 s	0.0003 s	0.0005 s	0.0014 s	0.0034 s
DispEn ($m = 3$)	0.0005 s	0.0007 s	0.0014 s	0.0041 s	0.0104 s
DispEn ($m = 4$)	0.0021 s	0.0029 s	0.0057 s	0.0188 s	0.0507 s
FDispEn ($m = 2$)	0.0002 s	0.0002 s	0.0005 s	0.0012 s	0.0029 s
FDispEn ($m = 3$)	0.0004 s	0.0006 s	0.0011 s	0.0031 s	0.0071 s
FDispEn ($m = 4$)	0.0021 s	0.0032 s	0.0079 s	0.0192 s	0.0457 s

between patterns.

We can also see that, while the differences in computational time are negligible for short signal segments, they become more severe for longer signals. Overall, LinCE, DispEn and FDispEn tend to be the faster algorithms in this comparison, followed by PermEn. This agrees with the computational costs stated above and with the fact that these methods does not need to neither sort the amplitude values of each embedded vector (like PermEn) nor calculate every distance between any two composite delay vectors with embedding dimensions m and $m+1$ (like SampEn and FuzzyEn). This makes these noticeably faster than PermEn, SampEn, and FuzzyEn.

It is important to note these results are based on straightforward implementations of the algorithms but there has been research to speed up the computation of the entropy estimators by developing more efficient algorithms. For example, Jiang *et al.* proposed a tree structure to compute SampEn.[196] Manis *et al.* speed up the algorithm for SampEn even further by devising algorithms that avoid unnecessary pattern comparisons.[197] They demonstrated speeding up factors of between 4 and 10 in comparison with the straightforward implementation, depending on the values of m and r. One must be aware that the optimal implementation of any method

may also depend on the programming language. The same algorithm may be very efficient in one but not in other languages, due to the correct usage of the most efficient data structures offered by the different programming languages.

Finally, it is worth noting that GPU-based implementations have been proposed for entropy analysis of multivariate signals.[198,199] Given the increasing popularity and relevance of GPU computing, this provides an avenue worth exploring when analysing long signals or large datasets.

6. Limitations and advantages

As mentioned before, ApEn and its improvements, i.e., SampEn and FuzzyEn, as well as corCE, are based on different principles compared to PermEn, DistEn, DipsEn, and FDispEn, meaning that ApEn, SampEn, and FuzzyEn denote the rate of information production (conditional entropy) whereas PermEn, DistEn, DispEn, and FDispEn quantify the total amount of information (Shannon entropy). Nevertheless, the comparison of these methods as different kinds of feature extraction approaches is meaningful and many studies based on one- and two-dimensional entropy methods showed their similar behaviors.

SampEn alleviates some shortcomings of ApEn.[29,200] First, ApEn inherently includes a bias towards regularity or complexity, as it counts a self-match of vectors while SampEn does not count a self-match and so eliminates the bias towards regularity. Second, ApEn lacks relative consistency, as the input parameters are changed, the value of ApEn, unlike SampEn, may "flip". For instance, white noise may have a much smaller ApEn value than a known periodic signal when one parameter of ApEn is set very small. Eventually this will "flip" and the ApEn value will become greater in the white noise time series as the input parameters are changed. Third, the parameters of ApEn should be fixed and comparing data should only be done when the input parameters are the same for both datasets due to the issue of relative consistency and also the overall sensitivity of the algorithm to the parameters of choice and to data length.[200]

Although SampEn is not sensitive to a noise with a low amplitude compared with the original signal, it is either undefined or unreliable for short signals and computationally expensive for real-time applications. SampEn is also sensitive to its parameters, especially to the tolerance factor r. FuzzyEn alleviates the problem of undefined values of SampEn. Additionally, FuzzyEn, compared with ApEn and SampEn, is less sensitive to its parameters and even data length. Nevertheless, the FuzzyEn algorithm is considerably slower than SampEn when dealing with a long time series or a large embedding dimension.

Among the MFs used for FuzzyEn,[43] when dealing with an equal value of the center of gravity, the Gaussian MF, as the fastest algorithm, results in the highest Hedges' g effect size for long signals. FuzzyEn based on exponential MF of order four better distinguishes short white, pink, and brown noises, and yields more significant differences for the short real signals based on Hedges' g effect size. The

triangular, trapezoidal, and Z-shaped MFs are not recommended for short signals as the FuzzyEn values may be undefined. FuzzyEn with Gaussian and exponential MF of order four for respectively characterization of short and long data are suggested.

The DistEn technique, unlike SampEn, does not lead to undefined entropy values. DistEn is also not sensitive to a noise with a low amplitude compared with the original signal. Nevertheless, it has two main limitations. First, since the total number of elements in distance matrix $\mathbf{D}$ is $(N-m)(N-m-1)$, for a long time series the computation of DistEn, compared with DispEn and PerEn, needs the storage of a large number of elements. More importantly, according to the DistEn algorithm, new signals created simply by random permutations of an original time series (shuffling data) have DistEn values close to that for the original time series. For example, if the elements of a signal are sorted, its DistEn value is not changed noticeably. However, as expected theoretically and intuitively, sorting leads to a lower entropy value (less irregularity).

PermEn is computationally fast, thus facilitating its use in real time applications. This approach also can be used for both short and long signals. Nevertheless, it has three main shortcomings since it considers permutation patterns of a signal. First, the original PermEn assumes a signal has a continuous distribution, thus equal values are infrequent and can be disregarded by ranking them based on the order of their emergence. For digitized signals with coarse quantization levels, yet, it may be imprecise to simply disregard them.[201] Second, when a time series is symbolized based on the permutation patterns (Bandt-Pompe procedure), only the order of amplitudes is considered and some information with regard to the amplitude values may be ignored.[34,201] Third, PermEn is sensitive to noise (even when the SNR of a data is high), because a small change in amplitude value may vary the order relations among amplitudes.[34]

DispEn and FDispEn, which are based on Shannon entropy, are computationally fast. These methods, ApEn, and SampEn have similar behavior when dealing with noise. In ApEn and SampEn, only the number of matches whose differences are smaller than a defined threshold is counted. Accordingly, a small change in the time series amplitude due to noise is unlikely to change ApEn or SampEn values. Similarly, in DispEn and FDispEn, a small change will probably not alter the index of class and so the entropy value will not change. Thus, ApEn, SampEn, DispEn, and FDispEn are relatively robust to noise (especially for signals with high SNR). DispEn and FDispEn, compared with ApEn, SampEn, FuzzyEn, and DistEn, needs to store a considerably smaller number of elements. For short signals, DispEn and FDispEn also do not result in undefined values. Nevertheless, DispEn and FDispEn are sensitive to signal length especially for short time series and high m or c values. Additionally, DispEn and FDispEn, like PermEn, are based on symbolic dynamics or patterns originated from a coarse-graining of the measurements, that is, the data are transformed into a new signal with only a few different elements. Therefore,

the study of the dynamics of time series is simplified to a distribution of symbol sequences. Although some of the invariant, robust properties of the dynamics may be kept, some of detailed information may be lost.[202–204]

7. Conclusions, and future directions

As nonlinearity and complexity are ubiquitous in living systems, there is little room for doubt that complex nonlinear dynamical systems better describe physiological regulations. Thus, there is an increasing interest in nonlinear approaches to characterize physiological signals generated by such physiological regulations. These signals may be exploited to detect physiological states, to monitor the health conditions over time, or to predict pathological events. One of the most popular and powerful nonlinear approaches used to assess the dynamical characteristics of time series is entropy.

In this chapter, we explained the basic concepts of probability and information theory used to define entropy metrics for biomedical signal processing. In addition to the nearest neighbors and parametric approaches, approximate, sample, fuzzy, permutation, distribution, dispersion entropies were detailed based on Shannon entropy, conditional entropy, and information storage. The entropy methods were then systematically compared from different theoretical, computational and practical views and their advantages and disadvantages were explained. We discussed how to set the parameters used in these entropy methods and their relationships. We also evaluated the dependencies between some nonlinear entropy measures due to their shared foundations. In order to explore the dependencies between the results of some entropy estimators empirically, we used three different kinds of biomedical times series (i.e., EEGs, RR interval data, and blood pressures).

The study done here has also the following implications for complexity or irregularity estimations. First, PermEn and DistEn values are not strongly correlated with the other entropy methods. The results of SampEn, and FuzzyEn, like those for DispEn and FDispEn, when applied to the same dataset, were strongly correlated. It was found that LinCE, FDispEn and DispEn are the fastest algorithms for white Gaussian noise with different lengths, followed by PermEn.

In spite of a large number of interesting studies about entropy-based approaches in the literature, there are still some challenges open to future investigation.

Firstly, although there are some suggestions about how to set the parameters used for the entropy methods (e.g., Refs[30,43,185,186]), there is still room to provide new insights and propose new techniques to find appropriate set of parameters for various data and applications. This is particularly important to avoid p-hacking. Even though the results in Sec. 5 are reassuring regarding the consistency of the results estimated with diverse entropy methods and parameters, one should define the parameters of the estimators *a priori* using the best guidance available and report clearly any process used to fine-tune the parameters.

Secondly, most biomedical signals, such as EEGs, HRVs, EMGs, and MEGs,

are usually non-stationary. This important issue needs to be taken into account before applying entropy measures that are strictly speaking only applicable to stationary time series. To this end, a biomedical signal is typically divided into short quasi-stationary segments and then the entropy of each segment is calculated. Nevertheless, there is a real need to propose entropy-based approaches to analyze non-stationary time series (e.g., any trends before applying an entropy method can be removed).

Thirdly, establishing the relationships between entropy estimators and other nonlinear methods, such as Lempel-Ziv complexity and fractal dimension, from empirical and theoretical perspectives is another potential feature of interest for future studies. For example, initial work has explored these relationships to understand the dependencies between complexity estimators and the synchrony of mean field models with simple oscillators coupled through a network.[205] Overall, the study of the interplay between complex networks, nonlinear analysis, and computational models can help us to understand the rules behind complex phenomena, and to monitor them.[206]

Finally, and in relation to the previous point and as Sec. 2.3 indicated, the hallmark of complex systems is the very large number of nonlinear interdependences among the elements that compose them.[59] This demands multivariate approaches to compute entropy measures from more than one signal. Overall, these approaches can be divided into methods focusing on the computation of directed information measures between two or more stochastic processes[55,60,69,207] and the simultaneous estimation of entropy from more than one signal component.[61–64] Both groups of techniques have already been used in the characterisation of biomedical recordings, and we expect the relevance of this area to increase in the coming years.

Overall, entropy-based metrics are now recognised as practical alternative to classical nonlinear analysis methods to study the dynamics of various kinds of systems, including biomedical signals. These approaches enable the evaluation of the degree of irregularity and complexity of such systems. The evidence gathered in this review shows the relevance of these approaches in multiple biomedical applications and their theoretical foundations and several promising areas for future research. We expect that entropy analysis will become an even more prominent field within biomedical signal analysis in the near future.

Acknowledgments

J.E. kindly acknowledges support by the Leverhulme Trust via a Research Project Grant (RPG-2020-158).
L.F. is supported by the Italian MIUR PRIN 2017 project, PRJ-0167, “Stochastic forecasting in complex systems”.

References

1. C. J. Stam, Nonlinear dynamical analysis of EEG and MEG: review of an emerging field, *Clinical Neurophysiology*. **116**(10), 2266–2301, (2005).
2. S. M. Pincus and A. L. Goldberger, Physiological time-series analysis: what does regularity quantify?, *Am J Physiol Heart Circ Physiol*. **266**(4), H1643–1656, (1994).
3. R. Sleimen-Malkoun, D. Perdikis, V. Müller, J.-L. Blanc, R. Huys, J.-J. Temprado, and V. K. Jirsa, Brain dynamics of aging: multiscale variability of EEG signals at rest and during an auditory oddball task, *Eneuro*. **2**(3), ENEURO–0067, (2015).
4. R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, *Phys. Rev. E*. **64**, 061907 (Nov, 2001). doi: 10.1103/PhysRevE.64.061907. URL `https://link.aps.org/doi/10.1103/PhysRevE.64.061907`.
5. D. Hoyer, U. Leder, H. Hoyer, B. Pompe, M. Sommer, and U. Zwiener, Mutual information and phase dependencies: measures of reduced nonlinear cardiorespiratory interactions after myocardial infarction, *Medical Engineering & Physics*. **24**(1), 33–43, (2002).
6. M. Palacios, H. Friedrich, C. Götze, M. Vallverdú, A. B. de Luna, P. Caminal, and D. Hoyer, Changes of autonomic information flow due to idiopathic dilated cardiomyopathy, *Physiological Measurement*. **28**(6), 677, (2007).
7. L. E. V. Silva, R. M. Lataro, J. A. Castania, C. A. A. Silva, H. C. Salgado, R. Fazan, and A. Porta, Nonlinearities of heart rate variability in animal models of impaired cardiac control: contribution of different time scales, *Journal of Applied Physiology*. **123**(2), 344–351, (2017). doi: 10.1152/japplphysiol.00059.2017.
8. M. Costa, A. L. Goldberger, and C.-K. Peng, Multiscale entropy analysis of complex physiologic time series, *Phys. Rev. Lett.* **89**(6), 068102, (2002).
9. A. L. Goldberger, C.-K. Peng, and L. A. Lipsitz, What is physiologic complexity and how does it change with aging and disease?, *Neurobiology of Aging*. **23**(1), 23–26, (2002).
10. A. Porta, L. Faes, V. Bari, A. Marchi, T. Bassani, G. Nollo, N. M. Perseguini, J. Milan, V. Minatel, A. Borghi-Silva, et al., Effect of age on complexity and causality of the cardiovascular control: comparison between model-based and model-free approaches, *PloS one*. **9**(2), e89463, (2014).
11. R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt, A comparison of waveform fractal dimension algorithms, *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*. **48**(2), 177–183, (2001).
12. A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, Determining Lyapunov exponents from a time series, *Physica D: Nonlinear Phenomena*. **16**(3), 285–317, (1985).
13. A. Lempel and J. Ziv, On the complexity of finite sequences, *IEEE Transactions on Information Theory*. **22**(1), 75–81, (1976).
14. C. E. Shannon, Communication theory of secrecy systems, *Bell Labs Technical Journal*. **28**(4), 656–715, (1949).
15. J. J. Heisz and A. R. McIntosh, Applications of EEG neuroimaging data: Event-related potentials, spectral power, and multiscale entropy, *Journal of visualized experiments: JoVE*. (76), (2013).
16. W. Xiong, L. Faes, and P. C. Ivanov, Entropy measures, entropy estimators, and their performance in quantifying complex dynamics: Effects of artifacts, nonstationarity, and long-range correlations, *Physical Review E*. **95**(6), 062114, (2017).

17. M. Rostaghi, M. R. Ashory, and H. Azami, Application of dispersion entropy to status characterization of rotary machines, *Journal of Sound and Vibration*. **438**, 291–308, (2019).
18. S. Pincus, Approximate entropy as an irregularity measure for financial data, *Econometric Reviews*. **27**(4-6), 329–362, (2008).
19. L. Efremidze, D. J. Stanley, and M. D. Kinsman, Stock market timing with entropy, *The Journal of Wealth Management*. **18**(3), 57–67, (2015).
20. L. Shuangcheng, Z. Qiaofu, W. Shaohong, and D. Erfu, Measurement of climate complexity using sample entropy, *International Journal of Climatology: A Journal of the Royal Meteorological Society*. **26**(15), 2131–2139, (2006).
21. G. Balasis, R. V. Donner, S. M. Potirakis, J. Runge, C. Papadimitriou, I. A. Daglis, K. Eftaxias, and J. Kurths, Statistical mechanics and information-theoretic perspectives on complexity in the earth system, *Entropy*. **15**(11), 4844–4888, (2013).
22.
23. S. Dasgupta, F. Wörgötter, and P. Manoonpong, Information dynamics based self-adaptive reservoir for delay temporal memory tasks, *Evolving Systems*. **4**(4), 235–249, (2013).
24. Y. Chen and T. D. Pham, Sample entropy and regularity dimension in complexity analysis of cortical surface structure in early alzheimer's disease and aging, *Journal of neuroscience methods*. **215**(2), 210–217, (2013).
25. P. Varotsos, N. Sarlis, E. Skordas, H. Tanaka, and M. Lazaridou, Entropy of seismic electric signals: Analysis in natural time under time reversal, *Physical Review E*. **73** (3), 031114, (2006).
26. L. Faes, A. Porta, and G. Nollo, Information decomposition in bivariate systems: theory and application to cardiorespiratory dynamics, *Entropy*. **17**(1), 277–303, (2015).
27. M. Rostaghi and H. Azami, Dispersion entropy: A measure for time-series analysis, *IEEE Signal Processing Letters*. **23**(5), 610–614, (2016).
28. S. M. Pincus, Approximate entropy as a measure of system complexity, *Proceedings of the National Academy of Sciences of the United States of America*. **88**(6), 2297–2301, (1991).
29. J. S. Richman and J. R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, *Am J Physiol Heart Circ Physiol*. **278**(6), H2039–2049, (2000).
30. A. Porta, G. Baselli, D. Liberati, N. Montano, C. Cogliati, T. Gnecchi-Ruscone, A. Malliani, and S. Cerutti, Measuring regularity by means of a corrected conditional entropy in sympathetic outflow, *Biological cybernetics*. **78**(1), 71–78, (1998).
31. W. Chen, Z. Wang, H. Xie, and W. Yu, Characterization of surface emg signal based on fuzzy entropy, *IEEE Transactions on neural systems and rehabilitation engineering*. **15**(2), 266–272, (2007).
32. C. Bandt and B. Pompe, Permutation entropy: a natural complexity measure for time series, *Phys. Rev. Lett.* **88**(17), 174102, (2002).
33. P. Li, C. Liu, K. Li, D. Zheng, C. Liu, and Y. Hou, Assessing the complexity of short-term heartbeat interval series by distribution entropy, *Med. Biol. Eng. Comput.* **53**, 77–87, (2015).
34. H. Azami and J. Escudero, Amplitude-and fluctuation-based dispersion entropy, *Entropy*. **20**(3), 210, (2018).
35. *Dimensions*, (2020 (accessed 28 September2020)). `https://app.dimensions.ai/discover/publication`.
36. V. Srinivasan, C. Eswaran, and N. Sriraam, Approximate entropy-based epileptic eeg detection using artificial neural networks, *IEEE Transactions on information*

Technology in Biomedicine. **11**(3), 288–295, (2007).
37. R. Hornero, D. Abásolo, J. Escudero, and C. Gómez, Nonlinear analysis of electroencephalogram and magnetoencephalogram recordings in patients with alzheimer's disease, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. **367**(1887), 317–336, (2009).
38. H. Gonçalves, T. Henriques-Coelho, J. Bernardes, A. P. Rocha, A. Nogueira, and A. Leite-Moreira, Linear and nonlinear heart-rate analysis in a rat model of acute anoxia, *Physiological Measurement*. **29**(9), 1133, (2008).
39. M. M. Platisa and V. Gal, Dependence of heart rate variability on heart period in disease and aging, *Physiological Measurement*. **27**(10), 989, (2006).
40. A. Porta, T. Gnecchi-Ruscone, E. Tobaldini, S. Guzzetti, R. Furlan, and N. Montano, Progressive decrease of heart period variability entropy-based complexity during graded head-up tilt, *Journal of applied physiology*. **103**(4), 1143–1149, (2007).
41. A. Porta, B. De Maria, V. Bari, A. Marchi, and L. Faes, Are nonlinear model-free conditional entropy approaches for the assessment of cardiac control complexity superior to the linear model-based one?, *IEEE Transactions on Biomedical Engineering*. **64**(6), 1287–1296, (2017).
42. A. U. Viola, E. Tobaldini, S. L. Chellappa, K. R. Casali, A. Porta, and N. Montano, Short-term complexity of cardiac autonomic control during sleep: Rem as a potential risk factor for cardiovascular system in aging, *PloS one*. **6**(4), e19002, (2011).
43. H. Azami, P. Li, S. E. Arnold, J. Escudero, and A. Humeau-Heurtier, Fuzzy entropy metrics for the analysis of biomedical signals: assessment and comparison, *IEEE Access*. **7**, 104833–104847, (2019).
44. H.-B. Xie, J.-Y. Guo, and Y.-P. Zheng, Fuzzy approximate entropy analysis of chaotic and natural complex systems: detecting muscle fatigue using electromyography signals, *Annals of biomedical engineering*. **38**(4), 1483–1496, (2010).
45. B. García-Martínez, A. Martínez-Rodrigo, R. Zangróniz Cantabrana, J. M. Pastor Garcia, and R. Alcaraz, Application of entropy-based metrics to identify emotional distress from electroencephalographic recordings, *Entropy*. **18**(6), 221, (2016).
46. X. Li, G. Ouyang, and D. A. Richards, Predictability analysis of absence seizures with permutation entropy, *Epilepsy research*. **77**(1), 70–74, (2007).
47. A. A. Bruzzo, B. Gesierich, M. Santi, C. A. Tassinari, N. Birbaumer, and G. Rubboli, Permutation entropy to detect vigilance changes and preictal states from scalp eeg in epileptic patients. a preliminary study, *Neurological sciences*. **29**(1), 3–9, (2008).
48. X. Li, S. Cui, and L. J. Voss, Using permutation entropy to measure the electroencephalographic effects of sevoflurane, *The Journal of the American Society of Anesthesiologists*. **109**(3), 448–456, (2008).
49. E. Olofsen, J. Sleigh, and A. Dahan, Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect, *British journal of anaesthesia*. **101**(6), 810–821, (2008).
50. D. Jarchi, J. Andreu-Perez, M. Kiani, O. Vysata, J. Kuchynka, A. Prochazka, and S. Sanei, Recognition of patient groups with sleep related disorders using bio-signal processing and deep learning, *Sensors*. **20**(9), 2594, (2020).
51. H. Azami, J. Escudero, A. Fernandez, and S. E. Arnold, Amplitude-and fluctuation-based dispersion entropy for the analysis of resting-state magnetoencephalogram irregularity in MCI and Alzheimer's disease patients, *Alzheimer's & Dementia*. **15**, 762–762, (2019).
52. A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. (Tata McGraw-Hill Education, 2002).
53. M. Cover Thomas and A. Thomas Joy, Elements of information theory, *New York:*

Wiley. **3**, 37–38, (1991).
54. C. Shannon, A mathematical theory of communication, bell system technical journal 27: 379-423 and 623–656, *Mathematical Reviews (MathSciNet): MR10, 133e*. (1948).
55. M. Wibral, R. Vicente, and J. T. Lizier, *Directed information measures in neuroscience*. (Springer, 2014).
56. A. N. Kolmogorov, A new metric invariant of transient dynamical systems and automorphisms in lebesgue spaces, *Dokl. Akad. Nauk SSSR*. **119**(5), 861–864, (1958).
57. Y. G. Sinai, On the notion of entropy of a dynamical system, *Dokl. Akad. Nauk SSSR*. **124**, 768–771, (1959).
58. M. Cover Thomas and A. Thomas Joy, Elements of information theory, *New York: Wiley*. pp. 74–78, (2006).
59. N. Boccara, *Modeling Complex Systems*. (Springer, 2004).
60. A. Porta and L. Faes, Wiener–granger causality in network physiology with applications to cardiovascular control and neuroscience, *Proceedings of the IEEE*. **104**(2), 282–309, (2015).
61. M. U. Ahmed and D. P. Mandic, Multivariate multiscale entropy: A tool for complexity analysis of multichannel data, *Phys. Rev. E*. **84**, 061918 (Dec, 2011). doi: 10.1103/PhysRevE.84.061918. URL `https://link.aps.org/doi/10.1103/PhysRevE.84.061918`.
62. H. Azami and J. Escudero, Refined composite multivariate generalized multiscale fuzzy entropy: A tool for complexity analysis of multichannel signals, *Physica A: Statistical Mechanics and its Applications*. **465**, 261–276, (2017).
63. Y. Zhang and P. Shang, Multivariate multiscale distribution entropy of financial time series, *Physica A: Statistical Mechanics and its Applications*. **515**, 72–80, (2019).
64. H. Azami, A. Fernández, and J. Escudero, Multivariate multiscale dispersion entropy of biomedical times series, *Entropy*. **21**(9), 913 (Sep, 2019). ISSN 1099-4300. doi: 10.3390/e21090913. URL `http://dx.doi.org/10.3390/e21090913`.
65. A. Goldberger, Giles f. filley lecture. complex systems, *Proc Am Thorac Soc*. **3**(6), 467–71, (2006).
66. Gell-Mann, Murray, Plectics: The study of simplicity and complexity, *Europhysics News*. **33**(1), 17–20, (2002). doi: 10.1051/epn:2002105. URL `https://doi.org/10.1051/epn:2002105`.
67. C. Hsu, S.-Y. Wei, H.-P. Huang, L. Hsu, S. Chi, and C.-K. Peng, Entropy of entropy: Measurement of dynamical complexity for biological systems, *Entropy*. **19**(10), 550 (Oct, 2017). ISSN 1099-4300. doi: 10.3390/e19100550. URL `http://dx.doi.org/10.3390/e19100550`.
68. L. E. V. Silva, B. C. T. Cabella, U. P. da Costa Neves, and L. O. Murta Junior, Multiscale entropy-based methods for heart rate variability complexity analysis, *Physica A: Statistical Mechanics and its Applications*. **422**, 143 – 152, (2015). ISSN 0378-4371. doi: https://doi.org/10.1016/j.physa.2014.12.011. URL `http://www.sciencedirect.com/science/article/pii/S0378437114010462`.
69. L. Faes and A. Porta. Conditional entropy-based evaluation of information dynamics in physiological systems. In *Directed information measures in neuroscience*, pp. 61–86. Springer, (2014).
70. A. M. Fraser and H. L. Swinney, Independent coordinates for strange attractors from mutual information, *Physical review A*. **33**(2), 1134, (1986).
71. G. A. Darbellay, An estimator of the mutual information based on a criterion for conditional independence, *Computational Statistics & Data Analysis*. **32**(1), 1–17, (1999).
72. A. Porta, L. Faes, M. Masé, G. D'addio, G. Pinna, R. Maestri, N. Montano,

R. Furlan, S. Guzzetti, G. Nollo, et al., An integrated approach based on uniform quantization for the evaluation of complexity of short-term heart period variability: application to 24 h holter recordings in healthy and heart failure humans, *Chaos: An Interdisciplinary Journal of Nonlinear Science*. **17**(1), 015117, (2007).

73. H. A. Sturges, The choice of a class interval, *Journal of the American Statistical Association*. **21**(153), 65–66, (1926). doi: 10.1080/01621459.1926.10502161.
74. D. P. Doane, Aesthetic frequency classifications, *The American Statistician*. **30**(4), 181–183, (1976).
75. L. Faes, A. Porta, G. Rossato, A. Adami, D. Tonon, A. Corica, and G. Nollo, Investigating the mechanisms of cardiovascular and cerebrovascular regulation in orthostatic syncope through an information decomposition strategy, *Autonomic Neuroscience*. **178**(1-2), 76–82, (2013).
76. M. Rezaeezadeh, S. Shamekhi, and M. Shamsi, Attention deficit hyperactivity disorder diagnosis using non-linear univariate and multivariate eeg measurements: a preliminary study, *Physical and Engineering Sciences in Medicine*. pp. 1–16, (2020).
77. V. Miskovic, K. J. MacDonald, L. J. Rhodes, and K. A. Cote, Changes in eeg multiscale entropy and power-law frequency scaling during the human sleep cycle, *Human Brain Mapping*. **40**(2), 538–551, (2019). doi: 10.1002/hbm.24393. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24393.
78. H. Azami, E. Kinney-lang, A. Ebied, A. Fernández, and J. Escudero. Multiscale dispersion entropy for the regional analysis of resting-state magnetoencephalogram complexity in alzheimer's disease. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3182–3185, (2017).
79. E. Kafantaris, I. Piper, T.-Y. M. Lo, and J. Escudero. Application of dispersion entropy to healthy and pathological heartbeat ecg segments. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2269–2272. IEEE, (2019).
80. J. J. Nicolet, J. F. Restrepo, and G. Schlotthauer, Classification of intracavitary electrograms in atrial fibrillation using information and complexity measures, *Biomedical Signal Processing and Control*. **57**, 101753, (2020).
81. X. Cheng, P. Wang, and C. She, Biometric identification method for heart sound based on multimodal multiscale dispersion entropy, *Entropy*. **22**(2), (2020). ISSN 1099-4300. doi: 10.3390/e22020238. URL https://www.mdpi.com/1099-4300/22/2/238.
82. H. Azami, M. Rostaghi, D. Abásolo, and J. Escudero, Refined composite multiscale dispersion entropy and its application to biomedical signals, *IEEE Transactions on Biomedical Engineering*. **64**(12), 2872–2879, (2017).
83. E. Kafantaris, I. Piper, T.-Y. M. Lo, and J. Escudero, Augmentation of dispersion entropy for handling missing and outlier samples in physiological signal monitoring, *Entropy*. **22**(3), 319 (Mar, 2020). ISSN 1099-4300. doi: 10.3390/e22030319. URL http://dx.doi.org/10.3390/e22030319.
84. H. Azami, S. E. Arnold, S. Sanei, Z. Chang, G. Sapiro, J. Escudero, and A. S. Gupta, Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases, *IEEE Access*. **7**, 68718–68733, (2019).
85. J. P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, *Rev. Mod. Phys.* **57**(3), 617–656, (1985).
86. P. Grassberger and I. Procaccia, Estimation of the kolmogorov entropy from a chaotic signal, *Phys. Rev. A*. **28**(4), 2591–2593, (1983).
87. M. Costa, A. L. Goldberger, and C.-K. Peng, Multiscale entropy analysis of biological

signals, *Physical Review E*. **71**(2), 021906, (2005).
88. S. Ahmad, T. Ramsay, L. Huebsch, S. Flanagan, S. McDiarmid, I. Batkin, L. McIntyre, S. R. Sundaresan, D. E. Maziak, F. M. Shamji, P. Hebert, D. Fergusson, A. Tinmouth, and A. J. E. Seely, Continuous multi-parameter heart rate variability analysis heralds onset of sepsis in adults, *PLOS ONE*. **4**(8), 1–10 (08, 2009). doi: 10.1371/journal.pone.0006642. URL https://doi.org/10.1371/journal.pone.0006642.
89. Y.-L. Ho, C. Lin, Y.-H. Lin, and M.-T. Lo, The prognostic value of non-linear analysis of heart rate variability in patients with congestive heart failure—a pilot study of multiscale entropy, *PLOS ONE*. **6**(4), 1–6 (04, 2011). doi: 10.1371/journal.pone.0018699. URL https://doi.org/10.1371/journal.pone.0018699.
90. P. R. Norris, S. M. Anderson, J. M. Jenkins, A. E. Williams, and J. A. J. Morris, Heart rate multiscale entropy at three hours predicts hospital mortality in 3,154 trauma patients, *Shock*. **30**(1), 17–22, (2008). ISSN 1073-2322.
91. L. E. V. Silva, R. M. Lataro, J. A. Castania, C. A. A. da Silva, J. F. Valencia, L. O. Murta, H. C. Salgado, R. Fazan, and A. Porta, Multiscale entropy analysis of heart rate variability in heart failure, hypertensive, and sinoaortic-denervated rats: classical and refined approaches, *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*. **311**(1), R150–R156, (2016). doi: 10.1152/ajpregu.00076.2016. PMID: 27225948.
92. A. Humeau-Heurtier, The multiscale entropy algorithm and its variants: A review, *Entropy*. **17**(5), 3110–3123 (May, 2015). ISSN 1099-4300. doi: 10.3390/e17053110. URL http://dx.doi.org/10.3390/e17053110.
93. W. Chen, J. Zhuang, W. Yu, and Z. Wang, Measuring complexity using FuzzyEn, ApEn, and SampEn, *Med. Eng. Phys.* **31**, 61–68, (2009).
94. P. Castiglioni and M. Rienzo, How the threshold R influences approximate entropy analysis of heart-rate variability, *Comp. Cardiol.* **35**, 561–564, (2008).
95. H. Xie, W. He, and H. Liu, Measuring time series regularity using nonlinear similarity-based sample entropy, *Phys. Lett. A*. **372**, 7140–7146, (2008).
96. G. Xiong, L. Zhang, H. Liu, H. Zou, and W. Guo, A comparative study on ApEn, SampEn and their fuzzy counterparts in a multiscale framework for feature extraction, *J. Zhejiang Univ.-Science A*. **11**, 270–279, (2010).
97. C. Liu, K. Li, L. Zhao, F. Liu, D. Zheng, C. Liu, and S. Liu, Analysis of heart rate variability using fuzzy measure entropy, *Comp. Biol. Med.* **43**, 100–108, (2013).
98. L. Ji, P. Li, K. Li, X. Wang, and C. Liu, Analysis of short-term heart rate and diastolic period variability using a refined fuzzy entropy method, *Biomed. Eng. Online*. **14**(1), 64, (2015).
99. C. Liu, J. Oster, E. Reinertsen, Q. Li, L. Zhao, S. Nemati, and G. D. Clifford, A comparison of entropy approaches for af discrimination. physiological measurement, *Biomed. Signal Process. Control*. **39**(7), 074002, (2018).
100. J. M. Girault and A. Humeau-Heurtier, Centered and averaged fuzzy entropy to improve fuzzy entropy precision, *Entropy*. **20**(4), 287, (2018).
101. S. He, K. Sun, and R. Wang, Fractional fuzzy entropy algorithm and the complexity analysis for nonlinear time series, *The European Physical Journal Special Topics*. **227** (7–9), 943–957, (2018).
102. R. Zhou, X. Wang, J. Wan, and N. Xiong, EDM-fuzzy: An Euclidean distance based multiscale fuzzy entropy technology for diagnosing faults of industrial systems, *IEEE Transactions on Industrial Informatics*. (doi: 10.1109/TII.2020.3009139), (2020).
103. A. M. S. Borin Jr, L. E. V. Silva, and L. O. Murta Jr, Modified multiscale fuzzy entropy: A robust method for short-term physiologic signals, *Chaos: An Interdisciplinary Journal of Nonlinear Science*. **30**(8), article id. 083135, (2020).

104. X. Zhu, J. Zheng, H. Pan, J. Bao, and Y. Zhang, Time-shift multiscale fuzzy entropy and Laplacian support vector machine based rolling bearing fault diagnosis, *Entropy*. **20**(8), article id. 602, (2018).
105. H. Azami, A. Fernández, and J. Escudero, Refined multiscale fuzzy entropy based on standard deviation for biomedical signal analysis, *Medical & Biological Engineering & Computing*. **55**(11), 2037–2052, (2017).
106. J. Zheng, H. Pan, and J. Cheng, Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines, *Mechanical Systems and Signal Processing*. **85**, 746–759, (2017).
107. L. Kozachenko and N. N. Leonenko, Sample estimate of the entropy of a random vector, *Problemy Peredachi Informatsii*. **23**(2), 9–16, (1987).
108. W. Gao, S. Oh, and P. Viswanath, Demystifying fixed k-nearest neighbor information estimators, *IEEE Transactions on Information Theory*. **64**(8), 5629–5661, (2018).
109. T. B. Berrett, R. J. Samworth, M. Yuan, et al., Efficient multivariate entropy estimation via k-nearest neighbour distances, *The Annals of Statistics*. **47**(1), 288–318, (2019).
110. J. D. Victor, Binless strategies for estimation of information from neural data, *Physical Review E*. **66**(5), 051903, (2002).
111. G. T. Heineman, G. Pollice, and S. Selkow, *Algorithms in a nutshell: A practical guide*. (" O'Reilly Media, Inc.", 2016).
112. A. Kraskov, H. Stögbauer, and P. Grassberger, Estimating mutual information, *Physical review E*. **69**(6), 066138, (2004).
113. R. Vicente, M. Wibral, M. Lindner, and G. Pipa, Transfer entropy—a model-free measure of effective connectivity for the neurosciences, *Journal of computational neuroscience*. **30**(1), 45–67, (2011).
114. D. Kugiumtzis, Direct-coupling information measure from nonuniform embedding, *Physical Review E*. **87**(6), 062918, (2013).
115. L. Faes, D. Kugiumtzis, G. Nollo, F. Jurysta, and D. Marinazzo, Estimating the decomposition of predictive information in multivariate systems, *Physical Review E*. **91**(3), 032904, (2015).
116. M. Wibral, J. Lizier, S. Vögler, V. Priesemann, and R. Galuske, Local active information storage as a tool to understand distributed neural information processing, *Frontiers in neuroinformatics*. **8**, 1, (2014).
117. M. Valente, M. Javorka, A. Porta, V. Bari, J. Krohova, B. Czippelova, Z. Turianikova, G. Nollo, and L. Faes, Univariate and multivariate conditional entropy measures for the characterization of short-term cardiovascular complexity under physiological stress, *Physiological measurement*. **39**(1), 014002, (2018).
118. L. Faes, M. Gómez-Extremera, R. Pernice, P. Carpena, G. Nollo, A. Porta, and P. Bernaola-Galván, Comparison of methods for the assessment of nonlinearity in short-term heart rate variability under different physiopathological states, *Chaos: An Interdisciplinary Journal of Nonlinear Science*. **29**(12), 123114, (2019).
119. A. B. Barrett, L. Barnett, and A. K. Seth, Multivariate granger causality and generalized variance, *Physical Review E*. **81**(4), 041907, (2010).
120. L. Barnett, A. B. Barrett, and A. K. Seth, Granger causality and transfer entropy are equivalent for gaussian variables, *Physical review letters*. **103**(23), 238701, (2009).
121. K. Hlavácková-Schindler, Equivalence of granger causality and transfer entropy: A generalization, *Applied Mathematical Sciences*. **5**(73), 3637–3648, (2011).
122. G. D. Hutcheson, Ordinary least-squares regression, *L. Moutinho and GD Hutcheson, The SAGE dictionary of quantitative management research*. pp. 224–228, (2011).
123. J. Durbin, Estimation of parameters in time-series regression models, *Journal of the*

Royal Statistical Society: Series B (Methodological). **22**(1), 139–153, (1960).
124. L. Faes, A. Porta, G. Nollo, and M. Javorka, Information decomposition in multivariate systems: definitions, implementation and application to cardiovascular networks, *Entropy*. **19**(1), 5, (2017).
125. A. Beda, D. M. Simpson, and L. Faes, Estimation of confidence limits for descriptive indexes derived from autoregressive analysis of time series: Methods and application to heart rate variability, *PloS one*. **12**(10), e0183230, (2017).
126. L. Faes, A. Porta, M. Javorka, and G. Nollo, Efficient computation of multiscale entropy over short biomedical time series based on linear state-space models, *Complexity*. **2017**, (2017).
127. L. Faes, M. A. Pereira, M. E. Silva, R. Pernice, A. Busacca, M. Javorka, and A. P. Rocha, Multiscale information storage of linear long-range correlated stochastic processes, *Physical Review E*. **99**(3), 032115, (2019).
128. A. Martins, R. Pernice, C. Amado, A. P. Rocha, M. E. Silva, M. Javorka, and L. Faes, Multivariate and multiscale complexity of long-range correlated cardiovascular and respiratory variability series, *Entropy*. **22**(3), 315, (2020).
129. R. Udhayakumar, C. Karmakar, P. Li, and M. Palaniswami, Influence of embedding dimension on distribution entropy in analyzing heart rate variability, *Conf. Proc. IEEE Eng Med Biol Soc.* pp. 6222–6225, (2016).
130. Y. Li, P. Li, C. Karmakar, and C. Liu, Distribution entropy for short-term QT interval variability analysis: a comparison between the heart failure and normal control groups, *Computing in Cardiology Conference (CinC)*. pp. 1153–1156, (2015).
131. P. Li, C. Karmakar, C. Yan, M. Palaniswami, and C. Liu, Classification of 5-S epileptic EEG recordings using distribution entropy and sample entropy, *Front. Physiol.* **7**, 136, (2016).
132. T. Zhang, W. Chen, and M. Li, Complex-valued distribution entropy and its application for seizure detection, *Biocybernetics and Biomedical Engineering*. **40**, 306–323, (2020).
133. R. Udhayakumar, C. Karmakar, P. Li, X. Wang, and M. Palaniswami, Modified distribution entropy as a complexity measure of heart rate variability (hrv) signal, *Entropy*. **22**(10), (2020). ISSN 1099-4300. doi: 10.3390/e22101077. URL `https://www.mdpi.com/1099-4300/22/10/1077`.
134. S. T. Aung and Y. Wongsawat, Modified-distribution entropy as the features for the detection of epileptic seizures, *Frontiers in Physiology*. **11**, article id. 607, (2020).
135. T. Zhang, W. Chen, and M. Li, Fuzzy distribution entropy and its application in automated seizure detection technique, *Biomedical Signal Processing and Control*. **39**, 360–377, (2018).
136. X. Zhao, J. Cheng, P. Wang, Z. He, H. Shao, and Y. Yang, A novelty detection scheme for rolling bearing based on multiscale fuzzy distribution entropy and hybrid kernel convex hull approximation, *Measurement*. **156**, article id. 107589, (2020).
137. Y. Dai, J. He, Y. Wu, S. Chen, and P. Shang, Generalized entropy plane based on permutation entropy and distribution entropy analysis for complex time series, *Physica A: Statistical Mechanics and its Applications*. **520**, 217–231, (2019).
138. P. Li, K. Li, C. Liu, D. Zheng, Z. Li, and C. Liu, Detection of coupling in short physiological series by a joint distribution entropy method, *IEEE Trans. Biomed. Eng.* **63**, 2231–2242, (2016).
139. Y. Wang and P. Shang, Complexity analysis of time series based on generalized fractional order cumulative residual distribution entropy, *Physica A: Statistical Mechanics and its Applications*. **537**, article id. 122582, (2020).
140. M. Xu, P. Shang, and S. Zhang, Multiscale analysis of financial time series by Rényi

distribution entropy, *Physica A: Statistical Mechanics and its Applications*. **536**, article id. 120916, (2019).
141. A. Porta, V. Bari, A. Marchi, B. De Maria, P. Castiglioni, M. Di Rienzo, S. Guzzetti, A. Cividjian, and L. Quintin, Limits of permutation-based entropies in assessing complexity of short heart period variability, *Physiological measurement*. **36**(4), 755, (2015).
142. L. Zunino, F. Olivares, F. Scholkmann, and O. Rosso, Permutation entropy based time series analysis: Equalities in the input signal can lead to false conclusions, *Phys. Lett. A*. **381**, 1883–1892, (2017).
143. A. Myers and F. A. Khasawneh, On the automatic parameter selection for permutation entropy, *Chaos: An Interdisciplinary Journal of Nonlinear Science*. **30**(3), article id. 033130, (2020).
144. X. Wang, S. Si, Y. Wei, and Y. Li, The optimized multi-scale permutation entropy and its application in compound fault diagnosis of rotating machinery, *Entropy*. **21** (2), 170, (2019).
145. C. Yang and M. Jia, Hierarchical multiscale permutation entropy-based feature extraction and fuzzy support tensor machine with pinball loss for bearing fault identification, *Mechanical Systems and Signal Processing*. **149**, article id. 107182, (2021).
146. J. Zheng, Z. Dong, H. Pan, Q. Ni, T. Liu, and J. Zhang, Composite multi-scale weighted permutation entropy and extreme learning machine based intelligent fault diagnosis for rolling bearing, *Measurement*. **143**, 69–80, (2019).
147. Y. Li, G. Li, Y. Yang, X. Liang, and M. Xu, A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy, *Mechanical Systems and Signal Processing*. **105**, 319–337, (2018).
148. Y. Zhang and P. Shang, Refined composite multiscale weighted-permutation entropy of financial time series, *Physica A: Statistical Mechanics and its Applications*. **496**, 189–199, (2018).
149. J. Zheng, H. Pan, S. Yang, and J. Cheng, Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis, *Mechanical Systems and Signal Processing*. **99**, 229–243, (2018).
150. H. Niu, J. Wang, and C. Liu, Analysis of crude oil markets with improved multiscale weighted permutation entropy, *Physica A: Statistical Mechanics and its Applications*. **494**, 389–402, (2018).
151. H. Shaobo, S. Kehui, and W. Huihai, Modified multiscale permutation entropy algorithm and its application for multiscroll chaotic systems, *Complexity*. **21**(5), 52–58, (2016).
152. H. Azami and J. Escudero, Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings, *Biomed. Signal Process. Control*. **23**, 28–41, (2016).
153. S. D. Wu, C. W. Wu, and A. Humeau-Heurtier, Refined scale-dependent permutation entropy to analyze systems complexity, *Physica A: Statistical Mechanics and its Applications*. **450**, 454–461, (2016).
154. A. Humeau-Heurtier, C. W. Wu, and S. D. Wu, Refined composite multiscale permutation entropy to overcome multiscale permutation entropy length dependence, *IEEE Signal Process. Lett.* **22**(12), 2364–2367, (2015).
155. B. Fadlallah, B. Chen, A. Keil, and J. Principe, Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information, *Physical Review E*. **87**(2), 022911, (2013).
156. Y. Yin and P. Shang, Weighted multiscale permutation entropy of financial time series, *Nonlinear Dynamics*. **78**(4), 2921–2939, (2014).

157. J. Xia, P. Shang, J. Wang, and W. Shi, Permutation and weighted-permutation entropy analysis for the complexity of nonlinear time series, *Communications in Nonlinear Science and Numerical Simulation*. **31**, 60–68, (2016).
158. H. Azami and J. Escudero, Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation, *Comput. Methods Programs Biomed.* **128**, 40–51, (2016).
159. C. Bian, C. Qin, Q. D. Ma, and Q. Shen, Modified permutation-entropy analysis of heartbeat dynamics, *Physical Review E*. **85**(2), 021906, (2012).
160. X. Liu, G. Wang, J. Gao, and Q. Gao, A quantitative analysis for eeg signals based on modified permutation-entropy, *IRBM*. **38**(2), 71–77, (2017).
161. D. Cuesta–Frau, M. Varela–Entrecanales, A. Molina–Picó, and B. Vargas, Patterns with equal values in permutation entropy: Do they really matter for biosignal classification?, *Complexity*. **2018**, 1324696, (2018).
162. D. Cuesta–Frau, Permutation entropy: Influence of amplitude information on time series classification performance, *Math. Biosci. Eng.* **5**, 1–16, (2019).
163. F. Traversaro, N. Ciarrocchi, F. P. Cattaneo, and F. Redelico, Comparing different approaches to compute permutation entropy with coarse time series, *Physica A: Statistical Mechanics and its Applications*. **513**, 635–643, (2019).
164. C. Bandt, A new kind of permutation entropy used to classify sleep stages from invisible EEG microstructure, *Entropy*. **19**, 197, (2017).
165. J. Yin, P. Xiao, J. Li, Y. Liu, C. Yan, and Y. Zhang, Parameters analysis of sample entropy, permutation entropy and permutation ratio entropy for RR interval time series, *Information Processing & Management*. **57**, article id. 102283, (2020).
166. Y. Zhang, C. Liu, S. Wei, Y. Liu, and H. Liu, Complexity analysis of physiological time series using a novel permutation-ratio entropy, *IEEE Access*. **6**, 67653–67664, (2018).
167. Y. Zhang and P. Shang, Permutation entropy analysis of financial time series based on Hill's diversity number, *Communications in Nonlinear Science and Numerical Simulation*. **53**, 288–298, (2017).
168. K. Xu and J. Wang, Weighted fractional permutation entropy and fractional sample entropy for nonlinear Potts financial dynamics, *Physics Letters A*. **381**, 767–779, (2017).
169. Y. Yin and P. Shang, Weighted permutation entropy based on different symbolic approaches for financial time series, *Physica A: Statistical Mechanics and its Applications*. **443**, 137–148, (2016).
170. S. Chen, P. Shang, and Y. Wu, Weighted multiscale Rényi permutation entropy of nonlinear time series, *Physica A: Statistical Mechanics and its Applications*. **496**, 548–570, (2018).
171. J. S. Cánovas, G. García-Clemente, and M. Muñoz Guillermo, Comparing permutation entropy functions to detect structural changes in time series, *Physica A: Statistical Mechanics and its Applications*. **507**, 153–174, (2018).
172. Y. Bai, Z. Liang, and X. Li, A permutation Lempel-Ziv complexity measure for EEG analysis, *Biomedical Signal Processing and Control*. **19**, 102–114, (2015).
173. Z. Liu, P. Shang, and Y. Wang, Multifractal weighted permutation analysis based on Rényi entropy for financial time series, *Physica A: Statistical Mechanics and its Applications*. **536**, article id. 120994, (2019).
174. M. Tao, K. Poskuviene, N. F. Alkayem, M. Cao, and M. Ragulskis, Permutation entropy based on non-uniform embedding, *Entropy*. **20**(8), article id. 612, (2018).
175. L. Tylová, J. Kukal, V. Hubata-Vacek, and O. Vysata, Unbiased estimation of permutation entropy in EEG analysis for Alzheimer's disease classification, *Biomedical*

Signal Processing and Control. **39**, 424–430, (2018).
176. Z. Chen, Y. Li, and J. Liang, H.and Yu, Improved permutation entropy for measuring complexity of time series under noisy condition, *Complexity*. **2019**, 1403829, (2019).
177. C. li and P. Shang, Multiscale Tsallis permutation entropy analysis for complex physiological time series, *Physica A: Statistical Mechanics and its Applications*. **523**, 10–20, (2019).
178. S. J. Watt and A. Politi, Permutation entropy revisited, *Chaos, Solitons & Fractals*. **120**, 95–99, (2019).
179. B. Yan, S. He, and K. Sun, Design of a network permutation entropy and its applications for chaotic time series and eeg signals, *Entropy*. **21**(9), article id. 849, (2019).
180. Y. Gao, H. Su, R. Li, and Y. Zhang, Synchronous analysis of brain regions based on multi-scale permutation transfer entropy, *Computers in Biology and Medicine*. **109**, 272–279, (2019).
181. X. Zhao, M. Ji, N. Zhang, and P. Shang, Permutation transition entropy: Measuring the dynamical complexity of financial time series, *Chaos, Solitons & Fractals*. **139**, article id. 109962, (2020).
182. Z. Zhang, Z. Xiang, Y. Chen, and J. Xu, Fuzzy permutation entropy derived from a novel distance between segments of time series, *AIMS Mathematics*. **5**(6), 6244–6260, (2020).
183. H. Haken. Pattern formation and pattern recognition—an attempt at a synthesis. In *Pattern formation by dynamic systems and pattern recognition*, pp. 2–13. Springer, (1979).
184. R. C. Gonzalez and R. E. Woods, Digital image processing. (2008).
185. D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, Sample entropy analysis of neonatal heart rate variability, *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*. **283**(3), R789–R797, (2002).
186. S. Lu, X. Chen, J. K. Kanters, I. C. Solomon, and K. H. Chon, Automatic selection of the threshold value r for approximate entropy, *IEEE Transactions on Biomedical Engineering*. **55**(8), 1966–1972, (2008).
187. F. Kaffashi, R. Foglyano, C. G. Wilson, and K. A. Loparo, The effect of time delay on approximate & sample entropy calculations, *Physica D: Nonlinear Phenomena*. **237**(23), 3069–3074, (2008).
188. S. Berger, G. Schneider, E. F. Kochs, and D. Jordan, Permutation entropy: Too complex a measure for eeg time series?, *Entropy*. **19**(12), 692, (2017).
189. R. Andrzeja. *Epileptologie Bonn / Forschung / AG Lehnertz / EEG Data Download*, (2007 (accessed 21 September, 2020)). `http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3&changelang=3`.
190. A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals, *circulation*. **101**(23), e215–e220, (2000).
191. *Physionet*, (2020 (accessed 30 September2020)). `http://www.physionet.org`.
192. N. Iyengar, C. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics, *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*. **271**(4), R1078–R1084, (1996).
193. D. Abásolo, J. Escudero, R. Hornero, C. Gómez, and P. Espino, Approximate entropy and auto mutual information analysis of the electroencephalogram in alzheimer's disease patients, *Medical & biological engineering & computing*. **46**(10), 1019–1028,

(2008).
194. B. García-Martínez, A. Martinez-Rodrigo, R. Alcaraz, and A. Fernández-Caballero, A review on nonlinear methods using electroencephalographic recordings for emotion recognition, *IEEE Transactions on A ective Computing*. (2019).
195. R. Hornero, J. Escudero, A. Fernández, J. Poza, and C. Gómez, Spectral and nonlinear analyses of meg background activity in patients with alzheimer's disease, *IEEE Transactions on Biomedical Engineering*. **55**(6), 1658–1665, (2008).
196. Y. JIANG, D. MAO, and Y. XU, A fast algorithm for computing sample entropy, *Advances in Adaptive Data Analysis*. **03**(01n02), 167–186, (2011). doi: 10.1142/S1793536911000775. URL https://doi.org/10.1142/S1793536911000775.
197. G. Manis, M. Aktaruzzaman, and R. Sassi, Low computational cost for sample entropy, *Entropy*. **20**(1), 61 (Jan, 2018). ISSN 1099-4300. doi: 10.3390/e20010061. URL http://dx.doi.org/10.3390/e20010061.
198. M. Martínez-Zarzuela, C. Gómez, F. J. Díaz-Pernas, A. Fernández, and R. Hornero, Cross-approximate entropy parallel computation on gpus for biomedical signal analysis. application to meg recordings, *Computer methods and programs in biomedicine*. **112**(1), 189–199, (2013).
199. P. Wollstadt, M. Martínez-Zarzuela, R. Vicente, F. J. Díaz-Pernas, and M. Wibral, Efficient transfer entropy analysis of non-stationary neural time series, *PLOS ONE*. **9**(7), 1–21 (07, 2014). doi: 10.1371/journal.pone.0102833. URL https://doi.org/10.1371/journal.pone.0102833.
200. J. M. Yentes, N. Hunt, K. K. Schmid, J. P. Kaipust, D. McGrath, and N. Stergiou, The appropriate use of approximate entropy and sample entropy with short data sets, *Annals of Biomedical Engineering*. **41**(2), 349–365, (2013).
201. M. Zanin, L. Zunino, O. A. Rosso, and D. Papo, Permutation entropy and its main biomedical and econophysics applications: a review, *Entropy*. **14**(8), 1553–1577, (2012).
202. J. Kurths, A. Voss, P. Saparin, A. Witt, H. Kleiner, and N. Wessel, Quantitative analysis of heart rate variability, *Chaos: An Interdisciplinary Journal of Nonlinear Science*. **5**(1), 88–94, (1995).
203. B.-l. Hao, Symbolic dynamics and characterization of complexity, *Physica D: Nonlinear Phenomena*. **51**(1-3), 161–176, (1991).
204. A. Voss, J. Kurths, H. Kleiner, A. Witt, and N. Wessel, Improved analysis of heart rate variability by methods of nonlinear dynamics, *Journal of Electrocardiology*. **28**, 81–88, (1995).
205. A. J. Ibáñez-Molina, S. Iglesias-Parro, and J. Escudero, Differential effects of simulated cortical network lesions on synchrony and eeg complexity, *International Journal of Neural Systems*. **29**(04), 1850024, (2019).
206. J. Cabral, E. Hugues, O. Sporns, and G. Deco, Role of local network oscillations in resting-state functional connectivity, *Neuroimage*. **57**(1), 130–139, (2011).
207. L. Faes, G. Nollo, and A. Porta. Information decomposition: A tool to dissect cardiovascular and cardiorespiratory complexity. In *Complexity and Nonlinearity in Cardiovascular Signals*, pp. 87–113. Springer, (2017).